

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Escola de Engenharia
Programa de Pós-graduação em Ciências e Técnicas Nucleares

José Péricles Freire

Explorando Modelos Híbridos para a Previsão de Vento: Uma Análise Multirregional

Belo Horizonte
2024

José Péricles Freire

Explorando Modelos Híbridos para a Previsão de Vento: Uma Análise Multirregional

Dissertação apresentada ao Programa de Pós-Graduação em Ciências e Técnicas Nucleares, do Departamento de Engenharia Nuclear da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do Mestrado em Ciências e Técnicas Nucleares. Área de Concentração: Engenharia Nuclear e de Energia.

Orientador: Prof. Dr. Carlos Eduardo Velasquez Cabrera

Coorientador: Lihki Jose Rubio Ortega

Belo Horizonte

2024

F866e	<p>Freire, José Pérciles. Explorando modelos híbridos para a previsão de vento [recurso eletrônico] : uma análise multirregional / José Pérciles Freire. – 2024. 1 recurso online (115 f. : il., color.) : pdf.</p> <p>Orientador: Carlos Eduardo Velasquez Cabrera. Coorientador: Lihki Jose Rubio Ortega.</p> <p>Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Engenharia.</p> <p>Bibliografia: f. 108-115.</p> <p>1. Engenharia nuclear – Teses. 2. Previsão do tempo – Teses. 3. Energia eólica – Teses. 4. Redes neurais (Computação) – Teses. I. Cabrera, Carlos Eduardo Velasquez. II. Ortega, Lihki Jose Rubio. III. Universidade Federal de Minas Gerais. Escola de Engenharia. IV. Título.</p>
	CDU: 621.039(043)



FOLHA DE APROVAÇÃO

Explorando Modelos Híbridos para a Previsão de Vento: Uma Análise Multirregional

JOSÉ PÉRICLES FREIRE

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIAS E TÉCNICAS NUCLEARES, como requisito para obtenção do grau de Mestre em CIÊNCIAS E TÉCNICAS NUCLEARES, área de concentração ENGENHARIA NUCLEAR E DA ENERGIA.

Aprovada em 29 de novembro de 2024, pela banca constituída pelos membros:

Documento assinado digitalmente
 **CARLOS EDUARDO VELASQUEZ CABRERA**
Data: 29/11/2024 15:31:56-0300
Verifique em <https://validar.iti.gov.br>

Prof. Carlos Eduardo Velasquez Cabrera – Orientador
Departamento de Engenharia Nuclear - UFMG

Prof. Lihki Jose Rubio Ortega
Universidad del Norte de Barranquilla

Documento assinado digitalmente
 **MILDE MARIA DA SILVA LIRA**
Data: 29/11/2024 15:44:01-0300
Verifique em <https://validar.iti.gov.br>

Prof.^a Milde Maria da Silva Lira
Universidade Federal de Pernambuco

Documento assinado digitalmente
 **AGNALDO JOSE DA ROCHA REIS**
Data: 02/12/2024 13:33:49-0300
Verifique em <https://validar.iti.gov.br>

Prof. Agnaldo Jose da Rocha Reis
Universidade Federal de Ouro Preto

Documento assinado digitalmente
 **CLARYSSON ALBERTO MELLO DA SILVA**
Data: 02/12/2024 15:15:22-0300
Verifique em <https://validar.iti.gov.br>

Prof. Clarysson Alberto Mello da Silva
Departamento de Engenharia Nuclear - UFMG

Belo Horizonte, 29 de novembro de 2024.

Agradecimentos

O autor expressa sua gratidão às agências brasileiras de fomento à pesquisa, Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (Brasil), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (Brasil) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG (MG/Brasil), bem como ao Departamento de Engenharia Nuclear da Universidade Federal de Minas Gerais, pelo inestimável apoio, sem o qual este trabalho não teria sido possível.

Resumo

A energia eólica é uma das principais tecnologias de baixo carbono para geração de eletricidade, e o Brasil possui um grande potencial para seu desenvolvimento, especialmente nas regiões Nordeste e Sul. Estudos recentes mostram que há uma complementaridade interessante entre a energia eólica e a hidráulica, pois em períodos de seca, a geração eólica tende a ser mais alta devido ao aumento da velocidade dos ventos, e vice-versa. Em 2022, a energia eólica se consolidou como a segunda fonte mais utilizada na matriz elétrica brasileira, correspondendo a 11,8% (ou 25,6 GW) da capacidade instalada, atrás apenas da fonte hidráulica, que responde por 61,9%. O Plano Nacional de Energia projeta que, até 2050, o Brasil terá uma capacidade instalada de energia eólica de 110 GW, consolidando ainda mais a relevância dessa fonte.

