

Universidade Federal Fluminense
Escola de Engenharia
Curso de Graduação em Engenharia Elétrica

Bernardo Albuquerque Domingues da Silva

**Estimação da Geração Hidrotermo-Eólica Utilizando
Redes Neurais e Variáveis do Fenômeno El Niño**

Niterói

2025

Bernardo Albuquerque Domingues da Silva

Estimação da Geração Hidrotermo-Eólica Utilizando Redes Neurais e Variáveis do Fenômeno El Niño

Projeto de Conclusão de Curso apresentado ao Corpo Docente do Departamento de Engenharia Elétrica da Escola de Engenharia da Universidade Federal Fluminense, como parte dos requisitos necessários à obtenção do título de Engenheiro Eletricista.

Orientador:
Prof. André da Costa Pinho, D.Sc.

Niterói

2025

Ficha Catalográfica elaborada pelo Aluno, acesse o site:

<https://bibliotecas.uff.br/bcg/fichacatalografica/>

Gere o pdf da ficha e substitua o arquivo ficha.pdf com a sua ficha.

As informações abaixo são ilustrativas. Os alunos devem inserir a Ficha Catalográfica nesta página.

M514 Tal, Fulano de

Título do Trabalho / Fulano de Tal. –Rio de Janeiro, RJ: [s.n.], 2022.

xx f.

Trabalho de Conclusão de Curso de Graduação em Engenharia Civil – Universidade Federal Fluminense, 2022.

1. Construção Civil. 2. Lean Construction. 3. Gerenciamento de Obra

Bernardo Albuquerque Domingues da Silva

Estimação da Geração Hidrotermo-Eólica Utilizando Redes Neurais e Variáveis do Fenômeno El Niño

Projeto de Conclusão de Curso apresentado ao Corpo Docente do Departamento de Engenharia Elétrica da Escola de Engenharia da Universidade Federal Fluminense, como parte dos requisitos necessários à obtenção do título de Engenheiro Eletricista.

Aprovado em: _____ de _____ de _____

BANCA EXAMINADORA

Prof. André da Costa Pinho, D.Sc. - UFF

Prof. Bruno Soares Borba, D.Sc. - UFF

Prof. Marcio Andre Ribeiro Guimaraens, D.Sc. - UFF

Niterói
2025

Em memória de minha querida avó Célia, por toda
a sua paciência, carinho e amor incondicionais.

Agradecimentos

Agradeço aos meus pais, Débora e Gilvan, por sempre me apoiarem nas minhas escolhas, terem me ajudado a superar os obstáculos que surgiram no caminho e sempre estarem presentes nos momentos mais importantes. Obrigado por acreditarem em mim, por terem abdicado de tantas coisas para me proporcionar uma educação de qualidade e me ensinado a importância da ética, esforço, estudo e trabalho. Serei eternamente grato pelos esforços e sacrifícios que fizeram por mim. Vocês sempre serão os meus maiores exemplos na vida.

À minha irmã, Letícia, por ser meu alívio cômico por todos esses anos. À minha família, por compreender minhas ausências e por todos os aprendizados que me proporcionaram.

À minha companheira, Juliana, por ter me apoiado, incentivado e compreendido durante essa trajetória. Por ter me ajudado a manter a calma e acreditar que eu era capaz de superar qualquer obstáculo. Obrigado por ser a minha maior incentivadora e por nunca me deixar desistir nos momentos de dificuldade. Seu amor e paciência foram essenciais para que eu pudesse concluir este trabalho.

Aos amigos que fiz durante a graduação, que sem dúvidas espero levar para a vida toda. Sem vocês o caminho teria sido muito mais difícil. Agradeço por todos os momentos de descontração, pelas risadas, pelos estudos em grupo e, principalmente, pelo revezamento de faltas nas disciplinas mais chatas. À Faraday E-Racing, que representou um marco na minha trajetória acadêmica e me proporcionou aprendizados e oportunidades essenciais para a minha formação.

Ao professor André Pinho, pela sua excelência, ética, maestria em ensinar e por ter me orientado de maneira exemplar durante o desenvolvimento deste projeto. Agradeço também aos professores Flávio Martins, Felipe Sass e Marcio Guimaraens, por me lembrarem em cada aula do motivo pelo qual escolhi a Engenharia Elétrica.

Resumo

A matriz energética brasileira é caracterizada por uma dependência significativa de fontes renováveis, especialmente a geração hidrelétrica. Essa dependência torna o sistema elétrico vulnerável a variações climáticas, como secas prolongadas que podem ser intensificadas por fenômenos como o El Niño (EN) e La Niña (LN), afetando a disponibilidade de água nos reservatórios e, conseqüentemente, a geração de energia. Com o crescimento da fonte eólica, também vulnerável a variações climáticas, é necessário investigar o impacto de variáveis climatológicas na operação do Sistema Interligado Nacional (SIN), com ênfase nas fontes hidráulica e térmica. Este trabalho analisa séries históricas disponibilizadas pelo Operador Nacional do Sistema Elétrico (ONS) e variáveis que definem o fenômeno El Niño, obtidas do ERA5, para avaliar os efeitos desses fenômenos na geração de energia elétrica. Diferentes modelos de regressão e aprendizado de máquina são aplicados para analisar o impacto dessas variáveis na geração de energia.

Palavras-chave: Geração de energia. Clima. Planejamento energético. Machine learning.

Abstract

The Brazilian energy matrix is characterized by a significant dependence on renewable sources, especially hydropower generation. This dependence makes the electric system vulnerable to climatic variations, such as prolonged droughts and phenomena such as El Niño (EN) and La Niña (LN), which can affect the availability of water in reservoirs and, consequently, energy generation. With the adoption of wind power, also vulnerable to climatic variations, it is necessary to investigate the impact of climatic variables on the operation of the National Interconnected System (SIN), with an emphasis on hydraulic and thermal sources. This work analyzes historical series provided by the National Electric System Operator (ONS) and variables that define the El Niño phenomenon, obtained from ERA5, to evaluate the effects of these phenomena on electricity generation. Different regression and machine learning models are applied to analyze the impact of these variables on energy generation.

Key-words: Energy generation. Climate. Risk mitigation. Machine learning.

