



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO E SISTEMAS
CURSO ENGENHARIA DE PRODUÇÃO ELÉTRICA

Marco Aurélio Ledur Porfírio

Auto Codificadores Variacionais Aplicados a Simulação da Geração Eólica e
Fotovoltaica: Um estudo de Caso de Prospecção de Cenários no Subsistema Sul
Brasileiro

Florianópolis

2026

Marco Aurélio Ledur Porfírio

Auto Codificadores Variacionais Aplicados à Geração Eólica/Fotovoltaica: Um estudo de Caso de Prospecção de Cenários no Subsistema Sul Brasileiro

Trabalho de Conclusão de Curso submetido ao curso de Engenharia de Produção Elétrica do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Bacharel em Engenharia Elétrica com habilitação em Engenharia de Produção.

Orientador: Prof. Ricardo Villarroel Dávalos, Dr.

Florianópolis

2026

Porfírio, Marco Aurélio Ledur

Auto Codificadores Variacionais Aplicados à Geração Eólica/Fotovoltaica: Um estudo de Caso de Prospeção de Cenários no Subsistema Sul Brasileiro :Auto Codificadores Variacionais Aplicados à Geração Eólica/Fotovoltaica: Um estudo de Caso de Prospeção de Cenários no Subsistema Sul Brasileiro /Marco Aurélio Ledur Porfírio ; orientador, Ricardo Villarroel Dávalos, 2026.

83 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Engenharia de Produção Elétrica, Florianópolis, 2026.

Inclui referências.

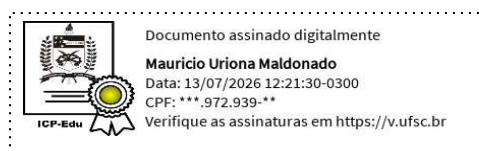
1. Engenharia de Produção Elétrica. 2. Energias renováveis. 3. Auto Codificadores Variacionais. 4. Prospeção de Cenários. 5. Subsistema Sul. I. Dávalos, Ricardo Villarroel. II. Universidade Federal de Santa Catarina. Graduação em Engenharia de Produção Elétrica. III. Título.

Marco Aurélio Ledur Porfírio

Auto Codificadores Variacionais Aplicados à Geração Eólica/Fotovoltaica: Um estudo de Caso de Prospeção de Cenários no Subsistema Sul Brasileiro

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de Bacharel em Engenharia Elétrica com habilitação em Engenharia de produção e aprovado em sua forma final pelo Curso Engenharia de Produção Elétrica.

Local Auditório João E. E. Castro – EPS/CTC, 03 de julho de 2026.



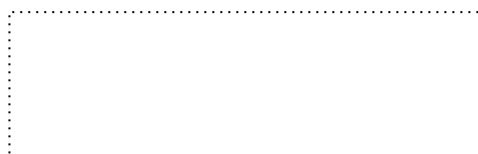
Coordenação do Curso

Banca examinadora



Prof. Ricardo Villarroel Dávalos, Dr.

Orientador



Prof. Dmontier Aragão, Dr.

Instituição Universidade Federal do Ceará



Prof. Sérgio Fernando Mayerle, Dr.

Instituição Universidade Federal de Santa Catarina

Florianópolis, 2026.

Dedico esse trabalho a todos meus familiares e amigos que contribuíram para minha formação como engenheiro.

AGRADECIMENTOS

Aos meus pais, agradeço por todo o apoio durante a minha graduação, sem vocês, nada disso seria possível.

À minha namorada, o meu muito obrigado por estar ao meu lado nos melhores e nos piores momentos desta trajetória.

Estendo meus agradecimentos à minha tia Valdete e aos meus primos, cujas dicas e incentivos foram fundamentais para a conclusão deste ciclo.

Por fim, expresso minha profunda gratidão ao meu orientador, Professor Ricardo Dávalos, pela paciência, pelas valiosas orientações e pela excelente condução deste trabalho desde o início.

RESUMO

A inserção crescente de fontes renováveis intermitentes, como eólica e fotovoltaica, no Sistema Interligado Nacional exige novas abordagens para a coordenação da operação elétrica. Este trabalho tem como objetivo desenvolver um modelo generativo baseado em Auto Codificadores Variacionais (VAE) para a prospecção de cenários sintéticos de desvios de geração no Subsistema Sul brasileiro. O método de pesquisa é aplicado em dados históricos do Operador Nacional do Sistema Elétrico (ONS), seguido pelo tratamento, normalização das informações e treinamento de uma arquitetura de rede neural profunda. Essa rede mapeia os desvios de geração em um espaço latente probabilístico para capturar as características estatísticas essenciais dos dados reais. O modelo gera perfis diários de 24 horas que auxiliam o planejamento da Programação Diária da Operação. O estudo demonstra a viabilidade do uso de VAEs como ferramenta de geração de dados sintéticos, visto que o modelo possibilita uma representação robusta das incertezas. A aplicação dessa técnica contribui para a segurança e a otimização do despacho hidrotérmico frente à variabilidade das fontes renováveis.

Palavras-chave: Energias renováveis; Auto Codificadores Variacionais; Prospecção de Cenários; Subsistema Sul.

ABSTRACT

The increasing integration of intermittent renewable energy sources, such as wind and photovoltaic, into the Brazilian National Interconnected System requires new approaches for coordinating the electrical operation. This work aims to develop a generative model based on Variational Autoencoders (VAE) to prospect synthetic scenarios of generation deviations in the Southern Brazilian Subsystem. The research methodology is applied to historical data from the Brazilian National Electric System Operator (ONS), encompassing data processing, normalization, and the training of a deep neural network architecture. This network maps generation deviations into a probabilistic latent space to capture the essential statistical characteristics of the real data. The model generates 24-hour daily profiles that assist in the planning of the Daily Operation Scheduling. The study demonstrates the feasibility of using VAEs as a tool for generating synthetic data, as the model provides a robust representation of uncertainties. The application of this technique contributes to the security and optimization of the hydrothermal dispatch in the face of the variability of renewable energy sources.

Keywords: Renewable Energy; Variational Autoencoders; Scenario Prospecting; Southern Subsystem.

LISTA DE FIGURAS

Figura 1 – Capacidade instalada no Brasil ao longo dos anos.....	22
Figura 2 – Etapas da previsão de geração eólica utilizada pelo ONS.....	23
Figura 3 – Evolução da geração fotovoltaica no Brasil.....	24
Figura 4 – Representação do Neurônio Artificial.....	27
Figura 5 – RNAs Unidirecionais (Feed-forward).....	29
Figura 6 – Arquitetura do VAE.....	31
Figura 7 – Distribuição posterior e de verossimilhança	34
Figura 8 – Fluxo metodológico do trabalho	38
Figura 9 – Fluxograma da arquitetura do modelo proposto.....	42
Figura 10 – Fluxograma dos dados do modelo	45
Figura 11 – Geração fotovoltaica prevista versus programada	49
Figura 12 – Geração eólica prevista versus programada.....	50
Figura 13 – Desvios de geração eólica	51
Figura 14 – Desvios de geração fotovoltaica.....	52
Figura 15 – Transformação da apresentação dos dados	53
Figura 16 – Evolução do erro do modelo ao longo das épocas de desvio de geração Eólica	67
Figura 17 – Evolução do erro do modelo ao longo das épocas de desvio de geração Fotovoltaica.....	67
Figura 18 – Comparação entre a evolução do treinamento em diferentes dimensões com dados de desvio de geração Eólica.....	68
Figura 19 – Comparação entre a evolução do treinamento em diferentes dimensões com dados de desvio de geração Fotovoltaica	69
Figura 20 – Cenários de desvio de geração eólica	70
Figura 21 – Cenários de desvio de geração fotovoltaica.....	71
Figura 22 – Comparação entre dados reais e cenários de desvios de geração eólica	72
Figura 23 – Comparação entre dados reais e cenários de desvios de geração fotovoltaica	73
Figura 24 – Comparação entre mediana dos dados reais fonte eólica	74
Figura 25 – Comparação entre mediana dos dados reais fonte fotovoltaica.....	75

Figura 26 – Comparação entre distribuição de probabilidades dos dados reais e gerados fonte eólica	76
Figura 27 – Comparação entre distribuição de probabilidades dos dados reais e gerados fonte fotovoltaica	77

LISTA DE QUADROS

Quadro 1 – Bibliotecas utilizadas no modelo.....	46
Quadro 2 – Funções de preparação dos dados para o modelo	54
Quadro 3 – Funções de construção do codificador	55
Quadro 4 – Funções de construção do Decodificador	56
Quadro 5 – Funções utilizadas no fluxo de dados do codificador	57
Quadro 6 – Funções utilizadas no fluxo de dados na reparametrização	58
Quadro 7 – Funções utilizadas no fluxo de dados decodificador	59
Quadro 8 – Funções utilizadas no fluxo da rede	60
Quadro 9 – Funções utilizadas na função de perdas	61
Quadro 10 – Funções utilizadas no otimizador	62
Quadro 11 – Funções utilizadas no treinamento do modelo	63
Quadro 12 – Funções utilizadas para a prospecção de cenários.....	65

SUMÁRIO

1	INTRODUÇÃO.....	14
1.1	CONTEXTUALIZAÇÃO.....	14
1.2	PROBLEMA.....	16
1.3	OBJETIVOS.....	17
1.3.1	Objetivo Geral.....	17
1.3.2	Objetivos Específicos.....	17
1.4	JUSTIFICATIVA.....	18
1.5	ESTRUTURA DO TRABALHO.....	19
2	REFERENCIAL TEÓRICO	21
2.1	FONTES GERADORAS DE ENERGIA.....	21
2.1.1	Usinas Eólicas.....	21
2.1.2	Usinas Fotovoltaicas	24
2.2	MÉTODOS COMPUTACIONAIS	26
2.2.1	Redes Neurais Artificiais	26
2.2.2	Auto codificadores Variacionais.....	30
2.3	CONSIDERAÇÕES FINAIS DO CAPÍTULO	35
3	METODOLOGIA DE PESQUISA.....	37
3.1	CARACTERIZAÇÃO DE TIPO DE PESQUISA.....	37
3.2	PROCEDIMENTO METODOLÓGICO	38
3.3	DELIMITAÇÕES DO TRABALHO.....	40
4	ESTUDO DE CASO	42
4.1	MODELO PROPOSTO	42
4.2	COLETA E TRATAMENTO DOS DADOS	46
4.2.1	Dados de Geração Eólica e Fotovoltaica	46
4.2.2	Desvios de Geração Prevista e Verificadas	47
4.2.3	Preparação dos Dados para o Modelo	52
4.3	APLICAÇÃO DO AUTO CODIFICADOR VARIACIONAL AOS DADOS	54
4.3.1	Construção da arquitetura do Auto Codificador Variacional	54
4.3.1.1	Codificador.....	55
4.3.1.2	Decodificador.....	56
4.3.1.3	Fluxo dos dados do modelo	57
4.3.2	Treinamento do modelo Variacional	60

4.3.2.1	<i>Função de perda do Auto Codificador Variacional</i>	60
4.3.2.2	<i>Otimizador do Modelo</i>	61
4.3.2.3	<i>Treinamento do Modelo</i>	62
4.4	GERAÇÃO DE CENÁRIOS	64
4.4.1	Criação dos cenários	64
5	RESULTADOS E DISCUSSÃO	66
5.1	TREINAMENTO DO MODELO	66
5.2	PROSPECÇÃO DE CENÁRIOS	70
6	CONCLUSÕES E RECOMENDAÇÕES	78
6.1	CONCLUSÕES	78
6.2	RECOMENDAÇÕES	79
	REFERÊNCIAS	80
	APÊNDICE	83

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Nos últimos anos, o setor elétrico global tem sido palco de uma transformação paradigmática, caracterizada pelo forte movimento de descarbonização e pela substituição de fontes tradicionais. Diferentemente do século anterior, quando a geração era majoritariamente baseada em fontes hídricas e térmicas, os países têm apresentado um crescimento acelerado na participação de fontes renováveis, como eólicas e fotovoltaicas. Este aumento na geração limpa não é apenas uma resposta a compromissos ambientais, mas também uma evolução econômica, redução dos custos de implementação e a oportunidade de abater o custo da energia elétrica por meio da mini e micro geração (Cortes, 2025).

Nesse cenário global, o Brasil se destaca por possuir um sistema elétrico de grande porte com uma matriz energética predominantemente renovável. Este feito é notável, pois aproximadamente 88% da eletricidade gerada no país provém dessas fontes, superando de forma significativa a média mundial, que é de apenas 26,6% (Candido, 2025). Essa alta proporção de fontes limpas posiciona o país como referência em sustentabilidade na produção de energia elétrica.

A composição da matriz renovável brasileira é diversificada. Segundo dados do Operador Nacional do Sistema (ONS, 2025), a participação da fonte hidrelétrica corresponde a 43,9%. As fontes eólica e solar (incluindo 18% da Mini e Micro Geração Distribuída - MMGD, e 7,4% de usinas centralizadas) somam, respectivamente, 13,9% e 25,4% da capacidade de geração. Em contrapartida, as usinas termelétricas (UTE), apesar de sua baixa participação na capacidade instalada (totalizando 9,6%, distribuídos entre Gás/GNL: 7,4%, Óleo/Diesel: 1%, e Carvão: 1,2%), exercem um impacto relevante no sistema. Diferentemente das UTEs, que incorrem em custos operacionais com a aquisição de combustíveis, as usinas renováveis, como solar, eólica e hídrica, utilizam recursos naturais livres de custos de aquisição.

A matéria de Frighetto (2025) aborda um tema extremamente relevante neste contexto de transição energética no uso e a expansão das termelétricas. O texto destaca que o principal argumento apresentado pelo Ministério de Minas e Energia a favor das UTEs é a segurança na geração. Diferentemente das fontes eólicas e fotovoltaicas, as UTEs podem ser acionadas a qualquer momento,

independentemente das condições climáticas. Dessa forma, ainda que sejam fontes poluidoras, as usinas térmicas oferecem um ganho de confiabilidade no suprimento da carga no sistema elétrico brasileiro.

Historicamente, o Sistema Interligado Nacional (SIN) apresentava uma complexidade operativa reduzida em comparação ao cenário atual. Antes da integração massiva de fontes não despacháveis e de alta volatilidade, a matriz elétrica brasileira era majoritariamente hidrelétrica, o que concentrava as incertezas operacionais quase exclusivamente na variabilidade das afluições e no nível de armazenamento dos reservatórios (do Nascimento, 2020). Entretanto, com o crescimento da geração eólica e fotovoltaica, essa incerteza permeou ainda mais o sistema, pois essas fontes dependem diretamente de fenômenos naturais como vento e irradiação solar, o que são variáveis não controláveis.

No cenário atual do setor elétrico brasileiro, as geradoras eólicas e fotovoltaicas possuem uma relação estratégica de suporte para as usinas hidrelétricas. De acordo com Paixão e Abaide (2026) em períodos de elevada disponibilidade de recursos renováveis, as hidrelétricas podem reduzir sua produção, conservando o volume de seus reservatórios (atuando como estocagem de energia). Em contrapartida, as usinas hídricas cooperam com as fontes intermitentes ao utilizarem a flexibilidade de suas turbinas para absorver a volatilidade característica dessas gerações.

Em um papel complementar à segurança operacional, as UTEs também assumiram um novo papel dentro do sistema. De acordo com (Castro, Brandão e Vardiero, 2017), as usinas térmicas passaram a compensar as intermitências das usinas solares e eólicas. Isso evidencia a necessidade estratégica desse tipo de geração para manter o sistema atendendo a demanda sem riscos de cortes de cargas.

Neste contexto a programação da operação não depende apenas dos valores reais determinísticos de geração, mas também do conhecimento das suas incertezas, quando se trata de fenômenos da natureza. Com a crescente inserção de fontes eólicas se faz necessário o conhecimento de incertezas relacionadas com a velocidade dos ventos, algo similar ao que ocorre nos estudos de afluições quando se trata de geração hidroelétrica (Pessanha *et al.*, 2021). Todavia, dada a configuração atual da matriz brasileira, tal preocupação deve estender-se à geração fotovoltaica, pois além de apresentar volatilidade intrínseca, já detém uma

participação significativa no suprimento eletro energético nacional, demandando métodos que sintetizem as incertezas de ambas as fontes de maneira integrada.

O ONS realiza a geração de cenários sintéticos de aflúências para usinas hidrelétricas utilizando modelos auto regressivos, com o objetivo de representar possíveis ocorrências futuras baseadas em valores observados no passado. No entanto, o setor já apontava para a evolução dessas técnicas, em 2015, quando já se discutia a utilização de inteligência artificial para tal finalidade (Cossich, Cataldi e Rotunno Filho, 2015).

1.2 PROBLEMA

O Operador Nacional do Sistema Elétrico (ONS) estrutura a programação da operação em diferentes horizontes de tempo, que incluem o longo, médio e curto prazo, esse último horizonte é conhecido como a Programação Diária da Operação (PDO). Cada etapa considera componentes e modelos específicos para lidar com as limitações e a complexidade computacional do Sistema Interligado Nacional (SIN).

A PDO, por sua vez, contempla um horizonte de uma semana com discretização temporal de 30 minutos. Nessa etapa, são considerados diversos elementos críticos, tais como, características das linhas de transmissão, previsões de aflúências hídricas, usinas em cascata, tempo de trânsito da água, e a disponibilidade prevista de geração das fontes eólica e solar (Operador Nacional do Sistema Elétrico, 2020).

A Programação Diária da Operação, emprega uma otimização inteira-mista, a qual trata variáveis intrinsecamente aleatórias, como as aflúências hídricas e as gerações eólica e solar, como variáveis determinísticas oriundas de estudos de previsão (Dagont, 2022). Esta abordagem evidencia que, mesmo com modelos sofisticados de previsão, o sistema está sujeito a desvios significativos motivados pela inerente aleatoriedade da natureza. Tais desvios nas previsões resultam na necessidade de despachar recursos de rápida resposta, sobretudo usinas termelétricas, para ajustar a oferta à demanda e, em especial, para mitigar as intermitências da geração eólica e solar (Castro, Brandão e Vardiero, 2017).

Entretanto, observa-se uma assimetria no tratamento das incertezas dentro do planejamento operativo. Enquanto as aflúências dos reservatórios hidrelétricos contam com metodologias consolidadas para a geração de cenários e análise de erros

baseados em séries históricas, as fontes eólica e fotovoltaica ainda carecem de ferramentas equivalentes em termos de integração sistêmica.

A aplicação da inteligência artificial nesse contexto é promissora, especialmente por meio do uso de modelos generativos. Estes modelos, baseados em aprendizado de máquina, são capazes de criar novos dados sintéticos a partir das propriedades estatísticas extraídas durante o treinamento. Um exemplo proeminente dessa abordagem são os Auto Codificadores Variacionais (VAE), cuja principal característica é a capacidade de reconstruir e gerar novas amostras a partir de parâmetros estatísticos, como média e variância, apreendidos do conjunto de dados original (IBM, 2024).

Diante desse cenário, o problema central desta pesquisa reside na seguinte questão. Como estruturar modelos capazes de gerar cenários sintéticos de desvios de geração para as fontes eólica e fotovoltaica para o Subsistema Sul brasileiro, de modo a oferecer uma base probabilística para a tomada de decisão no curto prazo?

1.3 OBJETIVOS

A seguir encontram-se descritos os objetivos deste trabalho.

1.3.1 Objetivo Geral

Propor um modelo matemático/computacional para geração de cenários sintéticos de desvios de geração Eólica/Fotovoltaica do Subsistema Sul Brasileiro utilizando Auto Codificadores Variacionais.

1.3.2 Objetivos Específicos

Os objetivos específicos encontram-se descritos a seguir.

- Concentrar bases de dados históricos de geração prevista e verificada das fontes eólica e fotovoltaica operantes no Subsistema Sul do Sistema Interligado Nacional (SIN);

- Calcular os desvios de geração previsto e verificado das fontes geração eólica e fotovoltaica a partir dos dados do Operador Nacional do Sistema;
- Aplicar o modelo generativo Auto Codificador Variacional aos dados históricos de desvios de geração;
- Determinar o conjunto de cenários de incerteza da geração renovável para uma prospecção de 2000 cenários;
- Avaliar o modelo generativo de redes neurais dos cenários de desvios de geração de acordo com as características dos dados históricos.

1.4 JUSTIFICATIVA

A crescente inserção de fontes renováveis não despacháveis, notadamente a eólica e a fotovoltaica, na matriz energética brasileira tem imposto desafios contínuos ao planejamento e à operação do Sistema Interligado Nacional (SIN). Embora o benefício ambiental dessas fontes seja indiscutível, a sua natureza estocástica exige ferramentas de modelagem avançadas para mitigar os impactos da intermitência no custo de produção e na estabilidade da rede. Como os desvios de previsão de geração impactam diretamente o despacho econômico e elevam o Custo Marginal de Operação (CMO), a análise detalhada dessas incertezas é fundamental para a gestão financeira e operativa do setor.

Fundamentando a busca por soluções, (Pessanha *et al.*, 2021) descrevem uma metodologia de previsão baseada na geração de 2000 cenários sintéticos para a fonte eólica. O referido estudo destaca que a quantificação das incertezas das fontes renováveis é hoje tão indispensável para a segurança do sistema quanto o monitoramento das afluências nas usinas hidrelétricas. Diante disso, este trabalho justifica-se pela aplicação de modelos de Redes Neurais a partir do modelo Auto Codificador Variacional para a prospecção de cenários de desvios de geração eólica e fotovoltaica no Subsistema Sul. A utilização de inteligência artificial permite capturar padrões complexos e não lineares que modelos estatísticos tradicionais podem não abranger plenamente, oferecendo ao operador uma visão mais robusta das incertezas.

Segundo Chen *et al.*, (2018), existem duas vantagens sobre a utilização de modelos de redes neurais generativas, a primeira é que as redes neurais podem gerar novos cenários a partir de dados históricos sem ajustar a distribuição de probabilidade

dos dados. O segundo é a utilização de aprendizagem não supervisionada, evitando alguma interferência humana direta, por vezes de difícil execução pela grande quantidade de dados históricos.

A motivação para a escolha do tema desta pesquisa reside na convergência de conhecimentos adquiridos ao longo da formação em Engenharia de Produção. Disciplinas cruciais como Pesquisa Operacional, que proporcionou o primeiro contato com os princípios de otimização, Operação de Sistemas de Energia Elétrica e Planejamento e Regulação de Mercados de Energia Elétrica demonstraram uma forte conexão entre os desafios da Engenharia Elétrica e as ferramentas de modelagem da Engenharia de Produção. Essa sinergia consolidou o interesse do autor em um tema interdisciplinar. Além disso, a literatura especializada, em destaque a obra Operação econômica e Planejamento (Brells, 1979), serviu como fonte fundamental, oferecendo o primeiro contato aprofundado com o tema de despacho econômico. Essa trajetória possibilitou a delimitação de um estudo diretamente relacionado ao eixo de Pesquisa Operacional da ABEPRO (Associação Brasileira de Engenharia de Produção), com enfoque nas subáreas de Modelagem, Simulação e Otimização e Programação Matemática.

