

**UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS**

Pedro Lucas Fernandes Carvalho

**Previsão de incertezas em séries temporais de geração de
energia fotovoltaica**

**São Carlos
2025**

Pedro Lucas Fernandes Carvalho

Previsão de incertezas em séries temporais de geração de energia fotovoltaica

Monografia apresentada ao Curso de Engenharia Elétrica com Ênfase em Sistemas de Energia e Automação, da Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Engenheiro Eletricista.

Orientador: Prof. Dr. Luís Fernando Costa Alberto

**São Carlos
2025**

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da
EESC/USP com os dados inseridos pelo(a) autor(a).

F331p	<p>Fernandes Carvalho, Pedro Lucas</p> <p>Previsão de incertezas em séries temporais de geração de energia fotovoltaica / Pedro Lucas Fernandes Carvalho; orientador Luis Fernando Costa Alberto. São Carlos, 2025.</p> <p>Monografia (Graduação em Engenharia Elétrica com ênfase em Sistemas de Energia e Automação) -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2025.</p> <p>1. Previsão de energia fotovoltaica. 2. Séries temporais. 3. Incerteza. 4. Redes Neurais em Grafos. 5. XGBoost. 6. Modelagem espaço-temporal. I. Título.</p>
-------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

FOLHA DE APROVAÇÃO

Nome: Pedro Lucas Fernandes Carvalho

Título: "Previsão de incertezas em séries temporais de geração de energia fotovoltaica"

Trabalho de Conclusão de Curso defendido e aprovado
em 05 / 12 / 2025,

com NOTA dez (10, 0), pela Comissão
Julgadora:

Prof. Dr. Luís Fernando Costa Alberto - Professor Titular -
Orientador - SEL/EESC/USP

Prof. Dr. Carlos Dias Maciel - Professor Titular UNESP/Campus de
Guaratinguetá

Dr. Pedro Inácio de Nascimento e Barbalho - Pós-Doutorando
EESC/USP

Coordenador da CoC-Engenharia Elétrica - EESC/USP:
Professor Associado José Carlos de Melo Vieira Júnior

AGRADECIMENTOS

Aos meus pais, João Maria e Maria Cristina, pelo suporte fundamental para a concretização do meu ciclo universitário, e aos meus irmãos, João Victor e Maria Júlia, pela presença e torcida, mesmo que à distância.

A todos os meus professores da graduação, pela passagem do conhecimento e ensinamentos.

Ao meu orientador Luis Fernando, pela supervisão e oportunidade de aprender fazer pesquisa científica.

A Vitória Augusti e Gabriela Ramalho, pela supervisão clínica psicológica, cuidados e escuta até esse resultado final.

Aos amigos e colegas que fiz durante a graduação, pelo conhecimento construído, ideias compartilhadas e memórias criadas em conjunto.

Wonder En Is Gheen Wonder
Stevin Pascal

RESUMO

CARVALHO, Pedro L. F. **Previsão de incertezas em séries temporais de geração de energia fotovoltaica.** 2025. 101 p. Trabalho de Conclusão de Curso - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2025.

O avanço da energia fotovoltaica na matriz energética brasileira impõe desafios à operação do Sistema Interligado Nacional (SIN) devido à sua natureza intermitente. Este trabalho aborda a previsão de incertezas na geração solar, propondo uma comparação sistemática entre modelos de aprendizado de máquina locais e globais. A metodologia investiga a hipótese de que a modelagem das interdependências espaciais entre usinas, representadas como um grafo, melhora a previsibilidade e a quantificação das incertezas. Foram utilizados dados de geração de 47 usinas, fornecidos pelo Operador Nacional do Sistema (ONS), e dados climáticos da NASA Power API. Como linha de base, foi implementado um modelo local robusto, o XGBoost, treinado individualmente para usinas-arquétipo (grupos de usinas com características estruturais e de conectividade similares). Em contrapartida, foram desenvolvidos modelos espaço-temporais baseados em Graph Neural Networks (GNNs), especificamente as arquiteturas Graph Attention Network-Long Short Term Memory (GAT-LSTM) e Graph Convolutional Network - 1 Dimension (GCN-1D), que processam simultaneamente todo o conjunto de usinas. Todos os modelos foram treinados com a função de perda *pinball* para gerar previsões quantílicas (0.1, 0.5 e 0.9), permitindo a construção de intervalos de predição e a análise das incertezas associadas. Os resultados demonstram que, embora os modelos locais apresentem boa aderência a eventos específicos, as abordagens baseadas em grafos são capazes de capturar parcialmente as dinâmicas de geração locais, oferecendo uma visão geral para a gestão do portfólio de geração e contribuindo para futuros estudos da segurança energética do sistema.

Previsão de energia fotovoltaica. Séries temporais. Incerteza. Redes Neurais em Grafos. XGBoost. Modelagem espaço-temporal.

Abstract

CARVALHO, Pedro L. F. **Undergraduate Thesis**. 2025. 101 f.

The expansion of photovoltaic energy in the Brazilian energy matrix poses challenges to the operation of the National Interconnected System (SIN) due to its intermittent nature. This work addresses uncertainty forecasting in solar power generation by proposing a systematic comparison between local and global machine learning models. The methodology investigates the hypothesis that modeling spatial interdependencies between power plants, represented as a graph, improves predictability and uncertainty quantification. Generation data from 47 power plants, provided by the National System Operator (ONS), and climate data from the NASA Power API were utilized. As a baseline, a robust local model, XGBoost, was implemented and trained individually for archetype plants (groups of power plants with similar structural and connectivity characteristics). In contrast, spatio-temporal models based on Graph Neural Networks (GNNs) were developed, specifically the GAT-LSTM and GCN-1D architectures, which simultaneously process the entire set of plants. All models were trained with the *pinball* loss function to generate quantile forecasts (0.1, 0.5, and 0.9), enabling the construction of prediction intervals and the analysis of associated uncertainties. The results demonstrate that while local models show good adherence to specific events, graph-based approaches are capable of partially capturing complex spatial dynamics. This offers a comprehensive view for generation portfolio management and contributes to the energy security of the system.

Photovoltaic energy forecasting. Time series. Uncertainty. Graph Neural Networks. XGBoost. Spatio-temporal modeling.

Undergraduate Thesis - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2025.

LISTA DE FIGURAS

Figura 1 – Distribuição Geográfica das 64 Usinas Fotovoltaicas Analisadas no Brasil, com destaque para a concentração no "cinturão solar".	30
Figura 2 – Deterioração dos dados horários de irradiância (ALLSKY_SFC_SW_DWN) e índice de claridade (ALLSKY_KT) para a usina Janaúba.	34
Figura 3 – Comportamento similar de deterioração dos dados meteorológicos (ALLSKY_SFC_SW_DWN e ALLSKY_KT) observado para a usina Belmonte.	34
Figura 4 – Visualização geográfica da rede de 47 usinas fotovoltaicas e suas interconexões baseadas no critério de k -vizinhos mais próximos ($k \approx 4$). Embora a escala de visualização continental possa sugerir a aparência de nós isolados em regiões de alta densidade (onde múltiplas usinas compartilham coordenadas muito próximas), a definição algorítmica do grafo via k -vizinhos assegura a conectividade global, mesmo naqueles nós cujas arestas são visualmente imperceptíveis devido à proximidade geográfica.	38
Figura 5 – Distribuições das métricas de centralidade para a rede de 47 usinas solares: (a) Centralidade de Grau, (b) Centralidade de Intermediação, (c) Coeficiente de Aglomeração Local e (d) PageRank.	39
Figura 6 – Visualização da rede de usinas solares destacando arquétipos baseados em sua influência estrutural (PageRank) e conectividade.	45
Figura 7 – Distribuição da geração para as usinas arquétipos - Hub (Lapa), Nó Isolado (Floresta) e Membro de Comunidade (Castilho). As estatísticas descritivas apresentadas referem-se à distribuição global da amostra observada na janela de análise. Ressalta-se que a série temporal bruta de geração solar é inerentemente não-estacionária devido à forte componente sazonal diária.	46
Figura 8 – Uma rede neural recorrente é composta por células idênticas e sequenciais. Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).	56
Figura 9 – A célula do LSTM é composta por múltiplas portas que regulam o fluxo de informação. Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).	57
Figura 10 – Porta de esquecimento: controla o que deve ser removido do estado de célula anterior.	57
Figura 11 – As portas de entrada e os candidatos de memória colaboram na atualização do estado interno da célula. Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).	58

Figura 12 – Porta de saída: define quais informações da célula serão utilizadas como saída h_t . Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).	58
Figura 13 – Porta de saída: define quais informações da célula serão utilizadas como saída h_t . Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).	59
Figura 14 – Exemplos de grafos não direcionado e direcionado. Figura de autoria própria.	62
Figura 15 – Comparação entre grafos não ponderado e ponderado, evidenciando a ausência ou presença de pesos nas arestas. Figura de autoria própria.	63
Figura 16 – Grafo conectado e não conectado. Figura de autoria própria.	63
Figura 17 – Ilustração do processo de construção do grafo via k-NN com $k = 2$: (a) disposição inicial dos nós representando usinas; (b) seleção unilateral dos k -vizinhos mais próximos (setas tracejadas); (c) grafo final não direcionado, onde uma aresta (i, j) existe se i está entre os k vizinhos de j ou vice-versa. Figura de autoria própria.	65
Figura 18 – O diagrama detalha as operações de agregação e atualização para um nó de interesse (nó 5).). Figura de autoria própria.	67
Figura 19 – O diagrama detalha como a representação do nó central (nó 1) é atualizada a partir de suas próprias características ($h_1^{(k-1)}$) e das características de seus vizinhos ($h_j^{(k-1)}$, para $j \in \{2, 3, 4, 5\}$). Figura de autoria própria.	69
Figura 20 – A cada instante de tempo, as características iniciais (X_t) de todas as usinas são processadas por uma camada GAT para capturar as dependências espaciais, gerando embeddings enriquecidos. Figura de autoria própria.	70
Figura 21 – Arquitetura do modelo GCN-1D. As entradas são processadas por uma camada GCNConv para agregação espacial, seguida por uma convolução 1D causal que extrai padrões temporais para gerar a previsão final . . .	72
Figura 22 – Comparativo de desempenho dos modelos GAT-LSTM, GCN-1D e XGBoost para o horizonte de previsão H+1.	81
Figura 23 – Comparativo de desempenho dos modelos GAT-LSTM, GCN-1D e XGBoost para o horizonte de previsão H+3.	82
Figura 24 – Comparativo de desempenho dos modelos GAT-LSTM, GCN-1D e XGBoost para o horizonte de previsão H+6.	83
Figura 25 – Comparativo de desempenho dos modelos GAT-LSTM, GCN-1D e XGBoost para o horizonte de previsão H+12.	83
Figura 26 – Comparativo de desempenho dos modelos GAT-LSTM, GCN-1D e XGBoost para o horizonte de previsão H+24.	84

LISTA DE TABELAS

Tabela 1 – Distribuição de dados ausentes da variável <code>val_geracao</code> por usina fotovoltaica.	33
Tabela 2 – Períodos com dados ausentes da variável <code>val_geracao</code> por usina fotovoltaica.	35
Tabela 3 – Lista final das 47 usinas fotovoltaicas selecionadas para análise.	36
Tabela 4 – Estatísticas descritivas comparativas para a geração horária (MW) das usinas arquétipo (Lapa, Floresta, Castilho) no período de análise.	46
Tabela 5 – Resumo estatístico por tipo de relação — Estrutura e ADF	48
Tabela 6 – Resumo estatístico por tipo de relação — Ljung-Box e Resíduos	48
Tabela 7 – Resumo estatístico por tipo de relação — Informação Mútua	50
Tabela 9 – Features utilizadas nos modelos GCN1D e GAT-LSTM, organizadas por categoria.	51
Tabela 8 – Features utilizadas no modelo XGBoost, organizadas por categoria.	52
Tabela 10 – Hiperparâmetros utilizados no modelo GAT-LSTM	78
Tabela 11 – Hiperparâmetros utilizados no modelo GCN-1D.	78
Tabela 12 – Comparativo de Métricas de Acurácia (RMSE e R^2) para a Previsão Mediana (Quantil 0.5) por Arquétipo e Horizonte de Previsão.	86
Tabela 13 – Validação dos arquétipos no modelo GAT-LSTM.	87
Tabela 14 – Validação dos arquétipos no modelo GCN-1D.	87
Tabela 15 – Desempenho sob cenários de estresse (rampas > 91.90 MW) - GAT-LSTM	87
Tabela 16 – Resultados do modelo GCN1D no conjunto de teste.	96
Tabela 17 – Resultados do modelo GCN1D para o quantil 0.5 no conjunto de teste.	96
Tabela 18 – Resultados do modelo GCN1D para o quantil 0.9 no conjunto de teste.	96
Tabela 19 – Métricas de desempenho do GAT_LSTM para o quantil 0.1.	97
Tabela 20 – Métricas de desempenho do GAT_LSTM para o quantil 0.5 (Mediana).	97
Tabela 21 – Métricas de desempenho do GAT_LSTM para o quantil 0.9.	97
Tabela 22 – Métricas de desempenho do XGBoost para o quantil 0.1 por Arquétipo.	97
Tabela 23 – Métricas de desempenho do XGBoost para o quantil 0.5 (Mediana) por Arquétipo.	98
Tabela 24 – Métricas de desempenho do XGBoost para o quantil 0.9 por Arquétipo.	98
Tabela 25 – Métricas de incerteza por usina no modelo GAT_LSTM, ordenadas por PICP.	99
Tabela 26 – Métricas de incerteza por usina no modelo GCN-1D, ordenadas por PICP.	99

LISTA DE ABREVIATURAS E SIGLAS

ADF	Augmented Dickey-Fuller
API	Application Programming Interface
GAT	Graph Attention Network
GCN	Graph Convolutional Network
GNN	Graph Neural Network
IPCC	Intergovernmental Panel on Climate Change
k-NN	k-Nearest Neighbors
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MI	Mutual Information
MW	Megawatt
ONS	Operador Nacional do Sistema Elétrico
PICP	Prediction Interval Coverage Probability
PINAW	Prediction Interval Normalized Average Width
RMSE	Root Mean Squared Error
SARIMA	Seasonal AutoRegressive Integrated Moving Average
SIN	Sistema Interligado Nacional
XGBoost	Extreme Gradient Boosting

LISTA DE SÍMBOLOS

G	Grafo
V	Conjunto de vértices (nós) de um grafo
A	Conjunto de arestas de um grafo ou Matriz de adjacência
\hat{A}	Matriz de adjacência normalizada com auto-conexões
y_t	Valor observado de uma série temporal no instante t
\hat{y}	Valor previsto pelo modelo
$H^{(k)}$	Matriz de representações (embeddings) dos nós na camada k de uma GNN
$W^{(k)}$	Matriz de pesos treináveis da camada k
$\sigma(\cdot)$	Função de ativação não linear (por exemplo, ReLU, sigmoide)
α_{ij}	Peso de atenção do nó j em relação ao nó i em uma GAT
L_τ	Função de perda quantílica (<i>pinball loss</i>)
τ	Quantil alvo (valor entre 0 e 1)
η	Taxa de aprendizado
λ	Parâmetro de regularização L2
γ	Parâmetro de regularização gamma (complexidade da árvore)

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Contextualização	21
1.2	Justificativa	22
1.3	Objetivos	23
1.4	Organização do trabalho	24
2	TRABALHOS RELACIONADOS	25
3	ANÁLISE EXPLORATÓRIA DE DADOS	28
3.1	Visão Geral	28
3.1.1	Dados do Operador Nacional do Sistema(ONS)	28
3.1.2	Dados Climáticos — NASA Power API	29
3.2	Cenário de Análise	32
3.3	Análise da estrutura em grafo	36
3.4	Análise de séries temporais e estimação dos resíduos	41
3.5	Análise funcional	44
3.6	Engenharia de variáveis	51
4	METODOLOGIA	53
4.1	XGBoost	53
4.2	LSTM	55
4.3	Modelos de Aprendizado Profundo em Grafos	60
4.3.1	Fundamentos	61
4.3.2	Construção do grafo	64
4.4	GAT-LSTM	68
4.5	GCN1D	70
4.6	Limitações na modelagem das GNNs	72
4.7	Previsão conforme e incertezas	73
4.8	Função de Perda	74
4.9	Métricas de Avaliação	76
4.10	Configuração Experimental	77
5	RESULTADOS E DISCUSSÃO	80
5.1	Desempenho geral e comparativo dos modelos	80
5.2	Análise da calibração e desempenho em rampas	85
5.3	Limitações Observadas	87

6	CONSIDERAÇÕES FINAIS	89
	REFERÊNCIAS	91
	APÊNDICES	95
	APÊNDICE A – RESULTADOS DETALHADOS E MÉTRICAS POR USINA	96

1 INTRODUÇÃO

1.1 Contextualização

O crescimento da energia solar na matriz energética brasileira impõe novos desafios ao planejamento e à operação do Sistema Interligado Nacional (SIN). Consequentemente, são necessárias adequações na infraestrutura de transmissão e na capacidade de controle operacional do sistema, devido às características próprias da geração solar, variabilidade da geração e caráter não despachável, conforme destacado no Plano da Operação Energética 2023–2027 do Operador Nacional do Sistema (ONS) (ONS, 2023).

Além disso, a intermitência da geração fotovoltaica exige maior capacidade de antecipação e controle. O mesmo documento ressalta que aspectos como o controle de tensão e a manutenção da segurança operacional do sistema são diretamente afetados pela natureza variável da geração, um desafio intensificado pela crescente integração de fontes não despacháveis ao SIN.

Diante desse cenário, é necessário desenvolver modelos preditivos das incertezas da geração fotovoltaica visando: a segurança energética, confiabilidade dos modelos e incertezas quantificáveis. Tendo em vista as dinâmicas de geração em comum das usinas, vide padrão de geração físico local, a incerteza depende da interação de variáveis desconhecidas entre as usinas vizinhas; ou seja, ela é correlacionada com a incerteza de suas vizinhas, um aspecto que modelos de múltiplos nós, como os baseados em grafos, são mais aptos a capturar.

Este estudo se propõe, portanto, a analisar a dinâmica da geração fotovoltaica e a desenvolver ferramentas de previsão de incerteza que considerem não apenas o comportamento individual das usinas, mas também suas interdependências espaciais.

As hipóteses consideradas neste trabalho são de que a conexão das usinas em grafos pode trazer representações estruturais do caráter operacional e de regimes climáticos similares, o que levaria a uma melhor previsão da geração solar a nível regional — as regressões feitas nos grafos acontecem em todas as suas usinas, o que permite prever todo um conjunto de usinas com um único modelo.

Assim, sob a perspectiva de grafos, as usinas fotovoltaicas são representadas como nós, aos quais se associam dados horários de geração e informações climáticas. Esses nós podem ser conectados com base em critérios de proximidade geográfica entre as usinas, permitindo investigar como a estrutura espaço-temporal da rede influencia na previsibilidade da geração.

A modelagem espaço-temporal do problema também encontra respaldo nas preocupações globais com as mudanças climáticas - o Intergovernmental Panel on Climate Change (IPCC) destaca que a variabilidade climática crescente afetará diretamente os sistemas de energia de baixo carbono, inclusive a previsibilidade da geração renovável. Além disso, a

transição para um paradigma de sistema de eletricidade digitalizável e controlável deve facilitar drásticas mudanças na segurança da oferta, levando as formas tradicionais de redundâncias dos ativos para um novo paradigma de controle inteligente. (IPCC, 2022).

A previsão da geração solar é fortemente associada a fatores climáticos — como a irradiância solar — e, por isso, o uso de variáveis exógenas climáticas é comum na literatura. Nesse contexto, o presente estudo aprofunda essa linha de pesquisa ao empregar modelos avançados que não apenas utilizam variáveis climáticas exógenas, mas as integram em uma perspectiva espaço-temporal, notadamente através de redes neurais em grafos.

Para contextualizar as contribuições deste trabalho e suas abordagens, a próxima seção apresenta um panorama dos estudos relevantes na literatura.

1.2 Justificativa

Apesar dos avanços na previsão de séries temporais, a análise da literatura revela que a abordagem predominante, tanto acadêmica quanto na prática industrial, foca em modelos independentes para cada usina. Modelos de alta performance, como o XGBoost, uma implementação otimizada para o algoritmo de árvore de decisão com aumento de gradiente (CHEN; GUESTRIN, 2016), frequentemente mantêm a premissa de tratar cada ativo como um sistema isolado. Embora eficaz para a acurácia local, esta abordagem ignora as dinâmicas compartilhadas de geração e impõe desafios operacionais significativos, como a complexa gestão de um portfólio crescente de modelos e a incapacidade de prover estimativas para novos ativos sem dados históricos.

Neste contexto, arquiteturas como Graph Neural Network(GNN) e Graph Convolutional Network(GCN) surgem como uma alternativa promissora, oferecendo um referencial unificado e escalável com potencial de generalização indutiva, i.e., aplicável em usinas que não foram utilizadas no treinamento de modelos de aprendizado. Contudo, apesar de seu potencial teórico, ainda carecem de comparações diretas com referências de base locais robustas sob condições equivalentes.

Essa carência é particularmente acentuada no contexto do SIN, um cenário complexo, de escala continental e diversidade climática, cuja exploração com tais técnicas ainda é incipiente.

É precisamente para preencher essa dupla lacuna — operacional e acadêmica — que este trabalho se justifica. Ele propõe uma comparação sistemática entre as duas frentes de modelagem. Ao confrontar o desempenho de um modelo local robusto com o de arquiteturas globais baseadas em grafos, busca-se oferecer um entendimento mais claro sobre as bases teóricas do problema e as compensações práticas de cada abordagem no contexto brasileiro.

Para guiar esta investigação, o trabalho parte da hipótese fundamental de que a modelagem explícita das interdependências espaciais entre as usinas fotovoltaicas, por meio de uma representação em grafo, oferece ganhos tanto na acurácia preditiva quanto na

qualidade da quantificação de incerteza, quando comparada a modelos puramente locais.

Essa melhoria proposta se desdobra em duas dimensões principais. No que tange à previsibilidade, postula-se que a informação contextual dos vizinhos, propagada através do grafo, permite aos modelos capturar padrões climáticos e operacionais regionais, resultando em uma maior capacidade de explicar a variância da geração (maior R^2), especialmente em horizontes de previsão de curtíssimo prazo.

Adicionalmente, no que se refere à incerteza, a hipótese é de que a modelagem conjunta do sistema de usinas permite uma melhor caracterização da incerteza espacialmente correlacionada, levando à construção de intervalos de predição mais informativos e adequados para a gestão de portfólio.

1.3 Objetivos

Para testar as hipóteses apresentadas e atender à justificativa do estudo, o trabalho foi estruturado em torno dos seguintes objetivos específicos:

- a) Coletar e realizar uma análise exploratória dos dados de geração fotovoltaica da ONS e de dados climáticos provenientes da NASA Power API;
- b) Definir o cenário de análise, compreender o comportamento temporal e espacial das usinas e identificar relações estruturais entre elas;
- c) Caracterizar o grafo das usinas e entender como a correlação espacial e a variabilidade local se relacionam com a geração de energia solar;
- d) Aplicar os fundamentos da modelagem clássica de séries temporais, utilizando modelos SARIMA e a análise de resíduos (incluindo decomposições em componentes brancos e não brancos), para investigar a dinâmica temporal das usinas e explorar relações de dependência e informação mútua entre os nós do grafo;
- e) Estudar e implementar arquiteturas de aprendizado de máquina capazes de modelar as dependências espaço-temporais da geração solar, incluindo modelos baseados em gradiente - Extreme Gradient Boosting (XGBoost) e redes em grafos (GAT-LSTM e GCN1D). Avaliar o potencial dessas arquiteturas para representar interações entre múltiplas usinas de geração solar;
- f) Implementar e comparar o desempenho dos modelos em diferentes horizontes de previsão (Hora+1, Hora+3, Hora+6, Hora+12 e Hora+24), utilizando métricas de erro e calibração;
- g) Avaliar tanto a acurácia das previsões dos quantis, quanto a estabilidade dos modelos sob diferentes escalas temporais;
- h) Estimar e comparar as incertezas associadas às previsões por meio de regressões quantílicas (função de perda pinball para os quantis 0.1, 0.5 e 0.9) e da análise dos resíduos dos modelos;

- i) Discutir a origem dessas incertezas, relacionando-as à variabilidade observada no conjunto de testes e às limitações de cada abordagem, com o intuito de compreender melhor o papel da incerteza nas previsões de de vários nós.

1.4 Organização do trabalho

O presente trabalho está estruturado em seis capítulos, organizados de forma a conduzir o leitor desde a contextualização do problema até a análise dos resultados obtidos.

No capítulo 1, são apresentados o contexto geral da pesquisa, a motivação, os objetivos e a justificativa do estudo, bem como a estrutura do trabalho.

O capítulo 2 aborda os trabalhos relacionados, discutindo os principais avanços na previsão da geração fotovoltaica, com ênfase em abordagens espaço-temporais, técnicas de aprendizado de máquina e métodos recentes de quantificação de incertezas, incluindo regressões quantílicas e redes neurais em grafos.

O capítulo 3 descreve os conjuntos de dados utilizados e as etapas de pré-processamento, incluindo a caracterização das usinas fotovoltaicas, o tratamento de séries temporais e a construção das estruturas em grafo que representam as relações espaciais entre as usinas.

O capítulo 4 apresenta a metodologia desenvolvida, detalhando os modelos empregados — estatísticos, de aprendizado de máquina e de aprendizado profundo —, a formulação das funções de perda quantílica, os critérios de avaliação e a estratégia de experimentação adotada.

No capítulo 5, são expostos e discutidos os resultados obtidos, com foco na análise comparativa entre os modelos, no comportamento dos resíduos, na calibração das previsões e na interpretação das incertezas associadas.

Por fim, o Capítulo 6 sintetiza as conclusões do trabalho, destacando as principais contribuições e limitações encontradas, além de propor direções para estudos futuros voltados ao aprimoramento da modelagem de incertezas em sistemas de geração solar de várias usinas.

2 TRABALHOS RELACIONADOS

A previsão da geração de energia solar é um problema consolidado na literatura da engenharia elétrica e em áreas correlatas que lidam com séries temporais. Com a crescente integração de fontes intermitentes, a criação e manutenção de modelos preditivos robustos tornou-se uma necessidade fundamental para a segurança e a estabilidade da operação do sistema energético.

A escolha dos algoritmos de estudo dependem de qual otimização e escala de estudo é feita, de modo que os critérios de escolha dos algoritmos devem ser definidos com base no escopo de análise e viabilidade prática em relação aos dados disponíveis.

Assim, a literatura existente pode ser categorizada por múltiplas dimensões, como o número de pontos de previsão, o horizonte temporal e as variáveis exógenas utilizadas. A análise pode focar em um único local ou em múltiplas usinas, embora a acurácia dos modelos tenda a diminuir à medida que a janela de predição se estende.

Com o intuito de garantir a robustez e modelagem física dos modelos, fatores como irradiância e temperatura são utilizados, geralmente obtidos de dados públicos, satélites ou previsões numéricas do tempo (NWP), sendo cruciais para o desempenho dos modelos (SOBRI; KOOHI-KAMALI; RAHIM, 2018).

Para além das variáveis escolhidas, as quais permitem a modelagem física do problema, tem-se de compreender a evolução dos métodos de análise. As abordagens metodológicas refletem uma clara evolução em termos de robustez, acurácia e interpretabilidade. Modelos estatísticos clássicos como a metodologia ARIMA (Média Móvel Integrada Autoregressiva) proposta por Box e Jenkins (BOX; JENKINS, 1970), são frequentemente o ponto de partida, mas demonstram limitações ao lidar com a não-linearidade inerente aos sistemas de geração de energia.