Lista de Figuras

Figura 1 - Geração centralizada anual por fonte	13
Figura 2 - Curva de carga diária do SIN em base horária	14
Figura 3 - Curva de carga do SIN em base mensal	15
Figura 4 - Geração hidráulica total em base mensal	16
Figura 5 - Índice ONI (Oceanic Niño Index)	16
Figura 6 - Regiões do fenômeno El Niño-Oscilação Sul (ENSO)	21
Figura 7 - Subsistemas do SIN	28

Lista de Tabelas

Tabela 1	- Bibliotecas utilizadas no projeto	25
Tabela 2	- Parâmetros dos dados de geração	26
Tabela 3	- Parâmetros dos dados de carga	26

Lista de Abreviaturas e Siglas

ONS Operador Nacional do Sistema Elétrico

EPE Empresa de Pesquisa Energética

SIN Sistema Interligado Nacional

Sumário

1	Introdução	13
1.1	Contexto	13
1.2	Motivação	14
1.3	Objetivo	17
1.4	Estrutura do Trabalho	18
2	Fundamentação Teórica	20
2.1	Impacto do ENSO na Geração de Energia Elétrica	20
2.2	O Modelo NEWAVE	21
2.2.1	Representação das Usinas	22
2.2.2	Dados de Entrada	22
2.3	Modelos Linear e Não-linear	23
2.3.1	Modelo Linear	23
2.3.2	Modelos Não-lineares: Random Forest e Gradient Boosting	23
2.4	Modelo Neural	23
3	Metodologia	24
3.1	Abordagem computacional	24
3.2	Obtenção e pré-processamento dos dados	25
3.2.1	Séries históricas do ONS	25
3.2.2	Séries históricas de variáveis climatológicas	27
3.2.3	Pré-processamento dos dados	28
3.3	Análise exploratória dos dados	28
3.4	Implementação dos modelos de regressão	28
3.4.1	Feature selection	28
3.4.2	Análise de componentes principais	28
3.4.3	Avaliação dos Modelos de regressão	28
3.5	Implementação do Modelo Neural	28

4 Resultados	29
5 Conclusão	30
Referências	31

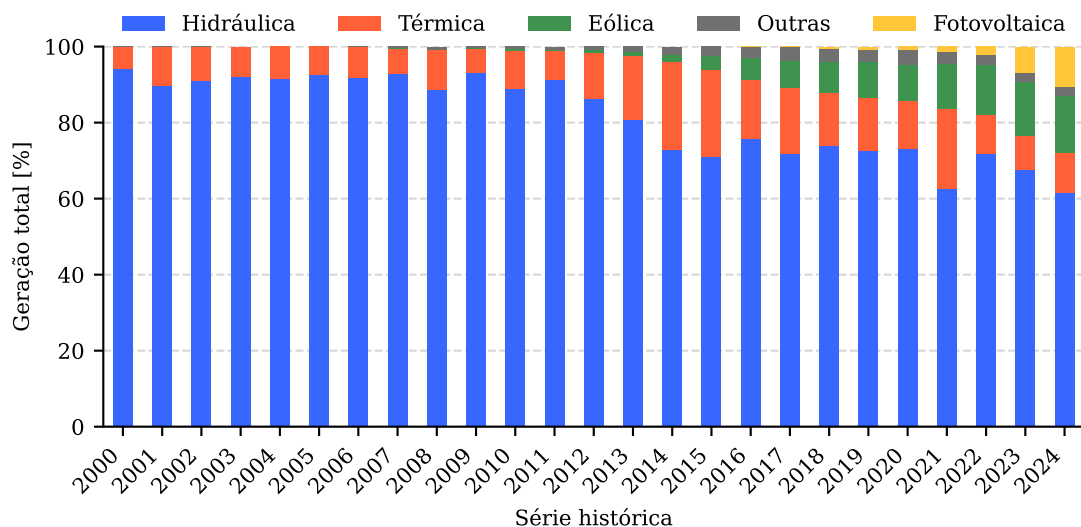
Capítulo 1

Introdução

1.1 Contexto

Historicamente, a matriz elétrica brasileira é considerada uma das mais limpas do mundo, com destaque para a fonte hidráulica, que é responsável pela maior parte da geração de energia elétrica no país. Nos últimos anos, outras fontes de geração vêm sendo incorporadas ao sistema, das quais destacam-se a eólica e solar fotovoltaica, conforme observado na Figura 1, elaborada a partir de dados brutos de geração centralizada obtidos do Operador Nacional do Sistema Elétrico (ONS), sem considerar a geração distribuída.

Figura 1 – Geração centralizada anual por fonte



Fonte: o autor.

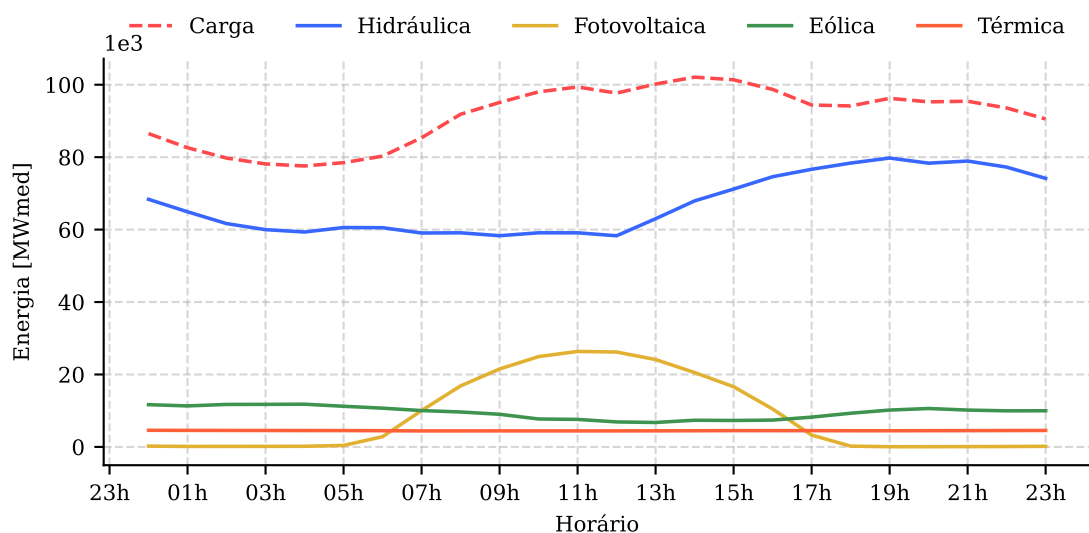
Nota-se, em especial, um crescimento significativo da geração eólica, observado a partir de 2015, e uma diminuição significativa da contribuição de geração térmica média no panorama geral nos anos seguintes. Em 2023, a fonte eólica foi responsável por 48% da expansão da capacidade instalada total de 10,19 GW (EPE, 2024). Essa expansão se dá em função do maior número de empreendimentos participantes nos Leilões de