Nesse sentido, o presente trabalho realiza um estudo de caso focado no Subsistema Sul brasileiro. A pesquisa propõe o emprego de redes neurais generativas a partir do modelo Auto Codificador Variacional, modelo esse que possui grande aplicabilidade na criação de novos dados mantendo suas características essenciais, fundamentadas em dados históricos de geração eólica e fotovoltaica do Operador Nacional do Sistema Elétrico (ONS), são capazes de sintetizar cenários para o horizonte de 24 horas.

1.5 ESTRUTURA DO TRABALHO

O presente Trabalho de Conclusão de Curso está organizado em seis capítulos que estabelecem uma progressão lógica desde a fundamentação teórica até a análise dos resultados e as conclusões finais.

Capítulo 1: Introdução. Apresenta a contextualização do tema no cenário do setor elétrico brasileiro, define o problema de pesquisa e estabelece o objetivo geral e os objetivos específicos.

Capítulo 2: Referencial Teórico. Compila e discute os conceitos e as ferramentas fundamentais para a pesquisa. Tais como fontes geradoras de energia não despacháveis, redes neurais artificiais, auto codificadores variacionais.

Capítulo 3: Procedimentos Metodológicos. Detalha a metodologia científica empregada, incluindo a classificação da pesquisa e o procedimento metodológico, onde descreve-se a trajetória do desenvolvimento da pesquisa.

Capítulo 4: Estudo de Caso. Apresenta a implementação da metodologia de forma detalhada.

Capítulo 5: Resultados e Discussão. Exibe os resultados obtidos com o modelo de prospecção de cenários e sua avaliação.

Capítulo 6: Conclusão. Sintetiza as principais descobertas do trabalho, verifica se os objetivos foram alcançados e aponta as contribuições do estudo, além de sugerir oportunidades para pesquisas futuras.

2 REFERENCIAL TEÓRICO

Este capítulo visa abordar os conceitos fundamentais para a compreensão e o desenvolvimento deste trabalho, consolidando o referencial teórico disponível na literatura.

2.1 FONTES GERADORAS DE ENERGIA

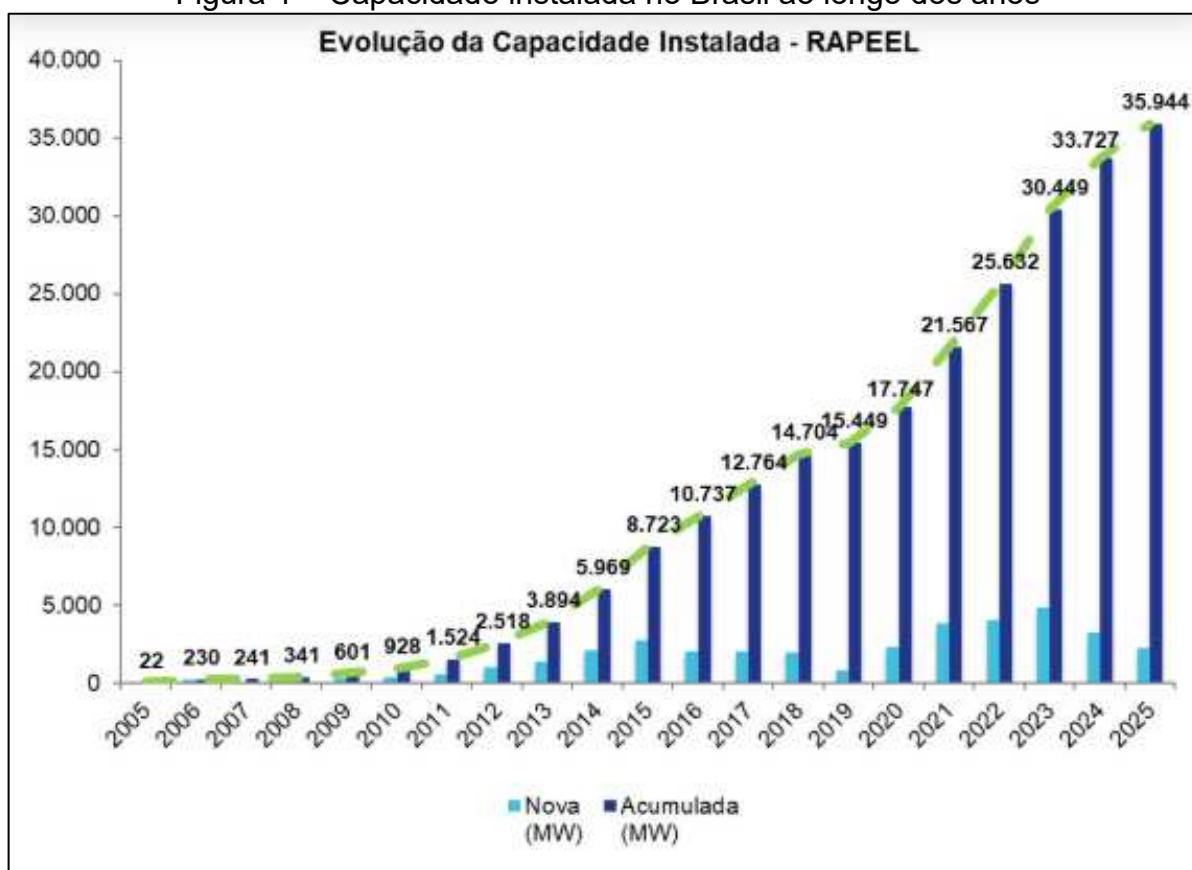
O item destaca a expansão das energias eólica e fotovoltaica no Brasil e o desafio de sua intermitência e os modelos de previsão utilizados pelo Operador Nacional do Sistema.

2.1.1 Usinas Eólicas

As usinas eólicas têm demonstrado um crescimento notável no Brasil, em 2005, a capacidade instalada era de apenas 22 MW, valor que em 2025 se encontra na faixa de 35.000 MW instalados, distribuídos em 1.143 parques eólicos e totalizando mais de 11.990 aerogeradores em 12 estados brasileiros (Associação Brasileira de Energia Eólica (ABEEólica), 2025). A Figura 1 ilustra o panorama do crescimento da capacidade instalada no país nas últimas duas décadas, evidenciando o forte incremento ocorrido no início da segunda década.

As usinas eólicas são classificadas como fontes de geração não despachável. Esta característica se deve ao fato de sua produção de energia depender intrinsecamente de fatores da natureza, que no caso da eólica, o vento, que não são passíveis de armazenamento direto. Consequentemente, a inserção dessa fonte na matriz energética impõe desafios operacionais ao Sistema Elétrico Brasileiro (SEB), principalmente em função de sua intermitência. Tais características tornam mandatório o desenvolvimento de modelos de previsão sofisticados para a gestão do sistema (Júnior *et al.*, 2024)

Figura 1 – Capacidade instalada no Brasil ao longo dos anos



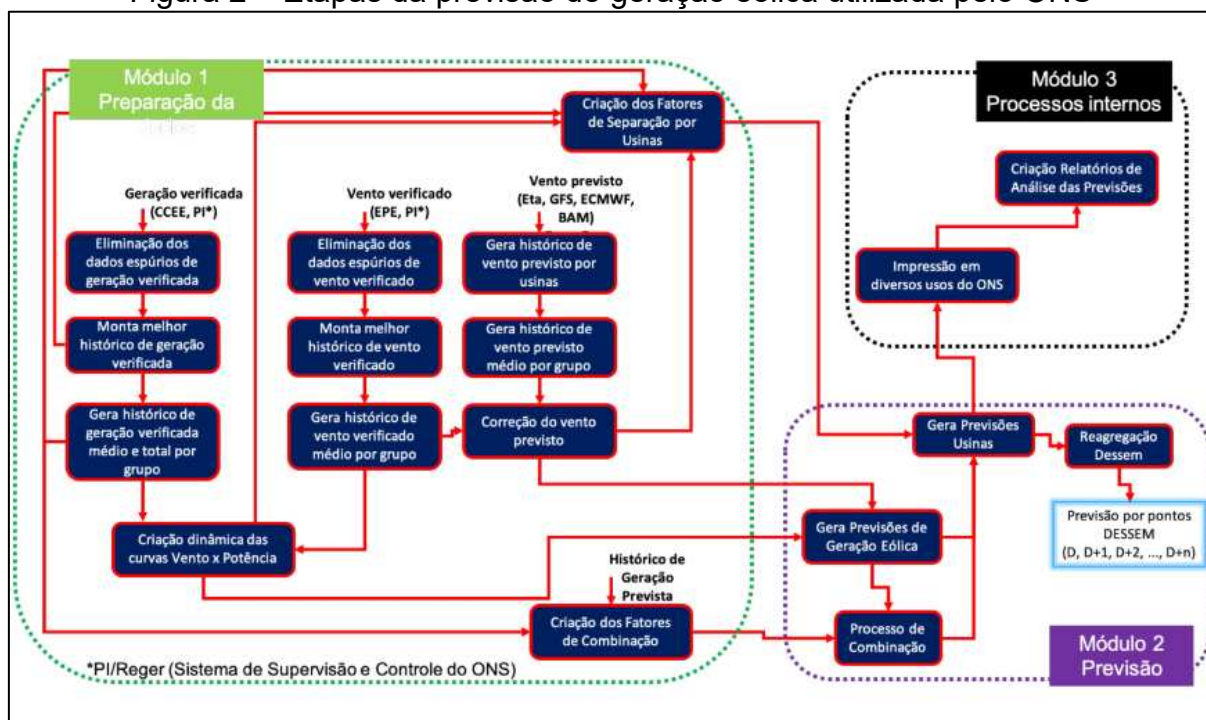
Fonte: (ABEEólica, 2025)

No ano de 2016, estudo do aumento da penetração da geração eólica e suas consequências no SEB já estava sendo realizadas. O trabalho de (El-Heri *et al.*, 2016) apresenta o estudo do impacto da variabilidade da geração eólica frente ao Custo Marginal de Operação com o aumento da participação da mesma. No mesmo ano o operador do sistema estava desenvolvendo modelos de previsão, porém contemplava apenas as fontes eólicas no Nordeste, e no ano de 2017 a região sul era inserida no modelo de previsão em tempo real (Operador Nacional do Sistema Elétrico (ONS), 2018).

Atualmente o sistema utiliza a previsão de geração eólica para operar a geração no SEB, a previsão é dividida em módulos de processamento de dados e cálculos computacionais (Operador Nacional do Sistema Elétrico, 2020).

A Figura 2 apresenta as etapas do processo de previsão de geração eólica utilizado pelo Operador Nacional do Sistema Elétrico (ONS). O modelo é estruturado em três módulos.

Figura 2 – Etapas da previsão de geração eólica utilizada pelo ONS



Fonte: Operador Nacional do Sistema Elétrico (2020).

O Módulo 1 (Preparação da Geração) é responsável pelo tratamento dos dados de entrada. Suas atividades incluem a eliminação de valores não genuínos da geração e do vento verificados, com o objetivo de formar um histórico de dados limpo e robusto. Em seguida, são criados históricos médios de geração e vento verificados, os quais são combinados para estabelecer a dinâmica das curvas Vento x Potência. Paralelamente, são gerados os históricos de vento previsto, e é realizada a correção do vento previsto a partir de sua correlação com o vento verificado.

O Módulo 2 (Previsão) recebe os dados de correção do vento previsto e as curvas Vento x Potência do Módulo 1 para realizar a previsão da geração eólica. As previsões resultantes são reagregadas e inseridas no Modelo DESSEM, que é utilizado tanto para a programação de tempo real (designado por D) quanto para os dias seguintes (representados por D+1, D+2, ..., D+n).

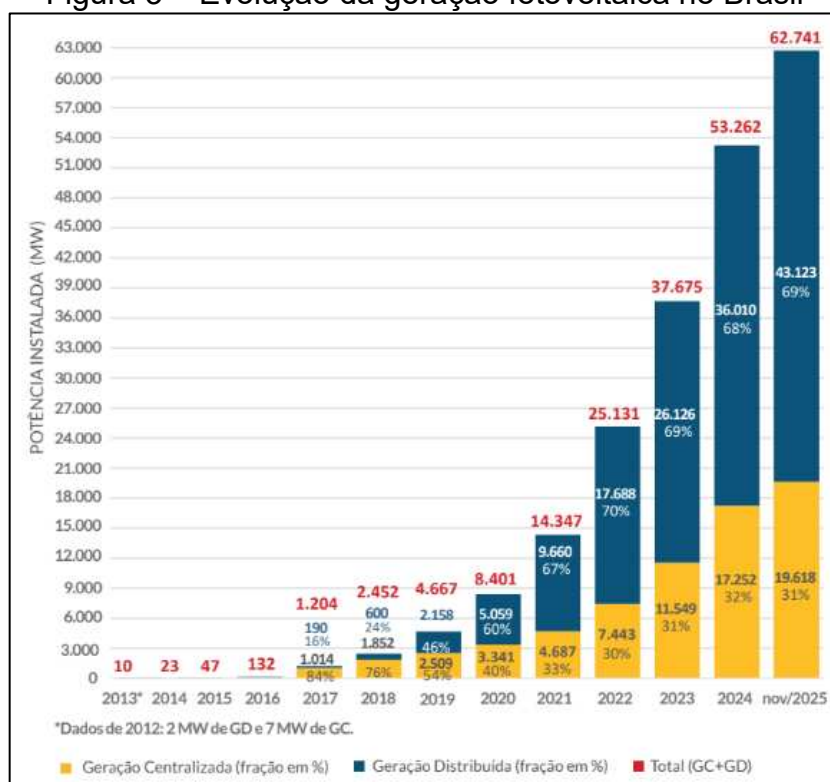
Por fim, o Módulo 3 (Processos Internos) é responsável por gerar relatórios de análise de previsões e outras informações para uso interno do Operador, a partir das saídas dos módulos anteriores (Operador Nacional do Sistema Elétrico, 2020).

2.1.2 Usinas Fotovoltaicas

A geração fotovoltaica tem se mostrado com uma fonte crescimento no Brasil nas últimas décadas. Segundo (Junior *et al.*, 2025) o país apresenta uma boa localização geográfica, onde possui uma das maiores potências de geração do mundo. O trabalho ainda acrescenta que apesar de ser promissora essa expansão ainda encontra dificuldades referentes a barreiras econômicas e outras burocracias que impedem o acesso da energia fotovoltaica para os consumidores residenciais e industriais.

Segundo a Associação Brasileira de Energia Solar Fotovoltaica, (2025) cerca de 96,7 milhões de toneladas de CO₂ foram evitados, dados da mesma fonte apresentam que o país possui aproximadamente 4 milhões de sistemas de geração distribuída e 18 mil usinas de geração centralizadas até início de novembro de 2025 (Associação Brasileira de Energia Solar Fotovoltaica, 2025). A Figura 3 apresenta a evolução da geração fotovoltaica no Brasil, evidenciando que a maior parcela dessa expansão provém da Geração Distribuída (GD).

Figura 3 – Evolução da geração fotovoltaica no Brasil



Fonte: (Associação Brasileira de Energia Solar Fotovoltaica, 2025)

A geração distribuída é composta majoritariamente por unidades de pequeno porte instaladas em residências e estabelecimentos comerciais, tendo apresentado um crescimento superior ao da Geração Centralizada (GC) constituída por usinas de grande porte.

Diferentemente da modalidade centralizada, o Operador Nacional do Sistema Elétrico (ONS) não detém o controle operativo direto sobre as unidades de geração distribuída. Por essa razão, torna-se imperativo o desenvolvimento de previsões de geração precisas para essa classe de ativos, uma vez que sua variabilidade impacta diretamente a carga líquida que o sistema precisa atender (Empresa de Pesquisa Energética (EPE) , 2022).

Dados consolidados pela ANEEL e ABSOLAR (2025), indicam que a Região Sul detém uma posição de destaque no cenário nacional de geração distribuída. O estado do Paraná ocupa a terceira posição no ranking, com uma potência instalada de 3.962,7 MW (9,2%), seguido imediatamente pelo Rio Grande do Sul, em quarto lugar, com 3.585,2 MW (8,3%). Somados à contribuição de Santa Catarina (1.513,0 MW), os três estados consolidam a região como um dos principais eixos de expansão dessa tecnologia no país. Essa expressiva participação reforça a necessidade de modelos de previsão robustos, uma vez que a variabilidade climática característica dessa localidade impacta diretamente a estabilidade da carga líquida regional.

A integração da fonte fotovoltaica introduz complexidades adicionais ao sistema elétrico, fundamentadas na variabilidade e na incerteza da carga líquida. Tais características decorrem da dependência direta de fenômenos meteorológicos, que podem sofrer alterações abruptas em curtos intervalos de tempo, além de estarem sujeitos à sazonalidade da geração conforme observado por (Milligan *et al.*, 2016). Embora essa tecnologia ofereça benefícios ambientais, como a mitigação de gases de efeito estufa, sua expansão exigiu adaptações estruturais no setor para manter a estabilidade da rede.

O Operador Nacional do Sistema Elétrico (2020) destaca que a precisão das previsões de geração fotovoltaica, especialmente em horizontes de poucas horas, é fundamental para a segurança do sistema, uma vez que eventuais erros de estimativa exigem o despacho imediato de outras fontes para manter o equilíbrio energético. Para viabilizar essas projeções, o monitoramento do Sistema Interligado Nacional integra diversas variáveis meteorológicas, tais como a irradiância solar na superfície, a pressão atmosférica, a umidade relativa e a temperatura a dois metros de altura. Um

fator determinante nesse processo é a análise da cobertura de nuvens em diferentes níveis da atmosfera (baixa, média e alta), que permite antecipar oscilações na produção de energia (Operador Nacional do Sistema Elétrico, 2020).

Para previsões de curto e médio prazo, que variam de algumas horas a vários dias de antecedência, a metodologia baseia-se prioritariamente em Modelos de *Numerical Weather Prediction*, Previsão Numérica do Tempo (NWP). O uso desses modelos é essencial pois eles fornecem dados meteorológicos precisos mesmo em áreas com pouca infraestrutura de medição física, garantindo a estabilidade da programação energética independentemente da presença de sensores locais (Operador Nacional do Sistema Elétrico, 2020).

Complementarmente, a estrutura de previsão para fontes intermitentes do ONS divide-se em duas abordagens metodológicas principais. A primeira utiliza modelos de séries temporais e técnicas estatísticas clássicas, como as famílias autorregressivas ARIMA e ARIMAX, para detectar padrões no histórico de geração. A segunda vertente utiliza métodos de Inteligência Artificial, incluindo Redes Neurais Artificiais, Máquinas de Vetores de Suporte e Sistemas Fuzzy. Esses modelos de aprendizado de máquina são particularmente eficazes para lidar com problemas não lineares e multivariados, características inerentes à natureza estocástica da radiação solar (Operador Nacional do Sistema Elétrico, 2020).

2.2 MÉTODOS COMPUTACIONAIS

Este tópico tem como objetivo apresentar os principais conceitos dos métodos computacionais utilizados no trabalho dentre eles as redes neurais artificiais e uma de suas formas particulares, o método de Auto Codificador Variacional.

2.2.1 Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) constituem uma forma de processamento de informação baseada no comportamento das ligações neurológicas do cérebro. No contexto computacional, são representadas por modelos matemáticos que trocam dados entre pontos de conexão, denominados neurônios. (Nelson, 2017).

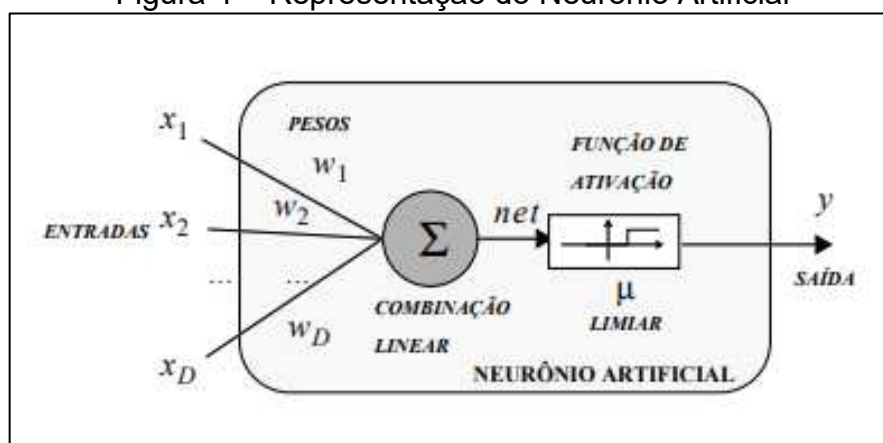
As bases das Redes Neurais Artificiais remontam à primeira metade do século XX. O marco inicial ocorreu em 1943, quando o neurologista Warren McCulloch e o

matemático Walter Pitts publicaram um modelo matemático que buscava modelar o funcionamento dos circuitos biológicos do cérebro através de circuitos elétricos. Como na época não era possível implementar tal ideia em computadores, a implementação prática em sistemas computacionais só foi concretizada pela IBM em 1959 (Dias *et al.*, 2023).

O componente mais elementar de uma rede neural é o neurônio artificial, chamado de perceptron, que foi apresentada inicialmente por Frank Rosenblatt em 1958 constituídas por neurônios McCulloch – Pitts, também chamados de neurônios MCP (Nied, 2007). O funcionamento de um perceptron fundamenta-se no modelo clássico de McCulloch-Pitts.

A Figura 4 apresenta essa representação clássica do perceptron, essa estrutura processa informações recebendo sinais de entrada, também referidos como sinapses, que são ponderados por meio de sua multiplicação por pesos específicos. Na etapa seguinte, os resultados dessas multiplicações são agrupados por um elemento somador. O valor resultante dessa soma serve como o sinal de entrada para a função de ativação, que detém a responsabilidade de determinar se o sinal de saída será propagado ou interrompido. (Rauber, 2005).

Figura 4 – Representação do Neurônio Artificial



Fonte: (Rauber, 2005).