Essa limitação impulsionou a pesquisa em direção ao estudo das dependências temporais, utilizando modelos de redes neurais recorrentes. Arquiteturas como Long Short-Term Memory (LSTM) (HOCHREITER; SCHMIDHUBER, 1997) e Gated Recurrent Unit (GRU), uma variação mais simples computacionalmente proposta por Cho et al. (CHO *et al.*, 2014), demonstraram grande potencial para capturar dependências de longo prazo. Embora arquiteturas como as LSTMs tenham demonstrado grande potencial para capturar dependências temporais, elas apresentam seus próprios desafios, como a necessidade de grandes volumes de dados e a alta sensibilidade ao ajuste arquitetural, que pode levar ao sobreajuste (SOBRI; KOOHI-KAMALI; RAHIM, 2018).

Posteriormente, o aprendizado de máquina, especialmente com algoritmos de gradient boosting como o XGBoost (CHEN; GUESTRIN, 2016), superou parte dessas barreiras, capturando com maior precisão as relações complexas entre as variáveis.

Contudo, mesmo em sua alta performance, esses modelos de aprendizado de má-

quina frequentemente mantêm a premissa de tratar cada usina como um sistema isolado, ignorando a possibilidade de modelar usinas em conjunto, as quais podem compartilhar dinâmicas de geração, fenômenos meteorológicos e perfis operacionais, em função das suas proximidades geográficas e da natureza de funcionamento.

Nesse contexto, GNNs e GCNs surgem como o estado da arte para problemas espaço-temporais (WU *et al.*, 2021). Seu uso em séries temporais já foi consolidado em revisões recentes, as quais se tornaram padrão para modelar a influência mútua entre entidades conectadas em uma rede (YE *et al.*, 2020) (SAHILI; AWAD, 2023). No campo de aprendizado de máquina e energia, a literatura aponta as GNNs como uma das abordagens mais promissoras para lidar com a interdependência entre múltiplos pontos de geração (YANG *et al.*, 2025) (AHMED; MOZO; KARAMCHANDANI, 2023) .

Apesar de seu potencial, mesmo estas arquiteturas avançadas enfrentam desafios, como a disponibilidade limitada de dados de séries temporais para pré-treinamento em larga escala, o que pode dificultar a generalização (HU *et al.*, 2020).

A análise da literatura revela, portanto, que apesar dos avanços, a maioria dos trabalhos ainda trata a previsão de cada usina de forma independente. São raras as comparações diretas entre modelos locais de alta performance e abordagens espaço-temporais sob condições equivalentes .

Essa carência é acentuada no contexto SIN, um cenário complexo, de escala continental e diversidade climática, cuja exploração com tais técnicas ainda é incipiente (SILVA *et al.*, 2020).

Visando endereçar essa questão, este trabalho propõe uma comparação sistemática entre as duas frentes. De um lado, estabelece-se uma referência base com regressões de XGBoost aplicadas a três usinas-arquétipo, definidos através de métricas de redes complexas, representativas de diferentes regimes climáticos e dinâmicas entre usinas. De outro, empregam-se as arquiteturas GAT-LSTM e GCN-1D que consideram simultaneamente todo o conjunto de usinas como um sistema interdependente.

A partir dos resultados das previsões para os cenários de hora futura, para entender a sensibilidade desses estimadores estudam-se as incertezas dos regressores no conjunto de testes, de modo a testar a confiabilidade dos resultados obtidos.

O objetivo é, portanto, oferecer um comparativo sobre métodos de previsão da geração solar em solo brasileiro, buscando compreender as bases teóricas do problema e entender as incertezas associadas.

Em síntese, a literatura aponta para as GNNs como a fronteira da pesquisa para problemas espaço-temporais, mas destaca a carência de estudos comparativos, especialmente no setor elétrico brasileiro.

Mais recentemente, a intersecção entre GNNs e quantificação de incertezas surgiu como uma fronteira ativa de pesquisa. O referencial teórico dessa segunda, que busca ir além das previsões pontuais para fornecer estimativas com confiabilidade, está bem estabelecido

em obras como a de (IZBICKI, 2025). A aplicação desses princípios às GNNs investiga como decompor a incerteza preditiva em suas fontes fundamentais: a incerteza aleatória, inerente ao ruído nos dados (como variações na irradiância), e a incerteza epistêmica, que reflete as limitações e o conhecimento do próprio modelo (MUNIKOTI *et al.*, 2023).

As abordagens para capturar essas incertezas em GNNs são diversas, incluindo métodos Bayesianos, *deep ensembles* (LAKSHMINARAYANAN; PRITZEL; BLUNDELL, 2017) e redes determinísticas projetadas para serem cientes das diferenças aos dados de treino (LIU *et al.*, 2020). Contudo, um desafio central, destacado em benchmarks recentes, é a dificuldade prática de desentrelaçar as estimativas de incerteza aleatória e epistêmica, que frequentemente se mostram altamente correlacionadas na prática (MUCSÁNYI; KIRCHHOF; OH, 2024).

Dado que a aplicação rigorosa dessas técnicas avançadas dessa interseção para o setor elétrico ainda é um campo em aberto e de alta complexidade, este trabalho adota uma abordagem fundamental e pragmática. Utiliza-se a regressão quantílica — um dos métodos basilares da quantificação de incertezas, conforme discutido em (IZBICKI, 2025, Capítulo 5) — como um primeiro passo essencial para a sua quantificação completa. Essa escolha estabelece uma referência inicial sobre a qual futuras pesquisas de decomposição de incerteza podem ser construídas.

Com essa perspectiva teórica estabelecida, o próximo capítulo se dedicará à análise exploratória dos dados específicos deste trabalho, detalhando o processo de tratamento e a construção da estrutura em grafo que servirá de base para os modelos propostos.

3 ANÁLISE EXPLORATÓRIA DE DADOS

Este capítulo apresenta a Análise Exploratória de Dados (AED) utilizada para definir o escopo do trabalho e fundamentar a metodologia subsequente. O processo inicia-se com a descrição e o pré-processamento das fontes de dados (ONS e NASA), assim, obtém-se a definição do conjunto de 47 usinas e da janela temporal que compõem o cenário de análise.

A seguir, a estrutura espacial da rede é investigada através da modelagem em grafos, com o cálculo de métricas de redes complexas para a identificação de arquétipos de usinas.

Por fim, a análise de séries temporais com modelos SARIMA é empregada para diagnosticar a dinâmica individual das usinas e testar a hipótese de suficiência de modelos temporais, através da análise de seus resíduos.

3.1 Visão Geral

3.1.1 Dados do Operador Nacional do Sistema(ONS)

Para a construção e validação dos modelos preditivos, foram utilizados dois conjuntos de dados principais: a geração de energia fotovoltaica por usina, proveniente do ONS, e dados climáticos obtidos da API da NASA Power. A integração e o pré-processamento desses dados foram etapas fundamentais para a análise espaço-temporal e a modelagem subsequente.

O conjunto de dados da geração solar de usinas brasileiras disponibilizado com atualização mensal pelo ONS, fornece o histórico horário de geração elétrica por usina. Neste trabalho, foi utilizada a série temporal referente ao ano de 2024, com as seguintes colunas principais:

- **din_instante**: timestamp (passo horário);
- **id_estado**: unidade federativa da usina;
- **nom_usina**: nome da usina;
- **val_geracao**: valor da geração (MW).

Durante a etapa de ETC (Extração, Transformação e Carga), foram selecionadas apenas usinas do tipo fotovoltaica. O *dataset* original do ONS inclui registros tanto de usinas centralizadas quanto de microgeração distribuída.

No entanto, a microgeração foi excluída da análise, uma vez que seus registros não permitem a identificação precisa da localização geográfica (apenas o estado de origem é informado), o que inviabilizaria o cruzamento com os dados climáticos da NASA e a

construção da rede de grafos baseada na localização espacial e para a análise das relações funcionais entre usinas.

Este processo resultou em um conjunto inicial de 64 usinas fotovoltaicas centralizadas. A inclusão de dados de microgeração distribuída com georreferenciamento preciso representaria um avanço significativo, permitindo uma análise mais granular da penetração de renováveis e dos desafios associados; contudo, a indisponibilidade atual desses dados em nível nacional com a resolução necessária para este estudo representou uma limitação.

A distribuição geográfica dessas 64 usinas pelo território nacional é apresentada na Figura 1. Observa-se uma concentração significativa na região Nordeste e em partes das regiões Sudeste e Centro-Oeste, áreas frequentemente referidas como o cinturão solar, devido à alta incidência de irradiação solar.

Essa distribuição heterogênea espacial, mas de irradiância forte na região das usinas no território nacional ressalta a complexidade de modelar a geração em um país de dimensões continentais e a variabilidade das condições ambientais regionais, reforçando a necessidade de abordagens que considerem a variabilidade espacial.

A localização exata das usinas selecionadas (latitude e longitude) foi obtida manualmente por meio de pesquisas online e cruzamento com bases de dados públicas, utilizando coordenadas representativas próximas às instalações reais.

3.1.2 Dados Climáticos — NASA Power API

Os dados climáticos foram obtidos via requisições HTTP à API da NASA Power ¹, que disponibiliza séries temporais horárias para aplicações em energia renovável. A grade espacial tem resolução de $0,5^\circ \times 0,5^\circ$ e cobre todo o território brasileiro.

Para cada usina, foram extraídas as variáveis climáticas correspondentes às suas coordenadas geográficas mais próximas. A seleção das variáveis considerou sua relevância para a geração fotovoltaica, com foco na caracterização da irradiância solar e das condições atmosféricas que afetam diretamente a produção de energia.

As principais variáveis selecionadas foram:

- **ALLSKY_SFC_SW_DWN**: irradiância solar global na superfície (W/m^2), a principal variável física indicadora da geração fotovoltaica.
- **T2M**: temperatura do ar a 2 metros ($^\circ\text{C}$), a eficiência dos módulos fotovoltaicos é inversamente afetada pela temperatura.
- **RH2M**: umidade relativa a 2 metros (%), incluída pela sua capacidade de influenciar a formação e dispersão de nuvens, além da atenuação da irradiância por partículas suspensas na atmosfera.

¹ <https://power.larc.nasa.gov/data-access-viewer/>

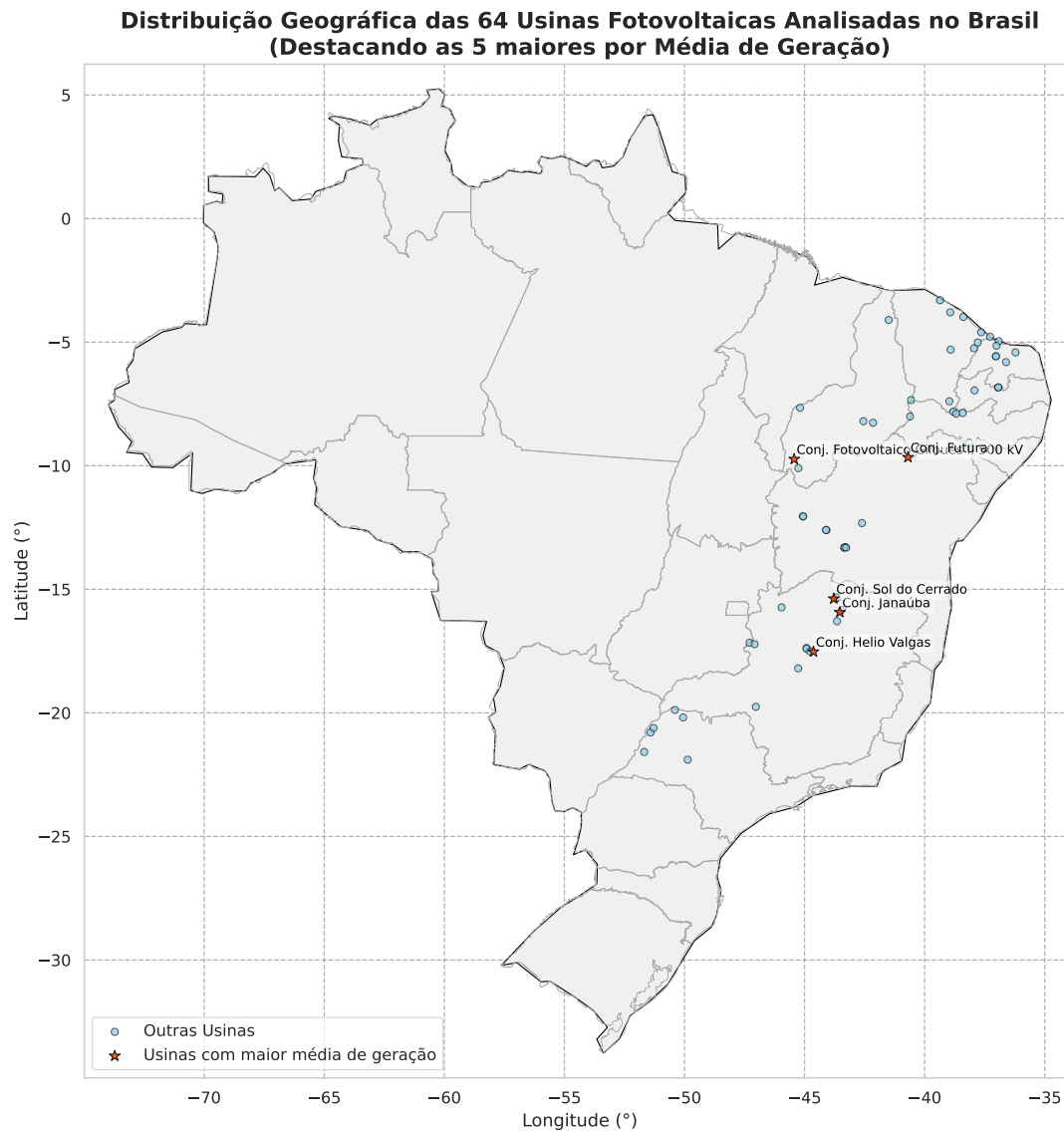


Figura 1 – Distribuição Geográfica das 64 Usinas Fotovoltaicas Analisadas no Brasil, com destaque para a concentração no "cinturão solar".

- **WS10M**: velocidade do vento a 10 metros (m/s), considerada como um indicador da dinâmica atmosférica local e seu impacto na dissipação de calor dos módulos.
- **ALLSKY_KT**: índice de claridade (adimensional), representa a razão entre a irradiância real e a irradiância potencial na atmosfera clara, servindo como uma medida da claridade atmosférica e da presença de nuvens.

A concatenação dos *timestamps* de ambas as fontes de dados (ONS e NASA) foi um passo fundamental do pré-processamento, garantindo a correspondência temporal precisa entre os dados de geração e as condições climáticas.

Após a coleta e a sincronização, todos os dados brutos foram limpos e integrados. O tratamento de valores ausentes foi realizado em duas etapas. Primeiro, para garantir que todas as usinas tivessem a mesma quantidade de passos de tempo, os passos horários

faltantes na geração de energia foram preenchidos com o valor NaN, conforme detalhado na Seção 3.2. Segundo, para um pequeno subconjunto de usinas com falhas pontuais, uma estratégia de imputação foi considerada para permitir sua inclusão na análise final. Este processo resultou em um formato unificado e balanceado, pronto para as etapas de engenharia de features e modelagem.

Conforme detalhado na seção 3.1.1, os dados de geração horária do ONS e os dados climáticos horários da API NASA Power foram coletados e preparados.

Na etapa de consolidação, as informações de geração de cada usina fotovoltaica foram associadas com as variáveis climáticas correspondentes à sua localização geográfica (utilizando o ponto de grade da NASA mais próximo) e ao respectivo instante de tempo.

Além disso, verificou-se a ausência de dados na geração solar, as usinas identificadas e a quantidade das entradas ausentes estão disponíveis na próxima subseção. Para garantir um conjunto de dados balanceado, os passos horários faltantes para cada usina foram preenchidos com o valor NaN na coluna de geração. Este procedimento assegura que todas as séries temporais tenham o mesmo comprimento e os mesmos passos de tempo, alinhadas às observações climáticas por localização geográfica.

Assim, garante-se que cada usina possui a mesma quantidade de passos horários, sendo esta também igual à quantidade de observações climáticas por ponto de grade.

O dataframe resultante, contém 562.176 observações (registros horários) e 17 colunas (variáveis). Abrange 64 usinas únicas, cada uma com sua série temporal para o ano de 2024 completo. As principais variáveis presentes no conjunto de dados são:

- **Identificação e Tempo:**

- `din_instante`: Timestamp da medição (passo horário).
- `id_subsistema`: Identificador do subsistema elétrico.
- `id_estado`: Sigla da unidade federativa da usina.
- `nom_tipocombustivel`: Tipo de combustível (sempre "Fotovoltaica" neste estudo).
- `nom_usina`: Nome da usina de geração.

- **Geração e Localização da Usina:**

- `val_geracao`: Valor da geração elétrica (MW).
- `latitude`: Latitude da usina.
- `longitude`: Longitude da usina.

- **Dados de Referência da NASA e Variáveis Climáticas:**

- `closest_latitude`: Latitude do ponto de grade da NASA mais próximo.
- `closest_longitude`: Longitude do ponto de grade da NASA mais próximo.

- `distance_km`: Distância (km) entre a usina e o ponto de grade da NASA.
- `location`: String combinada das coordenadas do ponto de grade da NASA.
- `ALLSKY_SFC_SW_DWN`: Irradiância solar global na superfície (W/m^2).
- `T2M`: Temperatura do ar a 2 metros ($^{\circ}\text{C}$).
- `RH2M`: Umidade relativa a 2 metros (%).
- `WS10M`: Velocidade do vento a 10 metros (m/s).
- `ALLSKY_KT`: Índice de claridade (adimensional).

Este conjunto de dados consolidado será utilizado nas subseções seguintes para uma análise exploratória detalhada, visando entender os padrões de geração, a influência dos fatores climáticos e outras características relevantes.

3.2 Cenário de Análise

A etapa de seleção de variáveis e tratamento de dados ausentes teve início com uma avaliação do conjunto de dados. Esta análise identificou um total de 58.104 valores omissos. Uma investigação subsequente revelou que a integralidade destes dados faltantes concentra-se exclusivamente na variável `val_geracao`.

Considerando que a ausência de dados de geração não se distribui uniformemente, torna-se relevante analisar a contribuição de cada usina para este total. A Tabela 1 detalha as usinas fotovoltaicas com registros de geração ausentes e a respectiva quantidade de ocorrências.

Como detalhado na Tabela 1, constatou-se que 22 das 64 usinas fotovoltaicas do conjunto de dados apresentam registros ausentes na variável de geração (`val_geracao`). O perfil dessa ausência varia consideravelmente, desde o Conjunto Banabuiu, com apenas 72 valores faltantes, até o Conjunto Fotovoltaico Simplicio, com 7.824 registros ausentes.

Para compreender os períodos específicos de interrupção no registro desta variável durante o ano de 2024, procedeu-se a uma investigação individualizada para cada usina afetada. Esta análise revelou que os valores faltantes ocorrem predominantemente em blocos contínuos, estendendo-se por múltiplos dias ou meses, comprometendo o registro da geração solar, conforme apresentado na Tabela 2.

A Tabela 2 detalha os períodos exatos de ausência de dados para cada uma dessas usinas. Esta informação é fundamental, pois levanta questões metodológicas importantes: quais usinas devem ser mantidas na análise e qual janela temporal será considerada.

A análise dos períodos de ausência de dados de geração (Tabela 2) revela heterogeneidade significativa. Algumas apresentam longos períodos contínuos de falha. No entanto, a definição da janela de eventos também considera a disponibilidade de variáveis exógenas.

Dados meteorológicos da NASA estão consistentemente disponíveis e validados apenas até 31 de julho de 2024 para o conjunto de usinas. Após este período, observam-se

Tabela 1 – Distribuição de dados ausentes da variável `val_geracao` por usina fotovoltaica.

Nome da Usina	Total de Dados Ausentes
Fotov. Simplicee	7.824
Jaíba 138 kV	7.488
Fotovoltaico Gilbués II 500 kV	6.288
Três Marias 3 138kV	5.448
Vista Alegre - Janaúba	5.424
Fotovoltaico Barreiras II 500 kV	4.032
Varzea da Palma	3.240
São Gonçalo	2.808
Bom Nome	2.712
Marangatu	2.160
Ribeiro Gonçalves	2.160
Boa Sorte	1.680
Conjunto fotovoltaico Santa Luzia II 500kV	1.392
Jaiba V	1.272
Monte Verde Solar	744
Fotov. Acu III 230KV	720
Fotovoltaico Abaiara 230 kV	720
Arinos 2 500 kV	720
Boa Hora	624
Pitombeira	360
Sol do Piauí	216
Banabuiu	72

valores degenerados (e.g., -999) em variáveis chave como `ALLSKY_SFC_SW_DWN` e `ALLSKY_KT`. Este fenômeno é ilustrado para as usinas Janaúba (Figura 2) e Belmonte (Figura 3). Após este período, observam-se valores degenerados nas variáveis climáticas (preenchidos com a flag de erro -999 pela API), o que inviabiliza o uso dessas séries para treinamento supervisionado a partir de agosto.

A análise dos padrões de falha sugere que a ausência de dados ocorre devido a problemas técnicos de comunicação ou manutenção dos sensores. Conforme as definições estabelecidas por (LITTLE; RUBIN, 2019), esse comportamento caracteriza-se predominantemente como Missing Completely at Random (MCAR) ou, no mínimo, independente dos valores de geração não observados, uma vez que não foram identificados vieses sistemáticos correlacionados à magnitude da geração.

Conforme a documentação da API, o valor -999 indica ausência de medição ou dados faltantes. Ao verificar os grids com dados climáticos, nota-se que a indisponibilidade após 31 de julho é homogênea.

Considerando a interseção entre a disponibilidade dos dados de geração das usinas com menor volume de falhas e o limite de confiabilidade dos dados meteorológicos da NASA, optou-se por definir a janela de eventos para este estudo como o período de 01 de Janeiro de 2024 a 30 de Junho de 2024. Esta janela busca maximizar a quantidade de

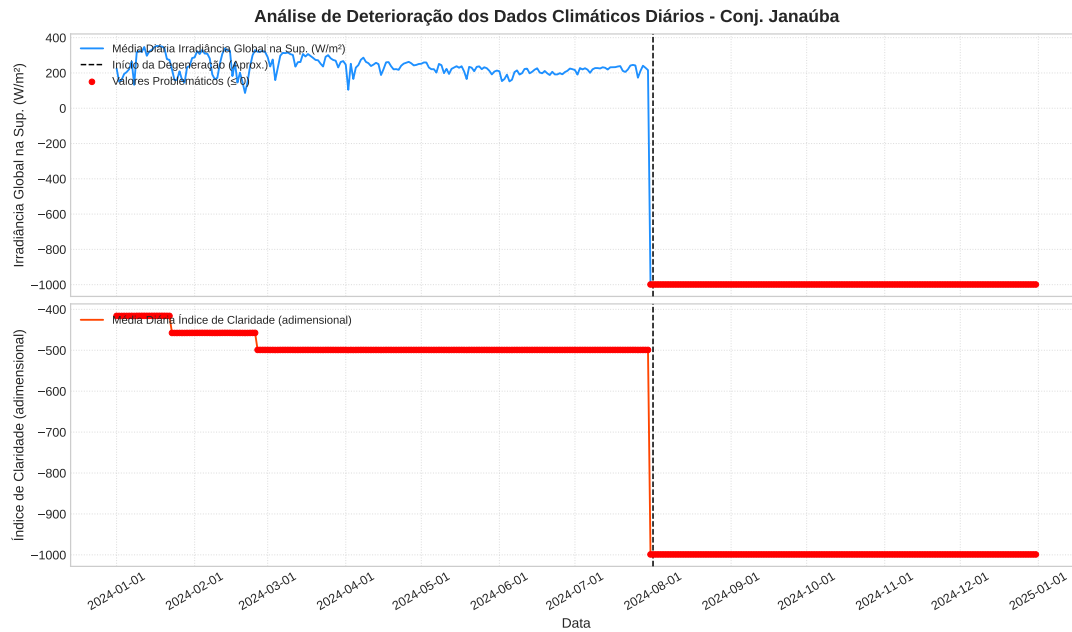


Figura 2 – Deterioração dos dados horários de irradiância (`ALLSKY_SFC_SW_DWN`) e índice de claridade (`ALLSKY_KT`) para a usina Janaúba.

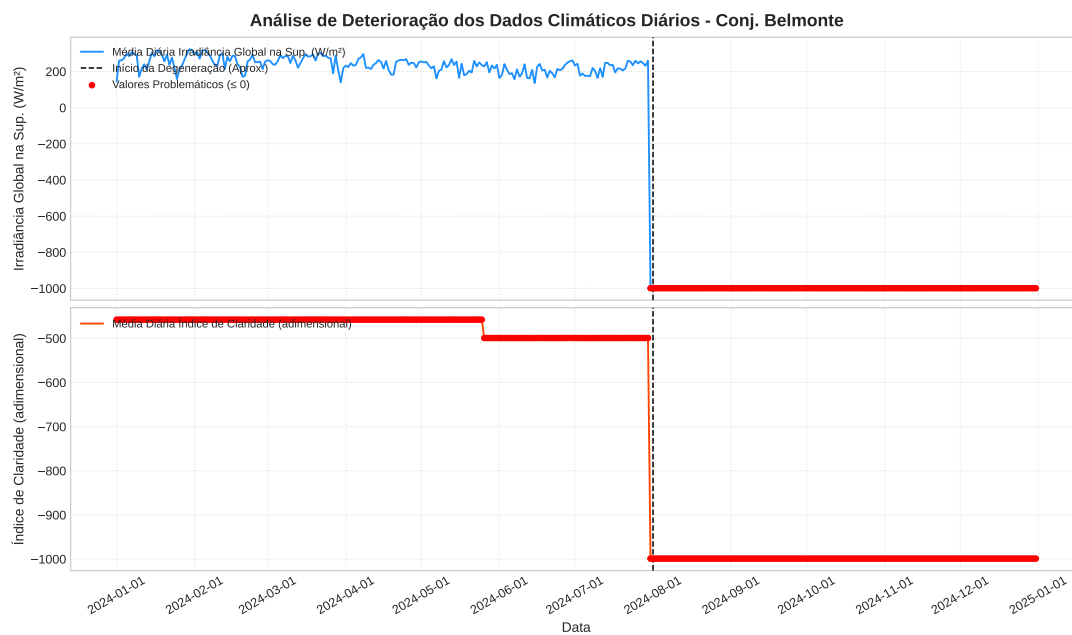


Figura 3 – Comportamento similar de deterioração dos dados meteorológicos (`ALLSKY_SFC_SW_DWN` e `ALLSKY_KT`) observado para a usina Belmonte.

Tabela 2 – Períodos com dados ausentes da variável `val_geracao` por usina fotovoltaica.

Nome da Usina	Períodos Ausentes (YYYY-MM-DD)
Fotov. Simplicite	2024-01-01 a 2024-11-21
Jaíba 138 kV	2024-01-01 a 2024-11-07
Fotovoltaico Gilbués II 500 kV	2024-01-01 a 2024-09-18
Três Marias 3 138kV	2024-01-01 a 2024-08-14
Vista Alegre - Janaúba	2024-01-01 a 2024-08-13
Fotovoltaico Barreiras II 500 kV	2024-01-01 a 2024-06-16
Varzea da Palma	2024-01-01 a 2024-05-14
São Gonçalo	2024-09-06 a 2024-12-31
Bom Nome	2024-01-01 a 2024-04-16; 2024-05-07 a 2024-05-12
Marangatu	2024-01-01 a 2024-02-07; 2024-02-09 a 2024-03-31
Ribeiro Gonçalves	2024-01-01 a 2024-02-07; 2024-02-09 a 2024-03-31
Boa Sorte	2024-01-01 a 2024-03-07; 2024-04-30 a 2024-05-02
Conjunto fotovoltaico Santa Luzia II 500kV	2024-01-01 a 2024-02-07; 2024-02-09 a 2024-02-28
Jaiba V	2024-01-01 a 2024-02-07; 2024-02-09 a 2024-02-23
Monte Verde Solar	2024-01-01 a 2024-01-31
Fotov. Acu III 230KV	2024-01-01 a 2024-01-30
Fotovoltaico Abaiara 230 kV	2024-01-01 a 2024-01-30
Arinos 2 500 kV	2024-01-01 a 2024-01-30
Boa Hora	2024-07-04 a 2024-07-29
Pitombeira	2024-01-01 a 2024-01-15
Sol do Piauí	2024-01-01 a 2024-01-09
Banabuiu	2024-01-01 a 2024-01-03

dados válidos e simultâneos, garantindo uma base íntegra para a modelagem. A seleção final das usinas será detalhada na próxima subseção.