No entanto, a natureza intermitente da energia eólica apresenta desafios para sua integração eficiente na rede elétrica nacional. Para lidar com esses desafios, a capacidade de prever e analisar as tendências de geração é fundamental, permitindo maior precisão na operação do Sistema Interligado Nacional (SIN) e possibilitando decisões mais informadas sobre o fornecimento de energia e a gestão dos reservatórios hídricos.

Dentre os principais fatores que influenciam a geração de um parque eólico estão a direção e a velocidade do vento. Este trabalho tem como objetivo estudar diferentes modelos para prever a geração de energia eólica no estado do Ceará e compará-los com outros estados brasileiros, como Bahia, Piauí, Rio Grande do Norte, Rio Grande do Sul e Santa Catarina. Foram avaliados modelos como ARIMA, CNN, LSTM, CNN-LSTM, CNN-ALSTM, ARIMA-CNN e ARIMA-LSTM, e os resultados evidenciaram a superioridade dos modelos híbridos em paralelo, especialmente o ARIMA-CNN e ARIMA-LSTM.

As previsões feitas pelo modelo ARIMA-CNN para o estado do Rio Grande do Sul foram aplicadas para estimar a produção de energia da usina Honda Energy Xangri-lá, utilizando duas abordagens: uma baseada em intervalos de potência e outra dependente da direção do vento. Ambas as abordagens mostraram-se promissoras, especialmente ao trabalhar com médias de produção em intervalos superiores a uma semana.

Palavras-chave: Energia eólica, Redes neurais, Previsão, ARIMA, Modelos Híbridos

Abstract

Wind energy is one of the leading low-carbon technologies for electricity generation, and Brazil has great potential for its development, especially in the Northeast and South regions. Recent studies show an interesting complementarity between wind and hydropower, as during drought periods, wind generation tends to be higher due to increased wind speeds, and vice versa. In 2022, wind energy became the second most used source in the Brazilian electricity matrix, accounting for 11.8% (or 25.6 GW) of the installed capacity, behind only hydropower, which represents 61.9%. The National Energy Plan estimates that by 2050, Brazil will have an installed wind energy capacity of 110 GW, further consolidating the relevance of this source.

However, the intermittent nature of wind energy presents challenges for its efficient integration into the national electricity grid. To address these challenges, the ability to predict and analyze generation trends is crucial, allowing greater precision in the operation of the National Interconnected System (SIN) and enabling more informed decisions regarding energy supply and reservoir management.

Among the main factors that influence the generation of a wind farm are wind direction and speed. This work aims to study different models to predict wind energy generation in the state of Ceará and compare them with other Brazilian states, such as Bahia, Piauí, Rio Grande do Norte, Rio Grande do Sul, and Santa Catarina. Models such as ARIMA, CNN, LSTM, CNN-LSTM, CNN-ALSTM, ARIMA-CNN, and ARIMA-LSTM were evaluated, and the results highlighted the superiority of the parallel hybrid models, especially ARIMA-CNN and ARIMA-LSTM.

The forecasts made by the ARIMA-CNN model for the state of Rio Grande do Sul were used to estimate the energy production of the Honda Energy Xangri-lá wind farm, using two approaches: one based on power intervals and the other dependent on wind direction. Both approaches proved to be promising, particularly when working with average production over intervals longer than a week.

Keywords: Wind energy, Neural networks, Forecasting, ARIMA, Hybrid models

Lista de ilustrações

Figura 1 – Hierarquia termodinâmica do planeta terra [1].	24
Figura 2 – Imagem da movimentação e aquecimento das massas de ar [2].	25
Figura 3 – Fotografia da primeira turbina eólica construída para geração de energia elétrica [3].	26
Figura 4 – Fotografia de uma fazenda eólica na Califórnia com aerogeradores HAWT [4].	27
Figura 5 – Curva de potência de um aerogerador [5].	28
Figura 6 – Validação cruzada para séries temporais de velocidade e direção do vento. Os parâmetros τ e β são selecionados de forma otimizada durante o processo de validação cruzada.	33
Figura 7 – Arquitetura da rede neural CNN para previsão de velocidade e direção do vento.	35
Figura 8 – Arquitetura da rede neural LSTM para previsão de velocidade e direção do vento.	37
Figura 9 – Ilustração da arquitetura do Autoencoder LSTM usada para previsão de velocidade e direção do vento.	38
Figura 10 – Arquitetura proposta do CNN-ALSTM para previsão de velocidade e direção do vento.	39
Figura 11 – Fluxograma das etapas para estimação da produção energética de uma usina.	41
Figura 12 – Bahia - Gráfico de correlação dos dados climáticos referentes ao estado da Bahia.	54
Figura 13 – Bahia - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.	55
Figura 14 – Bahia - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.	56
Figura 15 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.	57