O neurônio artificial pode ser formalizado por meio de uma estrutura matemática que relaciona as entradas recebidas com uma resposta de saída específica. A Equação (1) estabelece a relação fundamental entre a saída (y), a função de ativação (FA) e o sinal de entrada resultante (net):

$$y = FA(net) \quad (1)$$

Essa formulação indica que a ativação do neurônio depende estritamente do valor assumido por (net). Por sua vez, o sinal net corresponde à saída do somador das entradas ponderadas por seus respectivos pesos sinápticos, representando uma combinação linear das variáveis de entrada, conforme detalhado na Equação (2):

$$net = \sum_{i=1}^D (x_i * w_i) \quad (2)$$

Ao integrar esses conceitos, a Equação (3) descreve a representação completa do neurônio artificial de McCulloch-Pitts, consolidando todas as operações internas e as relações entre entrada e saída:

$$y = FA\left(\sum_{i=1}^D (x_i * w_i)\right) \quad (3)$$

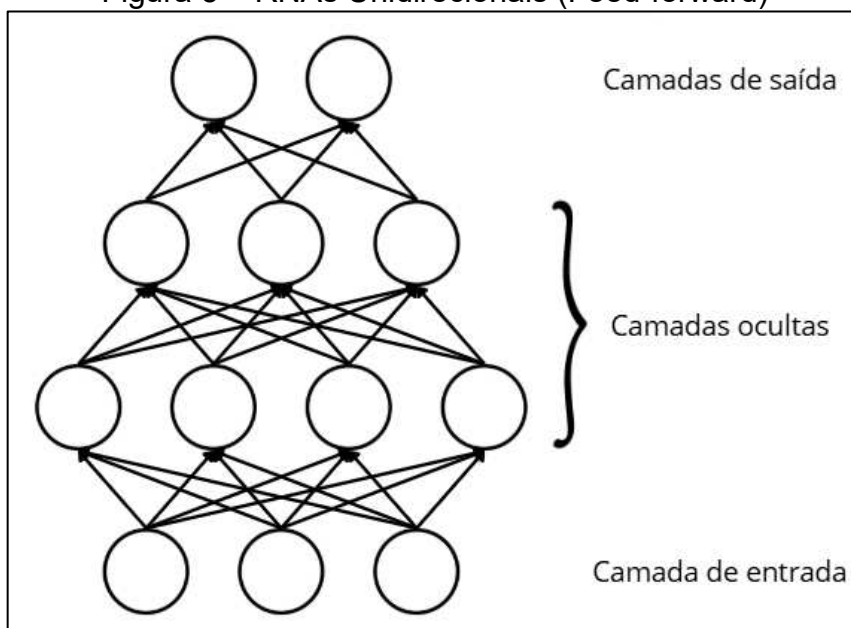
Adicionalmente, a Equação (4) apresenta uma formulação mais abrangente, que introduz uma variável suplementar denominada *Bias*. O *Bias* representa um valor intrínseco ao neurônio que, somado à combinação linear das entradas, permite o deslocamento da função de ativação para melhor ajuste dos dados. A inclusão dessa variável resulta na equação de saída final:

$$y = FA\left(\sum_{i=1}^D (x_i * w_i) + bias\right) \quad (4)$$

Conforme ilustrado na referida Figura 4, a função de ativação do modelo original é representada por uma função degrau (ou escada). Na prática, esse mecanismo opera de forma binária, assumindo apenas dois estados possíveis: ligado ou desligado. No entanto, conforme destaca (Rauber, 2005), existem outros tipos de funções de ativação. Estas podem variar desde modelos lineares até funções não lineares, que possibilitam à rede neural aprender padrões de comportamento que não se restringem a decisões puramente binárias.

Os modelos de Redes Neurais Artificiais (RNA) são estruturados em camadas, sendo que cada nível possui um número variável de neurônios. Em arquiteturas convencionais, o fluxo de informação ocorre de forma unidirecional, partindo da camada de entrada em direção à saída sem processos de retroalimentação, característica que define a arquitetura feed-forward (Nelson, 2017), conforme ilustrado na Figura 5. Esse processamento sequencial garante que a informação seja transformada gradualmente ao passar por cada estágio da rede.

Figura 5 – RNAs Unidirecionais (Feed-forward)



Fonte: Adaptado de (Nelson, 2017)

Para a construção de uma RNA é fundamental a definição de três conceitos estruturais básicos que determinam a capacidade de aprendizado do modelo. Primeiramente, deve-se estabelecer o número de entradas da rede, que corresponde à dimensão dos dados brutos a serem processados. Em seguida, define-se a configuração das camadas intermediárias, ou ocultas, especificando tanto a quantidade de camadas situadas entre a entrada e a saída quanto o número de neurônios contidos em cada uma delas. Por fim, determina-se a quantidade de neurônios na camada de saída, a qual é responsável por fornecer a resposta final ou a previsão gerada pelo sistema de inteligência artificial (Fleck *et al.*, 2016).

A estrutura de uma rede neural artificial, por si só, não é suficiente para a execução de tarefas complexas, é imperativo que os pesos sinápticos associados às entradas dos neurônios sejam ajustados para que, diante de um conjunto de dados, o

modelo forneça uma resposta coerente. De acordo com Fleck et al. (2016), a etapa dedicada ao ajuste desses pesos é denominada aprendizagem. Nesse estágio, o sistema é submetido a estímulos de entrada, e os resultados obtidos na saída são comparados aos valores reais esperados.

O processo de ajuste dos pesos ocorre de forma iterativa e é tecnicamente designado como treinamento. Durante as diversas iterações, o algoritmo busca a convergência dos parâmetros para um estado de erro mínimo. Contudo, a aprendizagem da rede só é considerada efetiva quando o modelo demonstra capacidade de generalização, ou seja, quando se torna suficientemente genérico para a finalidade proposta.

De acordo com Neid (2007), o treinamento é subdividido em duas etapas principais. Na primeira, denominada propagação, um sinal de estímulo é aplicado na entrada e se propaga pelas camadas até a saída, que retorna uma resposta ao estímulo inicial. Destaca-se que, nesta fase, os pesos sinápticos permanecem fixos. Após a propagação, ocorre a etapa de retropropagação, na qual o valor da saída da rede é comparado ao valor real esperado para o cálculo do erro. É especificamente nesta fase de retropropagação que os pesos sinápticos são ajustados, utilizando algoritmos de otimização para minimizar a discrepância entre o resultado retornado pela rede e o valor real.

Um dos algoritmos de otimização mais fundamentais e utilizados na literatura é o chamado Gradiente Descendente. O objetivo desse algoritmo é minimizar uma função de custo, definida como $J(\theta)$, na qual θ representa os parâmetros do modelo. Para isso, a otimização ocorre atualizando os parâmetros na direção oposta ao gradiente da função. Em termos visuais, o algoritmo atua "descendo" a superfície matemática de $J(\theta)$, guiando o modelo de forma iterativa até encontrar o ponto de mínimo da função de perda (Ruder, 2017).

2.2.2 Auto codificadores Variacionais

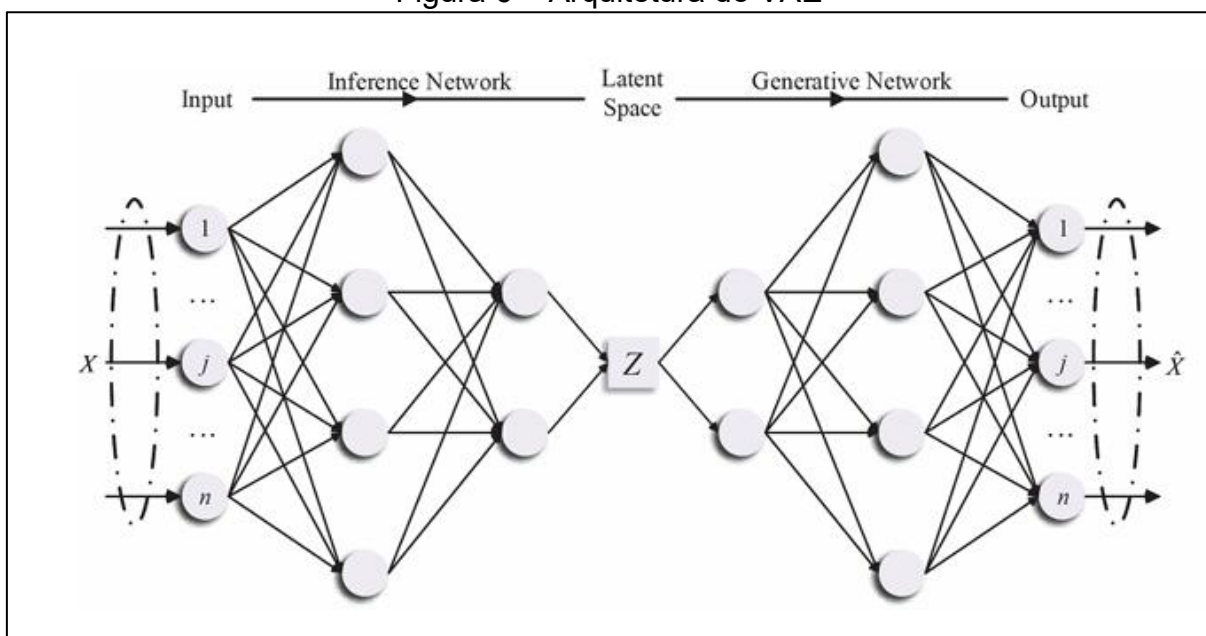
O Auto Codificador Variacional (VAE, em inglês *Variational Autoencoder*) é uma arquitetura de rede neural artificial baseada em aprendizado não supervisionado. O mecanismo principal consiste em aprender e extrair as características essenciais dos dados de entrada, promovendo a redução de dimensionalidade ao mapear esses dados de entrada em uma distribuição de probabilidade. A partir dessa representação

por dados com dimensões comprimidas o qual é chamado de espaço latente, a rede utiliza um decodificador para reconstruir os dados originais o mais fielmente possível a partir dos vetores do espaço latente (Liu *et al.*, 2022).

A viabilidade matemática dessa representação por distribuições de probabilidade foi estabelecida pelo trabalho de (Kingma e Welling, 2014), intitulado *Auto-Encoding Variational Bayes*. Nesse estudo, os autores introduziram o "truque de reparametrização", uma inovação crucial que isolou a aleatoriedade do modelo e permitiu a aplicação direta do algoritmo de retro propagação no treinamento da rede.

A arquitetura de um auto codificador variacional está dividida em três partes, o codificador, espaço latente e por fim o decodificador, sendo o codificador e decodificador redes neurais artificiais. A Figura 6 apresenta de forma simplificada a arquitetura da rede neural artificial de um auto codificador variacional a qual se divide em *Inference Network* (rede de inferência) sendo o codificador, o *Latente Space* (espaço latente) e por fim, *Generative Network* (rede generativa) referindo-se ao decodificador (Dairi *et al.*, 2020).

Figura 6 – Arquitetura do VAE



Fonte: (Liu *et al.*, 2022).

Codificador

O codificador é composto por uma rede neural artificial em que possui a característica de redução de número de neurônios a cada camada seguinte, que

possui a função de aprender as principais características sobre a distribuição de probabilidades das entradas X . Esse processo é caracterizado pela redução da dimensão dos dados de entrada, sua representação final são dados de média e desvio padrão, pois aproxima a distribuição de probabilidade dos dados de entrada a uma Gaussiana, dados esses presentes no espaço latente.

A notação que representa a operação matemática do codificador é apresentado pela Equação (5), ela representa a probabilidade de z , dado uma variável.

$$q_{\phi}(z|x) \approx p_{\phi}(z|x) = \frac{p(x, z)}{p(z)} \quad (5)$$

A distribuição $p_{\phi}(z|x)$ é chamada de distribuição posterior, a qual não é conhecida na prática, e é aproximada por $q_{\phi}(z|x)$, uma distribuição de probabilidade Gaussiana multivariada cujo o formato é conhecido, e permitindo ser descrita com a média e desvio padrão.

A aproximação $q_{\phi}(z|x)$ nos diz que dado uma entrada X ele produz uma distribuição Gaussiana sobre os possíveis valores de Z que podem ser gerados pela entrada X , como apresentado na Figura 7 (Kingma e Welling, 2014). Em outras palavras, a distribuição posterior $q_{\phi}(z|x)$ mostra as características potenciais do espaço latente Z representada pelos valores das amostras da entrada X (Liu *et al.*, 2022).

Espaço Latente

O espaço latente atua como o intervalo entre as duas redes neurais existentes, posicionando-se exatamente na transição entre o codificador e o decodificador. Ele é o ambiente matemático onde habitam as variáveis latentes. Essas variáveis formam um vetor de dimensionalidade reduzida comparada com os dados de entrada, sendo a representação mais comprimida dos dados em toda a rede neural, guardando apenas as características essenciais para descrever a entrada original (Dairi *et al.*, 2020).

Decodificador

Assim como o codificador, o decodificador também é uma rede neural artificial, porém, possui a função inversa do codificador. Nesta etapa o decodificador constrói as variáveis de saída “X chapéu” a partir do espaço latente Z , criado anteriormente pelo codificador. A sua arquitetura da rede neural segue a lógica inversa do codificador, que o número de neurônios aumenta ao passar pelas camadas da rede.

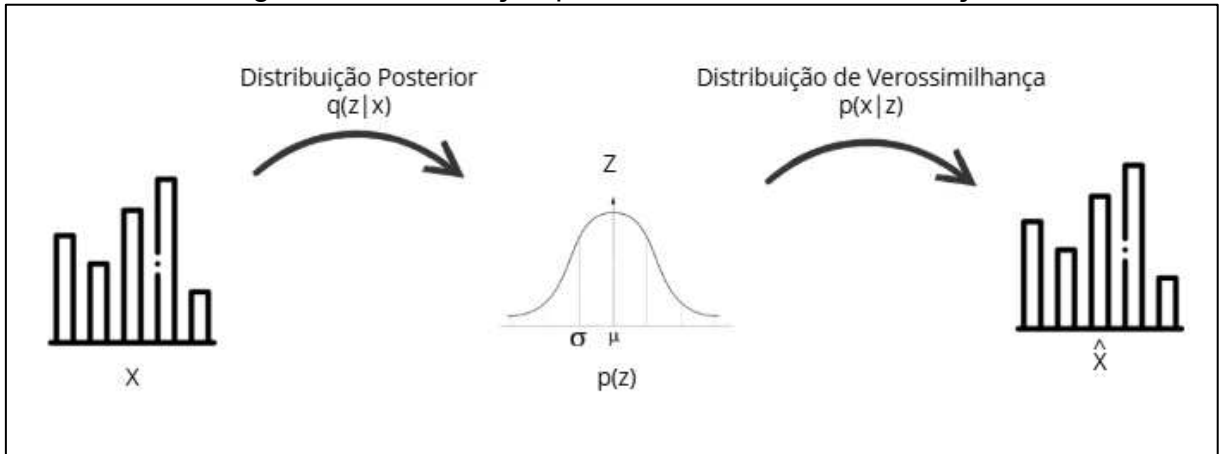
Matematicamente pode ser expresso pela Equação (6), onde é conhecida por distribuição de probabilidade de verossimilhança. A variável $p_{\theta}(x|z)$ mostra a distribuição de probabilidade de dado um vetor Z de reproduzir a entrada X (Kingma e Welling, 2014).

$$p_{\theta}(x|z) = \frac{p(x, z)}{p(x)} \quad (6)$$

O processo de geração de novos cenários acontece quando o modelo realiza uma amostragem aleatória nesse espaço latente, obtendo diversos valores para z . Para que esses dados gerados façam sentido, o treinamento do VAE foca em otimizar dois objetivos fundamentais. O primeiro é minimizar o erro de reconstrução do decodificador $p_{\theta}(x|z)$. O segundo é forçar a distribuição criada pelo codificador, $q_{\phi}(z|x)$, a se aproximar de uma distribuição Normal padrão matemática, conhecida como *a priori* que nada mais é que a distribuição $p(z)$ (Liu *et al.*, 2022).

A Figura 7 resume o fluxo completo de informação dentro da arquitetura. O processo se inicia com a leitura do conjunto de dados originais X pelo codificador, que mapeia suas características fundamentais das amostras vindos de X , como a média e o desvio padrão. Em seguida, ocorre a etapa de amostragem no espaço latente, dando origem ao vetor latente Z . É a partir dessa representação probabilística contínua que o decodificador entra em ação, realizando o caminho de reconstrução para projetar a os dados finais “X chapéu”. Essa estrutura ilustra o grande diferencial do VAE, a sua capacidade não apenas de comprimir informações, mas de mapear as características dos dados, permitindo a geração de cenários inteiramente novos e coerentes.

Figura 7 – Distribuição posterior e de verossimilhança



Fonte: O Autor

O treinamento de um VAE tem como objetivo aproximar o máximo possível os dados de entrada dos valores reconstruídos na saída. Para isso, busca-se a maximização da verossimilhança a partir da Equação (7). Essa equação é composta por duas partes: a primeira refere-se ao termo de reconstrução, e a segunda ao termo de regularização (Dairi *et al.*, 2020).

$$\mathcal{L}(x) = E_{q(z|x)}[\log p(x|z)] - KL(q(z|x)|P(z)) \quad (7)$$

O Termo de Reconstrução é responsável pela comparação entre os dados de entrada e a saída do Auto Codificador Variacional. Ele atua como uma medida de desvio (erro) entre os dados originais e os gerados, contudo, otimizar apenas esse termo não é suficiente para a geração de novos dados. Por isso, introduz-se o Termo de Regularização, proveniente da divergência de Kullback-Leibler (KL), cujo objetivo é comparar e medir a diferença entre duas distribuições de probabilidade. No caso do VAE, a divergência KL mede a distância entre a distribuição que o codificador está aprendendo a criar e uma distribuição predefinida, geralmente uma distribuição Normal padrão (Dairi *et al.*, 2020).

Entretanto, apenas a minimização do erro não é suficiente. As redes neurais aprendem por meio da retropropagação (*backpropagation*), ou seja, a partir do erro calculado no final da rede, um sinal (gradiente) é enviado no sentido inverso para ajustar os pesos dos neurônios, com o objetivo de minimizar a diferença entre os dados de entrada e saída. No entanto, em um VAE, esse sinal não consegue simplesmente retornar pela rede. Isso ocorre porque o vetor latente z é obtido por

meio de uma amostragem probabilística, o que torna a operação estocástica e não diferenciável, interrompendo o fluxo da retropropagação (Kingma e Welling, 2014).

Para contornar esse problema, (Kingma e Welling, 2014) propuseram o chamado "Truque da Reparametrização" (*Reparameterization Trick*). O método consiste em adicionar um ruído externo ϵ amostrado de uma distribuição Gaussiana padrão (com média 0 e desvio padrão 1). Dessa forma, o valor de z passa a ser gerado de forma determinística em relação aos parâmetros da rede, utilizando a média μ e o desvio padrão σ (produzidos pelo codificador) combinados com o ruído auxiliar ϵ conforme a Equação (8).

$$z = \mu + \sigma * \epsilon \quad (8)$$

Isto é, a amostra de z passa a ser construída a partir da multiplicação pelo ruído de distribuição conhecida seguida de uma soma. Como a multiplicação e a adição são operações matemáticas contínuas e perfeitamente diferenciáveis, o cálculo do gradiente torna-se possível. Essencialmente, o método "terceiriza" a etapa do sorteio aleatório, retirando a estocasticidade do caminho principal da rede e permitindo que a retro propagação flua do decodificador de volta para o codificador.

2.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

O aumento expressivo da contribuição das fontes renováveis, como a eólica e a fotovoltaica, impõe um desafio direto ao controle e à operação do Sistema Elétrico Brasileiro. Por se tratarem de fontes não despacháveis, cuja geração depende estritamente de variáveis climáticas, o operador do sistema não possui controle direto sobre a sua produção. Consequentemente, torna-se imprescindível a utilização de previsões de geração cada vez mais precisas, não apenas para garantir a segurança da rede, mas também para viabilizar o cálculo adequado dos custos de operação. Contudo, mesmo com o emprego de modelos de previsão avançados utilizados pelo operador, o sistema ainda está sujeito a desvios operacionais inevitáveis, que consistem justamente na diferença entre a energia que foi prevista e a que foi efetivamente realizada.

Diante dessas incertezas e da necessidade de aprimoramento, os métodos computacionais avançados ganham papel de destaque. O Auto Codificador Variacional, por exemplo, é um modelo de redes neurais que tem como objetivo a reconstrução de dados ou a geração de novos cenários a partir de informações históricas. Esse modelo é responsável por capturar as principais características das séries históricas em análise e representá-las em um espaço latente de dimensão menor que os dados originais. Ao mapear essas características de forma estatística, extraindo a média e o desvio padrão, o algoritmo torna-se capaz de reconstruir os dados com alta fidelidade e de simular novos cenários, oferecendo uma ferramenta promissora para lidar com a natureza estocástica e a intermitência inerentes às matrizes energéticas modernas.

3 METODOLOGIA DE PESQUISA

Este capítulo estabelece o caminho científico adotado para o alcance dos objetivos da pesquisa. Serão apresentadas a caracterização do estudo, o detalhamento dos procedimentos metodológicos em seu fluxo de execução, assim como, as delimitações que definem o escopo do modelo e as simplificações assumidas.

3.1 CARACTERIZAÇÃO DE TIPO DE PESQUISA

O presente trabalho classifica-se, quanto à sua natureza, como Pesquisa Aplicada, pois o objetivo central é gerar conhecimento científico que se destina à aplicação prática e à resolução de um problema específico do setor elétrico brasileiro. Em relação à abordagem do problema, adota-se o caráter quantitativo, com dados secundários do ONS e Câmara de Comercialização de Energia Elétrica (CCEE), visto que a investigação se fundamenta na coleta e tratamento de dados numéricos (séries históricas, desvios) e no emprego de técnicas estatísticas.