Definida a janela de eventos (01 de Janeiro a 30 de Junho de 2024), o próximo passo é a seleção das usinas fotovoltaicas para o conjunto de dados final, visando uma amostra robusta com dados de geração completos e consistentes.

O processo ocorreu em duas etapas. Primeiramente, identificaram-se 44 usinas com cobertura temporal completa e sem dados faltantes em `val_geracao` na janela de análise.

Em uma segunda etapa, reavaliaram-se as usinas descartadas, permitindo a reinclusão daquelas com até 10% de dados de geração ausentes dentro da janela de seis meses. Adotou-se como critério de inclusão a presença de, no máximo, 10% de dados ausentes na janela de análise. Usinas que ultrapassaram esse limiar de falhas foram descartadas definitivamente para evitar a introdução de ruído excessivo via métodos de imputação. Este critério equilibra representatividade e qualidade, permitindo o tratamento de pequenas

ausências por imputação sem vieses significativos.

Esta abordagem resultou na inclusão de 3 usinas adicionais, consolidando o escopo final em 47 usinas fotovoltaicas, visando refletir a realidade operacional com interrupções pontuais, mitigando o impacto de grandes volumes de dados ausentes(Tabela 3).

Tabela 3 – Lista final das 47 usinas fotovoltaicas selecionadas para análise.

Nome da Usina	Nome da Usina
Assú V	Lar do Sol
Alex	Lavras
Araxá	Luzia
BJL	Nova Olinda
Banabuiú	Paracatu
Belmonte	Pedranópolis
Boa Hora	Pereira Barreto
Bom Jesus	Pirapora 2
Calcário	Rio Alto
Castilho	Serra do Mato
Ciranda	Serra do Mel C
Dracena	Sertão Solar Barreiras
FV SJP	Sol do Cerrado
Floresta	Sol do Futuro
Francisco Sá	Sol do Sertão
Futura	São Basílio (Solar)
Guaimbê	São Gonçalo
Helio Valgas	São Pedro
Horizonte	Tacaratu (Solar)
Ituverava	Conjunto Jaíba 4 Dist
Janaúba	Flor de Mandacaru
Juazeiro Solar	Pitombeira
Juazeiro Solar 2	Sol do Piauí
Lapa	

3.3 Análise da estrutura em grafo

A Figura 4 ilustra a topologia espacial da rede. Considerando as 47 usinas selecionadas, uma análise estatística descritiva individual para modelagem de regressão seria extensiva e dificultaria a análise da influência espaço-temporal.

Por isso, A conectividade foi estabelecida utilizando o algoritmo dos k-vizinhos mais próximos, uma técnica em aprendizado de máquina para identificar vizinhanças em espaços métricos (BISHOP, 2006). O valor de k foi definido como k=4 (aproximadamente 10% do total de usinas), uma escolha que busca equilibrar a captura de relações relevantes e a esparsidade da rede. A metodologia completa para a construção do grafo é detalhada na

Seção 4.3.2. Visualmente, a rede resultante é globalmente conectada — uma propriedade importante para a propagação de informação nos modelos GNN.

Para analisar a influência estrutural das usinas fotovoltaicas, este trabalho emprega conceitos de Redes Complexas (COMIN *et al.*, 2020), uma área de pesquisa do estudo de grafos e suas estatísticas. Para este estudo, selecionaram-se quatro métricas visando a melhor compreensão do aspecto espaço-temporal do problema:

- **Centralidade de Grau:** Número de conexões diretas de um nó v , $\deg(v)$. Normalizada:

$$C'_D(v) = \frac{\deg(v)}{|V| - 1} \quad (3.1)$$

onde $|V|$ é o número total de nós. Indica o quão conectado é um nó em relação a sua rede, sendo igual a 0 se for isolado e igual a 1 se for conectado a todos os outros nós da rede, ou seja, mede a sua atividade local de um nó. Assim, a Centralidade de Grau é interpretada como uma medida da influência climática local de uma usina. Um grau elevado sugere que a usina está geograficamente cercada por muitas outras, aumentando a probabilidade de compartilharem regimes de irradiância e nebulosidade e, portanto, tornando-a um ponto focal para a análise de correlações espaciais.

- **Centralidade de Intermediação:** Frequência com que um nó v está nos caminhos mais curtos entre outros pares (s, t) :

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (3.2)$$

onde σ_{st} é o total de caminhos mais curtos entre s e t , e $\sigma_{st}(v)$ os que passam por v . No contexto de um grafo não-ponderado como o utilizado neste trabalho, um "caminho" é uma sequência de nós conectados, e o "caminho mais curto" é aquele com o menor número de arestas (ou "saltos") entre um nó de partida s e um nó de chegada t .

Mede a frequência com que um nó atua como intermediário nos caminhos mais curtos entre outros pares de nós. Nós com alta centralidade de intermediação são, portanto, cruciais para a coesão e o fluxo de informação ou recursos entre diferentes partes da rede, atuando como pontes.

- **Coefficiente de Aglomeração Local:** Conectividade dos vizinhos de um nó v com k_v vizinhos:

$$C_L(v) = \frac{2E_v}{k_v(k_v - 1)} \quad (3.3)$$

onde E_v é o número de arestas entre os k_v vizinhos de v . Assim, vizinhos não conectados entre si tem o coeficiente de aglomeração local nulo, e igual a 1 se todos os vizinhos de v estão conectados entre si.

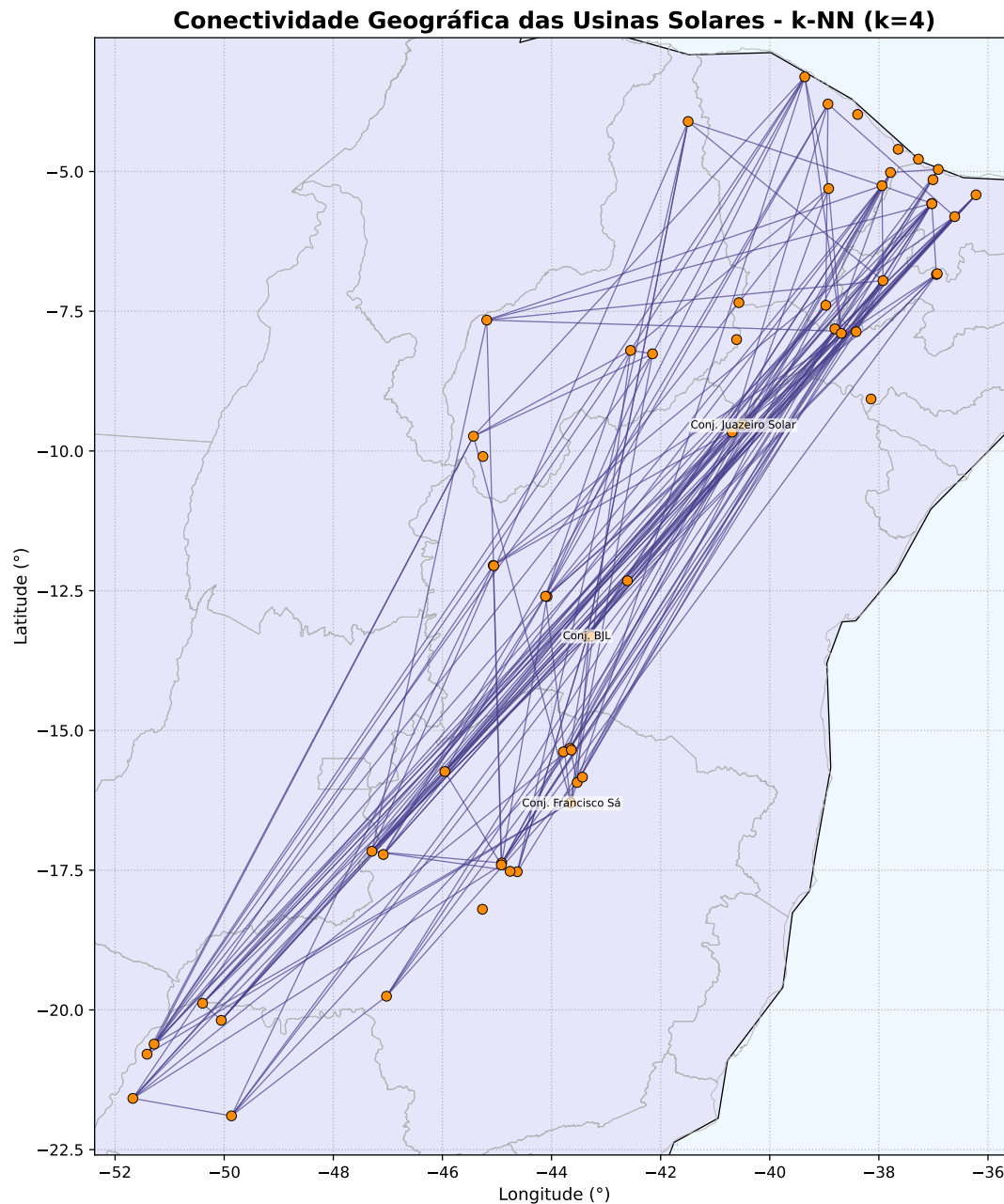


Figura 4 – Visualização geográfica da rede de 47 usinas fotovoltaicas e suas interconexões baseadas no critério de k -vizinhos mais próximos ($k \approx 4$). Embora a escala de visualização continental possa sugerir a aparência de nós isolados em regiões de alta densidade (onde múltiplas usinas compartilham coordenadas muito próximas), a definição algorítmica do grafo via k -vizinhos assegura a conectividade global, mesmo naqueles nós cujas arestas são visualmente imperceptíveis devido à proximidade geográfica.

Trata-se de uma métrica da densidade da vizinhança de um nó, i.e., Um alto coeficiente pode indicar a presença de uma comunidade de nós, cujos membros, por estarem densamente conectados, podem apresentar dinâmicas similares entre si.

- **PageRank:** Influência de um nó v_i considerando a importância de seus vizinhos $\mathcal{N}(v_i)$:

$$PR(v_i)^{(t+1)} = \frac{1-d}{|V|} + d \sum_{v_j \in \mathcal{N}(v_i)} \frac{PR(v_j)^{(t)}}{\deg(v_j)} \quad (3.4)$$

com d sendo o fator de amortecimento (usualmente 0.85).

É uma medida de influência que pode ser analogamente entendida como um voto de popularidade, onde a importância de um nó é recursivamente definida pela importância dos nós que se conectam a ele, ou seja, estar conectado a nós já influentes confere maior PageRank ao nó de interesse. Dessa forma, uma usina com alto PageRank é considerada estruturalmente influente na rede, podendo ser um nó central na disseminação de dinâmicas ou na robustez do sistema.

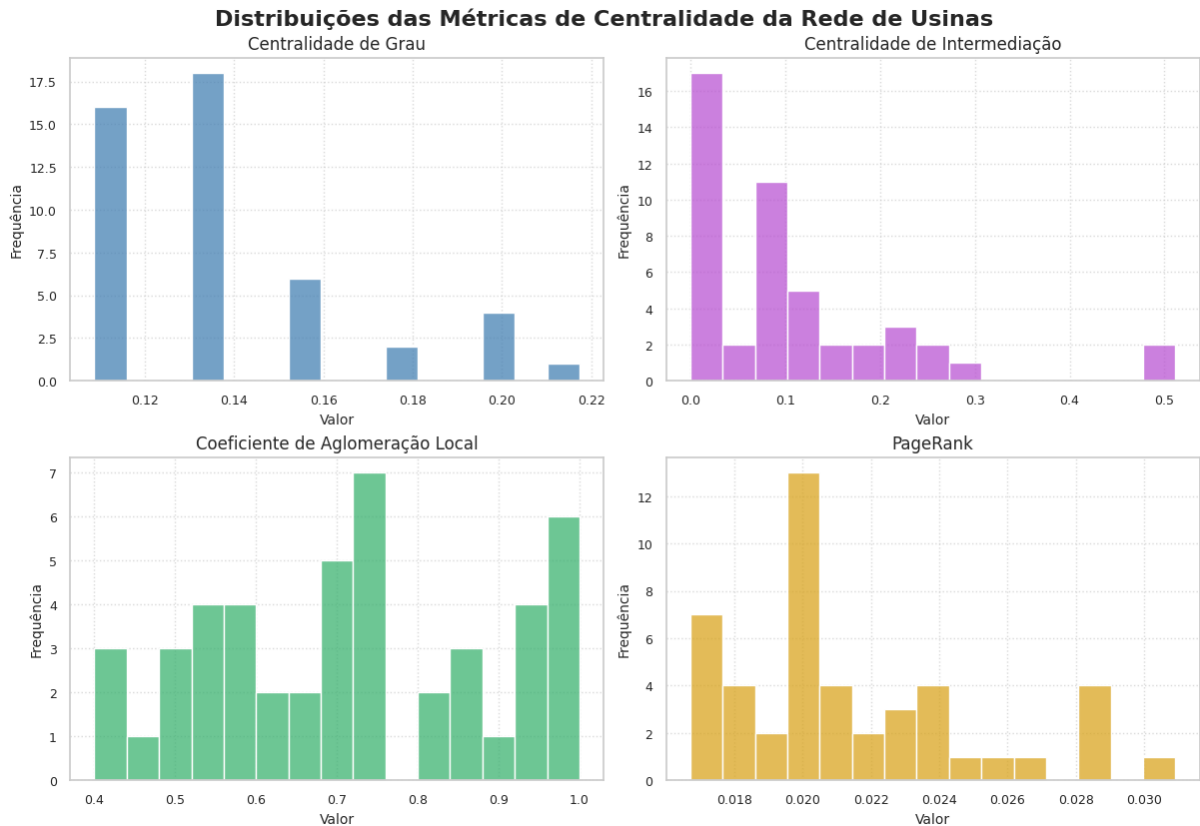


Figura 5 – Distribuições das métricas de centralidade para a rede de 47 usinas solares: (a) Centralidade de Grau, (b) Centralidade de Intermediação, (c) Coeficiente de Aglomeração Local e (d) PageRank.

Para compreender o perfil dessas métricas no conjunto de dados tem-se os histogramas das suas distribuições Figura 5 acima.

Percebe-se que a Centralidade de Grau se concentra predominantemente entre 0,12 e 0,22. A distribuição exibe uma assimetria positiva e um caráter bimodal, com modas em aproximadamente 0,13 e 0,19; sendo o primeiro modo mais proeminente e o segundo mais suave.

A assimetria positiva, juntamente com a bimodalidade, sugere que, embora a maioria das usinas possua um número similar e mais baixo de conexões diretas (primeiro modo), existe um grupo menor de usinas (segundo modo) que se destaca por possuir um número consideravelmente maior de conexões, caracterizando-as como potenciais hubs locais com grande interação direta.

Essa evidência da existência de diferentes papéis topológicos na rede é fundamental, pois sugere que a influência de uma usina no sistema não é homogênea. Para investigar o impacto dessas diferentes posições estruturais no desempenho dos modelos de previsão, definimos formalmente, na Seção 3.5, um conjunto de "usinas-arquétipo" (como Hub, Nó Isolado, etc.) para uma análise mais aprofundada.

Já a Centralidade de Intermediação também demonstra uma forte assimetria positiva, com a maioria dos nós apresentando valores baixos, concentrados em um modo em torno de 0,05. Isso indica que a maior parte das usinas não atua predominantemente como intermediária nos fluxos da rede.

Contudo, a presença de dois nós com valores excepcionalmente altos (próximos a 0,5) aponta para a existência de algumas poucas usinas que são cruciais como pontes, concentrando a função de intermediação e sendo vitais para a conectividade global e o fluxo eficiente entre diferentes partes do grafo.

O Coeficiente de Aglomeração Local apresenta uma distribuição com curtose média e um aspecto bimodal suave, com valores predominantemente concentrados em torno de 0,7.

Este valor, relativamente elevado, sugere que a rede possui uma tendência significativa à formação de comunidades locais, onde as usinas vizinhas a um nó também tendem a ser vizinhas entre si, indicando vizinhanças densas na estrutura.

A bimodalidade suave observada pode indicar a coexistência de regiões com diferentes níveis de coesão local na rede.

Por fim, o PageRank também apresenta uma distribuição bimodal, com um modo principal de maior amplitude em torno de 0,020 e um segundo modo menor, porém distinto, em aproximadamente 0,028. Esta bimodalidade indica uma diferenciação na influência estrutural das usinas.

A maioria delas possui um nível de influência similar (primeiro modo), enquanto um grupo restrito de cerca de seis usinas (segundo modo) se destaca com um PageRank significativamente maior, posicionando-as como candidatas a nós centrais ou hubs de

influência na rede, importantes para a disseminação de dinâmicas.

A alta curtose associada ao primeiro modo reforça a concentração da maioria das usinas em torno desse nível de importância basal.

Considerando que realizar uma análise minuciosa de cada uma das usinas é um trabalho intensivo e extensivo, busca-se definir modelos-padrão de análise das usinas, auxiliando na seleção de exemplos para análise temporal e na compreensão da relação entre posição estrutural e geração.

3.4 Análise de séries temporais e estimação dos resíduos

Uma série temporal é qualquer conjunto de observações indexadas temporalmente. Essas observações podem ocorrer de forma periódica ou irregular, podendo apresentar dependência entre valores passados, presentes e futuros.

Neste trabalho analisam-se múltiplas séries temporais multivariadas, onde cada série temporal tem várias variáveis exógenas que afetam a geração de energia solar, como a temperatura, velocidade do vento e umidade relativa.

As componentes mais comuns na modelagem de séries temporais são: tendência, sazonalidade, ruído e estacionariedade.

Existe uma tendência quando há uma alteração a longo prazo, seja de forma crescente ou decrescente da variável observada. E pode ser modelada como uma função $f[t]$, onde t é o tempo:

$$y_t = f[t] + \epsilon_t, \quad (3.5)$$

onde y_t é a variável observada no tempo t e ϵ_t é o erro aleatório ou ruído.

A sazonalidade é observada quando a série temporal é afetada por fatores que ocorrem em intervalos de tempo característicos. Para uma série y_t com sazonalidade s , isso pode ser modelado como:

$$y_t = \mu_t + s[t] + \epsilon_t, \quad (3.6)$$

onde $s(t)$ representa o efeito sazonal da série no tempo t e μ_t é a tendência ou média de longo prazo. A frequência f da sazonalidade é a quantidade de ciclos ou padrões que se repetem ao longo do tempo.

O ruído é a parte imprevisível da série, associada a flutuações aleatórias que não seguem padrões definidos. Pode-se modelar o ruído como um processo estocástico, que é descrito por uma sequência de erros aleatórios ϵ_t , tal que:

$$\epsilon_t \sim N(0, \sigma^2), \quad (3.7)$$

onde ϵ_t é o erro no tempo t e σ^2 é a variância do ruído.

A estacionariedade da série é uma propriedade que não depende do instante de tempo no qual foi observada. Para que uma série temporal seja estacionária no sentido amplo(ou fracamente estacionária), a média, a variância e a covariância da série devem ser

constantes ao longo do tempo. Logo, a série segue que y_t :

$$E[y_t] = \mu, \quad \text{Var}(y_t) = \sigma^2, \quad \text{Cov}(y_t, y_{t+h}) = \gamma_h, \quad (3.8)$$

onde $E[y_t]$ é a esperança (média) de y_t , $\text{Var}(y_t)$ é a variância, e $\text{Cov}(y_t, y_{t+h})$ é a covariância entre y_t e y_{t+h} , para qualquer h .

O SARIMA (Seasonal AutoRegressive Integrated Moving Average) é utilizado para a previsão de séries temporais univariadas, levando em consideração as técnicas que compõem o seu nome:

- S: Sazonalidade(S);
- AR: Autoregressão(AR);
- I: Integração(I);
- MA: Média Móvel(MA);

A sazonalidade nesse modelo é o número de observações que compõem um ciclo sazonal, podendo ser qualquer intervalo regular. Já a autoregressão modela o relacionamento entre o seu valor presente e os valores passados a partir da autocorrelação do sinal. Um processo $\text{AR}(p)$ pode ser expresso como: $y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t$, onde ϕ_i são os coeficientes autorregressivos e ϵ_t é o termo de erro.

A diferenciação busca capturar a quantidade de diferenciações necessárias para tornar os dados estacionários, fazendo isso através da aplicação de diferenças sucessivas nos dados até que a série se torne estacionária, eliminando tendências e sazonalidades. Além disso, a média móvel utiliza de resíduos de previsões anteriores para a previsão de valores futuros.

Um processo $\text{MA}(q)$ é modelado como: $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$, onde θ_i são os coeficientes da média móvel e ϵ_t são os erros passados.

$\text{SARIMA}(p, d, q)(P, D, Q)_s$ onde:

- p : ordem da parte autorregressiva (AR);
- d : número de diferenciações para tornar a série estacionária;
- q : ordem da média móvel (MA);
- P, D, Q : ordens sazonais dos componentes AR, I e MA, respectivamente;

Para a escolha do melhor modelo ARIMA, utilizam-se os critérios da metodologia Box-Jenkins (BOX; JENKINS, 1970), que propõe uma abordagem iterativa baseada na identificação, estimação e verificação do modelo.

Inicialmente, aplicam-se testes de estacionariedade para definir o número de diferenciações necessárias. Esse número é representado por d (não-sazonal) e D (sazonal),

conforme a estrutura da série. A verificação pode ser feita visualmente (presença de tendência ou sazonalidade) ou por testes estatísticos, como o teste de Dickey-Fuller aumentado (ADF). O ADF é baseado no seguinte modelo:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t \quad (3.9)$$

Para rejeitar a hipótese nula ($\gamma = 0$), que indica que a série é não-estacionária, precisa-se de valores pequenos do p-valor (normalmente $< 0,05$).

Funções de autocorrelação (ACF) também auxiliam na identificação da ordem de diferenciação. A ACF é definida por:

$$\rho_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (3.10)$$

Os correlogramas da ACF indicam os parâmetros p e q do modelo: um corte abrupto na PACF sugere um modelo $AR(p)$, enquanto um corte abrupto na ACF sugere um modelo $MA(q)$.

Como a identificação visual nem sempre é conclusiva, testa-se uma grade de combinações possíveis. A seleção se baseia em critérios de informação como o Akaike Information Criteria (AIC) e o Bayesian Information Criteria (BIC), que penalizam a complexidade do modelo:

$$AIC = -2 \ln(\hat{L}) + 2k \quad (3.11)$$

$$BIC = -2 \ln(\hat{L}) + k \ln(n) \quad (3.12)$$

em que \hat{L} é a verossimilhança, k o número de parâmetros, e n o número de observações.

O AIC mede a qualidade preditiva esperada de um modelo, penalizando modelos mais complexos a fim de reduzir o *overfitting*, porém esse ajuste de modelos é mais leve do que o BIC, o que pode fazer com que o AIC aponte para modelos mais complexos. Já o BIC mede comparativamente os modelos com base na sua verossimilhança, penalizada pela complexidade, o que tende a escolher modelos mais conservadores, podendo reduzir o *overfitting*. Apesar de rigoroso, o BIC assume que um dos modelos é o verdadeiro modelo e não mede a performance do modelo fora da amostra de treino.

Após a escolha do modelo, avalia-se seu desempenho pelos resíduos. Espera-se que estes se comportem como ruído branco, ou seja, sem autocorrelação ou estrutura sistemática. Essa verificação é feita por meio do teste de Ljung-Box e pela inspeção visual dos correlogramas dos resíduos.

Apesar de sua robustez à baixa quantidade de variáveis exógenas, o SARIMA apresenta limitações: modela apenas uma sazonalidade por vez, tem baixa adaptabilidade a eventos abruptos e costuma perder precisão em previsões de longo prazo.

3.5 Análise funcional

Esta seção aprofunda a análise dos perfis temporais de geração das usinas representativas dos arquétipos – Lapa, Floresta e Castilho – e como suas séries interagem com vizinhas e variáveis climáticas, buscando ligar topologia da rede e padrões funcionais.

Levando em consideração as propriedades das métricas de grafo e sua interpretação, definem-se três arquétipos de usinas para analisar a hipótese da influência estrutural do grafo:

- **Hub (Lapa):** Caracteriza-se por estar entre o 10% de usinas com maior centralidade de grau e entre o 10% com maior PageRank.
- **Nó Isolado (Floresta):** Caracteriza-se por estar entre os 10% de usinas com menor grau de centralidade e entre os 10% de menor PageRank.
- **Membro de Comunidade (Castilho):** Caracteriza-se por estar entre os 10% de usinas com maior coeficiente de aglomeração local, mas não estar entre os 10% de maior centralidade de grau.

A Figura 6 ilustra a atuação desses três nós na rede com base nos PageRanks, que variam entre 0.018 e 0.030. Observa-se que o Nó Isolado (Floresta) possui um PageRank consistentemente baixo, mesmo estando conectado à rede.

O Membro de Comunidade (Castilho) apresenta um PageRank intermediário e se integra a outros nós, refletindo sua participação em um sub-grafo interno com maior densidade. Por fim, o Hub (Lapa) exibe um PageRank elevado, inserido em sub-grafos com alta densidade e com conexões a outras usinas de alto PageRank, evidenciando sua posição central e influente na rede.

Para aprofundar a análise, selecionam-se Lapa, Floresta e Castilho como representantes de cada arquétipo. A Figura 7 apresenta as distribuições da geração horária (MW) para estas usinas.

Dada a sazonalidade diurna da geração de energia solar, essas distribuições são naturalmente bimodais, com um modo forte em torno de zero (correspondente à produção nula no período noturno) e outro modo representando a geração durante o dia. Cada uma dessas usinas apresenta um perfil de geração distinto durante a janela de análise, possivelmente influenciado por padrões operacionais específicos ou por variáveis exógenas em escala local. A usina Lapa exibe uma distribuição com bimodalidade pronunciada, onde o modo de geração diurna se concentra próximo a 50 MW, com frequência comparável à da geração nula. Isso sugere que, durante o período de produção, a usina tende a operar em um patamar de geração consistente.