Figura 16 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.	58
Figura 17 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.	58
Figura 18 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.	59
Figura 19 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.	59
Figura 20 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.	60
Figura 21 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.	60
Figura 22 – Piauí - Gráfico de correlação dos dados climáticos referentes ao estado do Piauí.	62
Figura 23 – Piauí - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.	63
Figura 24 – Piauí - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.	65
Figura 25 – Piauí - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.	65
Figura 26 – Piauí - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.	66
Figura 27 – Piauí - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.	66

Figura 28 – Piauí - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.	67
Figura 29 – Piauí - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.	68
Figura 30 – Piauí - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.	68
Figura 31 – Piauí - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.	68
Figura 32 – Rio Grande do Norte - Gráfico de correlação dos dados climáticos referentes ao estado do Rio Grande do Norte.	70
Figura 33 – Rio Grande do Norte - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.	72
Figura 34 – Rio Grande do Norte - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.	73
Figura 35 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.	74
Figura 36 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.	74
Figura 37 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.	74
Figura 38 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.	75
Figura 39 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.	76

Figura 40 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.	76
Figura 41 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.	76
Figura 42 – Santa Catarina - Gráfico de correlação dos dados climáticos referentes ao estado de Santa Catarina.	78
Figura 43 – Santa Catarina - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.	79
Figura 44 – Santa Catarina - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.	81
Figura 45 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.	81
Figura 46 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.	82
Figura 47 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.	82
Figura 48 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.	83
Figura 49 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.	84
Figura 50 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.	84
Figura 51 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.	84

Figura 52 – Ceará - Gráfico de correlação dos dados climáticos referentes ao estado do Ceará.	86
Figura 53 – Ceará - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.	88
Figura 54 – Ceará - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.	89
Figura 55 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.	90
Figura 56 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.	90
Figura 57 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.	90
Figura 58 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.	91
Figura 59 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.	92
Figura 60 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.	92
Figura 61 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.	92
Figura 62 – Rio Grande do Sul - Gráfico de correlação dos dados climáticos referentes ao estado do Rio Grande do Sul.	94
Figura 63 – Rio Grande do Sul - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.	96

Figura 64 – Rio Grande do Sul - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.	97
Figura 65 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.	98
Figura 66 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.	98
Figura 67 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.	98
Figura 68 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.	99
Figura 69 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.	100
Figura 70 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.	100
Figura 71 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.	100
Figura 72 – Curvas de potência interpoladas para os diferentes aerogeradores comparados com os dados reais de potência.	103
Figura 73 – As Figuras (a), (b), (c) e (d) mostram a projeção da potência utilizando a direção do vento e a potência média real sobre o intervalo projetado sem a direção do vento para os intervalos de tempo de 1 dia, 7 dias, 14 dias e 30 dias.	105

Lista de tabelas

Tabela 1 – Valores típicos do coeficiente de rugosidade z_0	28
Tabela 2 – Nomenclatura associada a cada variável física explicando a velocidade e direção do vento como alvos.	43
Tabela 3 – Valores de referência para o coeficiente KMO.	46
Tabela 4 – Bahia - Resultados de KMO e Teste de Esfericidade de Bartlett	54
Tabela 5 – Bahia - Variância Explicada por Fator	55
Tabela 6 – Bahia - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionaridade.	56
Tabela 7 – Bahia - Métricas de desempenho para CNN, LSTM e ARIMA (Velocidade e Direção do Vento).	59
Tabela 8 – Bahia - Métricas de desempenho para os modelos combinados (Velocidade e Direção do Vento).	61
Tabela 9 – Bahia - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM	61
Tabela 10 – Piauí - Resultados de KMO e Teste de Esfericidade de Bartlett	62
Tabela 11 – Piauí - Variância Explicada por Fator	62
Tabela 12 – Piauí - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionaridade.	64
Tabela 13 – Piauí - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).	67
Tabela 14 – Piauí - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).	69
Tabela 15 – Piauí - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM	69
Tabela 16 – Rio Grande do Norte - Resultados de KMO e Teste de Esfericidade de Bartlett	71
Tabela 17 – Rio Grande do Norte - Variância Explicada por Fator	71
Tabela 18 – Rio Grande do Norte - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionaridade.	73
Tabela 19 – Rio Grande do Norte - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).	75
Tabela 20 – Rio Grande do Norte - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).	77