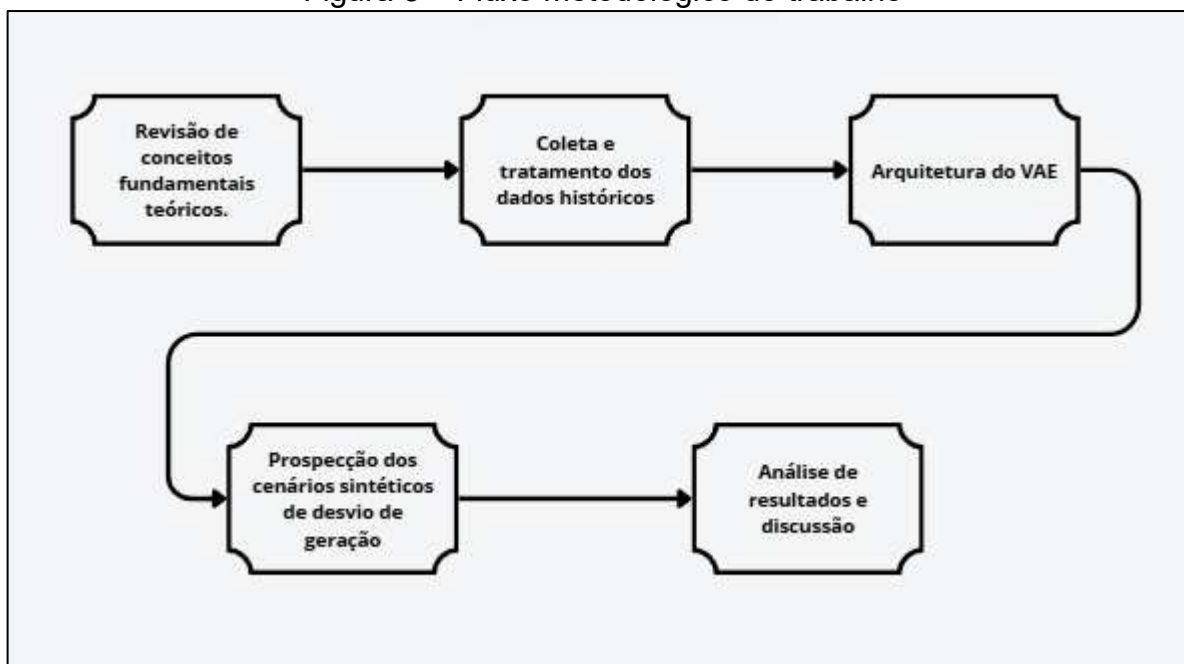
Quanto aos propósitos, a pesquisa possui um propósito descritivo e explicativo. É descritiva por buscar a caracterização detalhada do comportamento estatístico dos desvios de previsão no Subsistema Sul. É explicativa porque se propõe a identificar, quantificar e representar a geração e seus desvios das fontes eólica e fotovoltaica (Gil, 1994).

O procedimento técnico central desta pesquisa é o Estudo de Caso de Modelagem e Simulação. De acordo com Miguel et al. (2012), a modelagem quantitativa consiste em representar o sistema em análise de forma abstrata, utilizando a linguagem matemática e computacional para aplicar técnicas de simulação e calcular variáveis. No contexto do presente trabalho, este procedimento se manifesta na aplicação de técnicas de Aprendizado de Máquinas para a caracterização e prospecção de cenários de desvios de geração das fontes eólica e fotovoltaica do Subsistema Sul Brasileiro.

3.2 PROCEDIMENTO METODOLÓGICO

Este capítulo descreve a metodologia de pesquisa empregada a prospecção de cenários sintéticos de desvios de geração eólica e fotovoltaica no Subsistema Sul brasileiro. O procedimento utilizado na elaboração do trabalho está descrito no fluxograma apresentado na Figura 8.

Figura 8 – Fluxo metodológico do trabalho



Fonte: O Autor

Cada uma das etapas apresentadas no fluxograma da Figura 8 estão apresentadas nos tópicos a seguir.

Etapa revisão de conceitos fundamentais teóricos: É conduzida por meio de busca em bases de dados científicas como *Google Scholar* e repositórios institucionais. O foco são palavras-chave que envolvem incertezas, desvios de previsão, fontes renováveis não despacháveis, geração de cenários, redes neurais e Auto Codificadores Variacionais.

Etapa coleta, tratamento e caracterização dos dados: Esta etapa foca na aquisição e preparação dos dados de entrada necessários para a modelagem:

- **Dados de Geração e Previsão:** Serão coletadas séries históricas de geração verificada e prevista para as fontes eólica e fotovoltaica no Subsistema Sul, abrangendo o período de 01/10/2024 a 01/02/2026;
- **Cálculo dos Desvios:** Utilizando a biblioteca Pandas em Python, será realizada a diferença entre os dados verificados e previstos para obter os desvios de previsão. Estes desvios constituirão o objeto da análise estatística para caracterizar a incerteza;
- **Padronização dos dados de desvio de geração** na forma de data frame para forma matriz para a entrada no Auto Codificador Variacional.

Etapa da arquitetura do Auto Codificador Variacional: Esta etapa tem como objetivo descrever a arquitetura do modelo do Auto Codificador Variacional;

- **Características** Definição do número de neurônios, camadas e operações dentro do codificador, espaço latente e decodificador;
- **Fluxo de dados:** Definição do caminho e as manipulações dos dados em cada etapa dentro da rede neural do Auto Codificador Variacional;
- **Treinamento do modelo:** A definição da metodologia de treinamento do modelo, assim como sua função de perda para o ajuste dos pesos da rede.

Etapa Geração de Cenários de Incerteza: Esta seção detalha o uso de técnicas para simular a aleatoriedade:

- **Treinamento** com os dados da série histórica de geração eólica e fotovoltaica assim como seus valores de desvios de previsão, no horizonte de 01/10/2024 a 31/01/2026 com o método Auto Codificador Variacional;
- **Gerar 2000 cenários** de desvio de geração renovável que representem a incerteza inerente ao sistema, com base nas séries históricas de erros de previsão;
- **Apresentar os resultados** dos valores estatísticos de geração a partir dos cenários e comparar com dados do ONS para o dia 01/02/2026.

Análise de Resultados: Após a simulação de todos os cenários, será realizada a análise quantitativa final:

- Resultados do treinamento do modelo do Auto Codificador Variacional.
- Comparação entre os valores verificados escolhidos aleatoriamente com cenários gerados;
- Comparação entre a mediana dos valores verificados com a mediana dos cenários gerados;
- Comparação entre a distribuição de probabilidade dos valores verificados sorteados com a distribuição dos cenários gerados pelo modelo;
- Avaliação de representatividade dos cenários gerados com os valores reais verificados de acordo com a metodologia do espaço latente do modelo Auto Codificador Variacional.

3.3 DELIMITAÇÕES DO TRABALHO

O estudo restringe-se ao Subsistema Sul do Sistema Interligado Nacional (SIN), focando especificamente no comportamento das fontes eólica e fotovoltaica desta região.

A pesquisa concentra-se na geração de cenários de desvios de geração. Não faz parte do escopo a comparação entre diferentes arquiteturas de modelos generativos, limitando-se à aplicação e análise da rede neural Auto Codificador Variacional.

O modelo será treinado e validado utilizando exclusivamente séries temporais de geração verificada e geração prevista, disponibilizadas pelo Operador Nacional do Sistema Elétrico (ONS).

Não serão consideradas variáveis meteorológicas primárias (exógenas), tais como irradiância solar, temperatura ou velocidade do vento. O foco permanece estritamente nos dados de potência (MW) ao longo do tempo, tratando a incerteza a partir do comportamento histórico da produção.

O estudo limita-se ao horizonte de (24 horas à frente), não sendo aplicável para estudos de planejamento de longo prazo ou expansão da capacidade.

Assume-se que os dados de geração refletem a disponibilidade plena dos ativos, não sendo modeladas paradas para manutenção programada ou falhas forçadas de aereogeneradores e inversores fotovoltaicos.

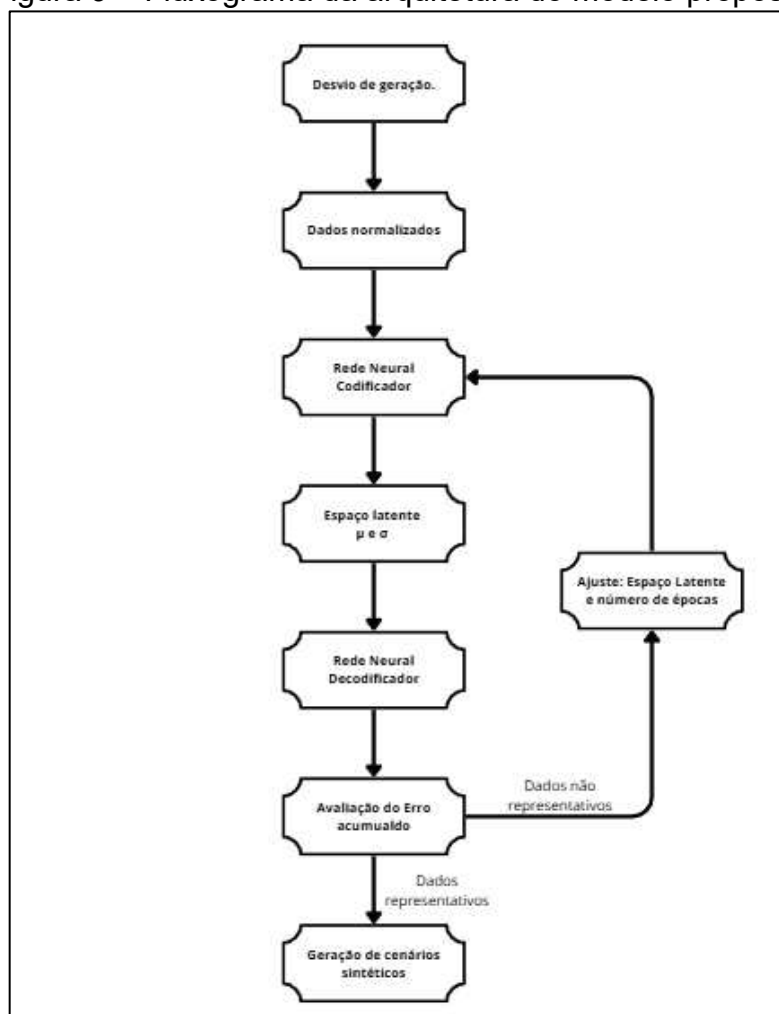
4 ESTUDO DE CASO

Nesta seção, são detalhadas as atividades práticas desenvolvidas conforme a metodologia estabelecida anteriormente.

4.1 MODELO PROPOSTO

O modelo tem como objetivo representar as características essenciais dos desvios entre a geração prevista e a verificada, calculados a partir dos dados oficiais obtidos junto ao Operador Nacional do Sistema Elétrico (ONS). A partir da rede neural, busca-se a prospecção de cenários sintéticos de desvio de geração eólica e fotovoltaica. A Figura 9 ilustra a arquitetura conceitual e o fluxo das técnicas que compõem o Auto Codificador Variacional implementado neste estudo.

Figura 9 – Fluxograma da arquitetura do modelo proposto



Fonte: O Autor

O funcionamento do modelo computacional engloba uma série de etapas sequenciais. A fase inicial consiste no pré-processamento dos dados brutos fornecidos pelo Operador Nacional do Sistema (ONS). Nessa etapa, realiza-se a filtragem da base de dados para excluir as usinas que não compõem o escopo de interesse deste estudo, seguida pela agregação (soma) dos montantes de geração de todas as usinas selecionadas, obtendo-se um valor totalizado para cada hora do dia. Esse procedimento de consolidação é aplicado de forma idêntica tanto para a fonte eólica quanto para a fotovoltaica. Após essa estruturação, realiza-se a normalização dos dados históricos, um tratamento matemático indispensável para assegurar a estabilidade e a convergência dos gradientes durante o treinamento da rede neural. Essa estrutura recebe as séries temporais e as comprime, extraindo os padrões e as principais características. O codificador mapeia essas informações para um espaço de menor dimensionalidade, o espaço latente. Nesta etapa, o codificador os mapeia em parâmetros de uma distribuição de probabilidade, definindo uma média e desvio padrão. Para que a rede consiga otimizar seus pesos através do algoritmo de retropropagação, aplica-se a técnica do truque de reparametrização, que introduz uma aleatoriedade controlada no processo.

O processo inverso fica em função do decodificador. Essa segunda rede neural coleta amostras do espaço latente e reconstrói os dados originais. Durante a fase de inferência, o decodificador atua diretamente como um gerador, convertendo um ruído aleatório do espaço latente em novos cenários temporais que representam de forma realista a geração de energia. Porém, os dados gerados podem ou não ser representativos, para isso, se faz a avaliação do erro em função do número de épocas e dimensionalidade do espaço latente.

Para uma melhor compreensão da metodologia, apresenta-se a seguir o fluxo de dados, detalhando desde o processamento inicial até a obtenção desses cenários.

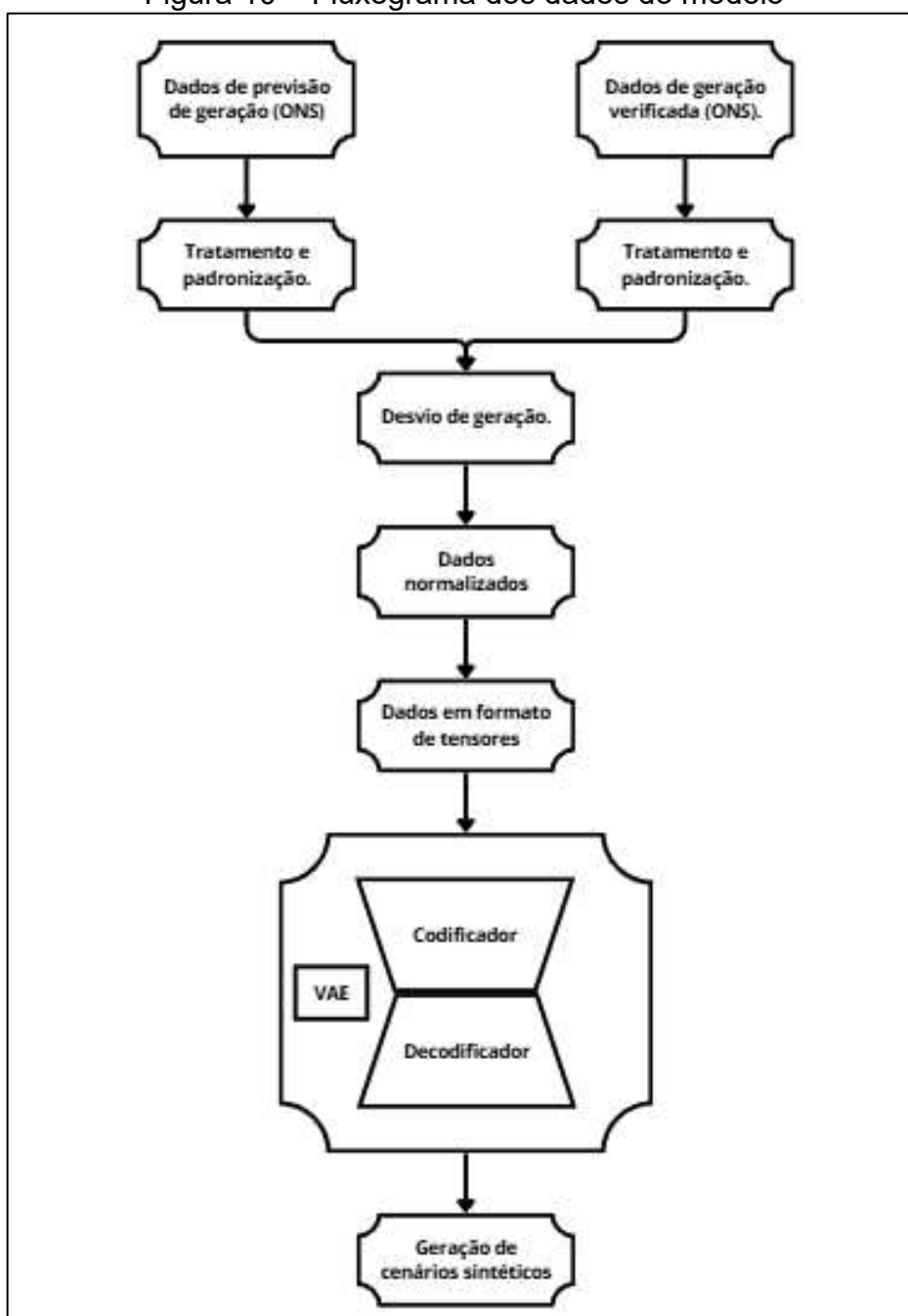
O fluxo de processamento de dados dentro do modelo de prospecção de cenários sintéticos de desvios de geração a partir do Auto Codificador Variacional é apresentado a partir da Figura 10. O processo inicia-se com a coleta de dados no site oficial do Operador Nacional do Sistema Elétrico, esses dados estão separados em planilhas que representam dias ou meses. Em seguida, esses dados são unificados em uma mesma planilha, e passam pelo tratamento das informações com o objetivo de isolar apenas os dados das usinas fotovoltaicas e eólicas do Subsistema Sul, assim como a padronização de datas da janela de análise.

Após a etapa de coleta e tratamento de dados, realiza-se o cálculo dos desvios de geração, por meio dos dados de geração prevista e geração real, os quais serão a base de dados para dar início ao modelo generativo.

Antes de realmente aplicar os dados no modelo VAE, realiza-se a padronização dos dados por meio da normalização e a conversão dos mesmos para o formato matricial, permitindo que a arquitetura de redes neurais do modelo possa processá-los adequadamente. Por fim, o fluxo é concluído com o treinamento do modelo Auto Codificador Variacional por meio dos dados matriciais de desvios de geração, e a subsequente prospecção de cenários sintéticos de desvios de geração.

Para cada uma das etapas do fluxo de dados será detalhada a seguir, respeitando a ordem cronológica do fluxograma, com o intuito de descrever as particularidades técnicas e as premissas consideradas em cada fase do desenvolvimento.

Figura 10 – Fluxograma dos dados do modelo



Fonte: O Autor

A validação do modelo baseia-se na comparação direta entre os dados observados e os dados gerados, bem como na análise de aderência entre as suas respectivas distribuições de probabilidade. Essa abordagem garante que o VAE não apenas replique perfis individuais de desvio, mas também preserve as propriedades estatísticas e a variabilidade inerente ao comportamento histórico das fontes eólica e fotovoltaica.

As bibliotecas utilizadas na coleta, tratamento dos dados, modelo e resultados estão apresentadas no Quadro 1, assim como sua aplicabilidade dentro do trabalho.

Quadro 1 – Bibliotecas utilizadas no modelo

Biblioteca	Objetivo
<i>Numpy</i>	Manipulação numérica e funções matemáticas.
<i>Pandas</i>	Manipulação e análise de dados estruturados, utilizada para o tratamento de séries temporais e organização dos <i>dataframes</i> .
<i>Torch</i>	Utilizado para a construção, treinamento e execução da arquitetura do Auto Codificador Variacional.
<i>Sklearn</i>	Pré-processamento utilizada para a normalização e dimensionamento dos dados (escala) antes da entrada no modelo.
<i>Matplotlib</i>	Criação de gráficos de visualização de dados.
<i>Seaborn</i>	Criação de gráficos estatísticos.

Fonte: O Autor

4.2 COLETA E TRATAMENTO DOS DADOS

A etapa foca a descrição do fluxo de processamento das informações, desde a extração das séries temporais nas plataformas oficiais do ONS até a consolidação de uma base de dados estruturada. Este processo é fundamental para garantir a integridade das análises sobre o comportamento das fontes renováveis e sua relação com o despacho termelétrico no Subsistema Sul.

4.2.1 Dados de Geração Eólica e Fotovoltaica

Os dados de geração eólica e fotovoltaica utilizados neste estudo são classificados em duas categorias distintas: geração prevista e geração verificada. Ambas as bases de dados são oriundas do Operador Nacional do Sistema Elétrico (ONS).

A base de dados de geração verificada compreende os valores reais de energia injetada no sistema pelas usinas, com resolução temporal horária (a cada uma hora). Os dados abrangem todos os subsistemas do Sistema Interligado Nacional (SIN) e estão organizados em planilhas de consolidação mensal. O recorte temporal definido para esta base compreende o período de 01/01/2023 a 31/01/2026.

Nesses registros, cada unidade geradora é caracterizada por atributos como localização (estado e subsistema), tipologia de fonte e combustível, além do respectivo montante de geração. Embora o documento original apresente classificações adicionais, estas foram desconsideradas por não integrarem o escopo deste trabalho. Para fins de análise, foram extraídos exclusivamente os dados referentes às fontes eólica e fotovoltaica pertencentes ao Subsistema Sul.

Complementarmente à base de dados verificados, foram obtidos os dados de geração prevista oriundos do Modelo de Despacho Hidrotérmico de Curto Prazo. Este modelo consiste na ferramenta oficial utilizada pelo ONS para o planejamento da operação, fornecendo a geração para as fontes eólica e fotovoltaica com base na disponibilidade dos recursos naturais e nas restrições elétricas do sistema.

Os resultados dessa programação constam na plataforma de dados abertos do ONS sob a denominação de Dados dos Valores do Planejamento Diário. Este conjunto de dados contempla as projeções de geração e demanda com discretização semi-horária, ou seja, com intervalos de 30 minutos, totalizando 48 postos diários. Esta resolução temporal difere daquela observada na base de geração verificada, que apresenta caráter horário.

A referida base de dados abrange o conjunto de informações de todo o SIN. Entretanto, para a execução desta pesquisa, realizou-se um tratamento para a filtragem exclusiva das fontes eólicas e fotovoltaicas pertencentes ao Subsistema Sul. O recorte temporal definido para a coleta do PDO compreende o período de 01/10/2024 a 24/02/2026, totalizando 498 planilhas processadas (uma para cada dia de operação). Esse volume substancial de dados foi necessário para garantir a representatividade estatística da análise frente às variações sazonais das fontes intermitentes.

4.2.2 Desvios de geração Prevista e Verificadas

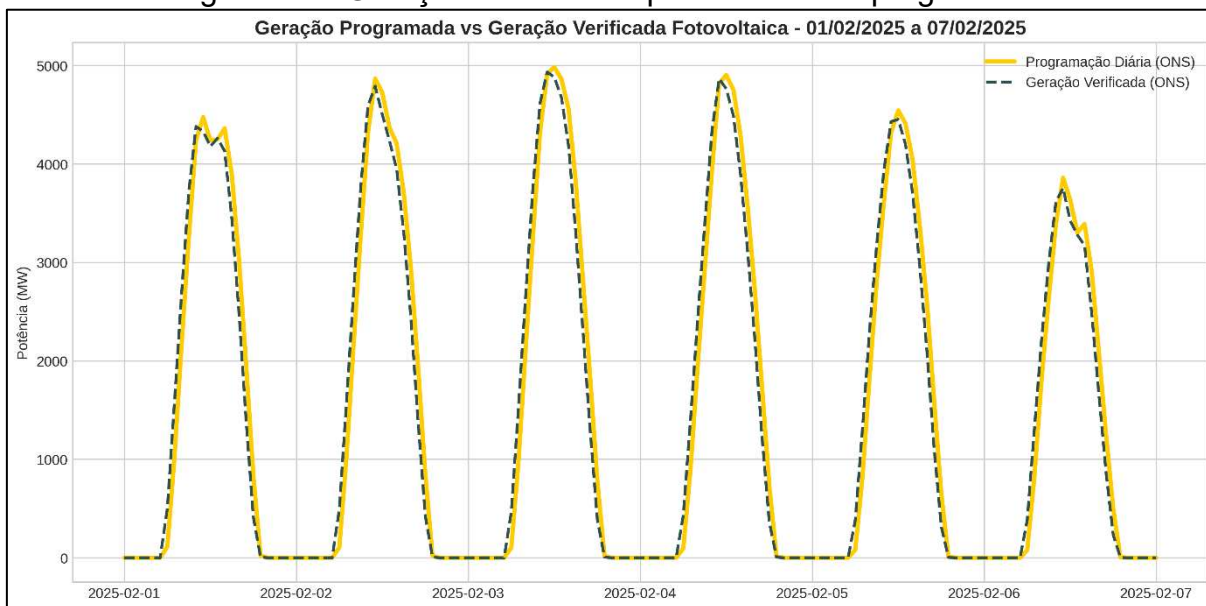
Para calcular as diferenças entre a geração de energia prevista e a que realmente aconteceu, foi desenvolvido um programa em Python focado em organizar e comparar os dois conjuntos de dados. O primeiro passo desse trabalho foi ajustar as marcações de tempo. Como os arquivos do ONS apresentam as datas e horas em formatos diferentes, foi necessário criar uma "chave única" para cada momento. No caso da geração verificada, as informações de dia e hora estavam em colunas separadas, então o código uniu esses dados em uma única coluna de data e hora. Isso permitiu que o programa identificasse exatamente qual previsão correspondia a cada registro real de geração.