Já a usina Floresta apresenta uma distribuição com uma cauda direita mais longa e um segundo modo mais disperso (achatado) em comparação com Lapa. Isso indica um

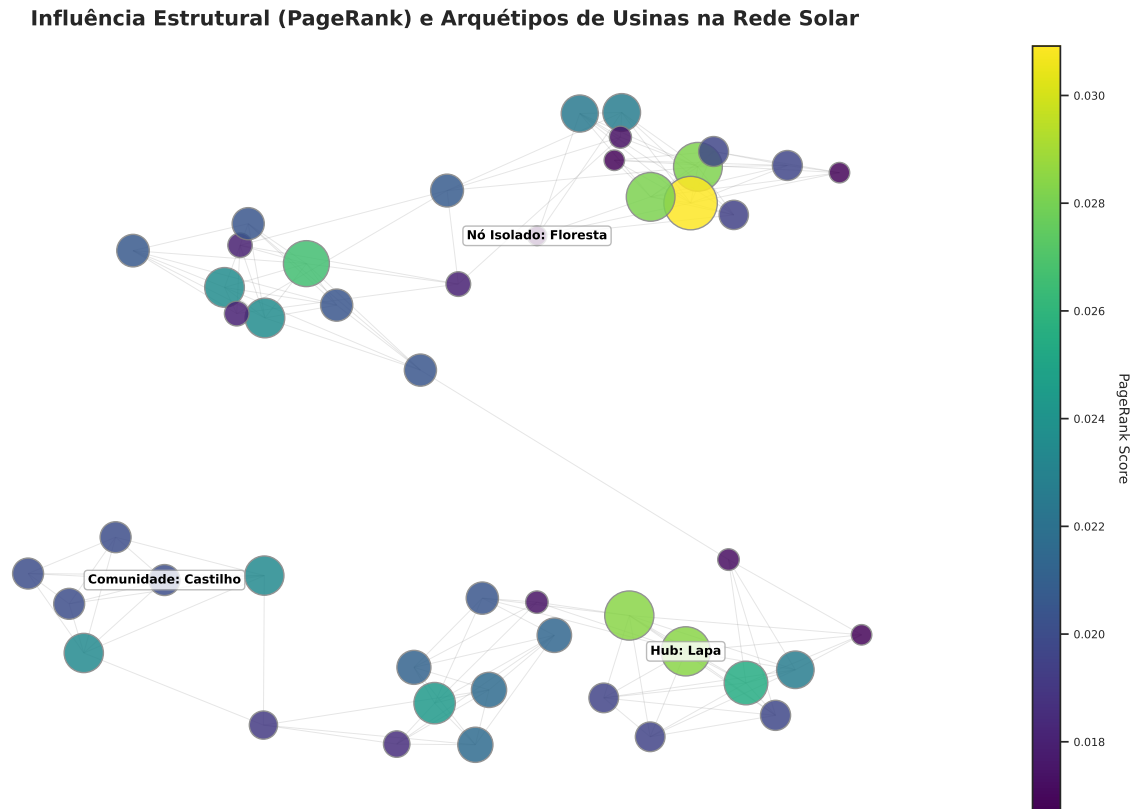


Figura 6 – Visualização da rede de usinas solares destacando arquétipos baseados em sua influência estrutural (PageRank) e conectividade.

perfil de geração com maior variabilidade durante as horas diurnas, onde a energia é produzida em uma faixa mais ampla (entre 20 MW e 75 MW) com frequências similares, mas sem um pico de geração tão concentrado quanto Lapa.

Por fim, a usina Castilho tem um perfil de distribuição com uma cauda direita acentuadamente longa, sugerindo picos de geração horária de maior intensidade. Contudo, a frequência de ocorrência desses picos de alta intensidade é menor.

O perfil de geração indica que, embora não gere energia de forma intensa continuamente, quando o faz, atinge patamares de geração elevados (frequência similar entre 50 MW e 200 MW), mas de forma mais intermitente do que os outros arquétipos.

A Tabela 4 apresenta o sumário estatístico dessas três usinas, detalhando suas características de geração horária (MW) ao longo do período de análise. É possível verificar que todas as usinas possuem registros para todos os passos horários e exibem uma geração efetiva mais concentrada a partir do percentil 75.

Os conjuntos Lapa e Floresta apresentam perfis de geração com estatísticas descritivas mais próximas entre si, enquanto o Castilho se destaca por sua geração mais intensa, com uma média e desvio padrão significativamente maiores.

Além disso, observa-se que a usina Castilho não apresenta valores nulos em sua

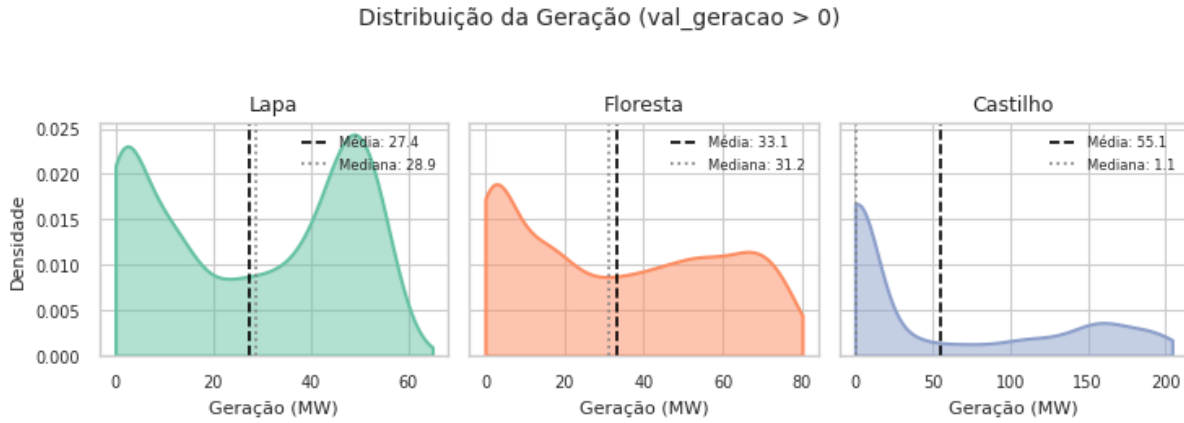


Figura 7 – Distribuição da geração para as usinas arquétipos - Hub (Lapa), Nó Isolado (Floresta) e Membro de Comunidade (Castilho). As estatísticas descritivas apresentadas referem-se à distribuição global da amostra observada na janela de análise. Ressalta-se que a série temporal bruta de geração solar é inerentemente não-estacionária devido à forte componente sazonal diária.

Tabela 4 – Estatísticas descritivas comparativas para a geração horária (MW) das usinas arquétipo (Lapa, Floresta, Castilho) no período de análise.

Métrica	Lapa	Floresta	Castilho
timestep	4368,00	4368,00	4368,00
Média	12,58	17,96	55,14
Std	19,36	25,15	71,95
Mínimo	0,00	0,00	0,05
1%	0,00	0,00	0,14
5%	0,00	0,00	0,14
10%	0,00	0,00	0,15
25%	0,00	0,00	0,15
50%	0,00	0,44	1,10
75%	23,95	35,28	126,89
90%	48,67	62,77	172,50
95%	52,17	70,34	190,24
99%	56,26	76,63	202,57
Max	65,00	80,04	204,16
Assimetria	1,18	1,09	0,84
Curtose	-0,30	-0,34	-0,98
Zeros (%)	54,12	45,76	0,00

distribuição de geração (0% de zeros), o que é uma particularidade notável. Mesmo com um valor mínimo de 0.05 MW, seu perfil indica que a geração, quando ocorre, é predominantemente de alta intensidade para os 25% superiores de sua produção.

Os comportamentos observados nessas três usinas, quando relacionados aos arquétipos definidos, sugerem que os papéis funcionais na rede não têm uma relação direta e exclusiva com a intensidade de energia gerada em condições normais.

A Figura 6 ilustra a atuação desses três nós na rede com base nos PageRanks, que variam entre 0.018 e 0.030. Observa-se que o Nó Isolado (Floresta) possui um PageRank consistentemente baixo, mesmo estando conectado à rede. O Membro de Comunidade (Castilho) apresenta um PageRank intermediário e se integra a outros nós, refletindo sua participação em um sub-grafo interno com maior densidade. Por fim, o Hub (Lapa) exibe um PageRank elevado, inserido em sub-grafos com alta densidade e com conexões a outras usinas de alto PageRank, evidenciando sua posição central e influente na rede.

Para aprofundar a análise, selecionam-se os arquétipos Lapa, Floresta e Castilho, cujas distribuições de geração horária (MW) são apresentadas na Figura 7. A usina Lapa (Hub) exibe uma distribuição com bimodalidade pronunciada, sugerindo um patamar de geração consistente em torno de 50 MW durante o período de produção. Em contraste, a usina Floresta (Nó Isolado) apresenta um perfil mais disperso e com uma cauda direita mais longa, indicando maior variabilidade durante as horas diurnas (entre 20 MW e 75 MW). Por fim, o perfil da usina Castilho (Comunidade) é marcado por uma cauda acentuadamente longa, sugerindo picos de geração de alta intensidade, embora menos frequentes.

Os comportamentos observados nessas três usinas, quando relacionados aos arquétipos definidos, sugerem que os papéis funcionais na rede não possuem uma relação direta e exclusiva com a intensidade média de energia gerada. A identificação e a caracterização inicial desses arquétipos fornecem uma base para investigar se diferentes papéis estruturais se traduzem em comportamentos temporais distintos.

Visando compreender as dinâmicas das usinas, foram realizadas estimações de modelos SARIMA para séries diferenciadas, considerando a forte sazonalidade diária presente em todas as séries. Para isso, utilizou-se a biblioteca ‘pmdarima’ para identificar parâmetros adequados a cada uma das 47 usinas, empregando o módulo ‘auto_arima’ para otimização automática. Além disso, os modelos estimados servem como base para obter os seus resíduos e realizar mais testes estatísticos.

Após as estimações, o teste ADF foi aplicado às séries resultantes para confirmar sua estacionariedade. Os resultados dessas análises são sumarizados na Tabela 5, que relaciona a estrutura do grafo com a estacionariedade das séries.

Tabela 5 – Resumo estatístico por tipo de relação — Estrutura e ADF

Relação	Nº Usinas	Hops (Média)	Hops (Mín)	Hops (Máx)	ADF
Arquétipo (Comunidade)	1	0,00	0	0	0,0
Arquétipo (Hub)	1	0,00	0	0	0,0
Arquétipo (Isolado)	1	0,00	0	0	0,0
Não Relacionado	9	7,6	2	14	0,0
Viz. Adj.	19	1,00	1	1	0,0
Viz. Func. (Hub)	4	6,00	1	8	0,0
Viz. Func. (Isolado)	4	4,25	2	7	0,0

A Tabela 5 também apresenta o número de hops (saltos entre nós) para diferentes tipos de relações, como vizinhos de adjacência e nós funcionalmente próximos aos arquétipos. Um hop representa o caminho mínimo necessário para uma informação atravessar o grafo, sendo um parâmetro relevante no design de arquiteturas de aprendizado de máquina ao denotar o campo de percepção de cada nó.

Observa-se que vizinhos de adjacência estão a 1 hop de distância, enquanto os vizinhos dos arquétipos exibem diferentes números de hops médios conforme sua função na rede, ilustrando a variabilidade da distância estrutural entre os diferentes grupos de usinas.

A Tabela 6 apresenta o resumo estatístico dos resíduos dos modelos SARIMA para cada tipo de relação, utilizando o teste de Ljung-Box, que avalia a presença de autocorrelação nos resíduos. Essa análise é feita para determinar se os modelos SARIMA capturaram toda a estrutura temporal ou se ainda há informações não explicadas, as quais podem ser atribuídas a influências não-temporais ou interações complexas entre as usinas.

Tabela 6 – Resumo estatístico por tipo de relação — Ljung-Box e Resíduos

Relação	LB (Média)	LB (Mín)	LB (Máx)	% Res. N/B
Arquétipo (Comunidade)	0,6477	0,6477	0,6477	0,00
Arquétipo (Hub)	0,0053	0,0053	0,0053	100,00
Arquétipo (Isolado)	0,8681	0,8681	0,8681	0,00
Não Relacionado	0,5108	0,0053	0,8681	11,11
Viz. Adj.	0,5014	0,0005	0,9985	15,79
Viz. Func. (Hub)	0,5109	0,1427	0,9169	0,00
Viz. Func. (Isolado)	0,4313	0,1262	0,8388	0,00

As séries temporais foram diferenciadas para remover a sazonalidade diária (24 passos horários), que se mostra muito forte em todas elas. Assim, a maioria das séries estimadas apresentou ordens típicas como (1,0,1)(1,1,0), com p-valor (que representa a probabilidade de observar os dados caso a hipótese nula seja verdadeira) do teste de Ljung-Box > 0.05 , o que confirma que os resíduos se assemelham a ruído branco.

No entanto, é importante salientar que as usinas Lapa (Hub), Pereira Barreto e Jaíba 4 Dist apresentaram p-valor de Ljung-Box < 0.05 , indicando que os modelos

SARIMA não foram capazes de remover completamente a autocorrelação residual nessas séries.

Essas evidências indicam que existem relações complexas em algumas usinas que não são plenamente capturadas por modelos de séries temporais clássicos univariados. Embora as estimações dos modelos SARIMA tenham sido realizadas de maneira computacionalmente exaustiva, testando diversas combinações de parâmetros, não foi possível obter resíduos puramente semelhantes a ruído branco para todas as séries, reforçando a limitação da abordagem clássica para o problema estudado.

Para complementar a análise de interdependência, a Tabela 7 sumariza a Informação Mútua (MI) entre os diferentes tipos de relação. A Informação Mútua é uma medida da dependência estatística entre duas variáveis, quantificando a quantidade de informação compartilhada entre elas.

Diferente da correlação linear, a MI é capaz de capturar relações não-lineares, indicando o quanto o conhecimento sobre uma variável reduz a incerteza sobre a outra. A MI entre duas variáveis aleatórias discretas X e Y é definida como:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (3.13)$$

onde $p(x, y)$ é a probabilidade de conjunta de X e Y , e $p(x)$ e $p(y)$ são as probabilidades marginais. Para variáveis contínuas, as somas são substituídas por integrais. Um valor de MI igual a zero indica independência total entre as variáveis, enquanto valores maiores indicam uma maior dependência, independentemente da natureza linear ou não linear dessa relação.

Neste trabalho, para viabilizar o cálculo computacional e capturar robustamente as não-linearidades, as séries temporais de resíduos foram discretizadas utilizando o método de quantis (quantile binning) com $k = 10$ intervalos. Dessa forma, os valores de MI obtidos expressam a redução de incerteza na unidade de nats (logaritmo natural). Assim, na Tabela 7, a auto-dependência (MI da usina consigo mesma) foi fixada como referência unitária (1,0000) para fins de comparação relativa, enquanto os demais valores refletem a magnitude absoluta calculada pelo estimador.

A partir da Tabela 7, verifica-se que os vizinhos de adjacência e os nós não relacionados diretamente ainda possuem uma presença de resíduos não-brancos em até 15% das séries. Isso serve como evidência adicional de que relações mais complexas persistem nos dados, não sendo passíveis de serem plenamente capturadas por métodos clássicos.

A Tabela 8 sumariza a Informação Mútua (MI) entre os diferentes tipos de relação, fornecendo uma medida de dependência entre as séries de geração das usinas, independentemente da causalidade.

Tabela 7 – Resumo estatístico por tipo de relação — Informação Mútua

Tipo de Relação	MI (Média)	MI (Mín)	MI (Máx)	% MI Alto (>0.5)
Arquétipo (Comunidade)	1,0000	1,0000	1,0000	100,00
Arquétipo (Hub)	1,0000	1,0000	1,0000	100,00
Arquétipo (Isolado)	1,0000	1,0000	1,0000	100,00
Não Relacionado (Distância)	0,5212	0,4226	0,6919	44,44
Viz. Adj.	0,5823	0,3268	0,8182	78,95
Viz. Func. (Hub)	0,5733	0,3760	0,8182	75,00
Viz. Func. (Isolado)	0,6298	0,4961	0,7265	75,00

As análises dos resultados das Tabelas 5, 6 e 7 revelam sobre a dinâmica da rede:

- A Tabela 6 mostra que, para o Arquétipo Lapa (Hub), os resíduos do SARIMA persistem como autocorrelacionados (100% de resíduos não-ruído branco e baixo p-valor LB de 0.0053). Isso indica que, mesmo após a modelagem temporal univariada, há uma estrutura residual significativa, possivelmente devido à sua posição de hub e a influências não capturadas pelo modelo SARIMA isolado, mas sim por interações complexas na rede.
- A Tabela 7, de Informação Mútua, evidencia uma forte dependência entre usinas com apenas 1 hop de distância (Viz. Adj.), com uma média de MI elevada (0.5823) e uma alta porcentagem de MI alto (>0.5), corroborando a relevância da proximidade geográfica.
- Usinas mais distantes, seja funcionalmente (Viz. Func. - Hub e Isolado) ou geograficamente (Não Relacionado), ainda podem apresentar uma MI significativa, sugerindo que interdependências existem mesmo para nós não diretamente adjacentes no grafo k-NN.

Essas análises fornecem a base empírica para a abordagem de modelagem espaço-temporal, pois demonstram que a dinâmica de geração de uma usina não é apenas uma função do seu próprio histórico, mas é influenciada por sua posição estrutural na rede e pelas interações com seus vizinhos.

3.6 Engenharia de variáveis

A performance dos modelos preditivos é impactada pela qualidade e pertinência das variáveis de entrada. A engenharia de features consiste na criação de novas variáveis a partir dos dados brutos e informações estruturais, visando enriquecer o conjunto de dados e otimizar a capacidade de aprendizado dos modelos.

Neste trabalho, a estratégia de engenharia de features foi segmentada em duas abordagens distintas, alinhadas às classes de modelos: uma abordagem extensiva para a referência XGBoost e outra, mais parcimoniosa, para as arquiteturas baseadas em grafos.

Para o modelo XGBoost, foi desenvolvido um conjunto exaustivo de 84 variáveis. Sua robusta arquitetura, que incorpora técnicas de gradient boosting e mecanismos de regularização L1 e L2, permite processar um conjunto de informações e inferir a relevância de cada feature, mitigando o risco de sobreajuste.

Este conjunto abrangeu as seguintes categorias: climáticas, sazonais, variáveis defasadas (lags), janelas móveis, interações entre variáveis e métricas de grafo, conforme detalhado na Tabela 8.

Em contrapartida, para as arquiteturas GAT-LSTM e GCN-1D, optou-se por um conjunto de features mais conciso. A justificativa para essa decisão é dupla: primeiramente, o elevado custo computacional associado ao treinamento de GNNs com um número excessivo de variáveis; e, fundamentalmente, a própria natureza desses modelos.

As GNNs são projetadas para aprender representações das interdependências espaciais a partir da topologia do grafo e do mecanismo de propagação de mensagens. Dessa forma, a inclusão explícita de métricas de rede como features de entrada é potencialmente redundante.

O conjunto de variáveis para os modelos em grafo, detalhado na Tabela 9, focou em fornecer os dados climáticos, sazonais e de lags de geração essenciais.

Tabela 9 – Features utilizadas nos modelos GCN1D e GAT-LSTM, organizadas por categoria.

Categoria	Features
Climáticas	ALLSKY_KT, ALLSKY_SFC_SW_DWN, RH2M, WS10M, T2M
Sazonais	dayofweek_cos, dayofweek_sin, dayofyear_cos, dayofyear_sin, hour_cos, hour_sin, month_cos, month_sin
Lags (históricos)	val_geracao_lag_1, val_geracao_lag_2, val_geracao_lag_3, val_geracao_lag_6, val_geracao_lag_9, val_geracao_lag_12, val_geracao_lag_24, val_geracao_lag_48

Tabela 8 – Features utilizadas no modelo XGBoost, organizadas por categoria.

Categoria	Features
Localização	nom_usina, din_instante, latitude, longitude, location, closest_latitude, closest_longitude, distance_km
Climáticas	ALLSKY_SFC_SW_DWN, T2M, RH2M, WS10M, ALLSKY_KT
Sazonais	hour_sin, hour_cos, dayofweek_sin, dayofweek_cos, month_sin, month_cos, dayofyear_sin, dayofyear_cos
Lags	val_geracao_lag_1, val_geracao_lag_2, val_geracao_lag_3, val_geracao_lag_6, val_geracao_lag_9, val_geracao_lag_12, val_geracao_lag_24, val_geracao_lag_48, ALLSKY_SFC_SW_DWN_lag_1, ALLSKY_SFC_SW_DWN_lag_2, ALLSKY_SFC_SW_DWN_lag_6, ALLSKY_SFC_SW_DWN_lag_12, T2M_lag_1, T2M_lag_2, T2M_lag_6, T2M_lag_12, RH2M_lag_1, RH2M_lag_2, RH2M_lag_6, RH2M_lag_12, WS10M_lag_1, WS10M_lag_2, WS10M_lag_6, WS10M_lag_12, ALLSKY_KT_lag_1, ALLSKY_KT_lag_2, ALLSKY_KT_lag_6, ALLSKY_KT_lag_12
Janelas móveis	val_geracao_rolling_mean12H_cleft, val_geracao_rolling_std12H_cleft, val_geracao_rolling_min12H_cleft, val_geracao_rolling_max12H_cleft, val_geracao_rolling_mean24H_cleft, val_geracao_rolling_std24H_cleft, val_geracao_rolling_min24H_cleft, val_geracao_rolling_max24H_cleft, val_geracao_rolling_mean48H_cleft, val_geracao_rolling_std48H_cleft, val_geracao_rolling_min48H_cleft, val_geracao_rolling_max48H_cleft, val_geracao_rolling_mean168H_cleft, val_geracao_rolling_std168H_cleft, val_geracao_rolling_min168H_cleft, val_geracao_rolling_max168H_cleft, ALLSKY_SFC_SW_DWN_rolling_mean6H_cleft, ALLSKY_SFC_SW_DWN_rolling_sum6H_cleft, ALLSKY_SFC_SW_DWN_rolling_mean12H_cleft, ALLSKY_SFC_SW_DWN_rolling_sum12H_cleft
Interações	T2M_x_RH2M, ALLSKY_SFC_SW_DWN_x_hour_cos, WS10M_x_val_geracao_lag_1, val_geracao_rolling_mean12H_cleft_x_ALLSKY_KT, val_geracao_lag_1_x_val_geracao_lag_24, T2M_lag_1_x_T2M_lag_6
Métricas de grafo	pagerank, degree centrality, local_clustering_coefficient, betweenness centrality, is_hub, is_isolated_node, is_community_member

4 METODOLOGIA

A Análise Exploratória de Dados consolidou o conjunto de dados para a modelagem e forneceu duas linhas de evidência que justificam a abordagem espaço-temporal do trabalho. Primeiramente, a análise da estrutura em grafo permitiu a identificação de arquétipos de usinas com base em seu papel topológico (Hub, Comunidade, Nó Isolado), fornecendo uma base para a avaliação granular dos resultados no Capítulo 5.

Em segundo lugar, a análise da dinâmica das séries revelou a insuficiência de modelos puramente univariados. Embora a maioria das séries tenha sido adequadamente modelada por ordens SARIMA, o teste de Ljung-Box indicou a persistência de autocorrelação nos resíduos de usinas estruturalmente importantes, como o Hub(Lapa), evidenciando uma complexidade temporal não capturada por completo. Esta observação é corroborada pela análise de Informação Mútua, que revelou dependências não-lineares significativas entre usinas adjacentes.

O conjunto desses achados — a complexidade interna de certas séries e a interdependência externa entre elas — constitui a justificativa empírica para a metodologia detalhada neste capítulo.

Para lidar com essa complexidade, estabelece-se como linha de base o XGBoost, um modelo amplamente validado na literatura para dados tabulares. Sua robustez deriva da combinação de técnicas de *gradient boosting* com mecanismos de regularização L1 e L2, o que o torna um *benchmark* de alta performance.

Em contrapartida, serão exploradas duas arquiteturas de grafos, o GAT-LSTM e o GCN-1D, projetadas para aprender diretamente a partir da estrutura do grafo e da evolução temporal dos dados.

Para além da previsão pontual, a metodologia visa quantificar a incerteza das estimativas. Para isso, todos os modelos serão treinados com a função de perda *Pinball Loss* para realizar regressão quantílica, permitindo a geração de intervalos de predição.

A seguir, serão detalhados os fundamentos de cada modelo, a construção do grafo, a função de perda, as métricas de avaliação e a configuração experimental.

4.1 XGBoost

XGBoost é um algoritmo que utiliza árvores de decisão otimizadas para a previsão da variável alvo em problemas de aprendizado supervisionado, amplamente testado na comunidade de inteligência artificial nos últimos anos, principalmente em dados tabulares ou estruturados.

Para explicar o funcionamento do algoritmo, vamos inicialmente dividir o seu funcionamento em três partes: árvores de decisão, aprendizado de conjuntos e gradient boosting.

As árvores de decisão utilizam regras decisórias para dividir as variáveis em folhas com base no critério de decisão estabelecido.

Com base nas variáveis disponíveis, estima-se o número mínimo de regras para verificar a probabilidade de fazer uma decisão correta - i.e. prever a decisão adequadamente com as melhores regras possíveis.

O boosting busca melhorar a previsão dos modelos a partir de agregações de modelos mais simples ou fracos a fim de obter modelos mais generalizados e robustos. Já para o gradient boosting essa ideia é preservada, porém utilizando uma função objetivo que tem o seu erro minimizado através do aprendizado do gradiente da função objetivo.

Nele, o modelo é construído de forma aditiva, onde cada nova árvore fraca é treinada para prever os resíduos do modelo combinado das árvores anteriores, seguindo a direção do gradiente descendente da função de perda.

A ideia do aprendizado de conjunto(ensemble) é de utilizar vários modelos de aprendizado de máquina para o mesmo conjunto de dados e agregar os seus resultados para obter um melhor modelo. O XGBoost utiliza de árvores de decisão com gradiente aumentado para realizar essa agregação dos modelos, essa técnica treina um conjunto de árvores e a cada iteração utiliza o erro residual da previsão do modelo anterior para o próximo modelo do conjunto.

Assim, a predição final é uma soma ponderada das previsões das árvores do conjunto. Isso proporciona uma minimização tanto do viés da previsão quanto do underfitting.

O XGBoost implementa essa técnica com otimizações adicionais: regularização, paralelização e controle de complexidade das árvores. Dessa forma, o modelo combina todas as árvores em uma soma ponderada que reduz tanto o viés quanto o underfitting. Finalmente, o algoritmo realiza a predição da variável alvo por meio de um conjunto de árvores de decisão. A predição final para a i -ésima amostra é dada por:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \quad (4.1)$$

onde cada f_k representa uma árvore de decisão, e \mathcal{F} é o espaço de funções das árvores possíveis. O treinamento do modelo busca minimizar a seguinte função objetivo:

$$L(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^K \Omega(f_k) \quad (4.2)$$

onde l é a função de perda (por exemplo, erro quadrático), e Ω é o termo de regularização definido por:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 + \alpha \sum_{j=1}^T |w_j| \quad (4.3)$$

com T sendo o número de folhas da árvore, e w_j os pesos atribuídos à j -ésima folha, γ a penalidade da complexidade (ou custo de adicionar uma folha à árvore), λ o termo de regularização L2 e α o termo de regularização L1. Essa estrutura promove regularização e controle da complexidade do modelo, contribuindo para evitar *overfitting*.

Os principais hiperparâmetros que influenciam o desempenho do XGBoost incluem:

- **Taxa de aprendizado (η):** define o impacto de cada nova árvore na predição final. A cada iteração obtém-se diretamente os novos pesos das novas *features*, de modo que esse parâmetro diminui os pesos das *features* de forma conservadora, isto é, cada árvore corrige o erro anterior de forma mais gradual. Assim, η funciona como um fator de amortecimento, controlando a agressividade das correções e diminuindo o risco de *overfitting*.
- **max_depth:** controla a profundidade máxima das árvores. Árvores mais profundas capturam mais complexidade, mas podem superajustar.
- **min_child_weight:** número mínimo de amostras necessárias em uma folha para que ela seja dividida. Ajuda a controlar o *overfitting*.
- **gamma (γ):** ganho mínimo de perda necessário para realizar uma divisão em uma árvore, quanto maior o γ , mais conservador será o modelo.
- **lambda (λ):** parâmetro para a regularização norma L2 nos pesos das *features*, aumentar esse parâmetro acarreta em um aumento da penalização da norma quadrática desses pesos.
- **alpha (α):** parâmetro para a regularização norma L1 nos pesos das *features*, aumentar esse parâmetro acarreta em um aumento da penalização absoluta desses pesos.
- **subsample e colsample_bytree:** frações de amostras e variáveis usadas para construir cada árvore, assim o primeiro evita que o modelo dependa de dados de treino específicos, o segundo evita que o modelo dependa demais de certas variáveis.

Além de ser eficiente e amplamente adotado, o XGBoost é robusto em uma grande gama de problemas de aprendizado supervisionado, assim como proporciona uma interpretação do modelo (e.g., importância das *features*), o que acrescenta o seu valor em aplicações práticas.