Tabela 21 – Rio Grande do Norte - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM	77
Tabela 22 – Resultados de KMO e Teste de Esfericidade de Bartlett	78
Tabela 23 – Santa Catarina - Variância Explicada por Fator	79
Tabela 24 – Santa Catarina - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionaridade.	80
Tabela 25 – Santa Catarina - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).	83
Tabela 26 – Santa Catarina - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).	85
Tabela 27 – Santa Catarina - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM	85
Tabela 28 – Ceará - Resultados de KMO e Teste de Esfericidade de Bartlett	86
Tabela 29 – Ceará - Variância Explicada por Fator	87
Tabela 30 – Ceará - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionaridade.	89
Tabela 31 – Ceará -Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).	91
Tabela 32 – Ceará - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).	93
Tabela 33 – Ceará - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM	93
Tabela 34 – Resultados de KMO e Teste de Esfericidade de Bartlett	94
Tabela 35 – Rio Grande do Sul - Variância Explicada por Fator	95
Tabela 36 – Rio Grande do Sul - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionaridade.	97
Tabela 37 – Rio Grande do Sul - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).	99
Tabela 38 – Rio Grande do Sul - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).	101
Tabela 39 – Rio Grande do Sul - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM	101
Tabela 40 – Dados de localização da usina e do ponto de captação de dados climáticos.	102
Tabela 41 – Características dos aerogeradores V112/3000 e V136/3450 utilizados no modelo de usina.	103

Tabela 42 – Características dos aerogeradores V112/3000 e V136/3450 utilizados no modelo de usina.	105
---	-----

Lista de abreviaturas e siglas

RNA	Rede Neural Artificial
LSTM	Long Short-Term Memory
CNN	Rede Neural Convolutacional
CNN-ALSTM	Convolutacional Autoencoder LSTM
ARIMA	Modelo Autorregressivo Integrado de Médias Móveis
BHO	Otimização Bayesiana de Hiperparâmetros
MAPE	Erro Médio Absoluto Percentual
MAE	Erro Médio Absoluto
RMSE	Root Mean Square Error
MSE	Mean Square Error
KMO	Kaiser-Meyer-Olkin
HAWT	Turbina Eólica de Eixo Horizontal

Sumário

1	INTRODUÇÃO	20
2	MARCO TEÓRICO	24
2.1	Energia eólica	24
2.2	Aerogeradores	26
2.3	Curva de Potência	27
2.4	Perfil logarítmico do vento	28
2.5	Interpolação por Splines de ordem superior	29
2.6	Imputação Múltipla por Equações Encadeadas	30
2.7	Otimização Bayesiana	30
2.7.1	Processo Gaussiano	30
2.8	Métricas de desempenho e Validação Cruzada	32
2.9	Modelos Preditivos	34
2.9.1	ARIMA	34
2.9.2	CNN	35
2.9.3	LSTM	36
2.9.4	Autocodificador LSTM	38
2.9.5	CNN-ALSTM	39
2.9.6	Modelos Acoplados por Regressão Linear Multivariada	39
3	METODOLOGIA	41
3.1	Dados Climáticos	42
3.2	Métodos de análise do conjunto de dados	43
3.2.1	Matriz de correlações amostrais	43
3.2.2	Análise de Fatores	44
3.2.2.1	Kaiser-Meyer-Olkin (KMO)	45
3.2.2.2	Teste de esfericidade de Bartlett	45
3.2.3	Kurtosis	46
3.2.4	Skewness	47
3.2.5	Jarque-Bera	48
3.2.6	Dickey-Fuller	48
3.2.7	Outras Métricas	48
3.3	Previsão dos dados climáticos e seleção do modelo preditivo	49
3.4	Modelo de usina	49
3.4.1	Estimativa da Curva de Potência para cada aerogerador	49
3.4.2	Potência estimada da usina	50

3.5	Aplicação computacional	51
3.5.1	Análise exploratória dos dados	51
3.5.2	Tratamento de dados	52
3.5.3	Otimização de hiperparâmetros e treinamento de modelos preditivos	52
4	RESULTADOS	53
4.1	Dados climáticos e modelos	53
4.1.1	Bahia	53
4.1.2	Piauí	61
4.1.3	Rio Grande do Norte	70
4.1.4	Santa Catarina	77
4.1.5	Ceará	86
4.1.6	Rio Grande do Sul	94
4.2	Considerações gerais	102
4.3	Modelo de Usína	102
5	CONCLUSÃO	107
	REFERÊNCIAS	109

1 Introdução

No cenário atual, onde países como China, Índia, Brasil, EUA e países da União Europeia buscam atingir emissões líquidas zero até 2050 [6], a energia eólica vem se apresentando como uma das grandes apostas para a transição energética global focada na redução de dependência de fontes de energia fósseis.