Após esse ajuste, o código resolveu o problema da diferença de tempo entre as bases. Enquanto a previsão (PDO) traz dados a cada 30 minutos, a geração real só é registrada a cada hora cheia. Para que a comparação fosse justa e direta, o programa filtrou a base de previsão, mantendo apenas os horários fechados (minuto "00") e descartando os horários intermediários. Assim, as duas listas de dados ficaram com 24 marcações por dia, permitindo uma comparação ponto a ponto.

Com os horários alinhados, as tabelas foram unidas em um único arquivo de trabalho. Durante esse processo, o código foi programado para manter apenas os horários que existissem nos dois arquivos ao mesmo tempo. Se por algum motivo houvesse uma falha de medição na usina e o dado real estivesse faltando, o programa automaticamente excluía a previsão daquele mesmo horário. Resultando no horizonte de análise de 01/10/2024 até 31/01/2026. Esse cuidado é importante para evitar erros de cálculo, garantindo que o resultado final refletisse apenas os momentos em que se tinha total certeza do que foi previsto e do que foi gerado.

A Figura 11 apresenta a comparação direta entre os montantes de geração fotovoltaica verificada e prevista. A partir da análise visual, percebe-se que os valores apresentam uma elevada aderência, mantendo-se bastante próximos ao longo do período. Ambas as curvas seguem o padrão de sazonalidade diária esperado para a fonte solar, caracterizado pelo ciclo de disponibilidade do recurso conforme o nascer e o pôr do sol.

Figura 11 – Geração fotovoltaica prevista versus programada

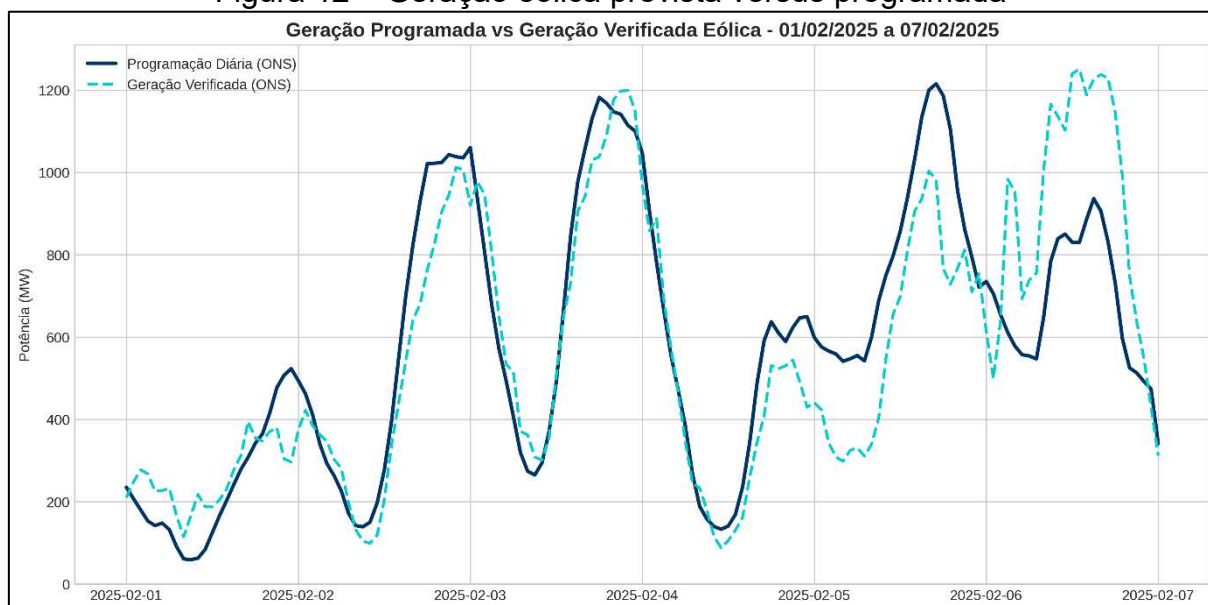


Fonte: O autor, dados (ONS,2026).

A Figura 12 apresenta o comportamento da geração eólica prevista e verificada ao longo de um período de sete dias. Diferentemente do que foi observado na fonte fotovoltaica, os valores de geração eólica apresentam uma variação significativamente maior entre o montante previsto pelo modelo e o que foi efetivamente gerado pelas usinas.

É importante ressaltar que os dados exibidos nos gráficos representam um recorte amostral do conjunto de dados total. Essa seleção foi realizada para permitir uma visualização mais detalhada da dinâmica entre as curvas, facilitando a identificação da precisão do modelo DESSEM frente à geração real verificada no Subsistema Sul.

Figura 12 – Geração eólica prevista versus programada



Fonte: O autor, dados (ONS,2026).

O cálculo do desvio entre a geração verificada e a prevista foi realizado por meio de uma subtração simples entre os dois montantes. Essa operação permite identificar tanto a magnitude da diferença quanto a sua direção, facilitando a análise. A Equação (9) descreve essa relação matemática entre o desvio de geração e as gerações verificadas e previstas.

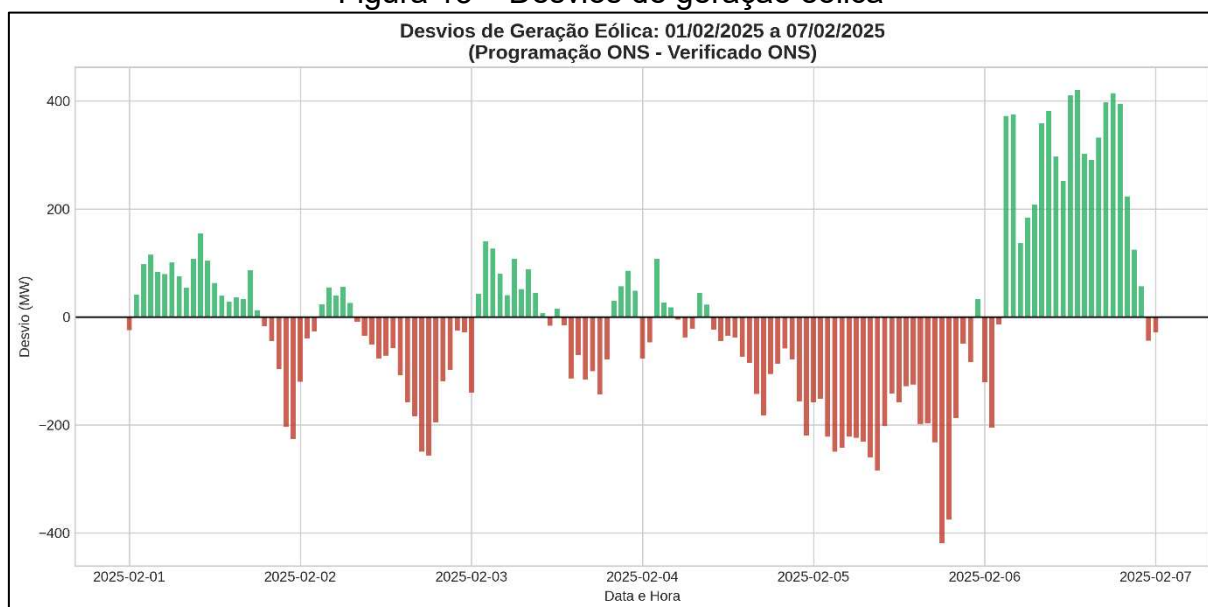
$$\text{Desvio} = \text{Geração}_{\text{verificada}} - \text{Geração}_{\text{prevista}} \quad (9)$$

Dentro dessa metodologia, o cálculo do desvio é feito subtraindo a previsão do valor real. Dessa forma, um resultado positivo indica que a previsão foi menor do que a geração verificada, ou seja, as usinas entregaram mais energia do que o ONS esperava. Por outro lado, um resultado negativo aponta que a geração verificada ficou abaixo do esperado, indicando que a previsão foi superior ao que foi efetivamente produzido.

Como pode ser observado na Figura 13 e Figura 14, apresentam os desvios em um recorte de uma semana escolhida aleatoriamente, as cores das barras facilitam essa distinção. As barras na cor verde correspondem aos períodos em que a geração verificada foi maior que a prevista, ou seja, desvios positivos. Já as barras na cor

vermelha mostram os momentos em que a geração prevista foi maior que a verificada, desvios negativos, permitindo identificar rapidamente as falhas de previsão para mais ou para menos.

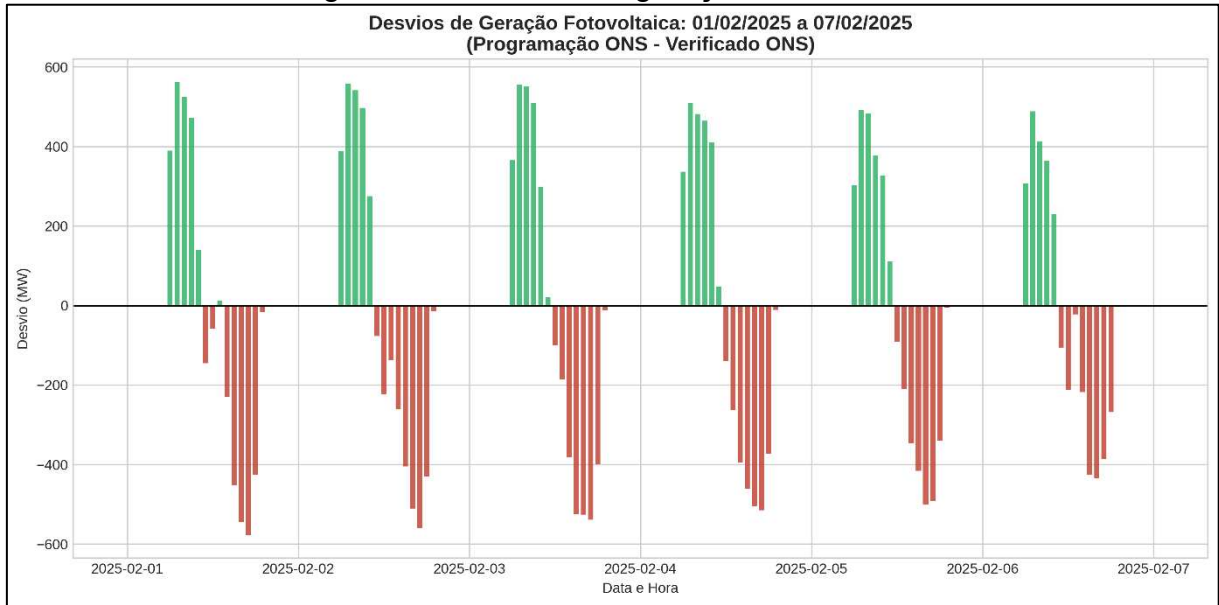
Figura 13 – Desvios de geração eólica



Fonte: O autor, dados (ONS,2026)

Percebe-se que os dados de desvios da geração eólica apresentam uma maior variabilidade e ausência de sazonalidade definida em seu comportamento. Diferentemente desse cenário, os desvios da geração fotovoltaica exibem uma forte sazonalidade em seus valores, refletindo a periodicidade característica do recurso solar.

Figura 14 – Desvios de geração fotovoltaica



Fonte: O autor, dados (ONS, 2026)

4.2.3 Preparação dos dados para o modelo

Com os cálculos dos desvios de previsão realizados, é necessário estruturá-los adequadamente para a entrada no Auto Codificador Variacional. Para o melhor aproveitamento do treinamento do modelo, é fundamental normalizá-los, isto é, garantir que os dados sigam um determinado ponderamento. Nesse processo, os valores são centralizados na média zero e escalonados a partir de um desvio padrão unitário, conforme demonstrado na Equação (10).

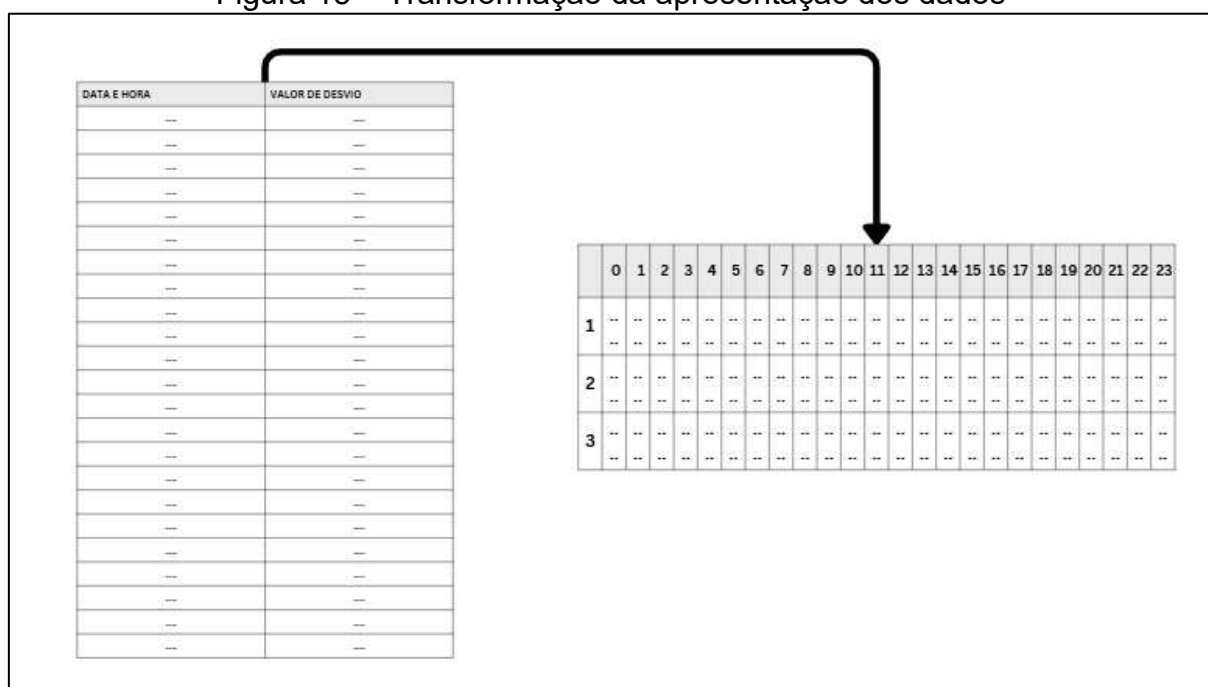
$$z = \frac{x - \mu}{\sigma} \quad (10)$$

Onde z representa o dado padronizado, x o valor original, μ a média e σ o desvio padrão do conjunto amostral. Essa transformação, conhecida como *Standard Score* ou *Z-score*, é essencial para que o modelo processe diferentes magnitudes de desvio de forma equilibrada, evitando que variáveis com maior amplitude numérica dominem indevidamente o ajuste dos pesos da rede neural.

O modelo foi projetado para processar um dia por vez, contemplando as 24 horas de forma integrada. Originalmente, os dados estão dispostos em um formato de

data frame de série temporal, onde as colunas registram a data, a hora e o valor do desvio de geração. Para que o modelo possa processar essas informações, o *data frame* deve ser transformado em uma estrutura matricial, na qual cada linha representa um dia específico e cada coluna corresponde ao desvio de geração de uma das 24 horas do dia. Essa organização permite que a rede neural capture a correlação entre os diferentes postos horários de um mesmo ciclo diário. A Figura 15 mostra essa transformação de apresentação entre os dados do *data frame* e na forma tensorial.

Figura 15 – Transformação da apresentação dos dados



Fonte: O autor

O processo inicia-se com a instanciamento do método *Z-score* e sua respectiva aplicação aos dados originais. Em seguida, realiza-se a transformação da estrutura de dados original para o formato matricial (tensorial), o que viabiliza o processamento matemático pela arquitetura da rede neural.

Por fim, define-se a estratégia de alimentação dos dados ao modelo. O algoritmo não recebe o conjunto de dados em sua totalidade de uma só vez, mas sim dividido em lotes (*batches*) de 30 dias. Esses lotes são embaralhados aleatoriamente a cada época de treinamento para evitar que o modelo memorize a sequência histórica dos dias. Dessa forma, garante-se que a rede concentre o aprendizado estritamente nos padrões de comportamento intradiário (a ordem das 24 horas), cuja sequência

permanece inalterada. O Quadro 2 apresenta as funções e métodos utilizados na preparação dos dados para o Auto Codificador Variacional.

Quadro 2 – Funções de preparação dos dados para o modelo

Função	Aplicação
<code>scaler = StandardScaler()</code>	Define o método do standard scaler o qual padronização os dados a partir da média e desvio padrão.
<code>dados_{norm} = scaler.fit_transform(dados_{24h})</code>	Aplica o método standard scaler aos dados.
<code>dataset = TensorDataset (torch.FloatTensor(dados_{norm}))</code>	Transforma a disposição dos dados de frame para o formato matricial/tensorial.
<code>dataloader = DataLoader(dataset, batch_{size} = BATCH_{SIZE}, shuffle = True)</code>	Define como será a entrega dos dados ao modelo, neste caso a entrega é realizada em pacotes de 30 dias por vez a partir dos dados do dataset.

Fonte: O Autor

4.3 APLICAÇÃO DO AUTO CODIFICADOR VARIACIONAL AOS DADOS

Essa etapa tem como objetivo descrever o processo de implementação do Auto Codificador Variacional no Python, a partir da biblioteca PyTorch por meio da plataforma Google Colab.

4.3.1 Construção da arquitetura do Auto Codificador Variacional

A primeira etapa da construção do Auto Codificador variacional se dá pela construção da arquitetura do modelo, isso é, a definição do número de camadas, neurônios por camadas e fluxo de dados dentro da rede neural. Os tópicos a seguir terão como objetivo a explanação da construção do modelo.

4.3.1.1 Codificador

Como a análise do estudo é baseada no perfil diário de 24 horas, a camada de entrada do codificador foi projetada com 24 neurônios, correspondendo a cada posto horário do dia. A arquitetura segue um fluxo de compressão onde esses dados são reduzidos inicialmente para uma camada de 16 neurônios, chamada de camada oculta, e para 4 neurônios para camada do espaço latente.

Diferente de um auto codificador comum, o estágio final do codificador mapeia essa representação para um espaço latente de 4 dimensões. Nesta etapa, o modelo gera dois vetores distintos, um representando a média e outro a log-variância da distribuição. No Python trabalha-se com a log-variância por conta da característica estabilidade numérica. Como a rede neural pode gerar valores negativos, prever a log-variância permite aplicar a função exponencial em seguida, assegurando que o resultado da variância seja sempre positivo e evitando erros de cálculo.

As funções utilizadas na construção da arquitetura do codificador estão apresentadas no Quadro 3.

Quadro 3 – Funções de construção do codificador

Função	Aplicação
<code>def __init__(self, input_dim = 24, hidden_dim = 16, latent_dim = 4): super(SimpleVAE, self).__init__()</code>	Responsável por inicializar a classe base <i>nn.Module</i> , permitindo que o modelo herde todas as funcionalidades de redes neurais do <i>PyTorch</i> .
<code>self.fc1 = nn.Linear(input_dim, hidden_dim)</code>	Define o atributo fc1 a camada linear que realiza a transformação dos 24 atributos de entrada para os 16 neurônios da camada oculta.
<code>self.fc_mu = nn.Linear(hidden_dim, latent_dim)</code>	Define o atributo média e sua relação linear entre a camada de oculta e a camada latente.
<code>self.fc_logvar = nn.Linear(hidden_dim, latent_dim)</code>	Define o atributo log-variância e sua relação linear para a dimensão latente.

Fonte: O Autor

A funções *nn.Linear* são responsáveis pela definição do tamanho da dimensão de cada camada do codificador onde em seus argumentos estão a dimensão da camada entrada e saída. Percebe-se que da camada de entrada até o

espaço latente ocorreu a diminuição da dimensão, característica clássica do codificador.

4.3.1.2 Decodificador

O decodificador recebe o vetor de 4 dimensões proveniente do espaço latente z gerado a partir das médias e log-variância e realiza o processo de expansão, simétrico ao codificador. Inicialmente, o vetor latente passa por uma camada oculta de 16 neurônios e posteriormente expandido para as 24 dimensões de saída do modelo.

A camada de saída do decodificador retorna à dimensão original de 24 neurônios, reconstruindo o perfil diário de desvios. As funções utilizadas para construção da arquitetura do decodificador estão apresentadas no Quadro 4.

Quadro 4 – Funções de construção do Decodificador

Função	Aplicação
<code>self.fc3 = nn.Linear(latent_dim, hidden_dim)</code>	Define o atributo fc3, referente a camada entre o espaço latente e a camada oculta, com dimensão de 4 e 16 neurônios respectivamente.
<code>self.fc4 = nn.Linear(hidden_dim, input_dim)</code>	Define o atributo fc4, referente a camada entre o espaço latente e a camada oculta, com dimensão de 16 e 24 neurônios respectivamente.
<code>self.relu = nn.ReLU()</code>	Define a função de ativação Não-Linear do tipo Unidade Linear Retificada, utilizada para permitir que a rede aprenda padrões complexos nos dados.

Fonte: O Autor

A função de ativação ReLU é um componente fundamental na arquitetura de redes neurais profundas, responsável por introduzir não-linearidade ao modelo. Matematicamente, a ReLU tem como objetivo manter valores positivos inalterados e converte qualquer valor negativo diretamente para zero.