4.2 LSTM

O método LSTM, introduzido por (HOCHREITER; SCHMIDHUBER, 1997), é uma arquitetura das redes neurais recorrentes (RNNs) projetada para capturar dependências de longo prazo em sequências temporais.

Um dos principais desafios das RNNs tradicionais é o problema do desaparecimento do gradiente, que ocorre durante o treinamento por retropropagação em sequências longas, dificultando o aprendizado de relações temporais distantes. O LSTM foi criado para mitigar esse problema, permitindo que informações importantes sejam mantidas por longos períodos.

A estrutura do LSTM baseia-se na recorrência, conforme descrito por (HOCHREITER; SCHMIDHUBER, 1997), que ocorre por meio de células conectadas sequencialmente, em que cada célula recebe a entrada atual x_t e a saída anterior h_{t-1} , gerando uma nova saída h_t enquanto atualiza seu estado interno. Essa arquitetura permite à rede adaptar suas estimativas com base tanto no contexto presente quanto no passado recente.

Assim, a saída oculta h_t representa a informação codificada naquele instante, enquanto que o estado C_t da célula contém as informações de longo prazo, sendo o estado interno mantido separado da saída oculta.

Na Figura 8, A representa a célula recorrente, x_t é o vetor de entrada no tempo t e h_t é a saída estimada. O diagrama mostra como uma única célula se repete ao longo do tempo, formando uma cadeia que processa as sequências temporais.

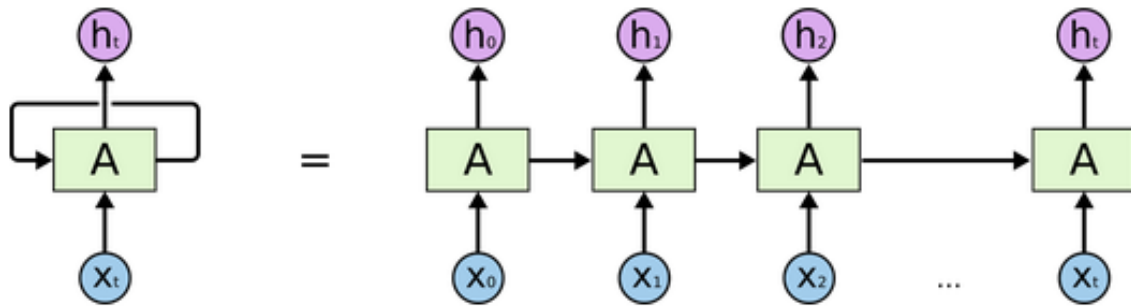


Figura 8 – Uma rede neural recorrente é composta por células idênticas e sequenciais. Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).

Diferentemente das RNNs tradicionais, que tem apenas uma célula simples, o LSTM incorpora múltiplas portas que regulam efetivamente o fluxo de informação.

A célula LSTM possui quatro componentes principais: porta de esquecimento, porta de entrada, estado da célula e porta de saída. O objetivo dessas portas é controlar quais informações devem ser esquecidas, armazenadas ou expostas, com base nas entradas e no histórico.

A Figura 9 ilustra a arquitetura interna da célula LSTM. A linha horizontal superior representa o fluxo contínuo de memória — chamado de estado da célula (C_t) — que carrega informações relevantes ao longo da sequência.

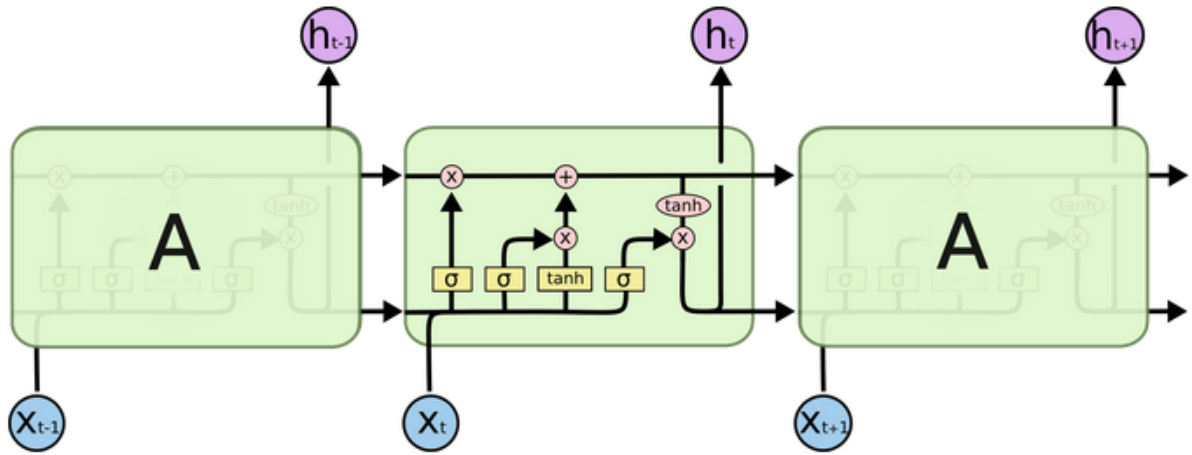


Figura 9 – A célula do LSTM é composta por múltiplas portas que regulam o fluxo de informação. Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).

A primeira parte da célula é a porta do esquecimento, mostrada na figura 10 abaixo. Essa porta decide quais informações do estado anterior devem ser descartadas. Ela recebe como entrada o vetor x_t e a saída da célula anterior h_{t-1} , e aplica uma função sigmoide - destacada pelo bloco em amarelo:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (4.4)$$

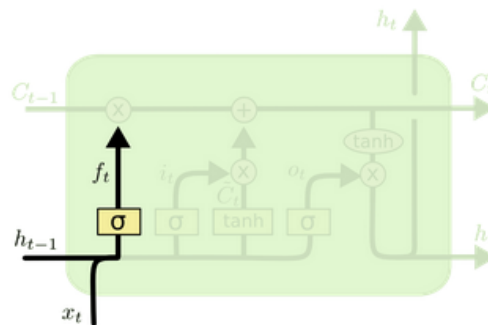


Figura 10 – Porta de esquecimento: controla o que deve ser removido do estado de célula anterior. Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).

O vetor f_t , com valores entre 0 e 1, atua como um filtro que controla o quanto do estado anterior será preservado.

Em seguida, tem-se a atuação da porta de entrada e a atualização do estado, aqui o modelo avalia quais novas informações devem ser adicionadas ao estado da célula. Esse processo ocorre em duas etapas, destacadas na figura abaixo:

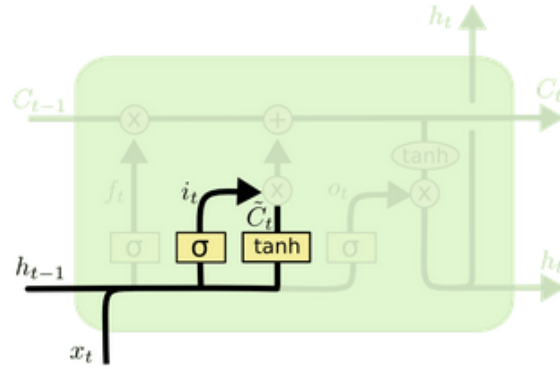


Figura 11 – As portas de entrada e os candidatos de memória colaboram na atualização do estado interno da célula. Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).

A primeira etapa trata da porta de entrada, destacada na figura como a sigmoide na esquerda, que determina quais valores serão atualizados a partir da entrada atual x_t : e da saída da célula anterior, h_{t-1} ,

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (4.5)$$

Já a segunda etapa, a tangente hiperbólica destacada à direita na figura, tem-se a geração de candidatos a serem armazenados:

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (4.6)$$

Com esses vetores, o estado da célula é atualizado de modo que parte do estado anterior é preservada via $f_t \odot C_{t-1}$ e parte é renovada com as novas informações relevantes $i_t \odot \tilde{C}_t$ da seguinte forma:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (4.7)$$

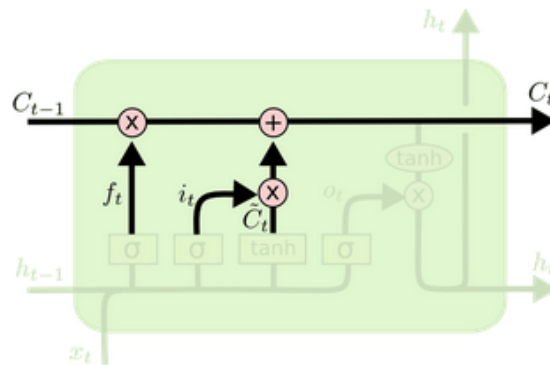


Figura 12 – Porta de saída: define quais informações da célula serão utilizadas como saída h_t . Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).

Assim, o estado atual da célula C_t a ser transmitido através da linha horizontal superior é calculado com base na memória C_{t-1} e da nova seleção de candidatos a terem informações relevantes.

Por fim, a porta de saída determina quais partes do estado da célula serão usadas como saída para o próximo tempo, como mostrado na figura abaixo:

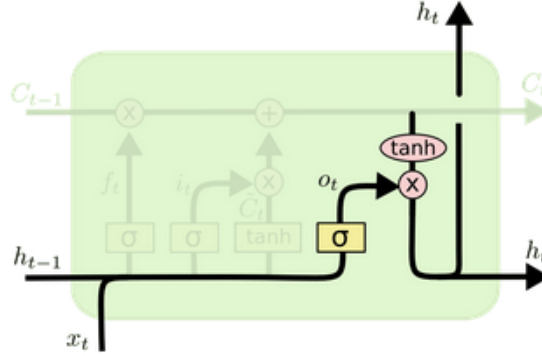


Figura 13 – Porta de saída: define quais informações da célula serão utilizadas como saída h_t . Fonte: (OLAH, 2015); estrutura baseada em (HOCHREITER; SCHMIDHUBER, 1997).

Primeiro, na sigmoide destacada em amarelo na figura, calcula-se:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (4.8)$$

E a saída final da célula h_t é obtida através do produto de Hadamard do resultado da sigmoide anterior com a tangente hiperbólica seletora das informações da célula atual C_t :

$$h_t = o_t \odot \tanh(C_t) \quad (4.9)$$

Essa saída h_t - memória a curto prazo - é utilizada tanto como output da célula atual C_t - memória a longo prazo - quanto como entrada para a célula seguinte, mantendo o encadeamento temporal.

Em suma, o funcionamento do LSTM é feito através das seguintes funções supracitadas:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad \text{Porta de esquecimento} \quad (2.8)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad \text{Porta de entrada} \quad (2.9)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad \text{Candidatos à memória} \quad (2.10)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad \text{Atualização do estado da célula} \quad (2.11)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad \text{Porta de saída} \quad (2.12)$$

$$h_t = o_t \odot \tanh(C_t) \quad \text{Saída da célula} \quad (2.13)$$

Com essa estrutura, o LSTM consegue armazenar informações por longos períodos, aprendendo a ignorar ou reforçar dados conforme necessário para a tarefa objetivo.

Em relação ao seu treinamento, o LSTM não difere significativamente de outras técnicas de aprendizado de máquina; este aspecto será detalhado na seção dedicada à função de perda e ao processo de treinamento.

A habilidade do algoritmo em preservar e manipular informações relevantes ao longo de extensas janelas temporais é fundamental para problemas de previsão no setor energético, nos quais as variáveis apresentam fortes correlações temporais e padrões sazonais complexos.

Essa característica permite ao modelo capturar relações de dependência de longo prazo que métodos tradicionais tendem a ignorar, aumentando a precisão e a robustez das previsões, conforme demonstrado por (KONG *et al.*, 2019) em cenários reais de previsão energética.

4.3 Modelos de Aprendizado Profundo em Grafos

Os métodos estudados constituem uma classe de modelos de aprendizado profundo projetados para processar dados organizados sob a forma de grafos. Diferentemente das redes neurais tradicionais, que operam majoritariamente sobre dados tabulares ou imagens, as GNNs exploram explicitamente a estrutura relacional intrínseca aos grafos, em que os nós representam entidades e as arestas, suas conexões ou relações.

O principal objetivo das GNNs é aprender representações (ou embeddings) dos nós, das arestas ou do grafo como um todo, preservando tanto a informação estrutural quanto os atributos associados aos componentes da rede.

Considerando o caráter espaço-temporal do problema tratado neste trabalho, as arquiteturas baseadas em GNNs têm se destacado como alternativas de ponta para a modelagem desse tipo de fenômeno, possibilitando capturar a dinâmica entre as usinas a partir de matrizes de adjacência, sejam elas estáticas ou dinâmicas (YU; YIN; ZHU, 2018; SONG *et al.*, 2020; WU *et al.*, 2021).

Assim, torna-se viável investigar como as correlações entre as usinas evoluem ao longo do tempo, inclusive levando-se em conta variáveis exógenas.

Para simplificar a construção das representações do grafo, uma estratégia possível é a definição dos vértices com base nas distâncias geográficas entre pares de usinas, o que reduz a complexidade espaço-temporal do modelo, mantendo, entretanto, sua capacidade representativa, conforme demonstrado na literatura especializada (LI *et al.*, 2018; ZHENG *et al.*, 2020).

A dinâmica temporal do problema é incorporada diretamente pelas operações fundamentais dessas arquiteturas, como agregação de informações e passagem de mensagens entre os nós, o que possibilita modelar interdependências complexas tanto no espaço quanto no tempo.

Essas propriedades diferenciam as GNNs dos métodos clássicos discutidos anteriormente, os quais são capazes de capturar padrões temporais, mas não consideram explicitamente as relações espaciais e tampouco permitem inferências sobre outras localidades além daquelas diretamente modeladas (WU *et al.*, 2021; SONG *et al.*, 2020).

Assim, a generalização proporcionada pela modelagem em grafos permite não apenas avaliar a contribuição individual de cada usina, mas sobretudo compreender como o sistema de geração solar, considerado em sua totalidade, impacta a operação e a confiabilidade da rede de distribuição elétrica (CHEN *et al.*, 2021; WU *et al.*, 2021).

Neste contexto, os modelos de redes neurais em grafos analisados nesta seção incluem: Graph Attention Networks (GAT-LSTM) e Graph Convolutional Networks (GCN-1D), (KIPF; WELLING, 2017; VELIČKOVIĆ *et al.*, 2018; YU; YIN; ZHU, 2018; SONG *et al.*, 2020).

4.3.1 Fundamentos

Um grafo é definido como uma estrutura $G = (V, A)$, em que V representa o conjunto de vértices (ou nós) e A o conjunto de arestas que conectam pares de vértices. As propriedades dessa estrutura de dados relevantes para esse trabalho são: direcionalidade, pesos das arestas de conexão e a conexão entre os nós.

Em um grafo não direcionado, as conexões entre os pares de nós são bidirecionais, i.e. a relação entre eles é simétrica, com influência ou associação mútuas logo $A_{ij} = A_{ji}$.

Em contraste, em um grafo direcionado essa relação não é recíproca, as conexões entre os nós possuem uma orientação específica, de modo que o nó A tem influência do nó B, mas a recíproca não é necessariamente verdade, assim caracteriza-se a assimetria das associações entre nós: $A_{ij} \neq A_{ji}$.

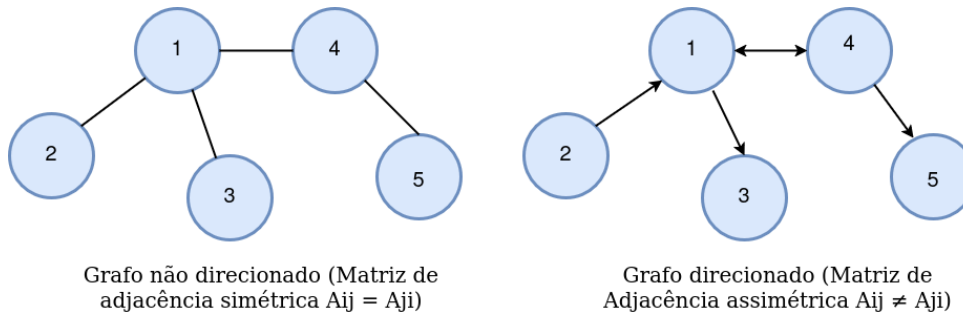


Figura 14 – Exemplos de grafos não direcionado e direcionado. Figura de autoria própria.

Essa propriedade é representada através da matriz de adjacência, que expressa as conexões entre os nós de um grafo. Nessa matriz, as linhas e colunas correspondem aos nós, e define-se que cada elemento $A_{ij} = 1$, se existe uma conexão entre o nó i e o nó j , e $A_{ij} = 0$ caso contrário.

Como exemplo ilustrativo, considere os dois grafos apresentados na Figura 14: um não direcionado e outro direcionado, com suas respectivas matrizes de adjacência.

• **Grafo não direcionado:**

Matriz de adjacência:

Nós: $\{1, 2, 3, 4, 5\}$

Conexões: $1 - 2, 1 - 3, 1 - 4, 4 - 5$

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

• **Grafo direcionado:**

Matriz de adjacência:

Nós: $\{1, 2, 3, 4, 5\}$

Conexões: $1 \rightarrow 4; 1 \rightarrow 3; 2 \rightarrow 1;$
 $4 \rightarrow 5;$

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Observe que, como esperado, a matrizes de adjacência do grafo não direcionado é simétrica, refletindo a reciprocidade das conexões entre os nós ($A_{ij} = A_{ji}$). Por outro lado, a matriz de adjacência do grafo direcionado é assimétrica, e nesse exemplo também é esparsa, visto que nem todas as conexões possíveis entre os nós estão presentes.

Para simplificação do problema de previsão de múltiplas séries temporais e considerando a relação mútua entre os regimes espaço-temporais das usinas, a matriz de adjacência escolhida para o trabalho será não-direcional.

A segunda propriedade refere-se aos pesos das arestas presentes em um grafo. Enquanto a primeira propriedade tratou da existência e direção das conexões entre os nós, agora busca-se entender a força dessa associação, ou seja, o peso atribuído às arestas. Na modelagem proposta neste trabalho, a força da associação entre os nós é baseada nas distâncias geográficas.

As arestas de um grafo podem ser classificadas em dois tipos: não-ponderadas ou ponderadas, o primeiro caso as arestas não possuem pesos, assim leva-se em consideração apenas se há uma conexão ou não. Já no segundo tipo, além de se considerar se existe ou não uma relação entre nós, também define-se o quão forte ela é.

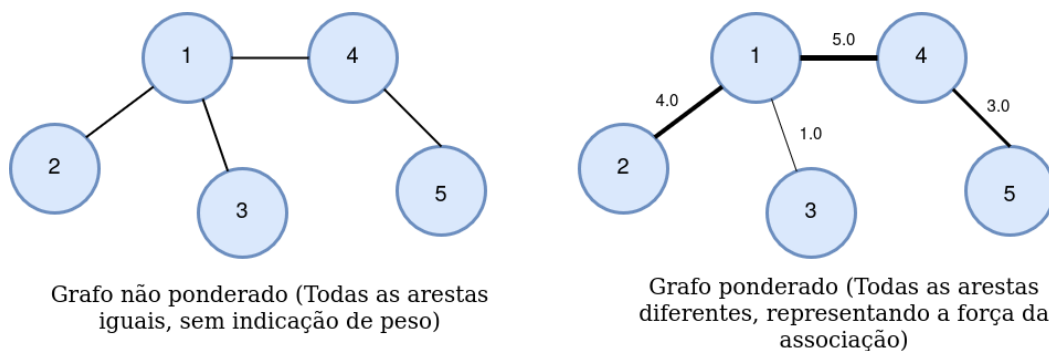


Figura 15 – Comparação entre grafos não ponderado e ponderado, evidenciando a ausência ou presença de pesos nas arestas. Figura de autoria própria.

A terceira propriedade é da conectividade dos nós, que podem estar conectados ou desconectados. Um nó é considerado conectado quando ao menos um outro nó compartilha uma aresta com ele, assim caso um nó esteja desconectado, ele não compartilha nenhuma aresta com outro nó do grafo, ou seja, está isolado da rede.

No grafo conectado, todos os nós possuem ao menos uma aresta que os liga a outros nós, permitindo a circulação de informações por toda a estrutura. No grafo não conectado, há nós isolados que não se conectam a nenhum outro, o que impede o fluxo de informações entre todas as partes do grafo.

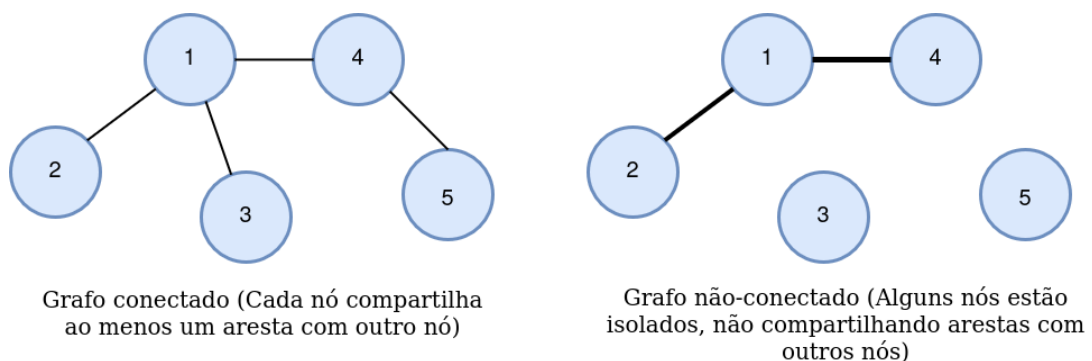


Figura 16 – Grafo conectado e não conectado. Figura de autoria própria.

Isso é importante no contexto de aprendizado em grafos, pois o isolamento de nós pode significar a ausência de compartilhamento de informações na rede, impactando a propagação de sinais e na efetividade desses modelos.

Considerando o aspecto espaço-temporal do trabalho, optou-se por construir um grafo $G = (V, E)$ não direcionado, não ponderado e conectado. Essas escolhas visam simplificar a análise e aumentar a interpretabilidade dos resultados.

4.3.2 Construção do grafo

A matriz de adjacência A foi construída a partir da distância geográfica entre as usinas, calculada pela fórmula de Haversine (4.10), gerando uma matriz quadrada de dimensão igual ao número de usinas, contendo as distâncias relativas entre todos os pares.

Essa matriz representa a força da associação entre as usinas com base na proximidade espacial, partindo da premissa de que usinas mais próximas tendem a compartilhar condições ambientais e operacionais similares.

A fórmula de Haversine calcula a distância entre dois pontos considerando a geometria da Terra, a partir das latitudes e longitudes (SINNOTT, 1984):

$$d = 2r \cdot \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta\phi}{2} \right) + \cos(\phi_1) \cos(\phi_2) \sin^2 \left(\frac{\Delta\lambda}{2} \right)} \right) \quad (4.10)$$

onde:

- d é a distância entre os dois pontos;
- r é o raio médio da Terra (≈ 6371 km);
- $\Delta\phi = \phi_2 - \phi_1$ é a diferença de latitude (em radianos);
- $\Delta\lambda = \lambda_2 - \lambda_1$ é a diferença de longitude (em radianos);
- ϕ_1, ϕ_2 são as latitudes dos pontos.

O resultado é uma matriz de distâncias D_H , quadrada, com dimensão igual ao número de usinas N , contendo as distâncias geodésicas entre todos os pares.

O uso das distâncias geográficas para estruturar o grafo é uma abordagem comum em estudos que utilizam GNNs para previsão de energia fotovoltaica, como exemplificado por (WOSCHITZ, 2023), que também emprega a distância de Haversine para modelar interações espaciais.

A partir da matriz D_H , define-se a estrutura do grafo aplicando o critério de k-vizinhos mais próximos (k-NN), técnica fundamental para identificar vizinhanças em espaços métricos (BISHOP, 2006, Capítulo 2).

Cada nó i é conectado aos seus k vizinhos mais próximos, com $k = \text{ceil}(0.1 \cdot (N - 1))$. A ordenação dos vizinhos é feita diretamente pelas distâncias Haversine, sem normalização nesta etapa.

Diferente de abordagens que utilizam pesos contínuos nas arestas, normalizando distâncias para definir características ponderadas (WOSCHITZ, 2023), aqui optou-se por manter a matriz de adjacência A binária. Essa decisão se baseia na premissa de que a topologia do k-NN já incorpora a informação de proximidade na estrutura da rede. Assim, A indica apenas a presença ou ausência da conexão, valorizando a influência da topologia.

Para garantir a simetria de A , necessária para grafos não direcionados, a conexão (i, j) é estabelecida ($A(i, j) = 1$) se a usina j está entre os k vizinhos mais próximos da usina i , ou se i está entre os k vizinhos de j . Caso contrário, $A(i, j) = 0$. Essa construção resulta em uma matriz simétrica e estática.

O processo de seleção k-NN e a simetrização são ilustrados na Figura 17.

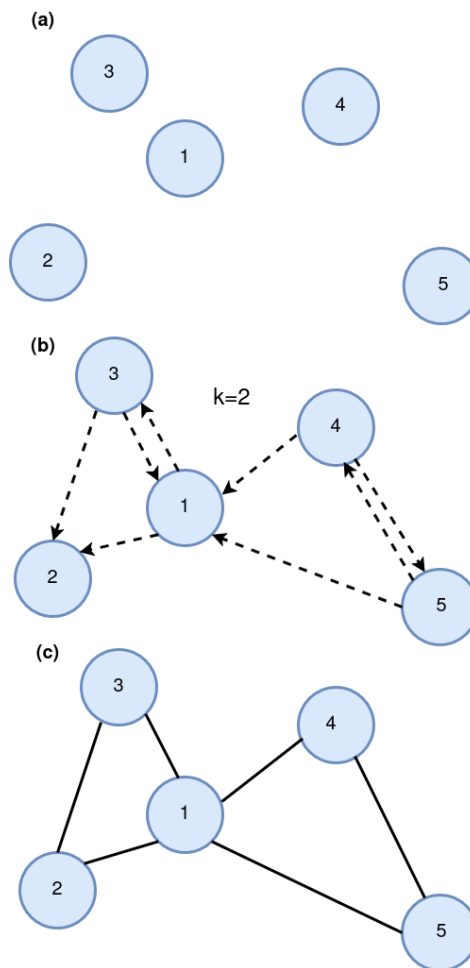


Figura 17 – Ilustração do processo de construção do grafo via k-NN com $k = 2$: (a) disposição inicial dos nós representando usinas; (b) seleção unilateral dos k -vizinhos mais próximos (setas tracejadas); (c) grafo final não direcionado, onde uma aresta (i, j) existe se i está entre os k vizinhos de j ou vice-versa. Figura de autoria própria.

Primeiro, adicionam-se auto-conexões para cada nó, gerando $\tilde{A} = A + I$, onde I é a matriz identidade. Isso permite que a informação do próprio nó seja considerada na agregação.

Depois, calcula-se a matriz diagonal de graus \tilde{D} , onde cada elemento $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ é o grau do nó i , incluindo sua auto-conexão. A normalização simétrica é então aplicada:

$$\hat{A} = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} \quad (4.11)$$

Essa normalização assegura que a agregação das informações dos vizinhos e do próprio nó seja equilibrada, evitando que nós com muitos vizinhos dominem o processo de propagação na GNN e contribuindo para a estabilidade do aprendizado.

Embora a topologia do grafo definido por A seja estática, os sinais nos nós apresentam variações temporais, capturadas pela passagem de mensagens da GNN. Isso permite aprender representações dinâmicas e identificar padrões espaço-temporais.

O critério de conexão — conectar cada nó aos seus k vizinhos mais próximos, com k cerca de 10% do número total de outras usinas — foi definido empiricamente, buscando um equilíbrio entre capturar interdependências espaciais e manter a rede parcimoniosa. Análises futuras podem investigar a sensibilidade dessa escolha e otimizar o grau médio da rede.