Energia Elétrica do Ambiente de Contratação Regulada (ACR) realizados pela Empresa de Pesquisa Energética (EPE). Isso ocorre, dentre outros fatores, devido à queda nos custos de aerogeradores e painéis fotovoltaicos, além do fator "combustível zero" dessas fontes, o que torna novos empreendimentos mais atrativos economicamente para os agentes.

Embora essa expansão seja positiva, poupando recursos hídricos, contribuindo para a diversificação da matriz elétrica e reduzindo o acionamento de usinas térmicas, essas fontes possuem características intrínsecas que as tornam intermitentes, como a incidência solar e a velocidade do vento. Sendo assim, uma alta dependência dessas fontes tem o potencial de tornar o sistema como um todo mais vulnerável.

Além disso, ao analisar a curva de carga do SIN, observa-se que, embora o seu pico ocorra no início da tarde, momento no qual a geração solar fotovoltaica apresenta significativa contribuição, o período noturno também apresenta carga considerável, conforme a Figura 2, que mostra a curva de carga do SIN para o dia 15 de março de 2024, dia em que registrou-se um recorde de demanda máxima instantânea de 102.478 MW, segundo o ONS, e como pode ser observado na Figura 3.

Figura 2 - Curva de carga diária do SIN em base horária



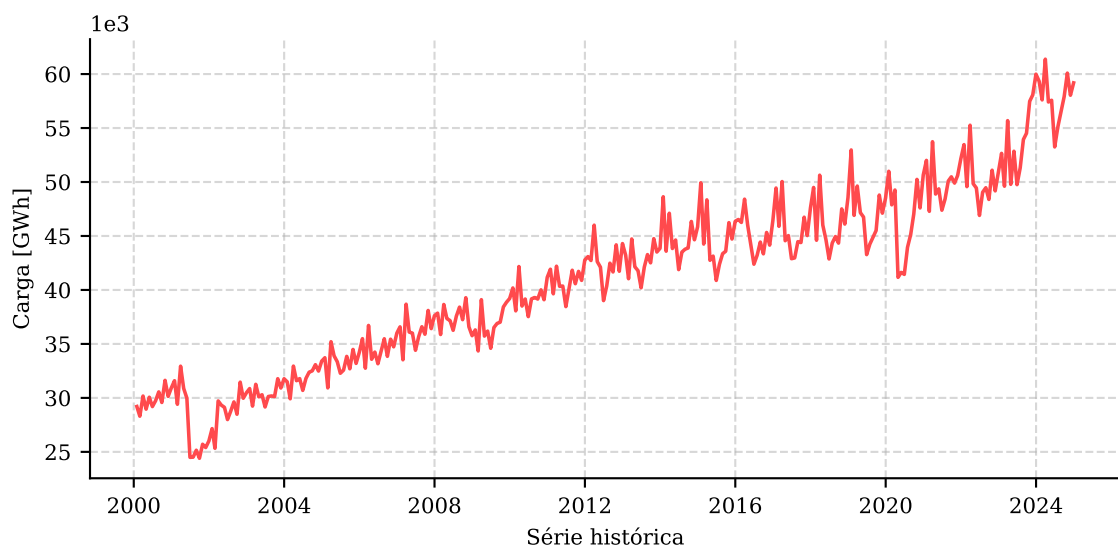
Fonte: o autor.

1.2 Motivação

Em um contexto no qual a implementação de sistemas de armazenamento de energia elétrica ainda é incipiente, a matriz segue bastante dependente da fonte hidráulica e, de

maneira complementar, das térmicas. A dependência da fonte hidráulica, por sua vez, torna o sistema elétrico vulnerável a eventos climáticos extremos ocasionados pelas mudanças climáticas. Por exemplo, em 2021, verificou-se um acionamento recorde de usinas térmicas e uma geração hidráulica percentual mínima. Isso se deve em razão da forte crise hídrica enfrentada pelo Brasil em 2021, a pior dos últimos 91 anos até então. (SOARES; COSTA, 2023)

Figura 3 – Curva de carga do SIN em base mensal



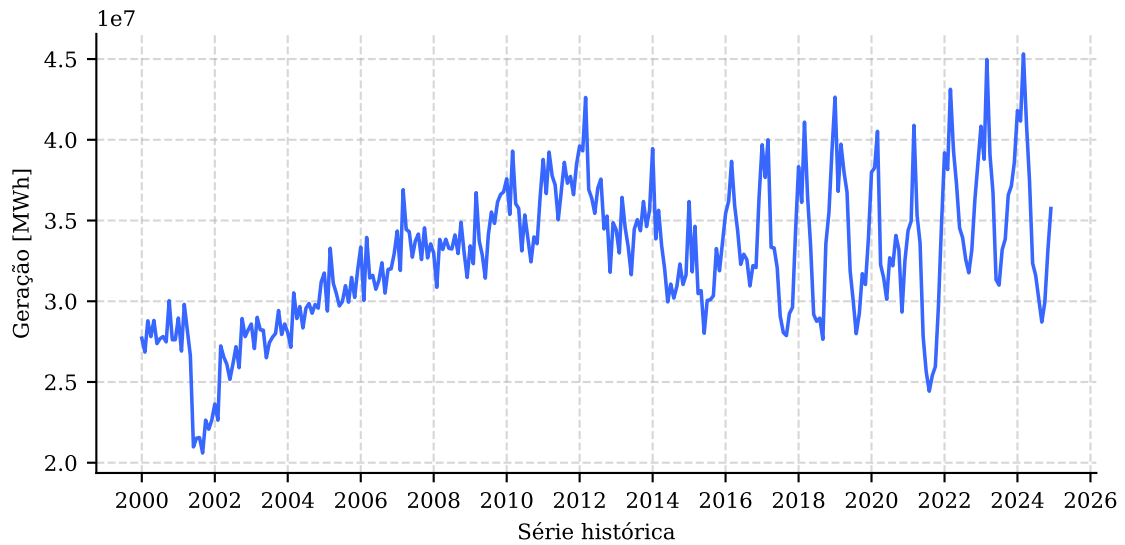
Fonte: o autor.