4.3.1.3 Fluxo dos dados do modelo

Após definir a arquitetura construtiva das dimensões do modelo, define-se agora o fluxo de informações pelo modelo, essa etapa é definida as conexões entre as variáveis e suas transformações dentro da rede neural, também conhecida como métodos dentro da linguagem Python.

No processo de codificação, define-se o método *encode* (codificador), que recebe como argumento os atributos do codificador a variável de entrada *x* (os dados de geração). Nesta etapa, a função de ativação ReLU é aplicada ao final da camada de entrada para processar as informações que chegam à camada oculta. A partir dessa camada, são extraídas as estimativas da média e da log-variância, que definem o espaço latente. Em resumo, o método *encode* processa os dados de entrada e retorna os parâmetros estatísticos necessários para a fase seguinte. As funções utilizadas para a criação do fluxo do codificador está apresenta no Quadro 5.

Quadro 5 – Funções utilizadas no fluxo de dados do codificador

Função	Aplicação
<i>encode(self, x)</i>	Função que define o método do fluxo de dados na rede neural do codificador.
<i>h = self.relu(self.fc1(x))</i>	Aplica a ativação ReLU sobre a primeira camada linear, introduzindo não-linearidade no processo de compressão dos dados.
<i>self.fc_mu(h1), self.fc_logvar(h1)</i>	Define que a camada oculta gerada a partir da função de ativação deve retornar valores de média de log-variância.

Fonte: O autor

Após a definição do fluxo do codificador, implementa-se a função do truque da reparametrização. Nesta fase, define-se uma variável de ruído, obtida por meio de uma amostragem de uma distribuição Gaussiana padrão. Em seguida, este ruído é multiplicado pelo desvio padrão derivado do espaço latente e somado à média,

resultando no vetor latente z que será decodificado. O Quadro 6 apresenta as funções utilizadas neste processo.

Quadro 6 – Funções utilizadas no fluxo de dados na reparametrização

Função	Aplicação
<i>reparameterize(self, mu, logvar)</i>	Define o método de reparametrização, onde tem como entrada a média e a log-variância.er
<i>std = torch.exp(0.5 * logvar)</i>	Operação matemática que converte a log-variância prevista pela rede em desvio padrão, garantindo estabilidade numérica.
<i>mu + torch.randn_like(std) * std</i>	Gera um vetor de ruído com a mesma dimensão do desvio padrão (std), seguindo uma distribuição normal padrão. E realiza uma combinação linear entre média, desvio padrão e o ruído.

Fonte: O Autor

O fluxo de dados no decodificador assemelha-se, de forma inversa, ao processo do codificador. Para sua implementação, define-se o método *decode* (decodificador), que tem como objetivo decodificar os valores provenientes do espaço latente para os dados de saída originais. Inicialmente, aplica-se a função de ativação *ReLU* à saída do espaço latente para a geração da camada oculta. Posteriormente, os valores dessa camada oculta são projetados para a camada de saída, retornando à dimensão de 24 horas.

É importante destacar que, na camada de saída, não deve ser aplicada a função de ativação *ReLU*, isso se deve ao fato de que os desvios de geração podem assumir valores negativos, enquanto a função *ReLU* os converteria obrigatoriamente em valores nulos. Portanto, a utilização de uma saída linear é fundamental para preservar a integridade física dos dados de desvio e permitir que o modelo represente corretamente. O detalhamento das funções utilizadas neste fluxo está apresentado no Quadro 7.

Quadro 7 – Funções utilizadas no fluxo de dados decodificador

Função	Aplicação
<i>decode(self, z)</i>	Define o método decodificador, que tem como entrada as os atributos da rede neural e os valores do espaço latente z, e retorna a saída dos dados reconstruídos.
<i>self.fc4(self.relu(self.fc3(z)))</i>	Recebe o vetor latente z (4 dimensões), expande para a camada oculta e aplica a ativação ReLU para processar as características extraídas. Por fim, transforma esses dados na representação para 24 horas.

Fonte: O Autor

Após a definição individual das etapas de codificação, reparametrização e decodificação, estabelece-se o fluxo geral da rede, que representa a visão macro do processamento de dados. Para isso, utiliza-se o método *forward*, responsável por interligar os métodos de codificação, reparametrização e decodificação descrita nas etapas anteriores em um ciclo contínuo.

Nesta etapa, os dados de entrada são inicialmente processados pelo método *encode*, que gera os parâmetros de média e log-variância. Estes parâmetros alimentam o método de reparametrização, que retorna o vetor latente z. Por fim, a função *decode* utiliza este vetor para reconstruir o perfil de desvios original. As funções que compõem esse fluxo estão apresentadas no Quadro 8.

Quadro 8 – Funções utilizadas no fluxo da rede

Função	Aplicação
<i>forward(self, x)</i>	Define o método de fluxo de dados dentro da rede neural, tendo como entrada os métodos e dados de entrada <i>x</i> .
<i>mu, logvar = self.encode(x)</i>	Aplica o método <i>encode</i> para retornar a média e a log-variância.
<i>z = self.reparameterize(mu, logvar)</i>	Aplica o método da reparametrização para alimentado pelos dados do <i>encode</i> para gerar o vetor do espaço latente <i>z</i> .
<i>self.decode(z), mu, logvar</i>	Aplica o método <i>decode</i> alimentado pelo vetor do espaço latente <i>z</i> vindo do método da reparametrização, retornando os valores reconstruído, assim como suas médias e log-variância.

Fonte: O autor

4.3.2 Treinamento do modelo Variacional

Após definir a arquitetura do modelo, a etapa de treinamento se faz necessária. Será a partir do treinamento que o modelo terá seus pesos ajustados para que a saída retorne as características dos valores mais próximos possíveis dos valores da entrada.

4.3.2.1 Função de perda do Auto Codificador Variacional

Nesta etapa, são implementados os códigos referentes à perda de reconstrução e à aproximação da distribuição de probabilidade real e gerada pelo modelo do Auto Codificador Variacional.

Define-se função denominada *vae_loss*, que possui como entrada as variáveis dos dados reconstruídos, os dados reais, a média e a log-variância. Dentro desta função, o algoritmo realiza o cálculo do erro de reconstrução, considerando o somatório das 24 horas, e o cálculo do erro da divergência KL. A função retorna a

perda total do Auto Codificador Variacional, que é o resultado da soma do erro de reconstrução com a divergência KL. As funções utilizadas nesta etapa para a construção da perda estão apresentadas no Quadro 9.

Quadro 9 – Funções utilizadas na função de perdas

Função	Aplicação
<i>mse</i> <i>= nn.functional.mse_loss(recon_x,</i> <i>x, reduction = 'sum')</i>	Calcula a diferença entre o dado original (x) com os reconstruídos (recon_x), o argumento <i>reduction = 'sum'</i> faz com que os erros de cada hora sejam somados.
<i>kld = -0.5 * torch.sum(1 + logvar</i> <i>– mu.pow(2)</i> <i>– logvar.exp())</i>	Calcula o valor da divergência KL, a partir das variáveis de média e log-variância.

Fonte: O Autor

4.3.2.2 Otimizador do Modelo

Para a otimização do modelo, optou-se pela utilização do algoritmo Adam (Adaptive Moment Estimation). Este otimizador é responsável por realizar o ajuste iterativo dos pesos sinápticos da rede neural, com o objetivo central de minimizar a função de perda, neste contexto, representa o erro acumulado na reconstrução do perfil diário de 24 horas.

Um dos hiperparâmetros fundamentais configurados nesta etapa é a taxa de aprendizado. Esse parâmetro controla a magnitude das atualizações dos pesos a cada iteração do algoritmo. O ajuste adequado do lr é essencial para a convergência do modelo. Um valor muito elevado pode impedir que o algoritmo encontre o ponto de erro mínimo, gerando instabilidade matemática, enquanto um valor demasiadamente baixo pode tornar o processo de treinamento excessivamente lento.

As principais funções utilizadas no código para operacionalizar o otimizador e o cálculo dos gradientes estão detalhadas no Quadro 10.

Quadro 10 – Funções utilizadas no otimizador

Função	Aplicação
<code>optimizer</code> <code>= torch.optim.Adam(modelo.parameters(),</code> <code>lr = LR)</code>	O otimizador Adam e define os parâmetros da rede a serem otimizados, juntamente com a taxa de aprendizado lr.
<code>modelo.train()</code>	Configura a arquitetura da rede neural para o modo de treinamento.

Fonte: O Autor

4.3.2.3 Treinamento do Modelo

O processo de treinamento do modelo desenvolve-se de maneira iterativa, sendo estruturado em ciclos contínuos denominados épocas (*epochs*). Uma época corresponde a uma passagem completa de todo o conjunto de dados históricos pela arquitetura da rede neural. A repetição desse ciclo não tem o objetivo de gerar novos dados sintéticos durante o treino, mas sim permitir que o algoritmo reveja a base histórica múltiplas vezes. Isso possibilita o ajuste progressivo dos pesos sinápticos, refinando a capacidade do modelo de extrair os padrões de comportamento da geração de energia e garantindo a sua convergência matemática.

Durante a execução de cada época, o conjunto de dados não é processado em sua totalidade de uma única vez. A alimentação da rede ocorre de forma fracionada, utilizando lotes (*batches*) compostos por 30 dias. Para cada lote processado, o modelo gera a reconstrução dos dados e as variáveis latentes para, em seguida, calcular a perda (*loss*) associada a esse pacote, avaliando a disparidade entre o dado gerado e o real. A partir dessa métrica, executa-se a retropropagação do erro (*backpropagation*), identificando a magnitude da correção necessária na rede. Imediatamente após essa etapa, o otimizador realiza o reajuste dos parâmetros com o intuito de minimizar as falhas nas iterações subsequentes.

Por fim, o erro de todos os lotes é acumulado. Ao término de cada época, calcula-se a perda média do período, métrica que é registrada em um histórico geral. Esse acompanhamento contínuo é vital para monitorar a evolução do aprendizado do

modelo ao longo do tempo, assegurando que o erro tenda a diminuir progressivamente ao longo do treinamento.

O Quadro 11 apresenta as funções utilizadas na etapa de treinamento do modelo do Auto Codificador Variacional.

Quadro 11 – Funções utilizadas no treinamento do modelo

Função	Aplicação
<i>modelo(x_batch)</i>	Realiza a propagação do lote de dados pela arquitetura do Auto Codificador Variacional, retornando à reconstrução dos dados e as variáveis do espaço latente média e log-variância.
<i>vae_loss()</i>	Calcula a função de perda específica do modelo, quantificando a divergência entre os dados reais de entrada e os dados reconstruídos.
<i>optimizer.zero_grad()</i>	Limpa os gradientes matemáticos acumulados da iteração anterior, preparando a rede neural para um novo cálculo livre de interferências do lote passado.
<i>loss.backward()</i>	Executa a retropropagação do erro, calculando a magnitude do ajuste necessário para cada peso sináptico da rede.
<i>optimizer.step()</i>	Efetua a atualização definitiva dos pesos e parâmetros do modelo com base nos gradientes calculados na etapa de retropropagação.
<i>loss.item()</i>	Extrai o valor numérico escalar do tensor computacional de erro, viabilizando o acúmulo algébrico para o cálculo da perda média ao final de cada época.

Fonte: O autor

Concluída a etapa de treinamento, o modelo encontra-se apto para iniciar sua operação generativa. Nesse estágio, os pesos da rede neural já estão devidamente ajustados, de modo que, teoricamente, os dados de entrada e de saída devem

apresentar comportamentos análogos. Dessa forma, a saída do modelo torna-se capaz de retornar os cenários sintéticos de desvio de geração.

4.4 GERAÇÃO DE CENÁRIOS.

Após a estruturação e o treinamento do modelo de redes neurais generativas, avança-se para o objetivo principal desta etapa: a geração de cenários com base nos dados históricos disponíveis. Nesse momento, o Auto Codificador Variacional já assimilou as características da base de treinamento, indicando que o espaço latente construído pelo codificador se encontra consolidado. Desse modo, o processo passa a operar exclusivamente por meio do decodificador. A partir de então, realiza-se a amostragem aleatória de valores dentro da distribuição de probabilidade do espaço latente, os quais são processados e reconstruídos pelo decodificador, resultando na geração de um novo cenário sintético a cada execução.

4.4.1 Criação dos cenários

Inicialmente, a rede neural é configurada para o seu modo de avaliação, desativando-se o cálculo de gradientes. Essa configuração otimiza o processamento e a memória, garantindo que os pesos do modelo não sejam mais alterados após o treinamento.

A amostragem é realizada por meio da extração de vetores aleatórios baseados em uma distribuição normal padrão, de acordo com as dimensões do espaço latente e a quantidade de 2000 cenários. Esses vetores são inseridos no decodificador, que projeta as curvas sintéticas. Como os dados de saída da rede neural encontram-se em uma escala normalizada, a etapa final consiste em aplicar a transformação inversa. Esse procedimento reverte a normalização, convertendo os cenários gerados de volta para a sua grandeza física original em Megawatts, resultando em uma matriz que correlaciona o volume de cenários com as horas operacionais.

O Quadro 12 apresenta as funções utilizadas para a prospecção dos cenários de desvios de geração a partir do espaço latente e reconstruído pelo decodificador.

Quadro 12 – Funções utilizadas para a prospecção de cenários

Função	Aplicação
<code>modelo.eval()</code>	Desativa o modo treino do modelo.
<code>with torch.no_grad()</code>	Informa que os pesos dos neurônios não devem ser mais modificados. Os gradientes não são mais calculados.
<code>z = torch.randn(N_CENARIOS, LATENT_DIM)</code>	Sorteia os valores do espaço latente a partir de uma distribuição normal padrão N-CENARIOS vezes, com o valor do espaço latente.
<code>cenarios_norm = modelo.decode(z).numpy()</code>	Reconstrói os dados sorteados do espaço latente pelo decodificador
<code>scaler.inverse_transform(cenarios_norm)</code>	Transforma os dados normalizados para a escala original.

Fonte: O autor

5 RESULTADOS E DISCUSSÃO

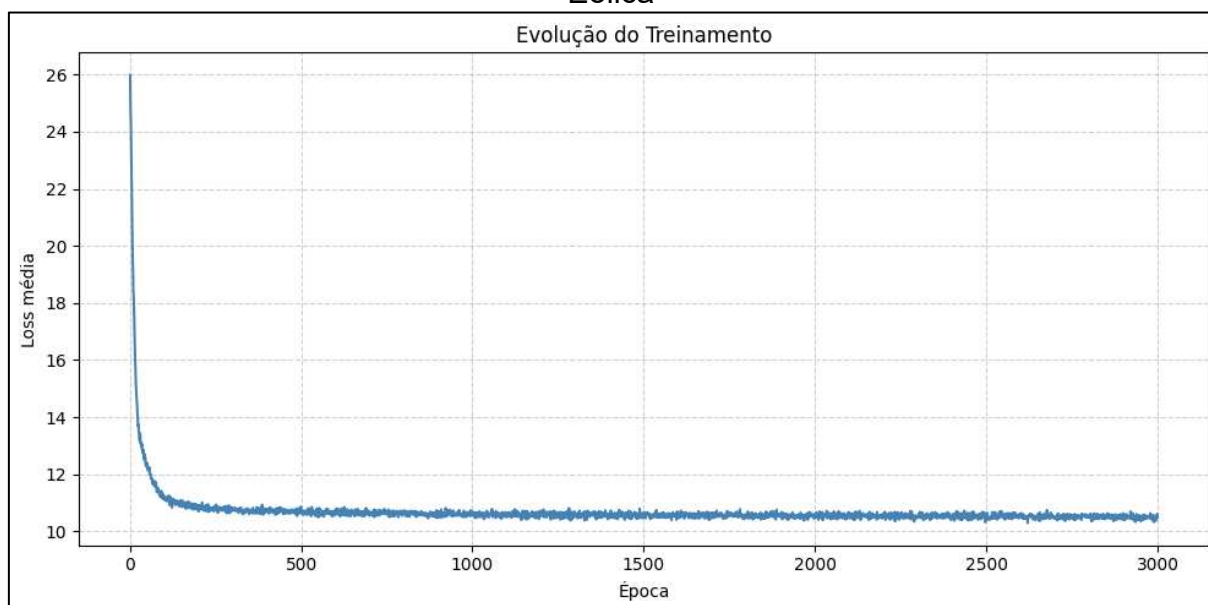
Nesta etapa estão apresentados os resultados obtidos após a construção do modelo, assim como a escolhas e avaliações dentro da arquitetura do Auto Codificador Variacional.

5.1 TREINAMENTO DO MODELO

A definição do número de épocas de treinamento foi baseada na análise da curva de erro. Essa curva apresenta uma característica assintótica, isso indica que, à medida que o número de épocas aumenta, a variação entre o erro atual e o erro da época anterior tende a diminuir progressivamente, até que o erro do modelo tente a estabilizar em um valor.

A Figura 16 apresenta a evolução do erro para os dados de desvio de geração eólica. Observa-se que, nas primeiras 100 épocas, o erro possui uma variação acentuada. Esse comportamento é normal, e reflete o ajuste inicial dos pesos da rede neural, que ainda está na fase de mapeamento e reconhecimento dos padrões dos dados. A partir da marca de 500 épocas, a curva atinge um patamar de estabilidade, e a redução do erro passa a ser muito pequena. Isso indica que treinar o modelo por 500 épocas já é suficiente para alcançar o resultado semelhante ao de 3.000 épocas, garantindo economia de tempo e maior eficiência computacional.

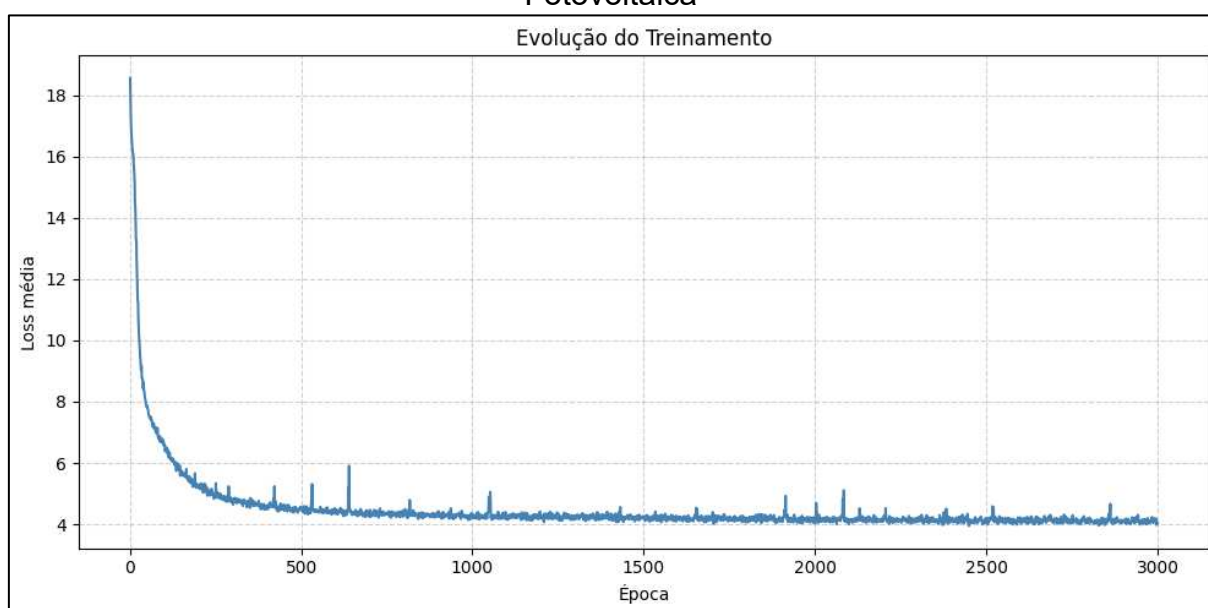
Figura 16 – Evolução do erro do modelo ao longo das épocas de desvio de geração Eólica



Fonte: O autor

Comportamento semelhante foi observado durante o treinamento do mesmo modelo com os dados de desvio de geração fotovoltaica. Embora se trate de um conjunto de dados com características distintas, a dinâmica de aprendizado do modelo seguiu o mesmo padrão, com convergência progressiva do erro ao longo das épocas, porém assumindo valores diferentes. Esse resultado pode ser verificado na Figura 17.

Figura 17 – Evolução do erro do modelo ao longo das épocas de desvio de geração Fotovoltaica

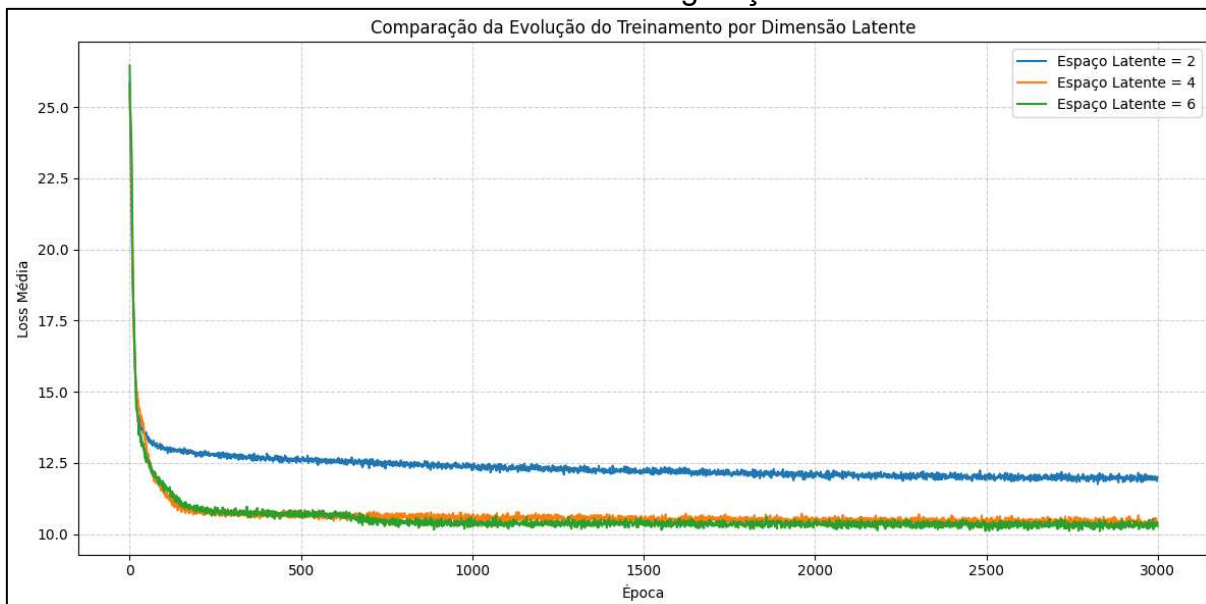


Fonte: O Autor.