Após a definição da topologia do grafo por meio da matriz \tilde{A} (matriz de adjacência com auto-conexões), que estabelece as conexões espaciais entre as usinas, incorporam-se as informações dinâmicas que variam no tempo. As GNNs operam com a representação inicial de cada nó através de um *feature vector* (vetor de características ou atributos), que encapsula suas informações em um dado instante.

O objetivo do aprendizado em grafos é transformar esses *feature vectors* em *embeddings* (representações de alta dimensionalidade) que capturam tanto os atributos do nó quanto seu contexto estrutural na rede. Esses *embeddings* representam a informação aprendida do nó, enriquecida por suas interações com a vizinhança.

O processo pelo qual as GNNs geram esses *embeddings* é denominado passagem de mensagens, um paradigma central que envolve duas operações fundamentais: agregação e combinação. Em GNNs genéricas, a atualização da representação de um nó v para a camada k é comumente expressa pela combinação de suas características anteriores $h_v^{(k-1)}$ com a informação agregada de seus vizinhos. Essa agregação é realizada por meio de uma operação que utiliza explicitamente a matriz de adjacência normalizada, \hat{A} .

Para uma camada GNN típica, a atualização dos *embeddings* dos nós pode ser formalizada da seguinte forma:

$$H^{(k)} = \sigma(\hat{A}H^{(k-1)}W^{(k)}) \quad (4.12)$$

onde $H^{(k)}$ é a matriz de *embeddings* dos nós na camada k (com $H^{(0)}$ sendo a matriz inicial de *feature vectors* X), \hat{A} é a matriz de adjacência normalizada do grafo, $W^{(k)}$ é a matriz

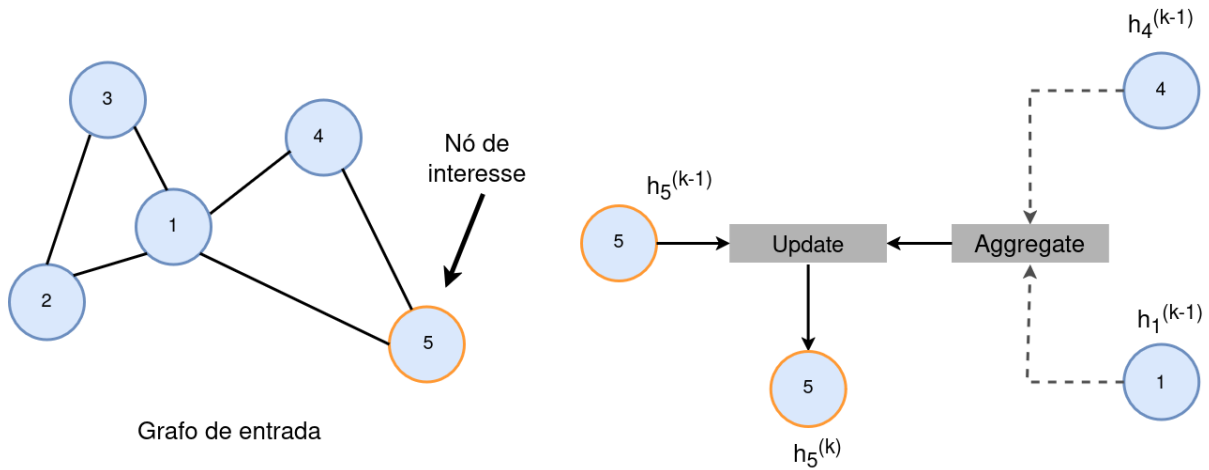


Figura 18 – O diagrama detalha as operações de agregação e atualização para um nó de interesse (nó 5).). Figura de autoria própria.

de pesos treináveis da camada k , e σ é uma função de ativação não linear (e.g., ReLU).

Nesta formulação, a operação de agregação é capturada pela multiplicação $\hat{A}H^{(k-1)}$, onde cada linha da matriz resultante representa a soma ponderada dos *embeddings* dos vizinhos (incluindo a auto-conexão) para cada nó.

A operação de combinação é realizada pela multiplicação pela matriz de pesos $W^{(k)}$ e aplicação da função de ativação σ , que transforma e refina essa informação agregada em novos embeddings. Este mecanismo iterativo permite que cada nó incorpore informações de sua vizinhança na construção de *embeddings* contextuais. Esse processo é ilustrado na Figura 18.

Iterando essas operações ao longo de k camadas, as GNNs capturam interações tanto locais quanto globais, permitindo que a informação flua por múltiplos hops no grafo, culminando na criação de embeddings contextualmente ricos para cada nó. Na figura, as representações da camada anterior ($h^{(k-1)}$) dos vizinhos (nós 1 e 4) são agregadas, e essa mensagem agregada é combinada com a representação do próprio nó 5 ($h_5^{(k-1)}$) para gerar sua nova representação na camada atual ($h_5^{(k)}$).

No contexto de aprendizado de máquina em grafos, a distinção entre aprendizado transdutivo e indutivo é fundamental (WU *et al.*, 2021; HAMILTON; YING; LESKOVEC, 2017). O aprendizado transdutivo foca na inferência sobre dados que já foram observados e são parte da estrutura do grafo de treinamento, enquanto o aprendizado indutivo capacita o modelo a generalizar para dados ou nós completamente novos, não presentes durante o treinamento (HAMILTON; YING; LESKOVEC, 2017). Para o problema deste trabalho essa distinção impacta a capacidade do modelo de lidar com a evolução da rede ou a adição de novas usinas (VELIČKOVIĆ *et al.*, 2018).

Neste trabalho, o foco principal é na previsão de séries temporais ao nível do nó, visando obter predições para todas as usinas (nós) do grafo em um horizonte futuro.

Para tanto, as séries temporais de geração e das variáveis exógenas de cada usina são transformadas em sequências fixas, utilizando janelas temporais de dados históricos como entrada para o modelo preditivo.

4.4 GAT-LSTM

A extração de padrões locais dentro dessas séries temporais é feita através de convoluções temporais, que atuam sobre as sequências para identificar características relevantes em diferentes dinâmicas. Assim, a incorporação de mecanismos de memória, como o LSTM, é utilizada para capturar dependências de curto e longo prazo nas séries temporais, garantindo que essas influências sejam capturadas na previsão.

Dessa forma, a escolha da arquitetura GAT-LSTM justifica-se pela sua capacidade de integrar o processamento espacial das relações de grafo com a modelagem temporal de dependências de curto e longo prazo.

Dentro das GNNs, a Graph Attention Network (GAT), representa uma arquitetura que se destaca por incorporar o mecanismo de atenção no processo de agregação de mensagens entre os nós.

Para um par de nós conectados (i, j) , a GAT calcula um coeficiente de atenção e_{ij} que calcula a relevância do nó vizinho j para a representação do nó i . Este coeficiente é derivado da concatenação das representações transformadas de ambos os nós, seguida por uma transformação linear e a aplicação de uma função de ativação não linear, como a LeakyReLU. O coeficiente e_{ij} pode ser expresso como:

$$e_{ij} = \text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}\mathbf{h}_i, ||, \mathbf{W}\mathbf{h}_j]) \quad (4.13)$$

onde h_i e h_j são os embeddings dos nós i e j , W é uma matriz de pesos compartilhada aplicada a todas as arestas, a é um vetor de pesos de atenção aprendível, e denota a operação de concatenação.

Para tornar esses coeficientes comparáveis e garantir que a soma das importâncias dos vizinhos seja unitária, os valores e_{ij} são normalizados por meio de uma função Softmax sobre todos os vizinhos do nó i . Isso gera os pesos de atenção α_{ij} :

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})} \quad (4.14)$$

onde \mathcal{N}_i representa o conjunto de vizinhos do nó i . Esses pesos α_{ij} indicam diretamente o quanto o nó j contribui para a atualização da representação do nó i . Este mecanismo permite que a GAT atribua uma importância diferenciada a cada vizinho, mesmo que todos estejam conectados, focando nos nós mais relevantes para a tarefa em questão.

Após o cálculo dos pesos de atenção normalizados, a nova representação do nó i para a próxima camada é obtida através de uma soma ponderada das características de seus vizinhos e, opcionalmente, de sua própria característica, utilizando os pesos de

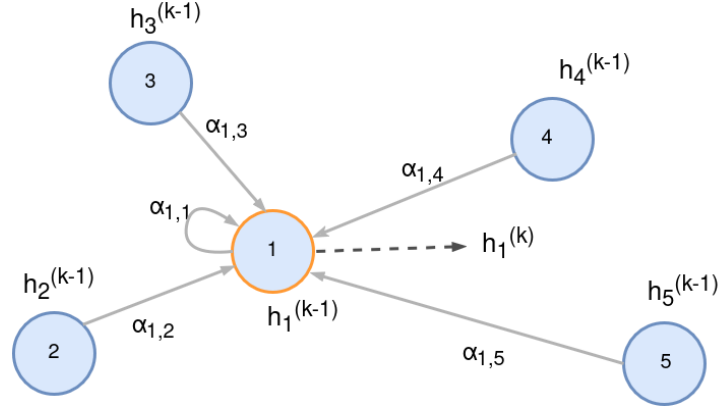


Figura 19 – O diagrama detalha como a representação do nó central (nó 1) é atualizada a partir de suas próprias características ($h_1^{(k-1)}$) e das características de seus vizinhos ($h_j^{(k-1)}$, para $j \in \{2, 3, 4, 5\}$). Figura de autoria própria.

atenção α_{ij} :

$$\mathbf{h}'_i = \sigma \left(\sum_{j \in \mathcal{N}i \cup i} \alpha_{ij} \mathbf{W} \mathbf{h}_j \right) \quad (4.15)$$

Essa operação permite que cada nó aprenda a focar nos vizinhos mais relevantes, tornando o processo de agregação mais flexível e adaptativo à estrutura local da rede.

O mecanismo de atenção pode ser empregado com a utilização de múltiplas cabeças de atenção, onde várias instâncias independentes do processo de atenção são executadas em paralelo. Cada cabeça opera com um conjunto diferente de parâmetros (matrizes de pesos \mathbf{W} e vetores \mathbf{a}), calculando sua própria atenção e agregação de forma independente.

Isso permite que o modelo capture diferentes tipos de relações ou importâncias entre os vizinhos. Os resultados dessas múltiplas cabeças de atenção são então concatenados (para formar uma representação mais rica e de maior dimensão) ou somados (para manter a dimensão), formando a representação final do nó para a próxima camada.

O princípio de funcionamento da GAT, incluindo o cálculo e a agregação dos coeficientes de atenção sob o mecanismo de multi-atenção, O GAT calcula um peso de atenção ($\alpha_{1,j}$) para cada vizinho, que é então utilizado para ponderar a contribuição dos vizinhos na agregação. O resultado dessa agregação ponderada é combinado com a representação do próprio nó para gerar a nova representação do nó 1 na camada k ($h_1^{(k)}$), visualizado na Figura 19. O funcionamento da GAT-LSTM ocorre da seguinte forma: a cada instante de tempo t da sequência de entrada, a camada GAT processa os atributos de todos os nós para aquele momento, utilizando a matriz de adjacência do grafo. O resultado dessa camada é um novo conjunto de *embeddings* para cada nó, $h_v^{(t, \text{GAT})}$, que encapsulam as informações espaciais ponderadas para o tempo t .

Essa sequência de *embeddings* espacialmente informados, $[h_v^{(1, \text{GAT})}, h_v^{(2, \text{GAT})}, \dots, h_v^{(T, \text{GAT})}]$, é então alimentada sequencialmente a uma camada LSTM. Essa, por sua vez, é responsável por capturar as dependências temporais de longo prazo entre esses *embeddings* espacialmente enriquecidos, produzindo um estado final que representa a dinâmica espaço-temporal

integrada do nó ao longo da janela de tempo. A sequência desses embeddings é então alimentada a uma camada LSTM, que modela as dependências temporais de longo prazo, obtendo a previsão da geração (Y_t).

Os hiperparâmetros-chave incluem o número de camadas GAT e LSTM, a dimensionalidade dos *embeddings*, o número de cabeças de atenção (para o GAT), e a taxa de aprendizado, que são ajustados para otimizar o desempenho do modelo. Uma ilustração detalhada desse fluxo de informações e da arquitetura do GAT-LSTM pode ser vista na Figura 20.

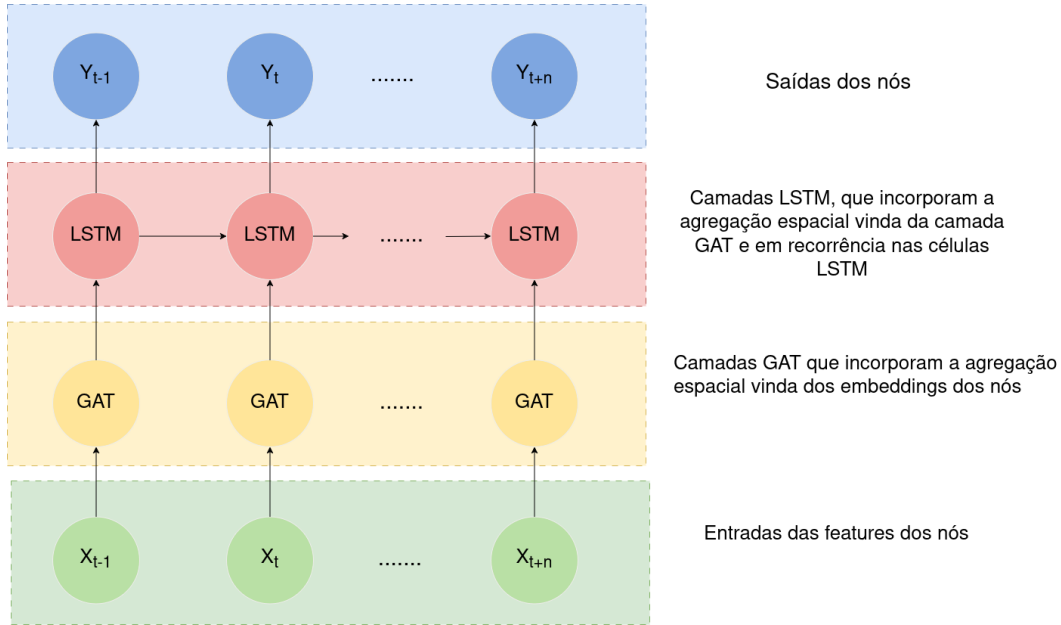


Figura 20 – A cada instante de tempo, as características iniciais (X_t) de todas as usinas são processadas por uma camada GAT para capturar as dependências espaciais, gerando embeddings enriquecidos. Figura de autoria própria.

4.5 GCN1D

A arquitetura Graph Convolutional Network 1-Dimension(GCN-1D) utiliza convoluções baseadas em polinômios de Chebyshev com modificações - feitas para adequar o operador à séries temporais, assim, ao invés de aplicar uma transformação linear entre o vetor de entrada e os pesos da transformação, aplica-se uma convolução de uma dimensão no vetor de entrada.

No modelo original proposto por (KIPF; WELLING, 2017) as convoluções em grafo, que aproximam o operador de filtragem espectral em torno do Laplaciano do grafo. Nessa formulação, a convolução em grafos(GraphConv) é definido por:

$$y = g_\theta * g_X = \sum_{k=0}^K P_k(\tilde{L}) \text{CNN}_{1D}(X) \quad (4.16)$$

Onde $\theta_k \in R^{CxO}$, C canais de entrada e O canais de saída são os filtros treináveis da camada de convolução. E $P_k(\tilde{L})$ são os polinômios de Chebyshev definidos como: $P_0 = 1; P_1 = \tilde{L}; P_k = 2\tilde{L}P_{k-1} - P_{k-2}$.

Os pesos θ_k carregam as informações dos coeficientes de Chebyshev para o operador de filtro inicial da arquitetura, assim a convolução em grafo é aplicada para o vetor de entrada $X \in R^{NxOxD}$ e retorna $y \in R^{NxOxW}$, sendo D e W passos anteriores e futuros. $P_k(\tilde{L}) \in R^{NxN}$ são os polinômios de Chebyshev de ordem k avaliados no laplaciano escalado $\tilde{L} = 2L/\Lambda_{max} - I_n$.

Entretanto, na implementação prática deste trabalho, o operador baseado em Chebyshev foi substituído por uma camada GCNConv da biblioteca PyTorch Geometric. Essa camada corresponde a uma simplificação da formulação espectral, equivalente ao caso K=1, em que a filtragem é feita diretamente pela matriz de adjacência normalizada:

$$\hat{A} = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} \quad (4.11)$$

$$H^{(k)} = \sigma(\hat{A}H^{(k-1)}W^{(k)}) \quad (4.12)$$

Dessa forma, a GCNConv realiza uma agregação de primeira ordem, propagando informações apenas entre vizinhos diretos. Isso reduz o custo computacional e simplifica o treinamento, mas também altera a natureza do filtro: o operador deixa de ser um polinômio de alta ordem do laplaciano e passa a atuar como um filtro passa-baixas suave no domínio espectral.

Essa convolução é aplicada com polinômios de ordem até K-1 do laplaciano escalado e é espacialmente localizado, dependendo somente de nós com até K-hops de distância do nó de interesse.

Tendo em vista que o problema de previsão de incertezas do trabalho trata de séries temporais, as convoluções utilizadas são causais, de modo a não violar a causalidade temporal necessária para prever passos futuros, i.e., a previsão $p(x_{t+1}|x_1, \dots, x_t)$ feita pelo modelo num passo de tempo τ não pode depender de nenhuma variável de passos futuros x_{t+1}, \dots, x_t . Para os dados unidimensionais, como os da geração solar, a convolução causal pode ser implementada transladando os dados de entrada - basta adicionar um pad na esquerda da sequência de entrada e isso assegura que a camada de convolução não recorre de vazamento de dados para realizar as suas predições.

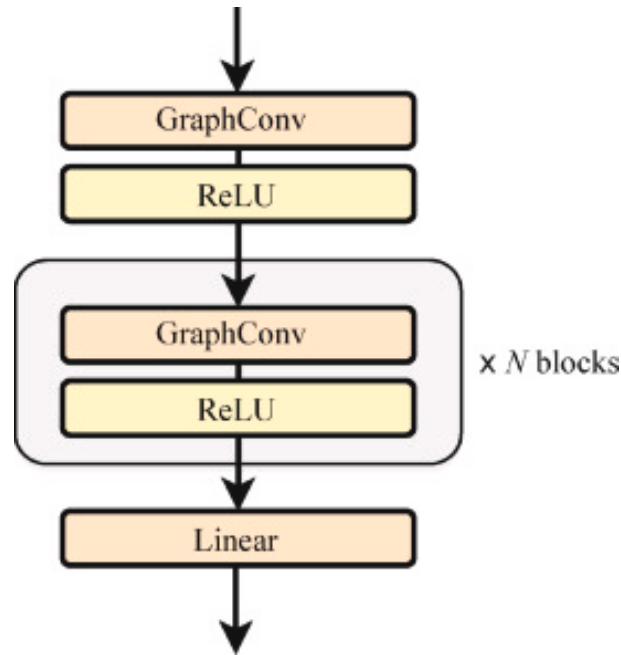


Figura 21 – Arquitetura do modelo GCN-1D. As entradas são processadas por uma camada GCNConv para agregação espacial, seguida por uma convolução 1D causal que extrai padrões temporais para gerar a previsão final

4.6 Limitações na modelagem das GNNs

A modelagem com GNNs envolve decisões de projeto que resultam em limitações e custo-benefício fundamentais. A primeira reside na dependência da topologia do grafo, que atua como um viés indutivo. Conexões mal especificadas podem introduzir ruído (no caso de arestas irrelevantes) ou impedir a propagação de informação (no caso de arestas ausentes), prejudicando o aprendizado.

Neste contexto, a escolha por uma topologia estática representa um equilíbrio deliberado entre fidelidade do modelo e viabilidade computacional.

Para a rede analisada, composta por 47 usinas geograficamente distantes dentro do cinturão solar brasileiro, a hipótese de interdependências fixas é uma aproximação razoável, pois as correlações são em sua maioria atrelada a fenômenos meteorológicos de macroescala. Essas correlações são em sua maioria decorrente da irradiância similar na região do cinturão solar brasileiro e das sazonalidades horária e diária.

A implementação de uma matriz de adjacência dinâmica, embora teoricamente superior para capturar variações funcionais, implicaria um aumento proibitivo no custo computacional em todas as etapas do projeto, desde a engenharia de features até o treinamento do modelo.

Outra limitação inerente a muitas arquiteturas, como a GCN, é a tendência à suavização, conforme analisado, operadores de agregação simples atuam como filtros passa-baixas, fazendo com que as representações dos nós se tornem mais similares a cada camada de propagação.

Embora o problema seja mais acentuado em redes profundas, essa característica pode atenuar sinais locais de alta frequência, como rampas de geração de energia solar, mesmo em arquiteturas rasas, representando um equilíbrio entre a captura de padrões regionais e a fidelidade a eventos locais.

4.7 Previsão conforme e incertezas

Os capítulos anteriores detalharam a construção de modelo para estimativas pontuais \tilde{y} ou quantílicas q_α . Porém, o objetivo da quantificação de incerteza é a construção de regiões de predição - intervalos $R(x)$ que devem conter o valor real y com uma probabilidade pré-definida.

Esta seção formaliza a passagem de estimativas de modelo para regiões de predição com garantias estatísticas, baseando-se no framework teórico da Previsão Conforme (IZBICKI, 2025).

Uma maneira de fazer isso é utilizar os quantis estimados pelo modelo treinado a partir da função de pinball (Seção 4.8), formando uma região de plug-in:

$$R_{\text{plug-in}}(x) = [\hat{q}_{\alpha/2}(x), \hat{q}_{1-\alpha/2}(x)] \quad (4.17)$$

Conforme discutido por (IZBICKI, 2025, Seção 5.2), embora essa região de plug-in capture a incerteza aleatória, ela não leva em conta a incerteza epistêmica na estimação dos quantis \hat{q} . Sendo esses quantis apenas estimativas a partir dos dados disponíveis, a região conforme resultante frequentemente falha em atingir a cobertura nominal desejada. Por exemplo, um intervalo de 90% pode conter a observação real em apenas 85% das vezes, tornando as garantias de incertezas não confiáveis.

Para superar a limitação dos métodos plug-in, a Previsão Conforme oferece um framework agnóstico ao modelo, projetado para produzir regiões de predição com cobertura marginal garantida (IZBICKI, 2025; VOVK; GAMMERMAN; SHAFER, 2005). A garantia é marginal por se aplicar em média sobre a distribuição de dados, e é obtida através de um processo de calibração. A metodologia *split conformal*, a mais comum, segue os seguintes passos:

- **Partição dos Dados:** O conjunto de dados é dividido em um conjunto de treino e um de calibração, disjuntos. O modelo preditivo (XGBoost ou GNN) é treinado exclusivamente no conjunto de treino.
- **Cálculo dos Escores de Não-Conformidade:** O modelo treinado é então usado para prever os resultados no conjunto de calibração. Para cada ponto i neste conjunto, calcula-se um escore de não-conformidade, que mede o quão atípica ou errada foi a predição. Para problemas de regressão, o escore mais simples é o erro absoluto residual:

$$s_i = |y_i - \hat{y}_i| \quad (4.18)$$

Esses escores formam uma distribuição empírica do erro que o modelo comete em dados que não viu durante o treino.

- **Calibração do Limiar:** Para um nível de erro α desejado, por exemplo, $\alpha = 0.1$ para 90% de cobertura), calcula-se o limiar $q = (1 - \alpha)$ quantil dos escores de não-conformidade. Este limiar representa, heurísticamente, o pior erro que o modelo cometeu em 90% dos casos no conjunto de calibração.
- **Construção da Região Conforme:** Para uma nova predição \hat{y}_{teste} , o intervalo de predição final é construído simetricamente:

$$R(\hat{y}_{teste}) = [\hat{y}_{teste} - q, \hat{y}_{teste} + q] \quad (4.19)$$

A teoria da Previsão Conforme garante que, sob a premissa de que os dados são trocáveis, o PICP resultante será no mínimo de $1 - \alpha$. Essa garantia é forte por ser independente da complexidade do modelo ou da distribuição dos dados.

A garantia teórica da Previsão Conforme depende da premissa de trocabilidade, que é violada por séries temporais devido à autocorrelação e não-estacionariedade. A distribuição dos erros de um modelo de previsão solar, por exemplo, não é constante ao longo do dia.

A aplicação do método previsão conforme padrão a dados de séries temporais, como feito neste trabalho, resulta em intervalos homoscedásticos, de largura constante $2q$, que não se adaptam à volatilidade local.

Consequentemente, os intervalos podem ser excessivamente largos em períodos estáveis, como no início da manhã, e demasiadamente estreitos durante períodos de alta variabilidade (rampas de geração), comprometendo a cobertura condicional.

Embora métodos mais avançados existam para lidar com dados não trocáveis, eles fogem ao escopo desta análise. Portanto, a Previsão Conforme será utilizada aqui como uma ferramenta para gerar uma referência inicial de incerteza com cobertura marginal, com a ressalva de que a obtenção de garantias condicionais em séries temporais constitui uma importante direção para trabalhos futuros.

4.8 Função de Perda

Para que os modelos de aprendizado de máquina possam gerar previsões de incerteza, em vez de apenas estimativas pontuais, é necessário um critério de otimização adequado. Esta seção apresenta a função de perda quantílica, o mecanismo que permite treinar os modelos para estimar diretamente os quantis da distribuição condicional da geração.

O processo de treinamento de um modelo consiste em minimizar uma função de perda, que quantifica o erro das previsões.

A função de perda adotada para os modelos foi a perda quantílica, também conhecida como perda de pinball. Ela permite que o modelo estime um quantil específico da distribuição de probabilidade da variável alvo, em vez de apenas a média. Sua fórmula é dada por:

$$L_{\tau}(y, \hat{y}) = \begin{cases} \tau(y - \hat{y}) & \text{se } y \geq \hat{y} \quad (\text{subestimação}) \\ (1 - \tau)(\hat{y} - y) & \text{se } y < \hat{y} \quad (\text{superestimação}) \end{cases} \quad (4.20)$$

Onde:

- y é o valor real.
- \hat{y} é o valor previsto pelo modelo para um determinado quantil.
- τ é o quantil alvo, um valor no intervalo $(0, 1)$.

A característica principal desta função é a sua assimetria, que permite direcionar o modelo para estimar diferentes pontos da distribuição condicional da variável alvo. Os modelos foram treinados para prever três quantis distintos: 0.1, 0.5 e 0.9, cada um com um objetivo específico:

- $\tau = 0.1$

Para este quantil, a função de perda penaliza mais a superestimação (quando $\hat{y} > y$) com um peso de $1 - \tau = 0.9$, enquanto a subestimação ($y \geq \hat{y}$) recebe um peso menor, de $\tau = 0.1$. Para minimizar essa perda assimétrica, o modelo aprende a gerar previsões conservadoras, com o objetivo de que o valor real seja superior à previsão em 90% das vezes, estabelecendo uma fronteira inferior para as estimativas.

- $\tau = 0.5$

Quando o quantil alvo é a mediana, os pesos para subestimação e superestimação são iguais: $\tau = 0.5$ e $1 - \tau = 0.5$.

- $\tau = 0.9$

De forma oposta ao quantil 0.1, aqui a função de perda penaliza mais a subestimação ($y \geq \hat{y}$) com um peso de $\tau = 0.9$. A superestimação ($\hat{y} > y$) recebe um peso de apenas $1 - \tau = 0.1$.

Logo, o modelo aprende a gerar previsões otimistas, estabelecendo uma fronteira superior onde se espera que o valor real seja inferior à previsão em 90% das vezes.

Assim, o uso da perda quantílica para os quantis 0.1 e 0.9 viabiliza a construção de um intervalo de predição de 80%. A qualidade e a calibração desses intervalos, bem como

a acurácia da estimativa mediana (quantil 0.5), serão avaliadas por meio das métricas descritas a seguir.