Para atingir esses objetivos, a União Europeia se comprometeu a aumentar sua capacidade de energia eólica de 204 GW em 2022 para 500 GW até 2030 [7]. Nos Estados Unidos, a meta é instalar 30 GW de energia eólica offshore até 2030 [8] [9], enquanto a China planeja atingir 1.200 GW de energia renovável, incluindo solar e eólica, até o mesmo ano [10]. Na Índia, o objetivo é construir 140 GW de energia eólica até 2030 [11], e o Japão espera adicionar 10 GW de energia eólica offshore até 2030 [12]. No Brasil, as expectativas são de triplicar a capacidade instalada, que atualmente está em torno de 30 GW [13], até 2030 [14, 15].

Nesse contexto das energias renováveis, o Brasil se apresenta como um dos grandes nomes possuindo, em 2021, 82,9% de sua rede elétrica composta de fontes renováveis [16]. Além disso, a indústria de energia eólica no Brasil experimenta um crescimento significativo desde 2015 ano no qual o Brasil já era o 8º maior produtor de energia eólica do mundo, possuindo uma capacidade de geração de 21,36 TWh e uma capacidade instalada de 8,7 GW [17]. Em 2022, o Brasil passou a ocupar a 6ª posição, com uma capacidade eólica de 25,6 GW, representando 13,4% de sua capacidade instalada total [18]. Atualmente, o Brasil ocupa o quarto lugar como maior gerador de energia eólica, com 72,24 TWh. A determinação do país em expandir a energia eólica é evidente, com a expansão anual de energia eólica passando de 2.014 MW/ano em 2016 para aproximadamente 4.100 MW/ano em 2022 [17, 18].

As mudanças expressivas na estrutura da matriz elétrica, como as que estão ocorrendo no Brasil, especialmente com a crescente participação de fontes intermitentes como a energia eólica, trazem desafios consideráveis de integração e operação para a gestão do sistema elétrico, tais como a variabilidade da geração, a necessidade de estabilização da rede, o armazenamento de energia, a integração com outras fontes, a previsão e planejamento mais precisos, além de investimentos em infraestrutura de transmissão nos locais de grande potencial eólico. Essas transformações exigem soluções mais complexas e uma adaptação contínua para garantir a estabilidade e eficiência da rede. Apesar de ser uma fonte de energia renovável com baixo carbono, a energia eólica enfrenta desafios significativos, especialmente em relação à sua integração nas redes elétricas devido à sua natureza intermitente. Essa intermitência pode causar congestionamentos nas linhas de

transmissão, que ocorrem quando a geração de energia excede a capacidade de transporte da rede, ou quando a distribuição da eletricidade entre os centros de consumo não é equilibrada de forma eficiente, exigindo uma coordenação eficaz entre diversos stakeholders [19].

Para enfrentar esses desafios de forma eficiente, os operadores de rede estão cada vez mais recorrendo a previsões e análises preditivas. Essas ferramentas permitem decisões mais precisas sobre a geração de energia e o gerenciamento de armazenamento, ajudando a equilibrar a oferta e demanda em tempo real. A qualidade dos recursos eólicos, como a velocidade do vento, é um fator crucial que afeta diretamente o desempenho dos parques eólicos e, por consequência, a operação da rede elétrica como um todo [20–24].

Com o intuito de superar esses desafios, pesquisadores ao redor do mundo têm explorado diversas abordagens com o intuito de melhorar a previsão da energia eólica. Dentre essas metodologias, as Redes Neurais Artificiais (RNA) emergem como uma solução promissora, uma vez que possuem a capacidade de capturar relações não lineares complexas e demonstram uma notável adaptabilidade e flexibilidade em relação às variáveis de entrada. As RNAs são especialmente eficazes em identificar padrões complexos nos dados, o que as torna uma ferramenta valiosa para prever os regimes do vento [25–28].

No Brasil, o uso de RNAs para previsão de energia eólica tem ganhado relevância, acompanhando o rápido crescimento do setor eólico e seu papel estratégico na matriz energética nacional. Com a expansão da capacidade de geração eólica, essas tecnologias de previsão são essenciais para garantir a eficiência e estabilidade do sistema elétrico brasileiro [25–28].

Dentre os modelos de RNAs que vêm sendo estudados, os modelos híbridos que combinam RNAs com outras técnicas de modelagem trazem vantagens ainda maiores em comparação aos métodos tradicionais. Além de capturar dinâmicas complexas e fornecer previsões mais precisas, esses modelos têm a capacidade de integrar o melhor de diferentes abordagens, combinando flexibilidade das RNAs para detectar padrões não lineares com a precisão de métodos tradicionais. Modelos híbridos como o ARIMA-LSTM, que une uma rede neural Long Short-Term Memory (LSTM) com um Modelo Autorregressivo Integrado de Médias Móveis (ARIMA), vem se demonstrando como ferramentas ainda mais poderosas para análises preditivas. Isso os torna mais eficazes na gestão da variabilidade e intermitência da geração eólica [29–32].