Portanto, o estudo da operação do sistema elétrico brasileiro, no contexto de cenários de eventos climatológicos extremos é altamente relevante para a segurança energética do país, considerando uma estimativa de crescimento médio anual da carga do SIN de 3,2%. (ONS, 2024)

Ao analisar a geração hidráulica bruta na Figura 4, evidenciam-se pontos nos quais a geração é reduzida. Isso ocorre devido à sazonalidade das vazões nas bacias hidrográficas, responsáveis pelo abastecimento dos reservatórios. Considerando a amostragem em base mensal, observa-se que a geração é reduzida nos meses de inverno, período caracterizado por menor ocorrência de precipitação e, conseqüentemente, menor vazão nos rios. Por outro lado, nos meses de verão, a geração atinge seus maiores valores.

Esse comportamento é natural e esperado, uma vez que a geração hidráulica é diretamente influenciada pelas condições que afetam a vazão dos rios. No entanto, a ocorrência de eventos climatológicos como o El Niño-Oscilação Sul (ENSO) pode favorecer condições que impactam diretamente no potencial de geração hidráulica. (GURJÃO et al.,

Figura 4 – Geração hidráulica total em base mensal

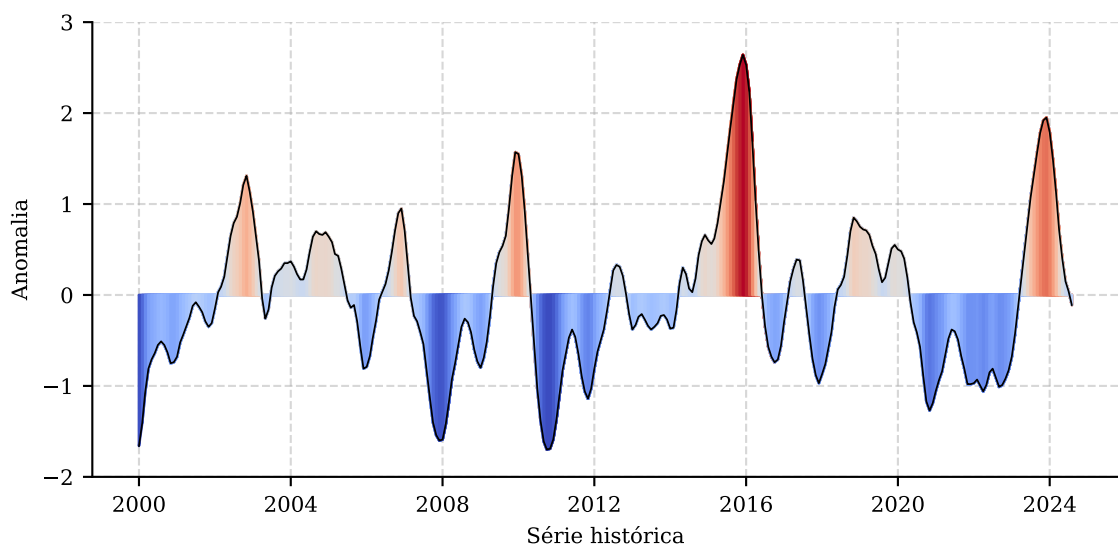


Fonte: o autor.

2012)

Fenômenos como o ENSO são caracterizados por anomalias na temperatura da superfície do mar no Oceano Pacífico Equatorial. Essas anomalias são monitoradas por meio de índices como o ONI (Oceanic Niño Index), que classifica os eventos em três categorias: *El-Niño* (EN), *La-Niña* (LN) e neutro. A Figura 5 mostra a classificação dos eventos de EN e LN ocorridos entre 2000 e 2024, sendo que a escala de cores indica a intensidade do evento. Ao analisar a geração hidráulica bruta no mesmo período, observa-se que a ocorrência de eventos de EN e LN pode estar associada a variações na geração.

Figura 5 – Índice ONI (Oceanic Niño Index)



Fonte: o autor.

Fundamentalmente, em sistemas interligados cuja fonte hidráulica constitui a base da matriz elétrica, é essencial, para um planejamento energético eficiente, otimizar o sistema de modo a considerar a operação de todas as usinas, considerando a incerteza associada às afluições futuras. Dessa forma, estima-se o valor da geração hidrelétrica que poderia substituir a geração térmica a curto ou longo prazo, de modo a minimizar os custos de operação do sistema e o risco de utilizar reservatórios de maneira desnecessária, garantindo assim o atendimento à demanda futura, principalmente em casos de escassez hídrica.

Os estudos de planejamento energético são realizados por meio de modelos computacionais como o NEWAVE, DECOMP e DESSEM, que consideram diferentes horizontes temporais: longo, médio e curto prazos, respectivamente. Também há outras soluções disponíveis, como o PSR SDDP, que engloba todos os horizontes temporais. Considerando o escopo deste trabalho, o modelo NEWAVE será brevemente apresentado no capítulo seguinte.

Embora esses modelos sejam amplamente utilizados pelo setor elétrico brasileiro, sendo consolidados como ferramentas confiáveis e essenciais para o planejamento energético, eles não consideram variáveis externas, como fenômenos climáticos como o EN e LN, que podem impactar a geração de energia elétrica. Esses modelos são baseados em dados históricos de vazões e velocidade do vento, que são obtidos a partir de medições locais, além de dados individuais de cada usina.

1.3 Objetivo

Sendo assim, o projeto tem como objetivo investigar um possível impacto de variáveis climatológicas externas na geração de energia elétrica no Brasil, com foco nas fontes hidráulica, térmica e eólica.

Para tanto, foram empregadas técnicas computacionais para relacionar as séries históricas de geração com as séries de variáveis climáticas, através de modelos lineares, não lineares e neurais. Além disso, foram utilizadas séries históricas de geração, carga e vazões disponibilizados pelo ONS, bem como séries históricas de variáveis climatológicas, como temperatura da superfície do mar, obtidas a partir de dados do ERA5, um projeto de reanálise atmosférica do *European Centre for Medium-Range Weather Forecasts* (ECMWF).