4.9 Métricas de Avaliação

Uma vez que o objetivo do trabalho abrange tanto a acurácia da previsão pontual quanto a qualidade dos intervalos de incerteza, é necessário um conjunto de métricas que avalie ambas as dimensões.

Esta seção define os critérios que serão utilizados para comparar o desempenho dos modelos, focando na calibração e precisão dos intervalos, bem como no erro das estimativas pontuais.

- **PICP (Prediction Interval Coverage Probability):** Mede a proporção de observações reais que se encontram dentro dos limites do intervalo de predição. Para um intervalo de confiança de $(1 - \alpha)100\%$, um PICP idealmente se aproximaria de $1 - \alpha$, avaliando a calibração e a confiabilidade da estimativa de incerteza.

$$\text{PICP} = \frac{1}{n} \sum_{i=1}^n c_i, \quad \text{onde } c_i = \begin{cases} 1, & \text{se } \hat{y}_i^L \leq y_i \leq \hat{y}_i^U \\ 0, & \text{caso contrário} \end{cases} \quad (4.21)$$

Nesta equação, \hat{y}_i^L e \hat{y}_i^U representam, respectivamente, os limites inferior e superior do intervalo de predição para a i -ésima observação.

- **PINAW (Prediction Interval Normalized Average Width):** Avalia a largura média dos intervalos de predição, normalizada pela amplitude da variável alvo. Esta métrica quantifica a precisão da incerteza; intervalos mais estreitos (PINAW menor) são desejáveis, desde que mantenham um nível de PICP adequado.

$$\text{PINAW} = \frac{1}{nR} \sum_{i=1}^n (\hat{y}_i^U - \hat{y}_i^L) \quad (4.22)$$

Onde R é a amplitude dos dados observados ($R = \max(y_i) - \min(y_i)$), garantindo que a métrica seja adimensional e comparável entre diferentes problemas.

- **RMSE (Root Mean Squared Error):** É a raiz quadrada do MSE. Sua vantagem é que a métrica fica na mesma unidade da variável alvo, tornando o erro mais interpretável.

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.23)$$

- **MAE (Mean Absolute Error):** Mede a média dos erros absolutos. É menos sensível a *outliers* que o RMSE, pois não eleva os erros ao quadrado.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.24)$$

- **R^2 (Coeficiente de Determinação):** Indica a proporção da variância dos dados que é explicada pelo modelo. Um valor próximo de 1 significa um bom ajuste.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.25)$$

Com estas métricas definidas, possui-se o ferramental necessário para realizar uma avaliação quantitativa e comparativa dos modelos propostos, o que será detalhado no capítulo de Resultados e Discussão.

4.10 Configuração Experimental

O conjunto de dados foi dividido seguindo o padrão 60% para treino, 20% para validação e 20% para teste. Os modelos foram treinados para realizar previsões multi-passo, sendo avaliados nos horizontes de H+1, H+3, H+6, H+12 e H+24.

Para o XGBoost, a previsão foi implementada para treinar um modelo por horizonte e quantil. O tuning de hiperparâmetros foi feito com a biblioteca Optuna, executando 100 trials com a técnica de *pruning* para descartar conjuntos de hiperparâmetros pouco promissores.

Para os modelos de aprendizado profundo (GAT-LSTM e GCN-1D), a função de perda pinball foi ponderada entre os horizontes de previsão. Adotaram-se os pesos [1.0, 1.25, 1.5, 2.0, 2.5] para os horizontes H+1, H+3, H+6, H+12 e H+24, respectivamente, atribuindo maior importância aos erros em previsões de longo prazo, que são inerentemente mais difíceis.

Para o GAT-LSTM, os hiperparâmetros da rede, detalhados na Tabela 10, foram definidos com base em testes preliminares. O treinamento utilizou o otimizador Adam com taxa de aprendizado de 1e-3 e parada antecipada com paciência de 15 épocas para monitorar a perda na validação e prevenir *overfitting*.

Para o GCN-1D, a arquitetura foi configurada com uma camada GCNConv de 64 canais, seguida por uma convolução 1D causal com 128 canais de saída e um kernel de tamanho 3. Os seus hiperparâmetros se encontram na tabela 11. Assim como no GAT-LSTM, utilizou-se o otimizador Adam com taxa de aprendizado de 1e-3 e parada antecipada com paciência de 15 épocas.

O ambiente computacional foi selecionado para otimizar o tempo de execução:

- O treinamento do XGBoost foi realizado em CPU.
- O treinamento do GAT-LSTM e GCN-1D, mais intensivos computacionalmente, foram acelerados em GPU utilizando a plataforma Google Colab.

Tabela 10 – Hiperparâmetros utilizados no modelo GAT-LSTM

Hiperparâmetro	Valor
Canais ocultos GAT	64
Cabeças de atenção	8
Dropout GAT	0.3
Tamanho da camada LSTM	128
Número de camadas LSTM	1
Dropout LSTM	0.0
Learning rate	0.001
Número de épocas	300
Paciência	15

Tabela 11 – Hiperparâmetros utilizados no modelo GCN-1D.

Hiperparâmetro	Valor
Canais ocultos GCN	64
Canais de saída Conv1D	128
Tamanho do kernel Conv1D	3
Taxa de aprendizado	0.001
Número de épocas	100
Paciência	15

Partindo das lacunas identificadas na Análise Exploratória de Dados, principalmente a persistência de autocorrelação espacial nos resíduos de modelos temporais clássicos, este capítulo construiu o arcabouço metodológico para investigar a hipótese central do trabalho.

Foi estabelecida uma comparação sistemática entre a abordagem de modelo local, representada pelo XGBoost, e a de modelos globais, materializada pelas arquiteturas GNN GAT-LSTM e GCN-1D. Detalhou-se o processo de construção do grafo, cuja topologia estática baseada em proximidade geográfica foi justificada como um equilíbrio deliberado entre a captura de correlações de macroescala e a viabilidade computacional.

Para atender ao objetivo de quantificar incertezas, a metodologia integrou a função de perda quantílica como meio de treinamento para a estimação de intervalos. Adicionalmente, foi introduzida a Previsão Conforme como um framework para a calibração desses intervalos, com uma análise crítica de sua aplicabilidade em séries temporais não trocáveis.

O desempenho dessas abordagens será avaliado por um conjunto duplo de métricas, capazes de mensurar tanto a acurácia da previsão pontual (RMSE, MAE) quanto a qualidade dos intervalos de incerteza (PICP, PINAW).

Essa metodologia, portanto, estabelece as bases para a avaliação empírica das hipóteses do trabalho, cujos resultados são apresentados e discutidos a seguir.

5 RESULTADOS E DISCUSSÃO

Neste capítulo, são apresentados e discutidos os resultados dos modelos de aprendizado de máquina propostos, com o objetivo de validar empiricamente a hipótese central do trabalho: a de que a modelagem explícita das interdependências espaciais, via GNNs e GCNs, pode oferecer ganhos de previsibilidade e uma melhor quantificação de incertezas em comparação a modelos puramente locais.

Conforme estabelecido na Análise Exploratória de Dados (Capítulo 3), a insuficiência dos modelos clássicos (SARIMA) e a presença de autocorrelação residual (Tabela 6) já indicavam a necessidade de arquiteturas mais complexas, capazes de lidar com a natureza espacial do problema.

As seções a seguir, portanto, avaliam a performance dos modelos locais (XGBoost) versus a abordagem de portfólio dos modelos globais (GNNs).

5.1 Desempenho geral e comparativo dos modelos

Uma análise agregada dos resultados, consolidada nas Tabelas 16 a 21 (Apêndice), oferece uma visão panorâmica do desempenho geral dos modelos GNN no conjunto de teste completo. Observa-se que o GCN-1D apresenta, de forma geral, RMSE consistentemente menor e R^2 superior ao GAT-LSTM em praticamente todos os horizontes e quantis. Este resultado sugere que, para a topologia de rede estudada, a arquitetura convolucional mais simples foi mais eficaz em generalizar os padrões espaço-temporais.

A análise dos resultados gráficos será feita para cada usina-arquétipo escolhida, refletindo sobre como cada modelo se comportou no horizonte em questão.

No horizonte $H+1$ (Figura 22), todos os modelos demonstraram alta capacidade de aderência à dinâmica de geração. Para o Conjunto Lapa (Hub), é notável a precisão da mediana do XGBoost. No entanto, no período de 29-30/05: enquanto os modelos previram uma geração amena, consistente com o comportamento esperado para as condições climáticas, os valores realizados foram nulos.

Este padrão sugere a ocorrência de *curtailment*, uma redução deliberada da geração de energia por ordem do operador do sistema, que os modelos em grafos, influenciados pela produção dos vizinhos, não conseguiram antecipar. O XGBoost, focado no histórico local, também foi induzido ao erro, mas a discrepância levanta uma questão fundamental sobre a sensibilidade dos modelos a fatores exógenos não-físicos.

Para o Conjunto Castilho (Comunidade), o XGBoost novamente se destaca pela precisão pontual, com um RMSE de 14.65 MW e R^2 de 0.68 (Tabela 12). Os modelos em grafos, embora visualmente razoáveis, apresentaram intervalos de incerteza mais largos. Contudo, o GCN-1D obteve um R^2 excepcional de 0.92, indicando que, para este arquétipo, a informação espacial dos vizinhos foi valiosa para explicar a variância da

geração, superando o modelo local neste quesito.

No Conjunto Floresta (Nó Isolado), todos os modelos apresentaram aderência satisfatória. O GCN-1D (RMSE de 7.77 MW; R^2 de 0.78) e o XGBoost (RMSE de 7.80 MW; R^2 de 0.29) foram precisos. A superioridade do R^2 do GCN-1D sugere que, mesmo para um nó isolado, a informação contextual dos poucos vizinhos melhora a capacidade de generalização do modelo em curtíssimo prazo.

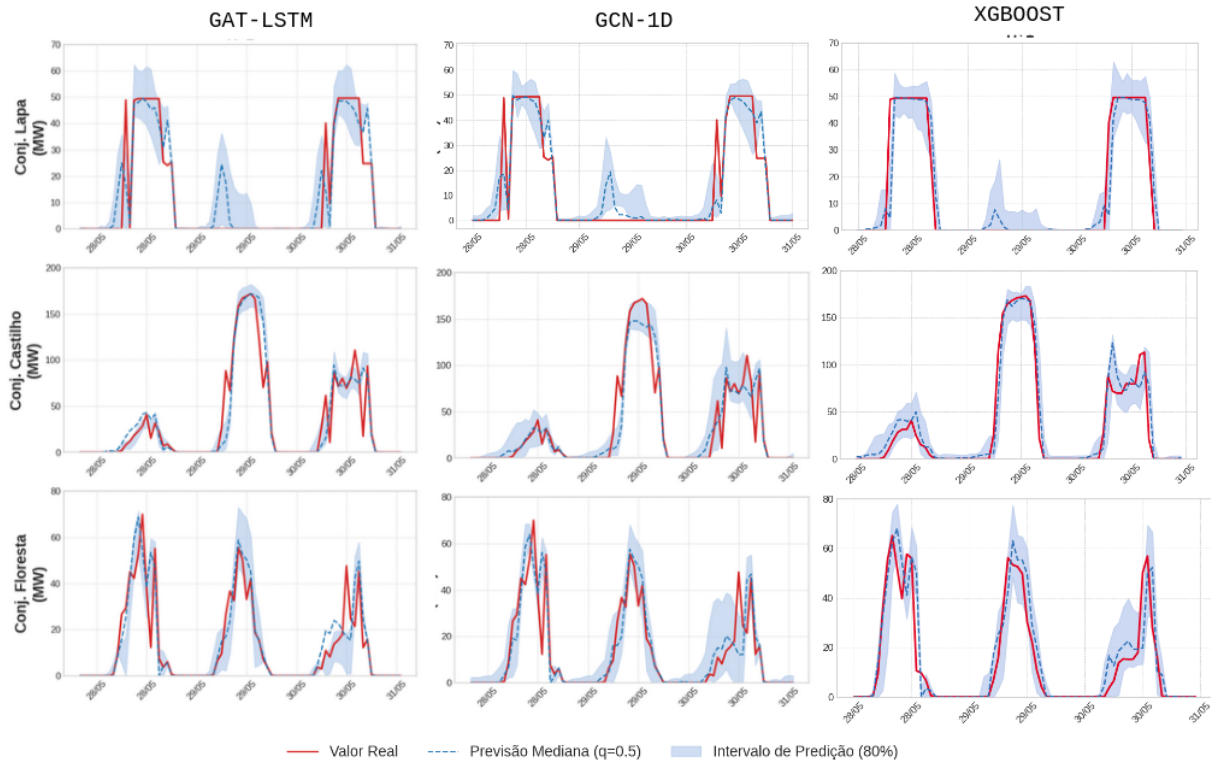


Figura 22 – Comparativo de desempenho dos modelos GAT-LSTM, GCN-1D e XGBoost para o horizonte de previsão H+1.

À medida que o horizonte de previsão se estende para H+3 (Figura 23) e H+6 (Figura 24), observa-se uma degradação clara na performance dos modelos GNN. Para o Conjunto Lapa, a capacidade de modelar o *curtailment* se perde, e a performance se deteriora drasticamente. A Tabela 12 quantifica este colapso: no H+3, o R^2 do GAT-LSTM torna-se negativo (-0.33), indicando que a previsão do modelo é pior do que uma simples média dos dados. O XGBoost, em contraste, mantém sua consistência, com um RMSE estável (7.70 MW) e um R^2 positivo, ainda que modesto (0.20).

Para o Conjunto Castilho, o XGBoost se firma como o modelo mais robusto, mantendo um RMSE em torno de 19 MW e R^2 de 0.46 em H+6. Os modelos em grafos, embora ainda capturem a morfologia geral do sinal, perdem precisão nos picos e vales, resultando em erros significativamente maiores (RMSE de 37.38 MW para GCN-1D em H+6).

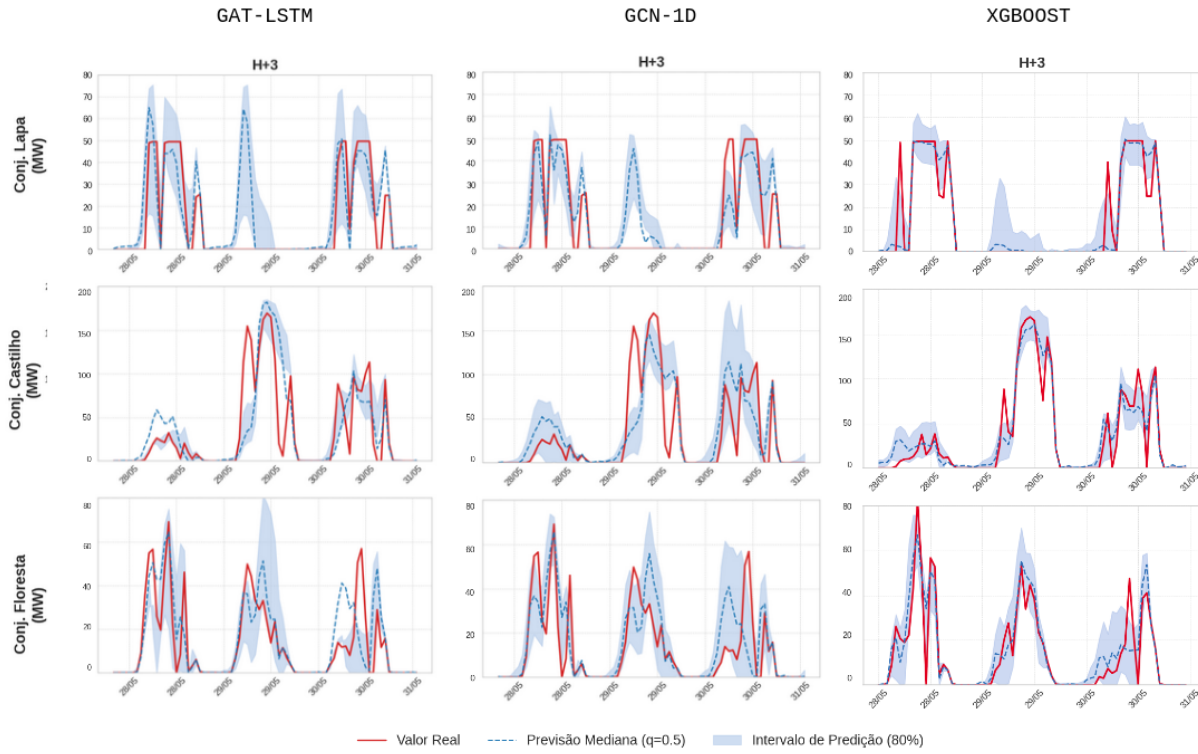


Figura 23 – Comparativo de desempenho dos modelos GAT-LSTM, GCN-1D e XGBoost para o horizonte de previsão H+3.

Nos horizontes mais longos (Figuras 25 e 26), a superioridade do modelo local (XGBoost) torna-se incontestável em termos de robustez. Para o Conjunto Lapa (Hub), os modelos GNN falham completamente em generalizar, com valores de R^2 consistentemente negativos e massivos (e.g., -2.33 para GAT-LSTM em H+12), tornando-os inutilizáveis para previsões operacionais. O XGBoost, por outro lado, mantém um desempenho estável, com RMSE em torno de 7-8 MW para todos os horizontes, um feito notável que evidencia sua capacidade de aprender padrões temporais robustos a partir de dados locais.

Para os arquétipos Castilho e Floresta, observa-se o mesmo padrão: o XGBoost mantém um erro (RMSE) relativamente constante, enquanto o erro dos modelos GNN escala significativamente com o horizonte. É digno de nota que, para o Conjunto Castilho, o GCN-1D ainda consegue explicar uma porção da variância em H+24 (R^2 de 0.66), um desempenho superior ao do XGBoost (R^2 de 0.62), apesar de um erro absoluto maior. Isso sugere que, mesmo em longo prazo, a informação espacial ajuda a capturar a tendência, mas falha em acertar a magnitude.

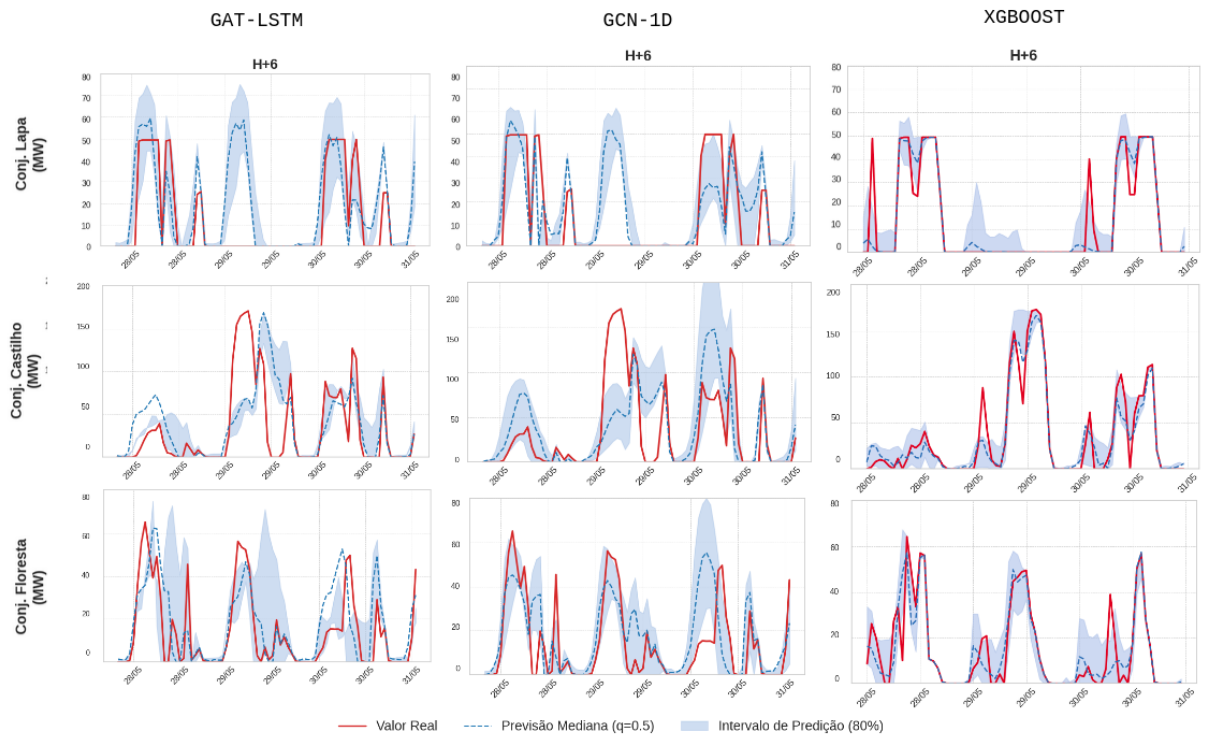


Figura 24 – Comparativo de desempenho dos modelos GAT-LSTM, GCN-1D e XGBoost para o horizonte de previsão H+6.

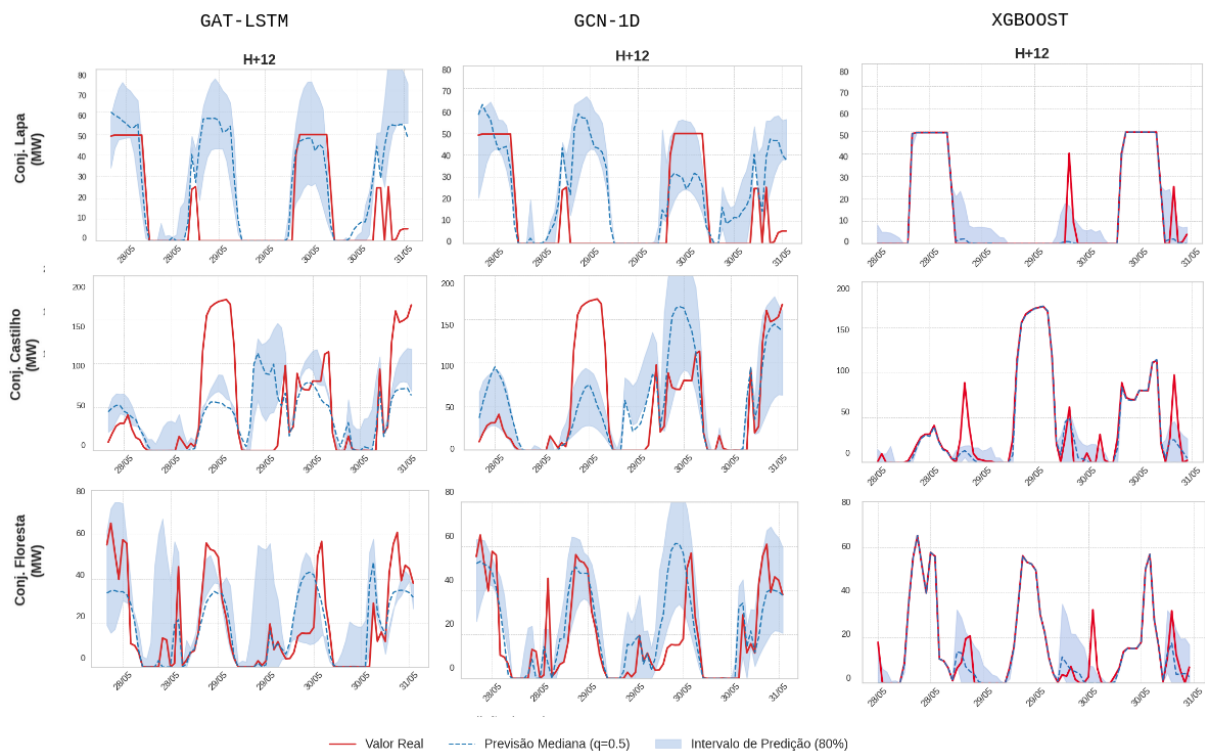


Figura 25 – Comparativo de desempenho dos modelos GAT-LSTM, GCN-1D e XGBoost para o horizonte de previsão H+12.

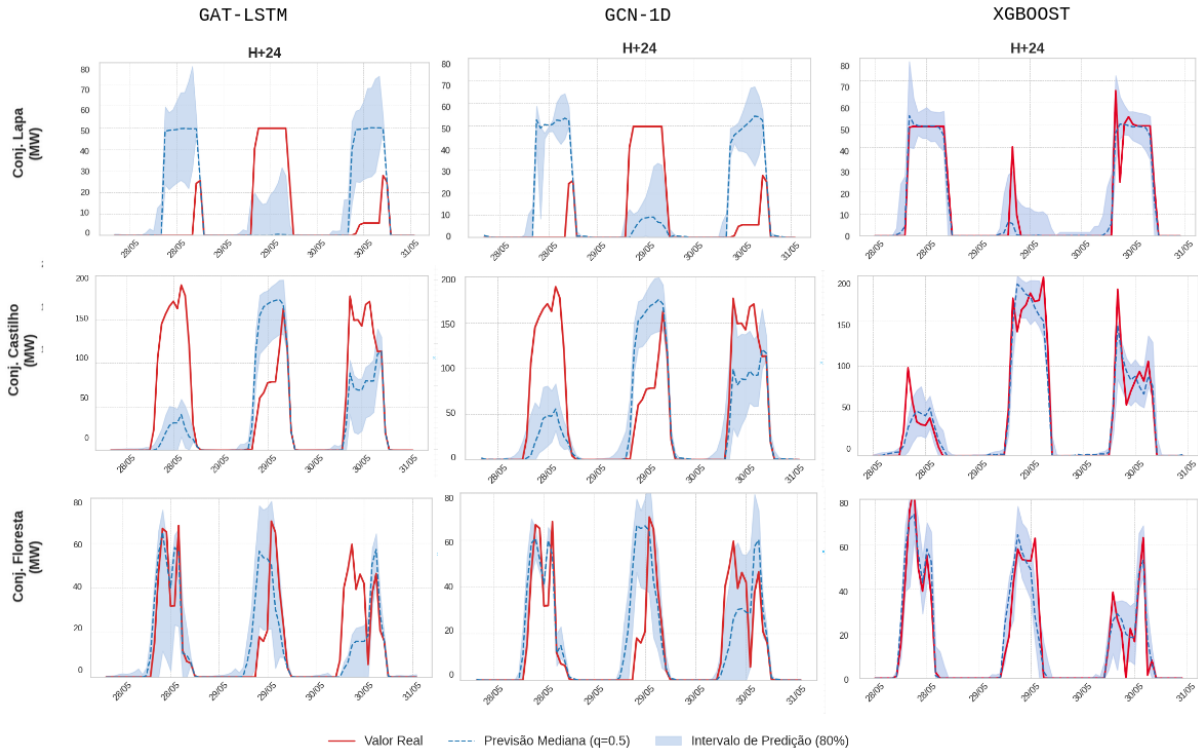


Figura 26 – Comparativo de desempenho dos modelos GAT-LSTM, GCN-1D e XGBoost para o horizonte de previsão H+24.

A análise comparativa entre os arquétipos revela a influência da topologia da rede no desempenho dos modelos, validando a hipótese central deste trabalho.

O Hub (Lapa) se mostrou o mais desafiador para os modelos GNN. A sua alta centralidade o torna suscetível não apenas a padrões climáticos regionais, mas também a dinâmicas operacionais complexas do sistema, como o *curtailment*. Os modelos GNN, ao agregarem informações dos vizinhos, foram influenciados por um sinal físico (geração esperada pelo clima) que conflitava com a realidade operacional (geração zerada por ordem externa). Este achado sugere que, para nós-chave do sistema, modelos puramente espaço-temporais podem ser frágeis se não incorporarem variáveis operacionais. A consistência do XGBoost neste cenário reforça o valor de modelos locais que aprendem os padrões únicos de um ativo específico.