Além dos modelos híbridos, novas técnicas de ajuste de hiperparâmetros, como a Otimização Bayesiana de Hiperparâmetros (BHO), e o uso de estruturas mais complexas baseadas em Redes Neurais Artificiais (RNAs), como os autocodificadores, têm mostrado resultados promissores no aumento da precisão de modelos preditivos [33–37].

Por fim, os modelos híbridos, que combinam a flexibilidade das RNAs para captar padrões não lineares com a precisão de métodos tradicionais, juntamente com técnicas

avançadas como as mencionadas, podem desempenhar um papel crucial na estabilidade e eficiência do sistema elétrico. Ao unir essas abordagens, é possível obter previsões mais confiáveis, tanto a curto quanto a longo prazo, o que contribui significativamente para o planejamento energético e para a redução das incertezas associadas à intermitência da geração eólica.

O objetivo deste trabalho é desenvolver uma ferramenta para prever a geração de energia elétrica em uma planta eólica, utilizando dados climáticos gerados por modelos preditivos.

Para alcançar esse objetivo, foram utilizados dados climáticos fornecidos pelo Instituto Nacional de Meteorologia (INMET) [38] de seis estados brasileiros: Bahia, Ceará, Piauí, Rio Grande do Norte, Rio Grande do Sul e Santa Catarina, que possuem grande relevância para a geração de energia eólica no país. Com esses dados, foram implementados diversos modelos de ajuste de séries temporais, incluindo Long Short-Term Memory (LSTM) [39], Redes Neurais Convolucionais (CNN) [40], o Modelo Autorregressivo Integrado de Médias Móveis (ARIMA), além de modelos híbridos como Convolutional Autoencoder LSTM (CNN-ALSTM) e modelos híbridos em paralelo como CNN-LSTM, ARIMA-LSTM e CNN-ARIMA para realizar previsões locais desses dados climáticos. Após isso, foi selecionada a melhor projeção para estimar a geração de uma planta eólica com as características de uma planta real em operação no ano de 2022, localizada próxima aos pontos de captação dos dados climáticos escolhidos. Espera-se que os resultados obtidos melhorem a confiabilidade e a precisão das previsões de energia eólica, contribuindo para a otimização do uso de energias renováveis no Brasil.

No Capítulo 1 – Introdução, foi fornecida uma visão geral sobre os desafios relacionados ao desenvolvimento e à operação da energia eólica, destacando como a intermitência das fontes eólicas pode gerar problemas. Também foi explorada a possível solução oferecida pelo uso de modelos preditivos para mitigar essas questões. Além disso, foram apresentados os métodos e objetivos deste trabalho, com o intuito de abordar esses desafios e otimizar a gestão da energia eólica.

No Capítulo 2 – Marco Teórico, serão abordados a história e funcionamento dos geradores eólicos, além da descrição dos modelos preditivos utilizados neste trabalho, assim como os métodos empregados no pré-tratamento de dados e as métricas utilizadas para a avaliação desses modelos.

No Capítulo 3 – Metodologia, será detalhada a metodologia utilizada neste trabalho, incluindo as etapas de construção do modelo de estimativa de geração de energia para uma usina e as diferentes abordagens adotadas para estimar o funcionamento de uma usina real.

No Capítulo 4 – Resultados, serão apresentados, separadamente para cada estado

estudado, as correlações entre os dados utilizados para treinamento das Redes Neurais Artificiais (RNAs), assim como as estatísticas gerais dos dados de principal interesse, que são a direção e a velocidade do vento. Além disso, os resultados das previsões geradas por cada modelo preditivo serão apresentados por meio de gráficos comparativos e tabelas com as principais métricas de avaliação. Por fim, será apresentado o resultado da previsão de geração de energia elétrica de uma usina eólica, comparada com uma usina real.

2 Marco Teórico

Este capítulo aborda os conceitos relacionados à energia eólica e apresenta os métodos computacionais aplicados para atingir os objetivos desta dissertação.

2.1 Energia eólica

A energia eólica é uma das formas em que a energia produzida pelo sol se manifesta no sistema termodinâmico da Terra, sendo esta uma parte importante do maquinário termodinâmico desse sistema. A energia solar ao adentrar na atmosfera, é absorvida de forma desigual aquecendo as massas de ar de forma desigual. A desigualdade de temperatura das massas de ar criam então zonas de pressão que, junto do movimento da terra, regem a movimentação dessas massas de ar resultando nos ventos de circulação global [1, 41].