A partir dessa investigação, espera-se poder avaliar o impacto dessas variáveis na

geração de energia elétrica, o que pode contribuir para o planejamento energético do país, especialmente em cenários de eventos climáticos extremos e tendências climáticas.

É importante salientar que outras variáveis externas poderiam ser incorporadas ao estudo, ou até mesmo uma combinação entre variáveis locais e externas. Também poderiam ser considerados indicadores econômicos e outros dados relevantes. No entanto, este trabalho considera apenas as variáveis relacionadas ao fenômeno EN e LN, uma vez que estudos indicam uma alta correlação entre esses fenômenos e o regime de chuvas no Brasil (GURJÃO et al., 2012; ANDREOLI et al., 2016), sendo a incorporação de outras variáveis reservada para trabalhos futuros.

1.4 Estrutura do Trabalho

No capítulo 1, é feita uma breve introdução apresentando o contexto, motivação, objetivo e a estrutura do trabalho. Uma breve análise da matriz elétrica é apresentada, com foco no histórico recente e no crescimento da geração eólica. Também são apresentadas as curvas de carga do SIN e de geração hidráulica. Uma breve contextualização acerca dos fenômenos do ENSO e a importância do modelo NEWAVE para o planejamento energético do país são apresentadas. Por fim, é apresentado o objetivo do trabalho e a estrutura do documento.

No capítulo 2, a fundamentação teórica necessária para a compreensão do projeto é apresentada. São abordados os fatores que fazem com que os fenômenos do EN e LN impactem o regime de chuvas no Brasil e a sua relação com a geração de energia elétrica. Também será feita uma breve introdução ao modelo NEWAVE, que é amplamente utilizado pelo setor elétrico brasileiro para planejamento energético. Por fim, serão apresentados os modelos de previsão de séries temporais implementados, com foco no modelo neural, uma implementação baseada na arquitetura *TSMixer* desenvolvida pela Google.

O capítulo 3 apresenta a metodologia utilizada para a realização do projeto. Serão apresentados os conjuntos de dados considerados e suas respectivas etapas de obtenção, tratamento e análise. Além disso, será apresentada a metodologia utilizada para implementação dos modelos de previsão, incluindo seus parâmetros e métricas de avaliação. *Snippets* de códigos serão apresentados para facilitar a compreensão.

O capítulo 4 apresenta os resultados obtidos a partir da implementação de cada modelo de previsão. Inicia-se com os resultados do modelo linear, seguido pelo modelo não

linear e, por fim, o modelo neural. Para cada modelo, serão apresentados os resultados de previsão, métricas de avaliação e uma análise crítica dos resultados obtidos.

Por fim, no capítulo 5, são apresentadas as considerações finais do trabalho, incluindo uma discussão final dos resultados obtidos, limitações observadas e sugestões para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

2.1 Impacto do ENSO na Geração de Energia Elétrica

O ENSO é um fenômeno que ocorre no Oceano Pacífico Equatorial, caracterizado por variações na temperatura da superfície do mar (TSM) em regiões específicas, como ilustrado na Figura 6. Segundo Andreoli et al. (2016), o fenômeno é um dos principais fatores que influenciam os padrões de vento e precipitação em diversas regiões da América do Sul e seus efeitos se estendem por todas as regiões do Brasil.

De acordo com Capozzoli, Cardoso e Ferraz (2017), diferentes fases do ENSO resultam em padrões distintos de precipitação sobre o território brasileiro, variando de acordo com a região e a estação do ano. Sendo assim, há um impacto direto sobre a disponibilidade de recursos hídricos para a geração hidrelétrica.

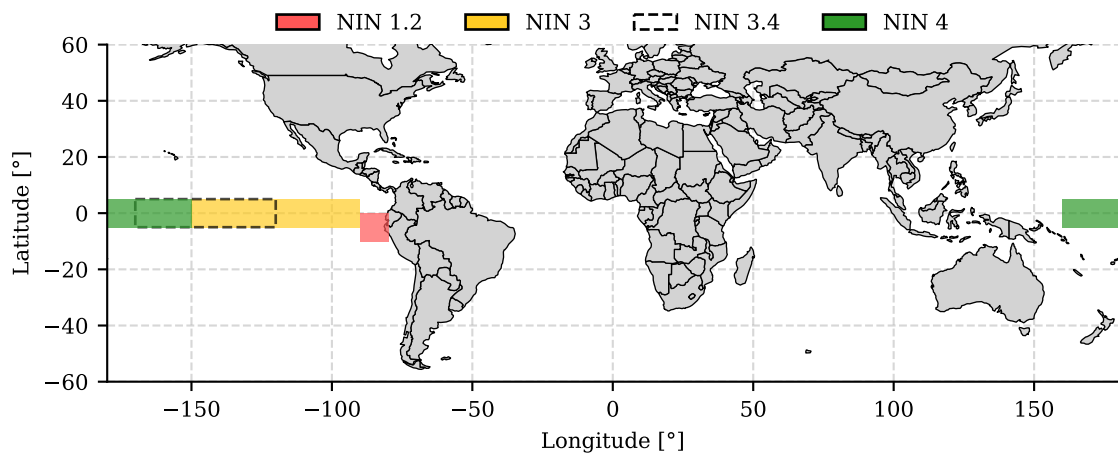
A região Sul é uma das mais consistentemente afetadas. Eventos de El Niño tendem a causar precipitação acima da média, particularmente durante a primavera e o verão, enquanto eventos de La Niña estão associados à condições de seca.

A região Sudeste apresenta uma resposta mais complexa, sendo considerada uma zona de transição. A bacia do Rio Paraná, em especial, apresenta sensibilidade aos fenômenos do ENSO, tendo apresentado tendência de aumento de vazão durante alguns eventos de El Niño.

Para as regiões Norte e Nordeste, eventos de El Niño estão associados a períodos de seca, enquanto eventos de La Niña tendem a trazer chuvas acima da média. No entanto, é importante destacar que outros fenômenos atmosféricos podem interferir com esses padrões, modulando os efeitos do ENSO.