A definição da dimensão do espaço latente também foi avaliada a partir da evolução do erro médio por época. A Figura 18 apresenta a comparação entre três diferentes dimensões latentes: 2, 4 e 6. Observa-se que todas as configurações apresentam um aprendizado estável, com curvas de erro que decrescem de forma consistente até a convergência. Nota-se, entretanto, que o modelo configurado com apenas duas dimensões estaciona em um patamar de erro ligeiramente superior às demais, o que sugere uma limitação na sua capacidade de mapear toda a variabilidade dos dados de desvio.

Em contrapartida, as dimensões 4 e 6 permitem que o modelo alcance valores de erro final inferiores, demonstrando uma melhor capacidade de reconstrução dos cenários. A escolha resultante foi pela dimensão 4, visto que o ganho de desempenho em relação à dimensão 6 mostrou-se marginal e pouco significativo. Dessa forma, a dimensão 4 foi selecionada por representar o melhor equilíbrio entre a simplicidade da arquitetura e a precisão na representação dos desvios de geração.

Figura 18 – Comparação entre a evolução do treinamento em diferentes dimensões com dados de desvio de geração Eólica



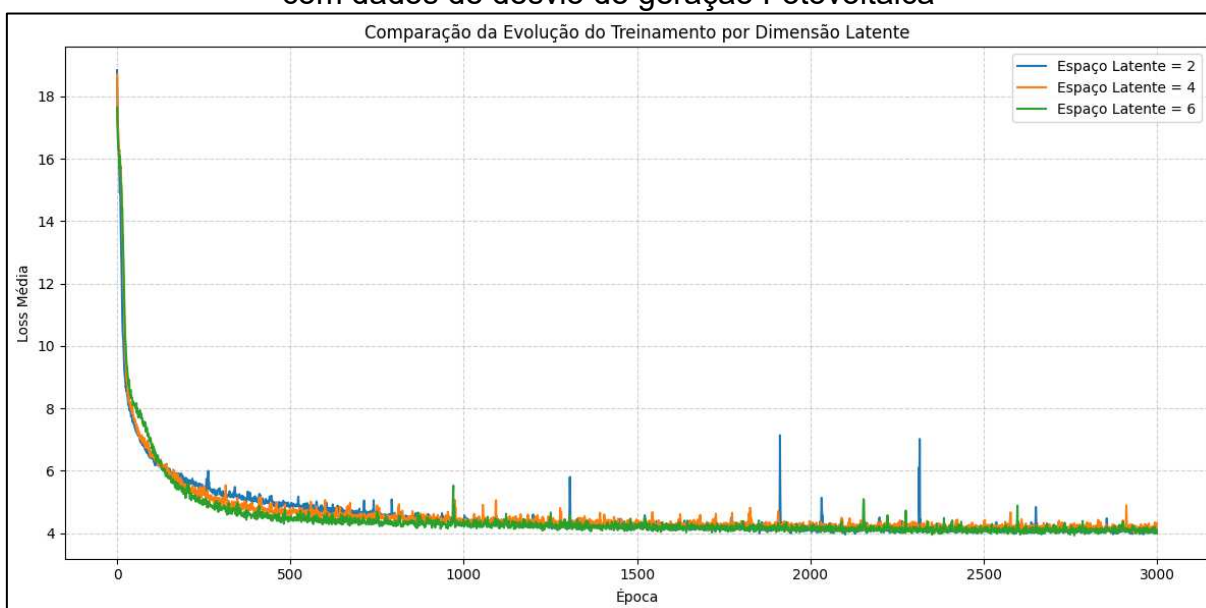
Fonte: O Autor

Para os dados de desvio de geração fotovoltaica, o comportamento geral do treinamento do modelo com diferentes dimensões é análogo ao observado para os dados de geração eólica. No entanto, nesse caso, a variação do erro médio entre as dimensões é menor, conforme ilustrado na Figura 19.

Um aspecto relevante que se destaca nessa análise é a estabilidade do processo de aprendizado em função da dimensão do espaço latente. O modelo configurado com duas dimensões latentes apresentou maior instabilidade, com oscilações no erro médio ao longo do treinamento, nas quais o valor do erro volta a crescer em determinadas épocas, desviando do comportamento assintótico esperado. Esse fenômeno foi observado com menor amplitude nos modelos com espaço latente de 4 e 6 dimensões, que demonstraram maior regularidade e estabilidade na convergência.

A análise comparativa revela que o aumento do espaço latente para 6 dimensões não resulta em melhorias expressivas no treinamento quando comparado à configuração de 4 dimensões. Ou seja, o ganho na precisão do modelo é muito pequeno e não justifica o acréscimo na complexidade da rede neural. Sendo assim, optou-se por manter a arquitetura com 4 dimensões, seguindo o mesmo critério adotado para a fonte eólica. Essa decisão baseia-se no princípio de escolher o modelo mais simples e com menos parâmetros, garante-se uma maior eficiência computacional e um treinamento mais rápido, sem abrir mão da qualidade e da estabilidade na geração dos cenários.

Figura 19 – Comparação entre a evolução do treinamento em diferentes dimensões com dados de desvio de geração Fotovoltaica



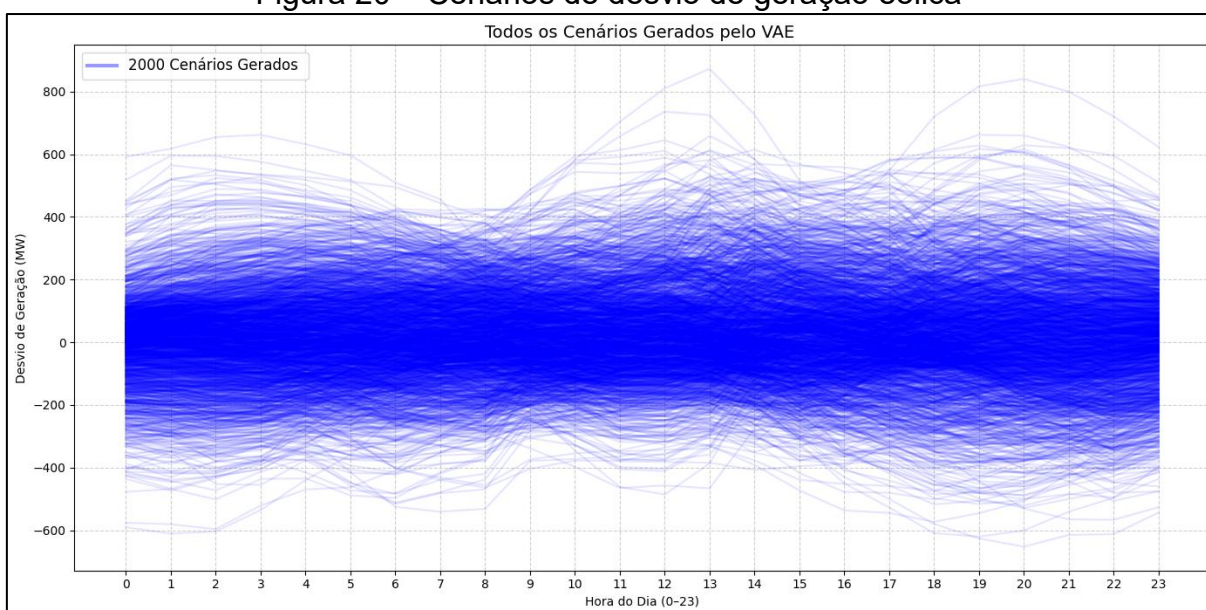
Fonte: O Autor.

5.2 PROSPECÇÃO DE CENÁRIOS

Com o modelo generativo devidamente estruturado e o treinamento concluído, inicia-se a etapa de prospecção de cenários. Neste estágio, a arquitetura do codificador, cuja função principal foi a construção e o mapeamento do espaço latente a partir dos dados históricos, deixa de ser utilizada. A geração de novos dados fundamenta-se, portanto, exclusivamente na interação entre o espaço latente consolidado e o decodificador.

Os cenários sintéticos de desvios de geração eólica, gerados pelo modelo, estão apresentados na Figura 20. Observa-se que as curvas produzidas exibem uma distribuição ampla, sem a repetição de padrões rígidos ou tendenciosidades visíveis. Esse comportamento indica uma maior variação dos valores, característica que corrobora a natureza da fonte eólica. Tal resultado é condizente com a análise preliminar dos dados históricos de sete dias, apresentada anteriormente na Figura 13.

Figura 20 – Cenários de desvio de geração eólica



Fonte: O autor

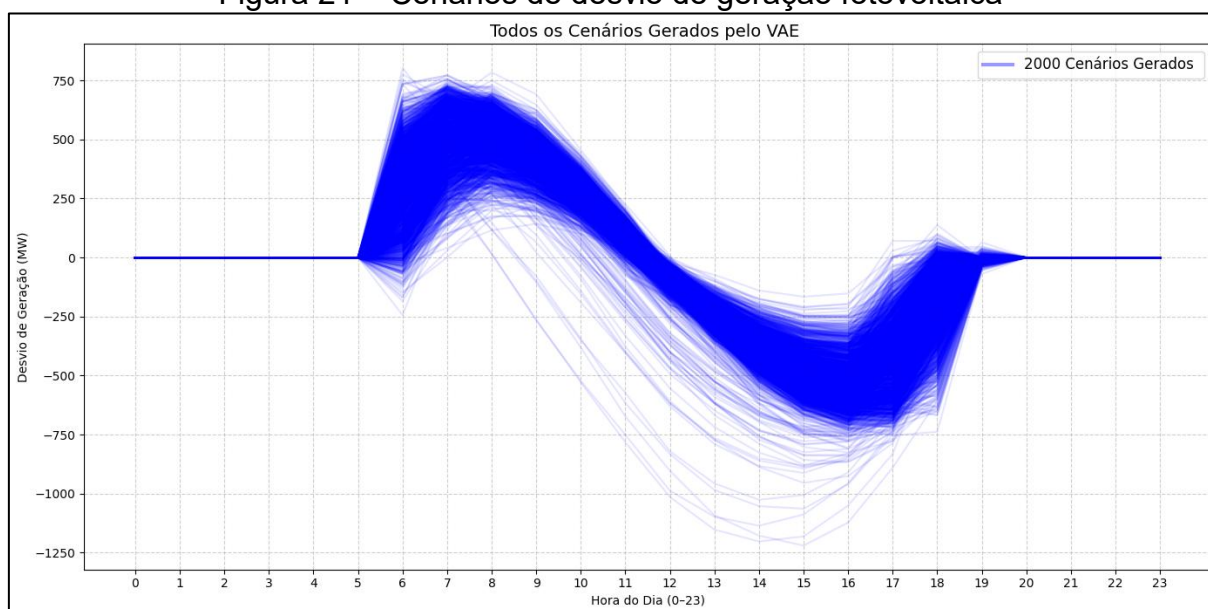
Diferentemente do comportamento observado na fonte eólica, os desvios de geração fotovoltaica apresentam uma forte característica de sazonalidade diária, conforme discutido previamente na Figura 14. Esse padrão é intrínseco à natureza da fonte, visto que a geração atinge seus níveis máximos nos períodos de maior irradiância solar. Consequentemente, qualquer oscilação na cobertura de nuvens ou

imprecisão na previsão meteorológica resulta em desvios de maior magnitude nessas janelas horárias.

Adicionalmente, nota-se que os valores máximos de desvio tendem a ocorrer com maior frequência nos períodos intermediários da manhã ou do meio da tarde. Esse comportamento evidencia que a previsão de geração é menos precisa durante as fases de transição operacional da usina, isto é, nas rampas de subida de geração e nas rampas de descida, momentos em que a taxa de variação da incidência solar exige maior sensibilidade dos modelos de previsão tradicionais.

Os cenários sintéticos gerados para a fonte fotovoltaica estão apresentados na Figura 21. Observa-se que o modelo foi capaz de reproduzir fielmente essa sazonalidade diária, apresentando curvas que se assemelham a um comportamento senoidal durante o período diurno e mantendo o desvio nulo nos períodos sem incidência solar, o qual tem-se total certeza do valor. Essa aderência do comportamento aos dados históricos indica que o Auto Codificador Variacional assimilou a dinâmica operacional da geração solar e eólica, validando a representatividade dos cenários produzidos frente à realidade física do sistema.

Figura 21 – Cenários de desvio de geração fotovoltaica



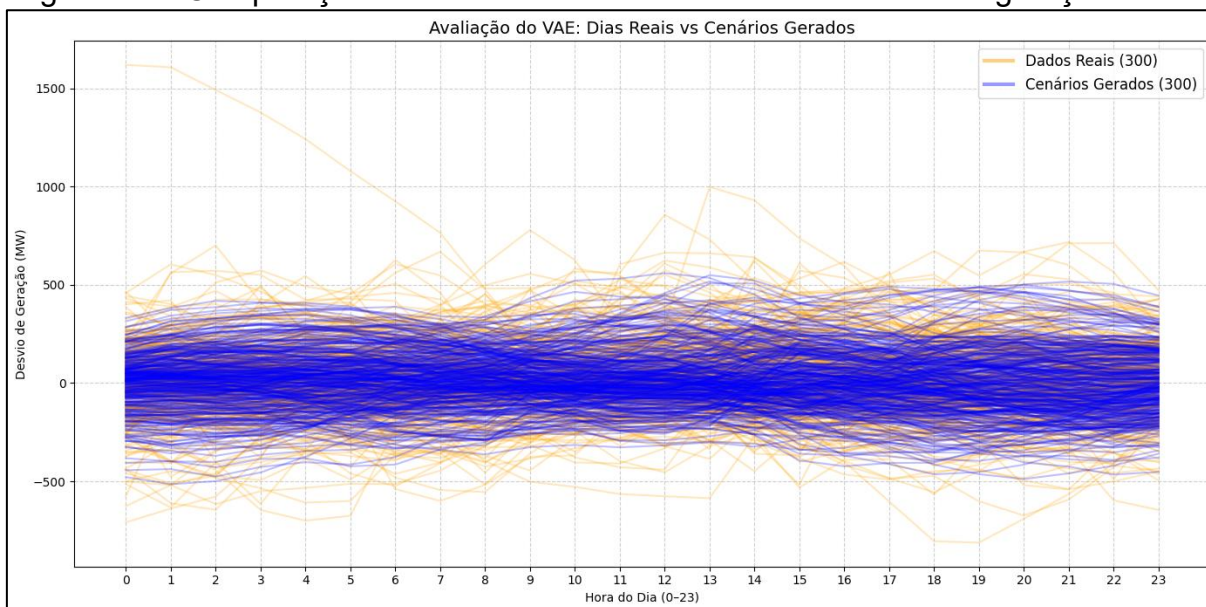
Fonte: O autor

Comparação entre os valores verificados e cenários gerados

Embora a análise individual dos cenários sintéticos indique uma coerência com as características físicas das fontes de geração, essa avaliação preliminar não é suficiente para atestar, de forma definitiva, que o modelo assimilou as propriedades essenciais dos dados históricos. Para validar se a magnitude dos cenários de desvio atende à realidade operacional do sistema, é necessário realizar um confronto visual direto, sobrepondo os dados sintéticos aos dados reais verificados.

A Figura 22 e a Figura 23 apresentam essa comparação para os desvios de geração eólica e fotovoltaica, respectivamente. Em ambas as análises, foram plotados 300 cenários gerados pelo modelo, representados pelas linhas em azul sobrepostos a 300 dias reais extraídos do histórico com as linhas em amarelo, permitindo avaliar a aderência das curvas geradas. Optou-se por representar apenas 300 cenários e dados reais no gráfico para melhor visualização, para evitar alguma tendência, os valores foram sorteados dentro do conjunto disponível de forma aleatória, assim como foi realizado com o sorteio dos vetores do espaço latente.

Figura 22 – Comparação entre dados reais e cenários de desvios de geração eólica

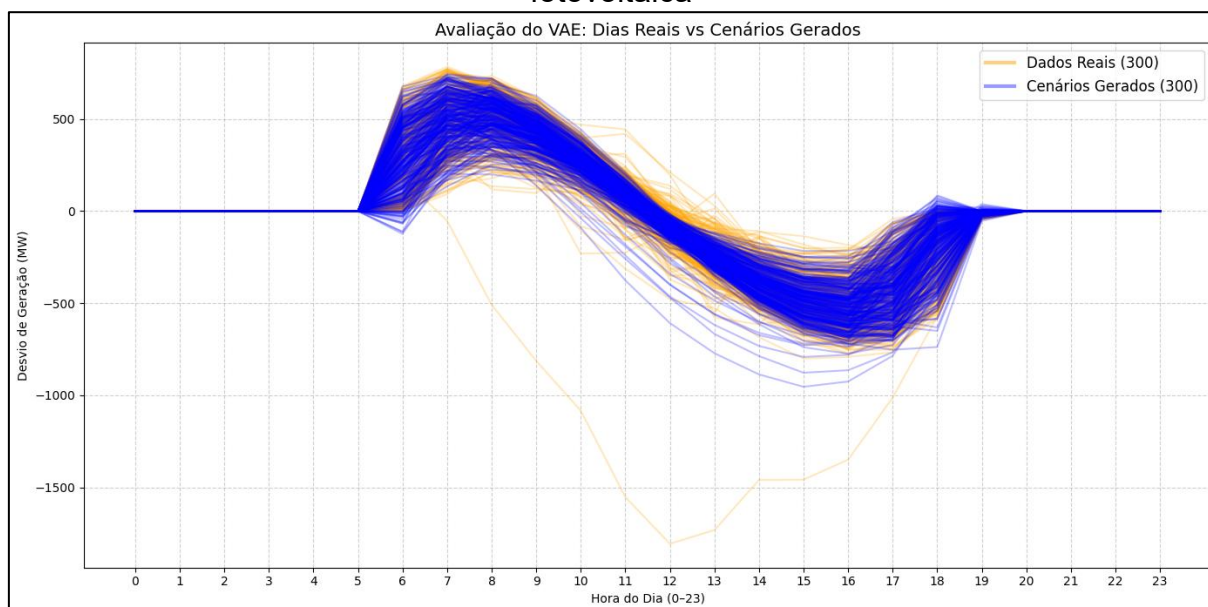


Fonte: O autor

Ao analisar a Figura 22, nota-se que as trajetórias dos cenários sintéticos eólicos se mantêm estritamente dentro da mesma ordem de grandeza e do mesmo envelope de dispersão dos dados reais. Isso indica que, além de capturar a característica estocástica do vento, o modelo foi capaz de mapear com precisão a

amplitude escalar dos desvios de geração eólica, sem extrapolar os limites físicos observados no histórico.

Figura 23 – Comparação entre dados reais e cenários de desvios de geração fotovoltaica



Fonte: O autor

Comportamento semelhante é verificado na Figura 23, referente à geração fotovoltaica. A sobreposição evidencia que o modelo não apenas respeita a sazonalidade diária com variações concentradas nos horários de incidência solar, mas também replica fielmente os limites inferiores e superiores dos desvios reais. A correspondência entre as áreas de maior densidade das curvas azuis e amarelas confirma a capacidade do Auto Codificador Variacional de gerar dados robustos e representativos para ambas as fontes.

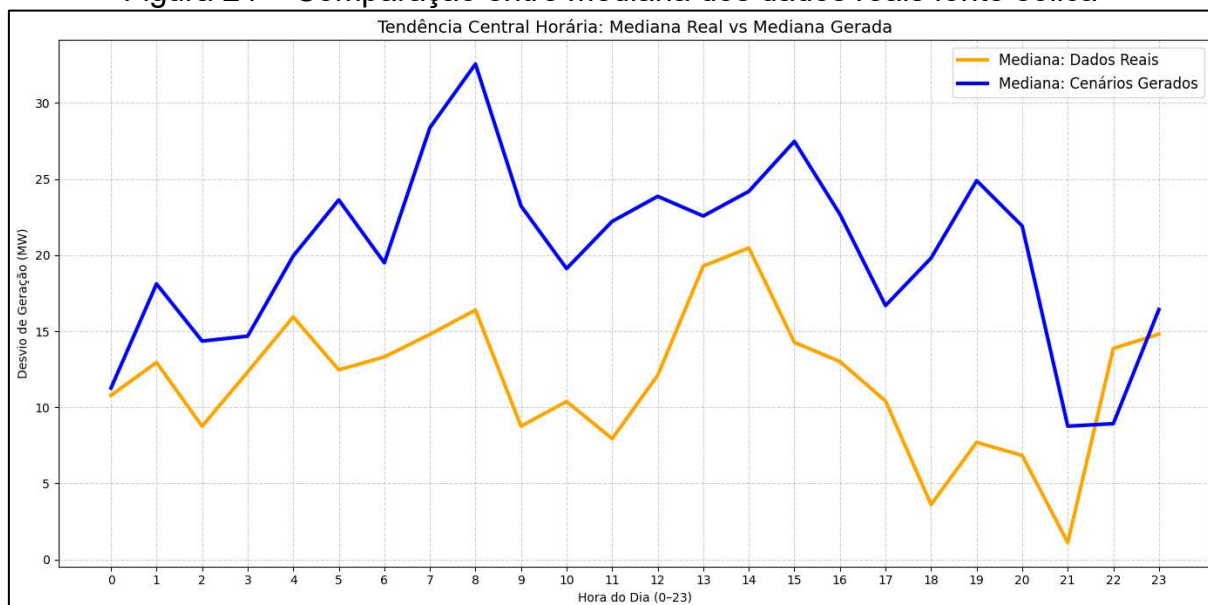
Comparação entre a mediana dos valores verificados com a mediana dos cenários gerados.

Uma métrica robusta para avaliar a proximidade entre os valores reais históricos e os cenários sintéticos é a mediana. Como medida de tendência central, a mediana apresenta a vantagem de não sofrer forte influência de valores extremos, representando o valor exato que divide a distribuição dos dados ao meio. Dessa forma, ela permite verificar se o comportamento central esperado para cada hora de operação foi adequadamente assimilado pelo modelo. O trabalho de (Pessanha *et al.*, 2021) que gerou cenários de geração eólica, e também apresentou a mediana para

validação dos cenários gerados em seu modelo para verificar a tendência central de dados.

A Figura 24 apresenta a evolução horária da mediana dos dados reais em comparação com a mediana dos cenários gerados para a fonte eólica. Observa-se que a curva apresenta alta variabilidade e ausência de um padrão cíclico definido, características inerentes à estocasticidade dos ventos. Ao analisar a curva gerada pelo modelo, constata-se que ela acompanha a dinâmica temporal dos dados reais, reproduzindo a mesma tendência de picos e vales ao longo do dia. Nota-se, contudo, a presença de um deslocamento vertical quase constante em relação aos dados reais, indicando que, embora o Auto Codificador Variacional tenha capturado o comportamento horário do fenômeno, ele superestima levemente a tendência central dos desvios eólicos.