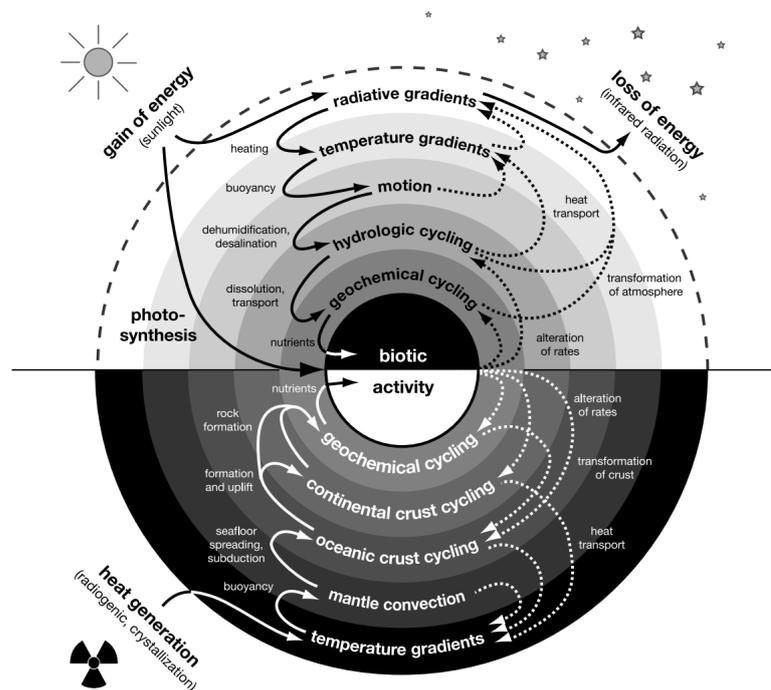


Figura 1 – Hierarquia termodinâmica do planeta terra [1].

Por causa da inclinação do planeta Terra em $23,5^\circ$ e dos movimentos de rotação e translação do planeta, as regiões tropicais e equatoriais possuem uma maior incidência de radiação solar. Por esse motivo, o ar frio, de mais alta pressão, tende a circular para os trópicos, de menor pressão, onde é aquecido e posteriormente deslocado em direção aos pólos. Os movimentos de rotação e translação afetam o funcionamento desse sistema ao deslocar os pontos de menor incidência de radiação solar ao passar do tempo. A influência

desses movimentos provoca efeitos sazonais no movimento das massas de ar provocando mudanças em sua direção e velocidade [41].

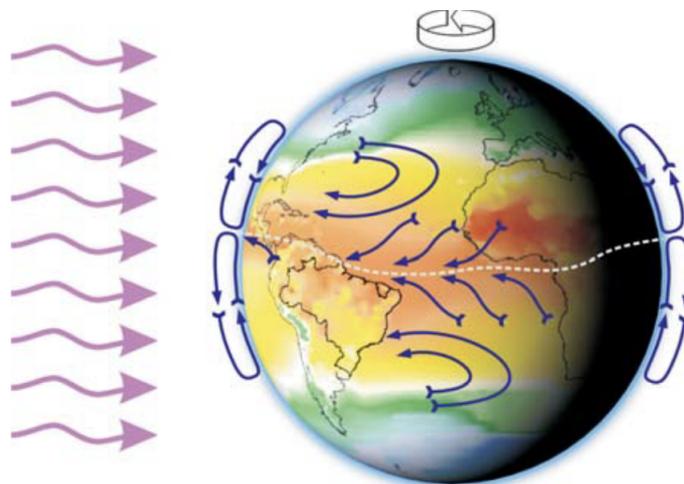


Figura 2 – Imagem da movimentação e aquecimento das massas de ar [2].

Outros fatores também influenciam na dinâmica das massas de ar. Efeitos globais como o ciclo de aquecimento das águas do Pacífico e características locais como altitude e relevo também influenciam nesse movimento em uma certa escala.

A história da energia eólica tem início especulado a mais de 3000 anos, onde especulações implicam na presença de mecanismos movidos a energia eólica na Grécia antiga, Babilônia, Gália antiga e Inglaterra saxônica. Heron de Alexandria se referiu na *Pneumática* a um *anemurion* que seria um cata-vento. No entanto, não existem registros confiáveis que possam confirmar a existência de um rotor eólico nesses períodos da história nos permitindo apenas especular sobre os primeiros passos dessa tecnologia [42, 43].

Embora dispositivos movidos pelo vento existam há milhares de anos, foi apenas há poucos séculos que a tecnologia eólica para a geração de trabalho começou a dar seus primeiros passos concretos. Os primeiros exemplos incluem os moinhos de vento de eixo vertical, usados nas fronteiras entre a Pérsia e o Afeganistão por volta de 200 a.C. Muito tempo depois, surgiram os moinhos de vento de eixo horizontal, mais conhecidos, que apareceram na Holanda e no Mediterrâneo entre os anos 1300 e 1875 d.C [44].