Sendo assim, verifica-se que as variações induzidas pelos fenômenos do ENSO traduzem-se diretamente em variações nas vazões dos rios que alimentam as bacias, que por sua vez impactam diretamente o potencial de geração da fonte hidráulica.

Figura 6 – Regiões do fenômeno El Niño-Oscilação Sul (ENSO)



Fonte: o autor.

2.2 O Modelo NEWAVE

Desenvolvido e mantido pelo Centro de Pesquisas de Energia Elétrica (CEPEL) e amplamente utilizado pelo setor elétrico brasileiro para definição de estratégias e tomada de decisão, o NEWAVE é um modelo de otimização que busca minimizar os custos de operação do sistema, considerando a incerteza das afluências futuras e a operação de um sistema hidro-térmico-eólico interligado. O modelo é utilizado para estudos como:

- Elaboração do Plano Decenal de Expansão de Energia (PDE), pela EPE;
- Elaboração do Programa Mensal de Operação (PMO) e Plano de Operação Energética (PEN), pelo ONS;
- Formação de preços, como no cálculo do Preço de Liquidação das Diferenças (PLD) pelo CCEE;
- Cálculo de Garantia Física e da Energia Assegurada para empreendimentos de geração participantes nos leilões de energia elétrica, pela EPE;
- Elaboração de diretrizes para os leilões de energia, pela EPE.

Em resumo, o modelo emprega a Programação Dinâmica Dual Estocástica (PDDE), uma técnica de otimização que permite lidar com as incertezas ligadas às afluências futuras sem que o modelo se torne computacionalmente impraticável, considerando múltiplos reservatórios, interconexões e o horizonte temporal de médio e longo prazos.

2.2.1 Representação das Usinas

O NEWAVE modela o sistema de geração hidrelétrico em Reservatórios Equivalentes de Energia (REEs), que são grupos de usinas associadas a um subsistema ou submercado de energia. Cada subsistema pode conter mais de um REE, possibilitando diferenciar bacias hidrográficas com regimes distintos, ainda que pertençam a um mesmo subsistema.

Além disso, cada REE é definido por um conjunto de parâmetros que são calculados a partir das características individuais de cada usina. Nas versões mais recentes do modelo, também é possível considerar todas as usinas individualmente ou operar de maneira híbrida, ou seja, considerando alguns REEs e outras usinas individualmente.

As usinas termelétricas são representadas no modelo através de classes térmicas. Cada classe agrupa usinas com custos semelhantes e está associada a um subsistema. Cada classe também é definida por um conjunto de parâmetros calculados a partir das características individuais de cada usina.

Nas versões mais recentes do modelo, a fonte eólica também é modelada. De maneira similar, os parques eólicos são agrupados em Parques Eólicos Equivalentes (PEE). O agrupamento é feito a partir de dados de cadastro de cada parque eólico, estado, submercado, função de produção (curva relacionando a velocidade do vento com a potência gerada), dados sobre torres de medição e séries históricas de velocidade do vento.

2.2.2 Dados de Entrada

O modelo requer um conjunto de dados de entrada que inclui as características das usinas, dados dos subsistemas, demanda, séries históricas de vazões e ventos, cronogramas de expansão, restrições operativas, dentre outros. Observa-se, portanto, que todos os dados de entrada são locais e, portanto, o modelo não considera variáveis externas, como fenômenos climáticos como o EN e LN, que podem impactar a geração de energia elétrica.

Ainda que as últimas versões do modelo apresentem campos previstos para a entrada de dados do ENSO, esses campos estão marcados como "não implementados". Dessa forma, entende-se que o modelo não considera diretamente o impacto dessas variáveis. No entanto, vale destacar que essas variáveis externas podem ser utilizadas para elaborar as séries históricas de vazões e velocidade de ventos utilizadas como dados de entrada.

2.3 Modelos Linear e Não-linear

fazer uma introdução sobre os modelos lineares e não lineares...

2.3.1 Modelo Linear

2.3.2 Modelos Não-lineares: Random Forest e Gradient Boosting

2.4 Modelo Neural

Atualmente, os modelos neurais mais avançados utilizam a arquitetura *Transformer*, que são modelos neurais com um mecanismo de atenção que permite ao modelo focar em diferentes partes da entrada de dados ao processar informações. Essa arquitetura foi introduzida no artigo "*Attention is All You Need*" de Vaswani et al. (2017) e hoje é a arquitetura por trás dos principais *Large Language Models* (LLMs) comercialmente disponíveis.

Inicialmente concebido para tarefas de tradução, modelos baseados nessa arquitetura demonstraram resultados superiores em outras aplicações e hoje são amplamente utilizados em diversas áreas, incluindo previsão de séries temporais. No entanto, esses modelos são altamente complexos e exigem um alto poder computacional para treinamento e inferência.

Além disso, conforme demonstrado por Zeng et al. (2022), modelos baseados em *Transformers* podem produzir resultados inferiores quando comparados a modelos mais simples.

Dessa forma, uma abordagem alternativa foi proposta por Chen et al. (2023), uma arquitetura mais simples em comparação aos *Transformers*, mas com resultados bastante superiores

Capítulo 3

Metodologia

3.1 Abordagem computacional

Considerando o escopo do projeto e o grande volume de dados a serem processados, não há outra abordagem viável senão a utilização de técnicas computacionais. Sendo assim, optou-se por desenvolver o projeto em Python, devido à familiaridade do autor com a linguagem e a vasta gama de bibliotecas disponíveis para manipulação de séries históricas, análise estatística e aprendizado de máquina.

Como o computador utilizado possui uma placa de vídeo da AMD e o projeto foi desenvolvido em Windows, não foi possível realizar o *offload* do processamento para a *Graphics Processing Unit* (GPU), como seria possível com placas de vídeo NVIDIA nativamente com a biblioteca *torch*. Para isso, seria necessário utilizar Linux, em que há suporte da biblioteca para a plataforma *Radeon Open Compute* (ROCm). Sendo assim, as etapas do projeto empregando modelos neurais foram processadas apenas no CPU. O sistema utilizado possui um processador AMD Ryzen 5900X e 32 GB de memória RAM. Todo o processamento intensivo foi realizado em paralelo, utilizando todas as 24 *threads* disponíveis.