Figura 24 – Comparação entre mediana dos dados reais fonte eólica

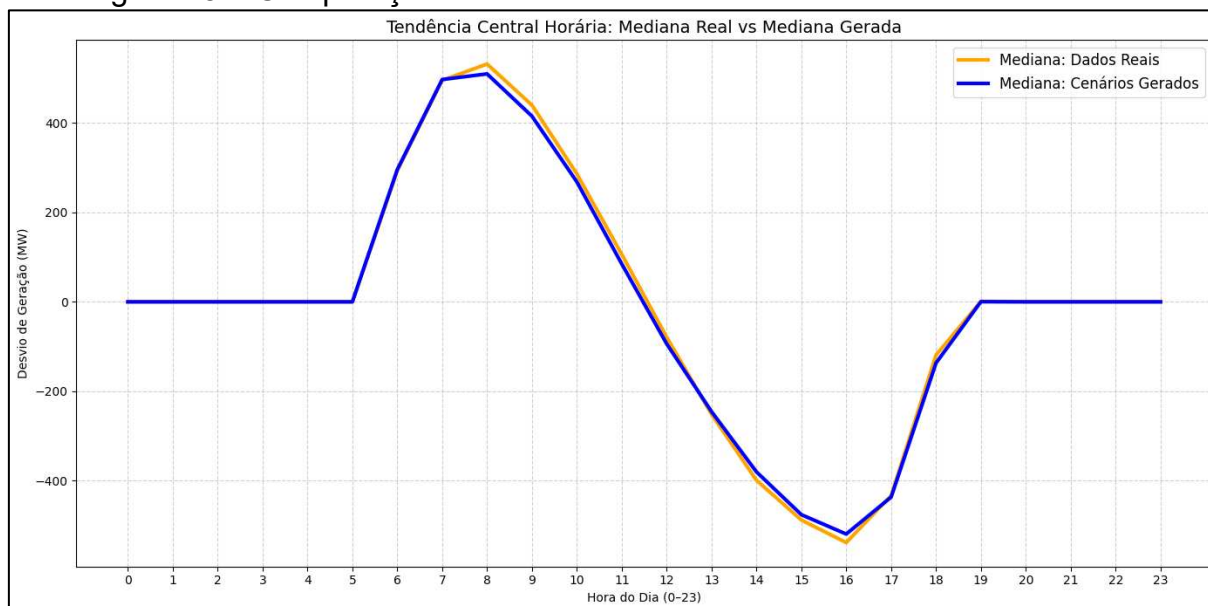


Fonte: O Autor

A mesma análise de tendência central foi aplicada aos dados da fonte fotovoltaica, com os resultados apresentados na Figura 25. Diferentemente da variabilidade observada na fonte eólica, constata-se uma aderência entre a mediana dos dados históricos e a mediana dos cenários sintéticos. Esse elevado grau de precisão já era esperado, por apresentar uma sazonalidade diária bem definida e um padrão estrutural fortemente atrelado ao ciclo solar, a dinâmica da geração

fotovoltaica possui um perfil mais previsível em termos de sua tendência central, assim, facilitando para que o modelo capture as características inerentes dos dados.

Figura 25 – Comparação entre mediana dos dados reais fonte fotovoltaica



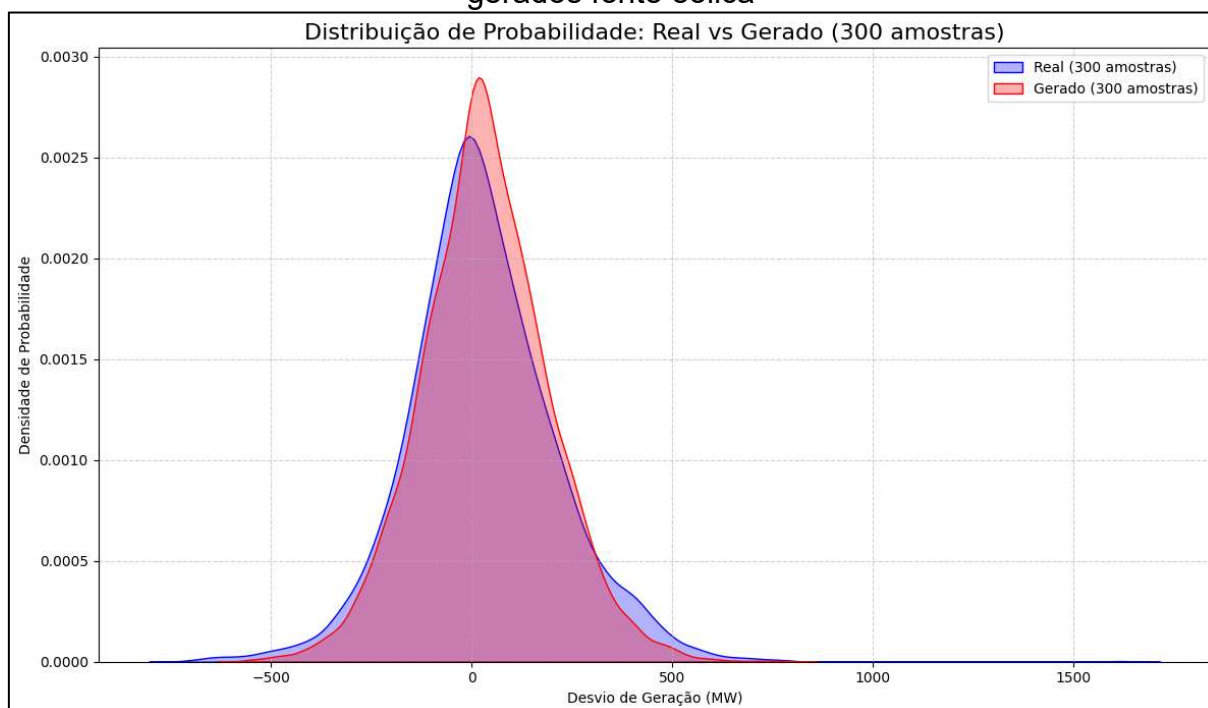
Fonte: O Autor

Comparação entre as distribuições de probabilidades dos cenários e dados reais.

De forma complementar às análises temporais, a validação da consistência estatística do modelo pode ser mensurada pela comparação das Funções de Densidade de Probabilidade dos dados reais frente aos cenários gerados. Essa análise permite visualizar a frequência de ocorrência e a dispersão das magnitudes dos desvios, confirmando se o modelo reproduz a macroestrutura estatística do fenômeno.

A Figura 26 ilustra essa comparação para a fonte eólica. Observa-se que a distribuição dos dados reais (em azul) possui uma semelhança com uma distribuição normal, apresentando um formato de sino, ainda que com uma leve assimetria. A distribuição dos dados sintéticos (em vermelho) segue fielmente a mesma característica, capturando as caudas da distribuição. Embora o modelo concentre uma densidade marginalmente maior na região central do que os dados reais, o nível de sobreposição atesta que a rede obteve um resultado satisfatório e próximo à realidade.

Figura 26 – Comparação entre distribuição de probabilidades dos dados reais e gerados fonte eólica

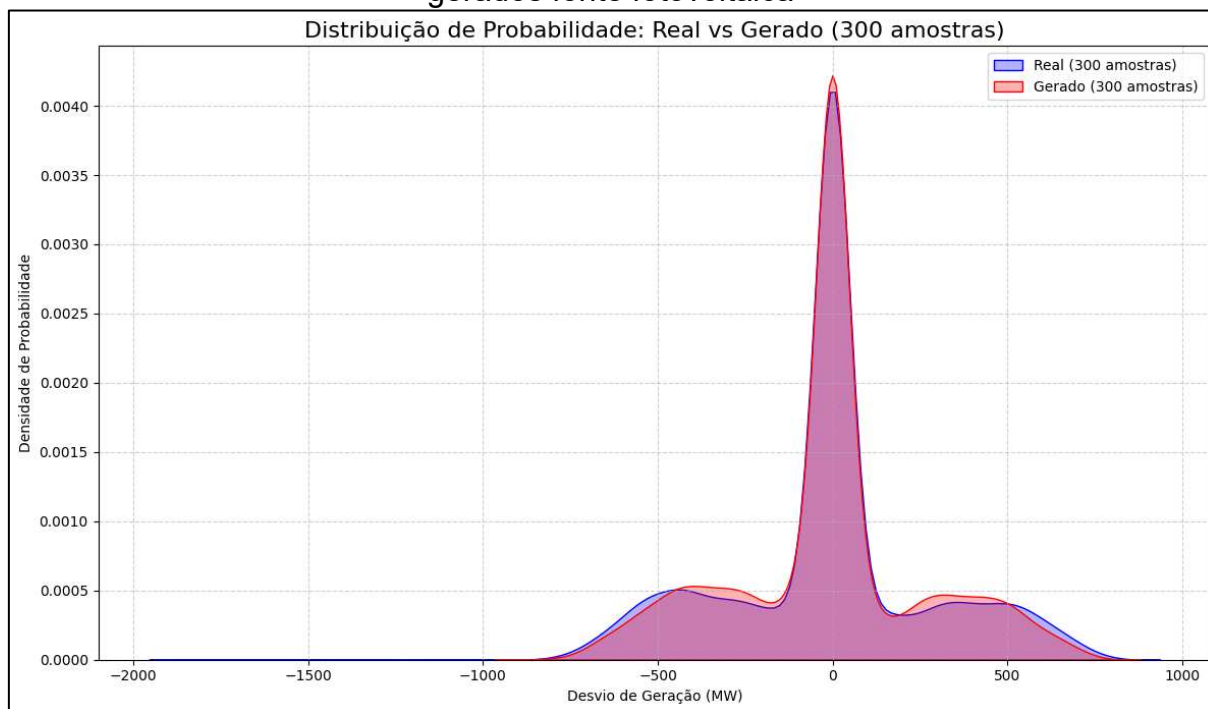


Fonte: O Autor

Um aspecto fundamental a ser destacado na Figura 27, que ilustra as distribuições de probabilidade dos desvios de geração fotovoltaica, é a capacidade do modelo de representar a forte característica não-normal desses dados. É muito interessante observar que, embora o Auto Codificador Variacional (VAE) realize a sua amostragem inicial a partir de uma distribuição normal padrão no espaço latente, o modelo não fica engessado por essa restrição matemática. Na prática, o decodificador consegue transformar esses dados, fazendo com que a distribuição de probabilidade final dos cenários gerados assuma uma forma geométrica totalmente distinta e fiel à complexidade dos dados históricos originais.

Esse fenômeno é explicado pela formulação matemática da função de perda do modelo, que não se baseia unicamente na comparação direta entre duas distribuições. O treinamento minimiza um erro composto, a Divergência de Kullback-Leibler atua regularizando o espaço latente para que se assemelhe a uma normal, enquanto o erro de reconstrução treina os pesos do decodificador para atuar como um aproximador universal. Assim, o decodificador aprende a mapear e distorcer as amostras da curva normal, transformando-as na complexa distribuição real dos dados históricos.

Figura 27 – Comparação entre distribuição de probabilidades dos dados reais e gerados fonte fotovoltaica



Fonte: O autor

6 CONCLUSÕES E RECOMENDAÇÕES

6.1 CONCLUSÕES

Este trabalho teve como objetivo geral o desenvolvimento e a avaliação de um modelo computacional baseado em Auto Codificadores Variacionais (VAE) para a prospecção de cenários sintéticos de desvios de geração eólica e fotovoltaica no Subsistema Sul do Sistema Interligado Nacional (SIN). A partir do cumprimento das etapas propostas e da análise dos resultados obtidos, conclui-se que a ferramenta generativa desenvolvida cumpriu com sucesso o seu propósito, mostrando-se viável, robusta e eficiente para a modelagem de incertezas de fontes renováveis.

Como primeiro passo metodológico, as bases de dados históricos de geração prevista e verificada disponibilizadas pelo Operador Nacional do Sistema (ONS) foram concentradas e tratadas. A partir desses dados, foi possível calcular com precisão as séries temporais dos desvios horários para ambas as fontes operantes na região Sul do país. Esse processo de estruturação dos dados foi fundamental para fornecer ao modelo generativo uma base de treinamento sólida e representativa da realidade física do sistema.

A aplicação prática do VAE sobre essa base de dados permitiu uma análise detalhada para definir a arquitetura da rede neural. O monitoramento das curvas de aprendizado ao longo das épocas revelou um comportamento assintótico estável para ambas as fontes. Constatou-se que um treinamento com 500 épocas garantiu a convergência do modelo com o mesmo nível de precisão de um treinamento estendido de 3000 épocas, representando uma economia de custo computacional. Além disso, os testes com diferentes tamanhos de espaço latente mostraram que a configuração com 4 dimensões entregou o equilíbrio ideal entre estabilidade e capacidade de reconstrução, evitando os picos de erro observados na dimensão 2 e a complexidade desnecessária da dimensão 6.

A partir dessa arquitetura otimizada, realizou-se a prospecção de 2000 cenários de desvio para cada fonte, cuja avaliação estatística e visual validou a aderência do modelo frente aos dados históricos. Para a fonte eólica, os cenários sintéticos gerados capturaram com precisão a natureza altamente estocástica e a grande variabilidade dos ventos, mantendo-se estritamente dentro da ordem de grandeza escalar e do envelope de dispersão real. Na análise de tendência central, a

mediana do modelo eólico acompanhou com fidelidade a dinâmica temporal dos picos e vales do histórico, apresentando apenas um viés positivo leve.

No caso da fonte fotovoltaica, o desempenho do VAE superou as expectativas iniciais. O modelo reproduziu fielmente a forte característica de sazonalidade diária, com erro zero no período noturno, e mapeou as maiores magnitudes de desvio justamente nas janelas de transição operacional, nas rampas de subida e descida da manhã e da tarde, respectivamente, onde os modelos tradicionais de previsão aparentemente enfrentam maior dificuldade. A mediana dos cenários fotovoltaicos apresentou uma sobreposição quase perfeita com os dados reais, confirmando a alta previsibilidade de sua tendência central estrutural.

Por fim, o grande destaque do VAE consolidou-se na análise das distribuições de probabilidade. Mesmo realizando sua amostragem inicial a partir de uma distribuição normal padrão no espaço latente, o decodificador da rede provou sua capacidade como aproximador ao transformar essas amostras e replicar a distribuição não-normal da fonte fotovoltaica. Portanto, os cenários prospectados não são meras repetições do histórico, mas sim dados sintéticos inéditos que respeitam as propriedades estatísticas do fenômeno real.

Diante do exposto, o modelo desenvolvido consolida-se como uma contribuição prática e de baixo custo computacional para o setor elétrico. A capacidade de gerar milhares de cenários realistas de desvio de forma rápida pode auxiliar agentes e planejadores em estudos de análise de risco, programação diária da operação, dimensionamento de sistemas de armazenamento e avaliação da confiabilidade do Subsistema Sul.

6.2 RECOMENDAÇÕES

Como trabalhos futuros, recomenda-se a inclusão de variáveis meteorológicas exógenas como temperatura, vento e nebulosidade, como dados de entrada junto aos desvios de geração, visando refinar o mapeamento das condições físicas no espaço latente. Além disso, sugere-se comparar o VAE com outras arquiteturas generativas, como GANs, avaliando tanto o rigor estatístico quanto o custo computacional. Por fim, propõe-se expandir a metodologia aos demais subsistemas do Sistema Interligado Nacional (SIN), a fim de validar a robustez e a capacidade de generalização do modelo frente a diferentes regimes climáticos e perfis de desvio de geração.

REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA DE ENERGIA EÓLICA (ABEEÓLICA). **ABEEólica - Associação Brasileira de Energia Eólica**. São Paulo: ABEEólica, 2025. Disponível em: <https://abeeolica.org.br>. Acesso em: 25 jan. 2025.
- ASSOCIAÇÃO BRASILEIRA DE ENERGIA SOLAR FOTOVOLTAICA (ABSOLAR). **Energia Solar Fotovoltaica no Brasil: infográfico ABSOLAR**. São Paulo: ABSOLAR, 2025.
- BRELLS, Walter F. **Operação econômica e planejamento**. Santa Maria: Editora da UFMS, 1979. v. 9.
- CANDIDO, Clarice. Você sabe como o sistema elétrico brasileiro funciona e por que é referência mundial? **Exame**, [S. l.], 31 mar. 2025. Disponível em: <https://exame.com/esg/voce-sabe-como-o-sistema-eletrico-brasileiro-funciona-e-por-que-e-referencia-mundial/>. Acesso em: 25 out. 2025.
- CASTRO, Nivalde de; BRANDÃO, Roberto; VARDIERO, Pedro. **Uma nova e estratégica função das usinas termoeletricas na matriz elétrica brasileira**. Rio de Janeiro: GESEL-UFRJ, 2017.
- CHEN, Yize *et al.* Model-free renewable scenario generation using generative adversarial networks. **IEEE Transactions on Power Systems**, Nova Iorque, v. 33, n. 3, p. 3265, maio 2018.
- CORTES, Pedro. Energia limpa já supera 40% da geração elétrica mundial. **CNN Brasil**, [S. l.], 2025. Disponível em: <https://www.cnnbrasil.com.br/blogs/pedro-cortes/economia/macroeconomia/energia-limpa-ja-supera-40-da-geracao-eletrica-mundial>. Acesso em: 25 out. 2025.
- COSSICH, William; CATALDI, Marcio; ROTUNNO FILHO, Otto Corrêa. Avaliação do desempenho da geração de cenários de aflúncias em reservatórios utilizando previsões de precipitação por conjunto. **Ciência e Natura**, Santa Maria, v. 37, n. 1, p. 55-62, 2015.
- DAGONT, Lucas Roberto. **Estimação de erros de previsão eólica do Sistema Interligado Nacional**. 2022. Trabalho de Conclusão de Curso/Dissertação – Universidade Federal de Santa Catarina, Florianópolis, 2022.
- DAIRI, Abdelkader *et al.* Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach. **Applied Sciences**, [S. l.], v. 10, p. 8400, 2020.
- DIAS, Lucas da Silva *et al.* Desenvolvimento e demonstração de uma [título incompleto]. **Revista Científica Faculdade Atenas**, Paracatu, v. 16, n. 1, 2023.
- EL-HERI, Yasmina S. *et al.* Análise do impacto energético da variabilidade da produção de energia eólica no sistema elétrico brasileiro. In: BRAZIL WINDPOWER, 2016, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: [s. n.], 2016.

EMPRESA DE PESQUISA ENERGÉTICA (EPE). **Plano Decenal de Expansão de Energia 2031 (PDE 2031)**. Rio de Janeiro: EPE, 2022.

FLECK, Leandro *et al.* Redes neurais artificiais: princípios básicos. **Revista Eletrônica Científica Inovação e Tecnologia**, Medianeira, v. 1, p. 47-57, 2016.

FRIGHETTO, Maurício. Por que o Brasil ainda aposta em termelétricas. **G1**, [S. l.], 30 maio 2025. Disponível em: <https://g1.globo.com/meio-ambiente/noticia/2025/05/30/por-que-o-brasil-ainda-aposta-em-termeletricas.ghtml>. Acesso em: 25 out. 2025.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 4. ed. São Paulo: Atlas, 1994.

IBM. **O que é um Autoencoder Variacional (VAE)**. IBM, 2024. Disponível em: <https://www.ibm.com/br-pt/think/topics/variational-autoencoder>. Acesso em: 24 abr. 2026.

KINGMA, Diederik P.; WELLING, Max. **Auto-encoding variational bayes**. arXiv preprint arXiv:1312.6114, 2014.

LIU, Chang *et al.* Intrusion detection system after data augmentation schemes based on the VAE and CVAE. **IEEE Transactions on Reliability**, [S. l.], v. 71, p. 1000, jun. 2022.

MAURICIO JUNIOR, Roberto Donisete *et al.* O mercado de energia solar no Brasil: uma análise dos desafios e oportunidades do setor. **Revista de Gestão e Secretariado - GeSec**, São José dos Pinhais, v. 16, p. 1-19, 2025.

MIGUEL, Paulo Augusto Cauchick *et al.* **Metodologia de pesquisa em engenharia de produção e gestão de operações**. 2. ed. Rio de Janeiro: Elsevier, 2012.

MILLIGAN, Michael *et al.* Wholesale electricity market design with increasing levels of renewable generation: revenue sufficiency and long-term reliability. **The Electricity Journal**, Nova Iorque, mar. 2016.

NASCIMENTO, Gabriel Araújo do. **Variabilidade na geração baseada em fontes renováveis: o potencial do Maranhão no fornecimento de flexibilidade ao Sistema Interligado Nacional**. 2020. Trabalho de Conclusão de Curso/Dissertação – Universidade Federal do Maranhão (UFMA), São Luís, 2020.

NELSON, David Michael Quirino. **Uso de redes neurais recorrentes para previsão de séries temporais financeiras**. 2017. Dissertação/Tese – Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, 2017.

NIED, Ademir. **Treinamento de redes neurais artificiais baseado em sistemas de estrutura variável com taxa de aprendizado adaptativa**. 2007. Tese (Doutorado) – Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO (ONS). **Desenvolvimento metodológico para previsão de geração de fonte eólica**. Rio de Janeiro: ONS, 2018. (ONS NT 151/2018).

OPERADOR NACIONAL DO SISTEMA ELÉTRICO (ONS). **O planejamento da operação**. 1. ed. São Paulo: Artliber, 2020. v. 1.

PAIXÃO, Joelson Lopes da; ABAIDE, Alzenira da Rosa. Oportunidades e desafios da exploração de energias renováveis no Brasil: uma análise integrada do potencial eólico e fotovoltaico na transição energética. *In*: PAIXÃO, Joelson Lopes da; ABAIDE, Alzenira da Rosa. **Pesquisas em engenharias e temas correlatos**. Curitiba: Aurum Editora, 2026. cap. 3, p. 29-49.

PESSANHA, José Francisco Moreira *et al.* Geração de cenários sintéticos de produção eólica de curto prazo a partir de um ensemble de previsões horárias ou semi-horárias de velocidade dos ventos. **Cadernos do IME - Série Estatística**, Rio de Janeiro, p. 55-74, 2021.

RAUBER, Thomas Walter. **Redes neurais artificiais**. Vitória: Universidade Federal do Espírito Santo (UFES), 2005.

RUDER, Sebastian. **An overview of gradient descent optimization algorithms**. arXiv preprint arXiv:1609.04747, [S. l.], 2017. Disponível em: <https://arxiv.org/abs/1609.04747>.

SOUSA JÚNIOR, Alvani Bomfim de *et al.* **Desafios da integração da energia**. Ponta Grossa: Atena Editora, 2024.

APÊNDICE

[Link para acesso aos códigos do modelo](#)