O arquétipo de nó Isolado (Floresta), contrariamente ao Hub, tem a dinâmica de geração do nó isolado mais autocontida. A menor influência de vizinhos torna o problema mais próximo de uma previsão de série temporal univariada clássica. Isso explica por que todos os modelos tiveram um bom desempenho inicial. No entanto, a superioridade do XGBoost em manter um RMSE baixo em todos os horizontes, mesmo com um R^2 inferior, indica que ele aprendeu um modelo de erro médio mais consistente, enquanto os GNNs, mesmo com menos vizinhos, ainda sofrem com a degradação da informação espacial ao longo do tempo.

Já o arquétipo de nó comunidade (Castilho), foi o cenário onde a arquitetura GNN, especialmente o GCN-1D, demonstrou seu maior potencial. A forte conexão com um cluster de vizinhos geograficamente próximos forneceu informação rica e altamente correlacionada, permitindo ao GCN-1D atingir um R^2 de 0.92 em H+1, superando amplamente o XGBoost. Este resultado corrobora a hipótese de que, em regiões de alta densidade de usinas com regimes climáticos homogêneos, a modelagem explícita das interdependências espaciais oferece uma vantagem competitiva significativa para previsões de curtíssimo prazo.

5.2 Análise da calibração e desempenho em rampas

A análise da qualidade dos intervalos de incerteza revela uma limitação crítica e sistêmica de ambos os modelos GNN: uma sistemática subcobertura. Este comportamento não se restringe aos arquétipos, mas se manifesta de forma generalizada em quase todas as 47 usinas, conforme detalhado nas Tabelas 15 e 16.

O objetivo era um PICP de 80% (correspondente aos quantis 0.1 e 0.9). No entanto, para o GAT-LSTM (Tabela 25), observa-se uma ampla dispersão na calibração, com valores de PICP que variam desde 48.82% até 86.64%, mas com a grande maioria das usinas apresentando cobertura bem abaixo do nível nominal. O GCN-1D (Tabela 26) apresentou uma calibração ainda mais pobre, com a maior parte das usinas concentrada em uma faixa de cobertura entre 50% e 65%.

Este padrão geral é bem exemplificado pela análise focada nos arquétipos (Tabelas 13 e 14). Neles, o GAT-LSTM alcançou, no máximo, 79.20% (Lapa), caindo para apenas 52.25% (Castilho), enquanto o GCN-1D variou entre 50.24% e 61.58%. Estes números indicam que os modelos são excessivamente confiantes em suas previsões, gerando intervalos de predição estreitos demais para capturar a verdadeira variabilidade dos dados. Entre os dois, o GAT-LSTM consistentemente produziu intervalos mais largos (maior PINAW), o que explica sua cobertura ligeiramente melhor em diversas usinas, porém ao custo de uma menor precisão.

O desempenho sob estresse, avaliado em cenários de rampa (Tabela 15), expõe ainda mais essa fragilidade. Para o GAT-LSTM, o PICP em eventos de rampa despencou de 69.12% para 39.68%. O GCN-1D teve uma queda similar, de 66.38% para 53.98%. Este resultado demonstra que os intervalos de predição, calibrados na totalidade dos dados, são inadequados para conter a volatilidade das rampas de geração. As rampas, por sua natureza de alta magnitude e frequência, representam os eventos que mais desafiam a capacidade preditiva e de quantificação de incerteza dos modelos.

Tabela 12 – Comparativo de Métricas de Acurácia (RMSE e R^2) para a Previsão Mediana (Quantil 0.5) por Arquétipo e Horizonte de Previsão.

Arquétipo: Hub (Lapa)					
Métrica	H+1	H+3	H+6	H+12	H+24
<i>GCN-1D</i>					
RMSE (MW)	7.04	12.80	16.51	27.60	20.64
R^2	0.84	0.47	0.28	-0.83	-0.32
<i>GAT-LSTM</i>					
RMSE (MW)	8.70	20.29	20.29	37.24	29.53
R^2	0.75	-0.33	-0.33	-2.33	-1.71
<i>XGBoost</i>					
RMSE (MW)	7.19	7.70	7.72	7.93	7.33
R^2	0.30	0.20	0.20	0.15	0.28
Arquétipo: Comunidade (Castilho)					
Métrica	H+1	H+3	H+6	H+12	H+24
<i>GCN-1D</i>					
RMSE (MW)	16.06	28.79	37.38	46.22	36.83
R^2	0.92	0.73	0.56	0.53	0.66
<i>GAT-LSTM</i>					
RMSE (MW)	19.17	37.77	47.98	58.97	36.68
R^2	0.88	0.53	0.27	0.23	0.66
<i>XGBoost</i>					
RMSE (MW)	14.65	18.61	19.20	16.99	16.09
R^2	0.68	0.49	0.46	0.58	0.62
Arquétipo: Nó Isolado (Floresta)					
Métrica	H+1	H+3	H+6	H+12	H+24
<i>GCN-1D</i>					
RMSE (MW)	7.77	14.38	18.23	20.48	17.76
R^2	0.78	0.33	0.01	0.01	0.19
<i>GAT-LSTM</i>					
RMSE (MW)	9.87	20.60	26.80	24.61	16.96
R^2	0.65	-0.38	-1.14	-0.43	0.26
<i>XGBoost</i>					
RMSE (MW)	7.80	8.05	8.03	8.66	8.59
R^2	0.29	0.24	0.25	0.13	0.14

Tabela 13 – Validação dos arquétipos no modelo GAT-LSTM.

Arquétipo	Usina	PICP (%)	PINAW
Hub	Lapa	79.20	0.19
Isolado	Floresta	76.00	0.14
Comunidade	Castilho	52.25	0.09

Tabela 14 – Validação dos arquétipos no modelo GCN-1D.

Arquétipo	Usina	PICP (%)	PINAW
Hub	Lapa	50.24	0.08
Isolado	Floresta	62.17	0.11
Comunidade	Castilho	61.58	0.13

Tabela 15 – Desempenho sob cenários de estresse (rampas > 91.90 MW) - GAT-LSTM

Cenário	PICP (%)	PINAW	Nº de Pontos
Não Rampa (Normal)	69.12	0.03	37 776
Rampa (Estresse)	39.68	0.09	1 986

Desempenho sob cenários de estresse (rampas > 91.90 MW) - GCN-1D

Cenário	PICP (%)	PINAW	Nº de Pontos
Não Rampa (Normal)	66.38	0.03	37 776
Rampa (Estresse)	53.98	0.11	1 986

5.3 Limitações Observadas

A análise dos resultados revela limitações intrínsecas às abordagens de modelagem adotadas, cujas causas podem ser rastreadas até seus fundamentos teóricos. Uma limitação primária, e talvez a mais crítica do ponto de vista da quantificação de incertezas, reside na má calibração dos intervalos de predição gerados pelos modelos GNN. Conforme evidenciado pelos valores de PICP consistentemente abaixo do nível nominal de 80% (Tabelas 13 e 14), os modelos se mostraram excessivamente otimistas. Este fenômeno está diretamente ligado à natureza da regressão quantílica via função de perda *pinball*, como discutido na Seção 4.7.

Tal abordagem é eficaz em estimar a incerteza aleatória, inerente à variabilidade dos dados, mas falha em incorporar a incerteza epistêmica, associada ao próprio modelo. A

ausência de um framework de calibração mais robusto, como a Previsão Conforme, resulta em intervalos que não alcançam a cobertura estatística desejada.

Adicionalmente, os modelos GNN demonstraram uma vulnerabilidade a eventos operacionais não-físicos, como o *curtailment* observado no arquétipo Hub. A arquitetura GNN opera sob o pressuposto de que nós espacialmente próximos devem exibir correlações em seu comportamento, princípio fundamental para a agregação de informação via passagem de mensagens (Seção 4.3).

Este mecanismo permite capturar fenômenos climáticos regionais, mas se torna uma falha quando um evento local, como uma ordem de despacho do ONS, desacopla a geração de uma usina das condições de seus vizinhos. O modelo continua a prever geração baseado no consenso da vizinhança, expondo uma relação custo-benefício da modelagem espacial: sua força na captura de padrões regionais é também sua fraqueza diante de dinâmicas que não respeitam a topologia geográfica.

A degradação acentuada do desempenho em horizontes de previsão estendidos, como H+12 e H+24, representa outra limitação significativa das arquiteturas GNN, conforme ilustrado na Tabela 12 com valores de R^2 que se tornam negativos. Este comportamento pode ser atribuído ao acúmulo de erros em previsões multi-passo, um problema exacerbado no contexto dos grafos.

Neles, erros não se acumulam apenas no tempo, mas também no espaço, pois uma previsão incorreta para um nó no instante t é propagada como uma característica errônea para seus vizinhos no instante $t+1$, contaminando a rede. Em contrapartida, a notável estabilidade do XGBoost, treinado sob uma estratégia de previsão direta (um modelo por horizonte, conforme Seção 4.10), demonstra a robustez desta abordagem para mitigar o acúmulo de erros.

Finalmente, a análise de desempenho em rampas (Tabela 15) evidenciou a dificuldade dos modelos GNN em capturar eventos de alta frequência. Esta limitação é inerente à mecânica da convolução em grafos que, conforme discutido na Seção 4.6, atua como um filtro passa-baixas. O processo de agregação, ao atualizar a representação de um nó a partir de uma média ponderada de seus vizinhos, inerentemente suaviza as características do sinal. Picos e rampas de geração, que são eventos locais e de alta frequência, são atenuados neste processo, resultando em previsões mais conservadoras. Este é o equilíbrio da modelagem em grafos: a capacidade de capturar padrões regionais se dá ao custo da fidelidade a eventos locais e transientes.

6 CONSIDERAÇÕES FINAIS

Este trabalho realizou uma comparação sistemática entre modelos de previsão locais (XGBoost) e globais baseados em grafos (GNNs) aplicados à geração fotovoltaica no SIN, com o objetivo de investigar como a modelagem das interdependências espaciais entre usinas pode contribuir para a previsibilidade e quantificação das incertezas em séries temporais multi-site.

Os resultados revelaram uma relação de compromisso entre acurácia determinística e calibração probabilística. O modelo local XGBoost apresentou desempenho mais estável e previsões mais bem calibradas em horizontes variados, especialmente em usinas com características operacionais únicas — como eventos de *curtailment* ou regimes climáticos atípicos. Já as arquiteturas baseadas em grafos (GAT-LSTM e GCN-1D), embora tenham capturado relações espaciais e melhorado a explicação da variância em horizontes de curtíssimo prazo, apresentaram degradação mais rápida da performance ao longo do tempo e subestimação sistemática dos intervalos de incerteza, refletindo desafios ainda abertos na literatura de modelagem espaço-temporal probabilística.

A principal contribuição deste estudo está em integrar conceitos de aprendizado de máquina, redes complexas e quantificação de incertezas (UQ) em um framework coerente e aplicável ao contexto brasileiro. A utilização de uma metodologia de arquétipos — baseada em métricas de centralidade e influência estrutural — permitiu interpretar os resultados não apenas em termos de erro médio, mas também em função da posição topológica das usinas e de sua correlação climática, evidenciando que a incerteza da previsão é em parte estrutural e espacialmente correlacionada, e não apenas fruto de ruído estocástico.

Do ponto de vista científico, o trabalho contribui ao trazer para o domínio da previsão solar nacional uma análise comparativa rigorosa entre abordagens locais e globais sob a ótica da incerteza preditiva. Ao empregar previsões quantílicas e métricas de calibração como PICP e PINAW, a pesquisa avança no entendimento de como diferentes arquiteturas representam a dispersão preditiva, conectando-se diretamente aos desafios atuais de *Uncertainty Quantification* em energias renováveis: distinguir incertezas aleatórias, epistêmicas e estruturais em sistemas de alta variabilidade.

Importa ressaltar que as arquiteturas em grafos foram aqui implementadas de forma fundamental — com topologia estática e agregação de primeira ordem — como um esforço deliberado para isolar o papel da informação espacial. Essa decisão metodológica foi crucial para compreender os limites e benefícios reais da conectividade antes da introdução de mecanismos mais sofisticados.

Para trabalhos futuros, propõe-se a extensão para filtros espectrais de ordem superior (Chebyshev, $k > 1$) e mecanismos de atenção dinâmica, capazes de ajustar as relações espaciais e temporais em tempo real. Além disso, a integração com Redes

Neurais Informadas pela Física (PINNs) e dados operacionais do sistema elétrico — como perfis de despacho, controle de tensão e condições meteorológicas de alta resolução — representa uma via promissora para incorporar de forma explícita a natureza física das variáveis exógenas e não apenas o padrão histórico de geração local. De igual importância, a transição das previsões quantílicas para abordagens totalmente probabilística, como inferência Bayesiana ou *Deep Ensembles*, poderá permitir a decomposição explícita das incertezas aleatória e epistêmica, fortalecendo o papel da previsão solar como ferramenta de apoio à operação e ao planejamento do SIN.

Em síntese, ao combinar fundamentos de séries temporais, aprendizado profundo e teoria de grafos, este trabalho oferece uma visão integrada sobre a origem e a estrutura das incertezas na geração fotovoltaica, contribuindo para a construção de modelos mais interpretáveis, calibrados e úteis para a gestão segura e eficiente do sistema elétrico brasileiro.

REFERÊNCIAS

OLAH, C. Understanding LSTMs. 8 2015. Blog post. Disponível em: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997.

ONS. **Plano da Operação Energética 2023–2027**. [S.l.], 2023. Disponível em: <https://www.ons.org.br/AcervoDigitalDocumentosEPublicacoes/NT-ONS%20DPL%200112-22023%20-%20PEN%202023%20-%20Condi%C3%A7%C3%B5es%20de%20Atendimento.pdf>.

IPCC. **Climate Change 2022: Mitigation of Climate Change**: Contribution of working group iii to the sixth assessment report of the intergovernmental panel on climate change. [S.l.], 2022. Disponível em: <https://www.ipcc.ch/report/ar6/wg3/chapter/chapter-6/>.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. *In*: **ACM. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. [S.l.: s.n.], 2016. p. 785–794.

SOBRI, S.; KOOHI-KAMALI, S.; RAHIM, N. A. Solar photovoltaic generation forecasting methods: A review. **Energy Conversion and Management**, v. 156, p. 459–497, 2018.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1970.

CHO, K. *et al.* Learning phrase representations using rnn encoder-decoder for statistical machine translation. *In*: **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. [S.l.: s.n.], 2014. p. 1724–1734.

WU, Z. *et al.* A comprehensive survey on Graph Neural Networks. **IEEE Transactions on Neural Networks and Learning Systems**, v. 32, n. 1, p. 4–24, 2021.

YE, J. *et al.* Coupled layer-wise graph convolution for transportation demand prediction. **Proceedings of the AAAI Conference on Artificial Intelligence**, v. 34, n. 1, p. 352–359, 2020.

SAHILI, Z. A.; AWAD, M. Spatio-temporal graph neural networks: A survey. **arXiv preprint arXiv:2301.10569**, 2023.

YANG, Y. *et al.* Dest-gnn: A double-explored spatio-temporal graph neural network for multi-site intra-hour pv power forecasting. **Applied Energy**, Elsevier, v. 378, p. 124637, 2025.

AHMED, M. J.; MOZO, A.; KARAMCHANDANI, A. A survey on graph neural networks, machine learning and deep learning techniques for time series applications in industry. **PeerJ Computer Science**, v. 9, p. e1694, 2023.

HU, W. *et al.* Strategies for pre-training graph neural networks. *In: Proceedings of the 8th International Conference on Learning Representations (ICLR)*. Addis Ababa, Ethiopia: [S.l.: s.n.], 2020.

SILVA, R. *et al.* Short-term solar power forecasting in brazil: challenges and opportunities. *In: 2020 IEEE Power & Energy Society General Meeting (PESGM)*. [S.l.: s.n.], 2020. p. 1–5.

IZBICKI, R. **Machine Learning Beyond Point Predictions: Uncertainty Quantification**. 1. ed. [S.l.: s.n.], 2025. Version from January 13, 2025. Disponível em: <https://rafaelizbicki.com/uq4mlpt/>.

MUNIKOTI, S. *et al.* A general framework for quantifying aleatoric and epistemic uncertainty in graph neural networks. **Neurocomputing**, v. 521, p. 1–10, 2023.

LAKSHMINARAYANAN, B.; PRITZEL, A.; BLUNDELL, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *In: Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2017.

LIU, J. Z. *et al.* Simple and principled uncertainty estimation with deterministic deep learning via distance awareness. *In: Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2020.

MUCSÁNYI, B.; KIRCHHOF, M.; OH, S. J. Benchmarking uncertainty disentanglement: Specialized uncertainties for specialized tasks. *In: Conference on Neural Information Processing Systems (NeurIPS)*. [S.l.: s.n.], 2024.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical analysis with missing data**. 3. ed. [S.l.: s.n.]: John Wiley & Sons, 2019.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. New York: Springer, 2006.

COMIN, C. H. *et al.* Complex systems: Features, similarity and connectivity. **Physics Reports**, v. 861, p. 1–41, 2020.

KONG, W. *et al.* Short-term residential load forecasting based on LSTM recurrent neural network. **IEEE Transactions on Smart Grid**, v. 10, n. 1, p. 841–851, 2019.

YU, B.; YIN, H.; ZHU, Z. Spatio-temporal Graph Convolutional Networks: A deep learning framework for traffic forecasting. *In: Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*. Stockholm, Sweden: [S.l.: s.n.], 2018.

SONG, G. *et al.* Spatial-temporal synchronous Graph Convolutional Networks: A new framework for spatial-temporal network data forecasting. *In: Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*. New York, NY, USA: [S.l.: s.n.], 2020.

LI, Y. *et al.* Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *In: Proceedings of the 6th International Conference on Learning Representations (ICLR)*. Vancouver, BC, Canada: [S.l.: s.n.], 2018.

-
- ZHENG, C. *et al.* GMAN: A graph multi-attention network for traffic prediction. *In: Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*. New York, NY, USA: [S.l.: s.n.], 2020.
- CHEN, J. *et al.* Forecasting fine-grained power usage with Graph Neural Networks. **ACM Transactions on Knowledge Discovery from Data (TKDD)**, v. 15, n. 5, p. 1–23, 2021.
- KIPF, T. N.; WELLING, M. Semi-supervised classification with Graph Convolutional Networks. *In: Proceedings of the 5th International Conference on Learning Representations (ICLR)*. Toulon, France: [S.l.: s.n.], 2017.
- VELIČKOVIĆ, P. *et al.* Graph Attention Networks. *In: Proceedings of the 6th International Conference on Learning Representations (ICLR)*. Vancouver, BC, Canada: [S.l.: s.n.], 2018.
- SINNOTT, R. W. Virtues of the haversine. **Sky and Telescope**, v. 68, n. 2, p. 159, 1984.
- WOSCHITZ, M. **Spatio-temporal PV forecasting with graph neural networks**. 2023. Dissertação (Master's Thesis) — University of Applied Sciences Upper Austria, 2023.
- HAMILTON, W.; YING, Z.; LESKOVEC, J. Inductive representation learning on large graphs. *In: Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2017. v. 30.
- VOVK, V.; GAMMERMAN, A.; SHAFER, G. **Algorithmic Learning in a Random World**. New York, NY, USA: Springer, 2005. ISBN 978-0-387-00152-2.

APÊNDICES

APÊNDICE A – RESULTADOS DETALHADOS E MÉTRICAS POR USINA

Tabela 16 – Resultados do modelo GCN1D no conjunto de teste.

Horizonte	Pinball Loss	MAE	MSE	RMSE	R ²
H+1	2.8044	13.0675	1037.2212	32.1963	0.8738
H+3	5.2657	25.4463	3346.7427	57.8467	0.5877
H+6	7.2134	33.1230	4401.4473	66.3331	0.4739
H+12	9.4119	44.4776	5709.8481	75.5680	0.4813
H+24	5.7261	23.3264	3620.4336	60.1658	0.6378

Tabela 17 – Resultados do modelo GCN1D para o quantil 0.5 no conjunto de teste.

Horizonte	Pinball Loss	MAE	MSE	RMSE	R ²
H+1	4.9532	9.9063	553.8802	23.5404	0.9326
H+3	9.2064	18.4128	1575.0679	39.6857	0.8060
H+6	12.3528	24.7056	2439.8599	49.3890	0.7084
H+12	15.9471	31.8943	3179.4846	56.3988	0.7112
H+24	9.0317	18.0634	2200.9451	46.9344	0.7798

Tabela 18 – Resultados do modelo GCN1D para o quantil 0.9 no conjunto de teste.

Horizonte	Pinball Loss	MAE	MSE	RMSE	R ²
H+1	2.7181	18.8594	1405.3160	37.4955	0.8290
H+3	4.7595	29.6587	3597.8274	59.9836	0.5568
H+6	6.3240	36.5882	4763.7910	68.9963	0.4306
H+12	8.2468	43.7967	5758.4966	75.9046	0.4769
H+24	5.6353	24.1675	3486.3477	59.0377	0.6512

Tabela 19 – Métricas de desempenho do GAT_LSTM para o quantil 0.1.

Horizonte	Pinball Loss	MAE	MSE	RMSE	R ²
H+1	3.2991	15.7422	1331.4901	36.49	0.8380
H+3	7.5196	29.9540	4501.1875	67.09	0.4455
H+6	12.1564	43.8447	7609.0098	87.25	0.0905
H+12	17.9105	57.0348	8681.8252	93.19	0.2113
H+24	5.0644	28.8002	4056.9126	63.71	0.5941

Tabela 20 – Métricas de desempenho do GAT_LSTM para o quantil 0.5 (Mediana).

Horizonte	Pinball Loss	MAE	MSE	RMSE	R ²
H+1	6.0241	12.0482	867.2909	29.46	0.8945
H+3	12.8022	25.6044	2936.9878	54.17	0.6382
H+6	18.3426	36.6852	4295.4448	65.57	0.4866
H+12	23.9936	47.9872	5479.6895	74.00	0.5022
H+24	8.2591	16.5182	2316.4758	48.13	0.7682

Tabela 21 – Métricas de desempenho do GAT_LSTM para o quantil 0.9.

Horizonte	Pinball Loss	MAE	MSE	RMSE	R ²
H+1	3.7688	18.6266	1460.3765	38.21	0.8223
H+3	7.7157	34.5262	4707.4668	68.60	0.4200
H+6	11.7499	45.8079	6929.7617	83.25	0.1717
H+12	14.3510	53.1434	7622.6514	87.34	0.3075
H+24	5.9373	23.4732	2882.8921	53.72	0.7116

Tabela 22 – Métricas de desempenho do XGBoost para o quantil 0.1 por Arquétipo.

Usina	H+1	H+3	H+6	H+12	H+24
<i>Pinball Loss</i>					
Conj. Castilho	2.2581	2.6422	3.3385	2.6616	2.4720
Conj. Floresta	1.4598	1.2363	1.1066	1.1142	1.0727
Conj. Lapa	0.7800	0.9871	1.0300	1.0660	1.1249

Tabela 23 – Métricas de desempenho do XGBoost para o quantil 0.5 (Mediana) por Arquétipo.

Métrica	H+1	H+3	H+6	H+12	H+24
Arquétipo: Comunidade (Conj. Castilho)					
RMSE (MW)	14.65	18.61	19.20	16.99	16.09
R ²	0.68	0.49	0.46	0.58	0.62
Arquétipo: Nó Isolado (Conj. Floresta)					
RMSE (MW)	7.80	8.05	8.03	8.66	8.59
R ²	0.29	0.24	0.25	0.13	0.14
Arquétipo: Hub (Conj. Lapa)					
RMSE (MW)	7.19	7.70	7.72	7.93	7.33
R ²	0.30	0.20	0.20	0.15	0.28

Tabela 24 – Métricas de desempenho do XGBoost para o quantil 0.9 por Arquétipo.

Usina	H+1	H+3	H+6	H+12	H+24
<i>Pinball Loss</i>					
Conj. Castilho	2.0942	2.6863	3.0945	2.9556	2.4149
Conj. Floresta	1.1158	1.2781	1.2012	1.0978	1.0966
Conj. Lapa	1.0136	1.0602	1.0843	1.1003	1.0874

Tabela 25 – Métricas de incerteza por usina no modelo GAT_LSTM, ordenadas por PICP.

Usina	PICP (%)	PINAW	Usina	PICP (%)	PINAW
Sol do Sertão	48.82	0.06	Banabuiú	66.90	0.07
Serra do Mel C	49.76	0.06	Luzia	70.09	0.14
Sol do Cerrado	50.12	0.12	Francisco Sá	70.33	0.19
Belmonte	51.18	0.09	Calcário	71.04	0.06
Castilho	52.25	0.09	Pedranópolis	73.52	0.17
Janaúba	52.72	0.11	Lavras	74.23	0.10
Ciranda	53.90	0.11	Jaíba 4 Dist	75.41	0.17
Futura	54.85	0.10	Dracena	75.77	0.20
Rio Alto	56.26	0.12	Floresta	76.00	0.14
Paracatu	56.62	0.17	Juazeiro Solar	77.07	0.15
Helio Valgas	56.74	0.16	Sol do Piauí	77.54	0.24
Lar do Sol	56.97	0.16	Horizonte	77.78	0.15
Pirapora 2	56.97	0.18	Pitombeira	78.96	0.17
Guaimbê	57.09	0.11	Lapa	79.20	0.19
São Gonçalo	59.57	0.12	S. Solar Barreiras	80.38	0.14
Ituverava	61.58	0.11	Serra do Mato	80.50	0.14
Pereira Barreto	62.41	0.10	Sol do Futuro	80.85	0.16
S. Basílio (Solar)	64.42	0.16	Bom Jesus	81.09	0.20
FV SJP	64.66	0.13	Tacaratu (Solar)	81.56	1.08
Nova Olinda	64.78	0.18	Assú V	81.91	0.26
Juazeiro Solar 2	65.60	0.14	São Pedro	82.03	0.26
Boa Hora	66.08	0.11	BJL	85.11	0.28
Araxá	66.08	0.15	Flor de Mandacaru	86.64	3.58
Alex	66.08	0.07			

Tabela 26 – Métricas de incerteza por usina no modelo GCN-1D, ordenadas por PICP.

Usina	PICP (%)	PINAW
Sol do Piauí	48.46	0.13
Lapa	50.24	0.08
Boa Hora	51.18	0.07
Araxá	51.18	0.12
Pedranópolis	52.01	0.13

Tabela 26 – – continuação

Usina	PICP (%)	PINAW
Dracena	53.31	0.15
Bom Jesus	54.14	0.09
Francisco Sá	54.61	0.12
Guaimbê	56.50	0.11
Calcário	56.74	0.05
São Pedro	56.86	0.10
BJL	57.45	0.12
Horizonte	57.80	0.08
Jaíba 4 Dist	57.80	0.10
Sertão Solar Barreiras	58.27	0.11
Flor de Mandacaru	58.63	3.13
Castilho	61.58	0.13
Paracatu	61.94	0.12
Pereira Barreto	62.17	0.12
Floresta	62.17	0.11
Assú V	63.12	0.15
Banabuiu	63.83	0.06
Juazeiro Solar	63.95	0.11
FV SJP	65.01	0.08
Ituverava	65.13	0.09
Alex	66.43	0.08
Nova Olinda	66.78	0.09
Juazeiro Solar 2	66.78	0.12
Pirapora 2	68.79	0.13
S. Basílio (Solar)	69.62	0.12
Ciranda	70.33	0.10
Tacaratu (Solar)	71.04	0.85
Luzia	71.28	0.11

Tabela 26 – – continuação

Usina	PICP (%)	PINAW
Pitombeira	71.39	0.11
Rio Alto	72.46	0.11
Serra do Mato	73.17	0.14
Serra do Mel C	73.52	0.11
Lar do Sol	76.00	0.12
Belmonte	77.07	0.12
Sol do Sertão	77.30	0.11
Helio Valgas	77.42	0.12
Sol do Cerrado	77.66	0.13
Sol do Futuro	79.31	0.16
Janaúba	83.57	0.13
Futura	84.28	0.13
São Gonçalo	84.75	0.12
Lavras	87.71	0.14