Além disso, toda a base de código foi desenvolvida em Python versão 3.12 e está disponível em um repositório público no GitHub. O projeto foi organizado em módulos, cada um responsável por uma etapa do processo, desde a obtenção dos dados até a implementação dos modelos de aprendizado de máquina. A tabela 1 mostra as bibliotecas utilizadas no projeto, suas finalidades e versões.

Tabela 1 – Bibliotecas utilizadas no projeto

Biblioteca	Descrição	Versão
numpy	Cálculos numéricos e manipulação de arrays	1.26.4
pandas	Manipulação e análise de dados (DataFrames)	2.2.3
requests	Requisições HTTP	2.32.3
urllib3	Gerenciamento de conexões HTTP	2.2.3
alive_progress	Barra de progresso para loops	3.2.0
netCDF4	Leitura de arquivos NetCDF	1.7.2
cdsapi	API para download de dados do ECMWF	0.7.5
geopandas	Manipulação de dados geoespaciais	1.0.1
matplotlib	Visualização de dados	3.9.2
scikit-learn	Aplicação de modelos iniciais	1.5.2
scipy	Ferramentas e algoritmos matemáticos	1.14.1
transformers	Modelos Neurais Pré-treinados	4.52.3
torch	Processamento de Redes Neurais	2.7.0

Fonte: o autor.

3.2 Obtenção e pré-processamento dos dados

3.2.1 Séries históricas do ONS

A primeira etapa do projeto consiste na consolidação das séries históricas de geração, carga e variáveis hidrológicas, que são disponibilizadas publicamente no portal Dados Abertos do ONS, a partir do ano 2000. As séries referentes às variáveis hidrológicas são disponibilizadas em base diária, e os dados de geração e carga são disponibilizados em base horária.

Os dados de geração são disponibilizados em Mega Watt médio (MWmed) por fonte de energia, subsistema, estado, modalidade de operação, entre outras variáveis. Os dados de carga também são disponibilizados em MWmed e contêm informações sobre a carga em cada subsistema do SIN.

Para as séries de geração, os dados de 2000 a 2021 são agrupados pelos respectivos anos, e a partir de 2022, as informações estão agrupadas em arquivos por mês e ano. Para as séries de carga, os dados são disponibilizados por ano. Como o ONS não disponibiliza *Application Programming Interface* (API) para a obtenção dos dados diretamente, foi necessário uma outra abordagem, a fim de evitar o download manual dos dados.

Após identificar o padrão de nomenclatura utilizado pelo ONS para os arquivos, foram desenvolvidas as funções *GetGeracao*, *GetCarga* e *GetVazao* para obtenção dos dados de geração, carga e hidrológicos, respectivamente. Os arquivos foram baixados por meio de requisições HTTP, utilizando a biblioteca *requests* e a biblioteca *urllib3* para gerenciar as conexões. Além disso, as funções realizam o download dos arquivos em paralelo, utilizando

todas as threads disponíveis do sistema. Ao todo, cerca de 10 GB de dados em arquivos Comma Separated Values (CSV) foram consolidados.

A fim de obter uma amostragem representativa e suficiente para aplicação dos modelos computacionais, optou-se por fazer uma reamostragem dos dados em base horária para a base diária. Sendo assim, considerando o período de 2000 a 2024, foram consolidados ao todo 9132 amostras diárias para cada variável. Caso fosse considerada a amostragem mensal, o número de amostras seria de apenas 300, o que poderia não ser suficiente para a aplicação dos modelos. Ainda assim, análises em base mensal poderão ser realizadas posteriormente, através de outro processo de reamostragem.

O código ?? mostra a função *GetGeracao*, que recebe como argumentos a lista de anos de interesse e a URL base em que os dados estão hospedados. A mesma abordagem foi utilizada para as funções *GetCarga* e *GetVazao*.

Tabela 2 - Parâmetros dos dados de geração

Parâmetro	Descrição	Tipo
din_instante	Instante de aferição	Datetime
nom_subsistema	Subsistema da usina	String
id_estado	Estado onde a usina está localizada	String
nom_tipousina	Tipo de usina	String
nom_tipocombustivel	Tipo de combustível	String
nom_usina	Nome da usina	String
val_geracao	Geração de energia (MWmed)	Float

Fonte: ONS (2024)

Nota: Variáveis não utilizadas foram omitidas.

Tabela 3 - Parâmetros dos dados de carga

Parâmetro	Descrição	Tipo
din_instante	Instante de aferição	Datetime
nom_subsistema	Subsistema da usina	String
val_cargaenergiahomwmed	Carga de energia (MWmed)	Float

Fonte: ONS (2024)

Verifica-se, a partir da tabela 2, que os dados de geração contém informações que permitem uma análise detalhada da operação do SIN em diferentes níveis de granularidade. Dessa maneira, possíveis impactos em diferentes escalas geográficas e temporais poderão ser avaliados. Por outro lado, a tabela 3 mostra que os dados de carga não possuem a mesma granularidade que os dados de geração, limitando a análise a nível de subsistema, conforme a figura 7.

3.2.2 Séries históricas de variáveis climatológicas

A fim de maximizar o potencial de análise do projeto, é essencial considerar um grande número de dados climatológicos, como anomalias de temperatura, precipitação, pressão, radiação solar, entre outros. No entanto, a obtenção de todos esses dados em base diária só é possível através do portal Climate Data Store (CDS) do ECMWF (European Centre for Medium-Range Weather Forecasts), através do ERA5, que é um reanálise climática de alta resolução espacial e temporal. (C3S, 2024)

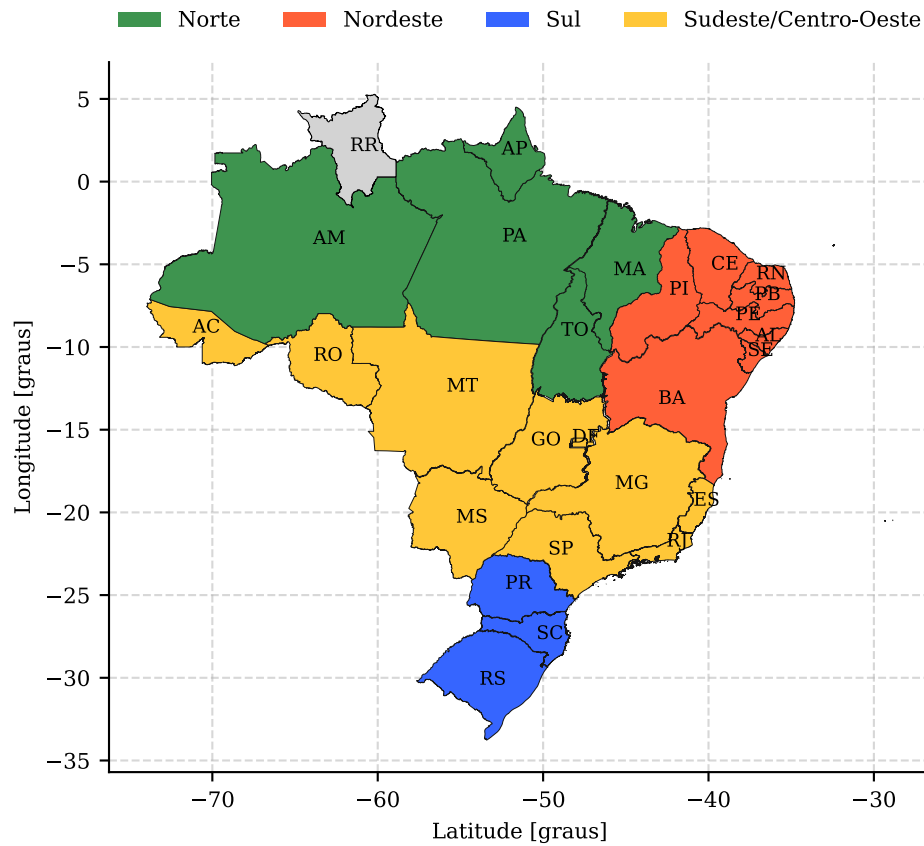
Embora os dados do ERA5 possam ser obtidos através da API do CDS, sendo possível especificar a escala temporal e a área geográfica de interesse, há um limite de requisições por usuário e uma file de espera para o processamento das requisições. Sendo assim, apenas os dados de temperatura da superfície do mar (TSM) e precipitação das regiões do ENSO foram obtidos em base diária. O código ?? mostra como as solicitações à API do CDS foram realizadas, utilizando a biblioteca *cdsapi*.

Os dados do ECMWF foram obtidos através da API do CDS (Climate Data Store), que permite o download de dados meteorológicos de alta resolução, sendo possível especificar a escala temporal e a área de interesse, especificando as coordenadas geográficas. Os dados do NOAA foram obtidos através de requisições HTTP.

É importante ressaltar que, embora a maioria dos dados climatológicos seja disponibilizada em base mensal, alguns possuem escala temporal diária, como alguns dados do ECMWF. Considerando que os dados de energia são disponibilizados em base horária, diferentes análises poderão ser realizadas, considerando as escalas temporais diária e mensal. Para isso, durante a etapa de pré-processamento dos dados, será realizada a consolidação dos dados de geração e carga de modo que estejam na mesma escala temporal dos dados climatológicos.

Além disso, nas próximas etapas do projeto, será realizado a seleção das variáveis climatológicas mais relevantes para a análise, considerando a correlação com a geração de energia. Para isso, serão utilizadas técnicas de *feature selection* e redução de dimensionalidade, como a análise de componentes principais. A tabela ?? mostra as variáveis climatológicas utilizadas no projeto.

Figura 7 - Subsistemas do SIN



Fonte: o autor.

3.2.3 Pré-processamento dos dados

3.3 Análise exploratória dos dados

3.4 Implementação dos modelos de regressão

3.4.1 Feature selection

3.4.2 Análise de componentes principais

3.4.3 Avaliação dos Modelos de regressão

3.5 Implementação do Modelo Neural

Capítulo 4

Resultados

Capítulo 5

Conclusão

Referências

ANDREOLI, Rita Valéria et al. The influence of different el niño types on the south american rainfall. **International Journal of Climatology**, Wiley, v. 37, n. 3, p. 1374-1390, maio 2016. ISSN 1097-0088.

C3S. **ERA5 post-processed daily statistics on single levels from 1940 to present**. [S.l.]: ECMWF, 2024.

CAPOZZOLI, Caluan Rodrigues; CARDOSO, Andrea de Oliveira; FERRAZ, Simone Erotildes Teleginski. Padrões de variabilidade de vazão de rios nas principais bacias brasileiras e associação com Índices climáticos. **Revista Brasileira de Meteorologia**, FapUNIFESP (SciELO), v. 32, n. 2, p. 243-254, jun. 2017. ISSN 0102-7786.

CHEN, Si-An et al. Tsmixer: An all-mlp architecture for time series forecasting. arXiv, 2023.

EPE. **Balanço Energético Nacional 2024**. [S.l.], 2024. Disponível em: <<https://www.ons.org.br/AcervoDigitalDocumentosEPublicacoes/Relato%CC%81rio%20PEN%202024%20VF.pdf>>. Acesso em: 12.01.2025.

GURJÃO, Carlos Diego de Sousa et al. Influência do enos (el niño-oscilação sul) no regime hidrológico do rio são francisco: uma análise em regiões com fortes pressões antrópicas (influence of enso (el niño-southern oscillation) in the hydrological regime são francisco river: an analysis...). **Revista Brasileira de Geografia Física**, v. 5, n. 4, p. 774-790, 2012.

ONS. **Plano da Operação Energética (PEN) 2024**. [S.l.], 2024. Disponível em: <<https://www.ons.org.br/AcervoDigitalDocumentosEPublicacoes/Relato%CC%81rio%20PEN%202024%20VF.pdf>>. Acesso em: 10.12.2024.

SOARES, Marcos de Abreu; COSTA, Hirdan Katarina de Medeiros. A crise hídrica enfrentada em 2021: uma avaliação das principais medidas adotadas pelo governo. **DELOS: DESARROLLO LOCAL SOSTENIBLE**, South Florida Publishing LLC, v. 16, n. 43, p. 489-503, abr. 2023. ISSN 1988-5245.

VASWANI, Ashish et al. **Attention Is All You Need**. [S.l.]: arXiv, 2017.

ZENG, Ailing et al. **Are Transformers Effective for Time Series Forecasting?** [S.l.]: arXiv, 2022.