Apesar da longa existência da tecnologia eólica, a geração de energia elétrica por meio destes dispositivos começou há pouco mais de 120 anos, quando o Prof. James Blyth, na Escócia, construiu pela primeira vez um moinho de vento para gerar eletricidade em 1887 (Figura 3). Desde então, o esforço para desenvolver a energia eólica tem variado de acordo com as oscilações nos preços do petróleo, intensificando-se durante os momentos de crise dessa indústria. A tecnologia da energia eólica foi impulsionada pela primeira vez durante a crise do petróleo dos anos 1970, mas decaiu logo depois disso [3, 43].

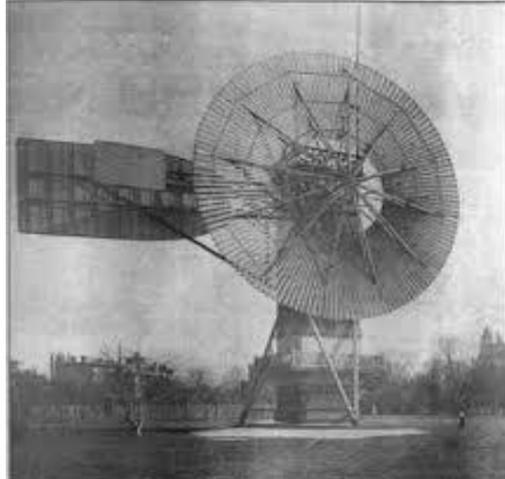


Figura 3 – Fotografia da primeira turbina eólica construída para geração de energia elétrica [3].

2.2 Aero geradores

O princípio fundamental utilizado para o funcionamento de um gerador eólico consiste na transformação da energia cinética do vento em energia elétrica. A energia cinética do vento está estritamente ligada à velocidade do ar e à densidade do mesmo. A Equação 2.1 descreve a energia cinética do ar.

$$E = \frac{1}{2}v^2M \quad (2.1)$$

Onde v é a velocidade do vento e M é a massa do ar.

O equipamento construído com o propósito de realizar essa conversão recebe o nome de Aero gerador e normalmente é composto por duas partes principais: as pás, que são os componentes responsáveis pela captação da energia motriz disponível nos ventos e conversão desta para energia mecânica no eixo do gerador, e o gerador, que é o componente responsável por converter a energia mecânica do eixo de transmissão conectado ao mancal da ligação com as pás [4, 45, 46].

Nos dias de hoje, os geradores de eixo horizontal, Horizontal Axis Wind Turbine (HAWT) (representados na Figura 4), aqueles no qual o eixo principal é posto no sentido horizontal de forma paralela ao solo, é o modelo mais utilizados pelos grandes produtores eólicos. Para aero geradores a massa do ar depende diretamente do tamanho das pás que efetivamente limitam a quantidade de ar com a qual a superfície do aero gerador interage. Sendo assim, para esses objetos a equação de energia fica sendo descrita como na Equação 2.2.

$$E_{aerogerador} = \frac{1}{2}A\rho v^2C_p \quad (2.2)$$



Figura 4 – Fotografia de uma fazenda eólica na Califórnia com aerogeradores HAWT [4].

Onde A é a área correspondente ao círculo de raio igual ao comprimento das pás do aerogerador, ρ é a densidade do ar e C_p é o coeficiente de potência do aerogerador. O coeficiente de potência é definido como a razão entre a energia disponível no vento e a energia efetivamente convertida pelo aerogerador (Equação 2.3). Esse coeficiente quantifica a fração da energia contida no vento que é convertida em energia útil pelo aerogerador. O valor teórico máximo de C_p é cerca de 59,3%, conhecido como o limite de Betz.

$$C_p = \frac{E_{aerogerador}}{E_{vento}} \quad (2.3)$$

2.3 Curva de Potência

Nos geradores atuais, a ação do vento sobre a superfície das pás em movimento e a atuação dos vários mecanismos de controle que atuam nesses equipamentos geram uma resultante de força complexa impedindo essa de ser estudada de uma forma estática. O resultado prático desta ação complexa sobre as pás pode ser quantificado pela energia gerada pelo gerador. Essa relação é chamada de curva de potência do aerogerador e fica sendo apresentada como relação entre velocidade do vento e a energia gerada [45]. Uma imagem representativa da curva de potência de um aerogerador é apresentada na Figura 5.

Essa curva é disponibilizada pelos fabricantes e apresentam três novas grandezas que são de grande importância para a geração da energia eólica, sendo essas as velocidades de *cut-in*, ideal e *cut-off*. Essas grandezas representam os limites dos diferentes regimes de geração, sendo a velocidade de *cut-in* a velocidade mínima para a geração de energia daquele aerogerador, a ideal representa o início do *plateau* onde o aerogerador tem seu

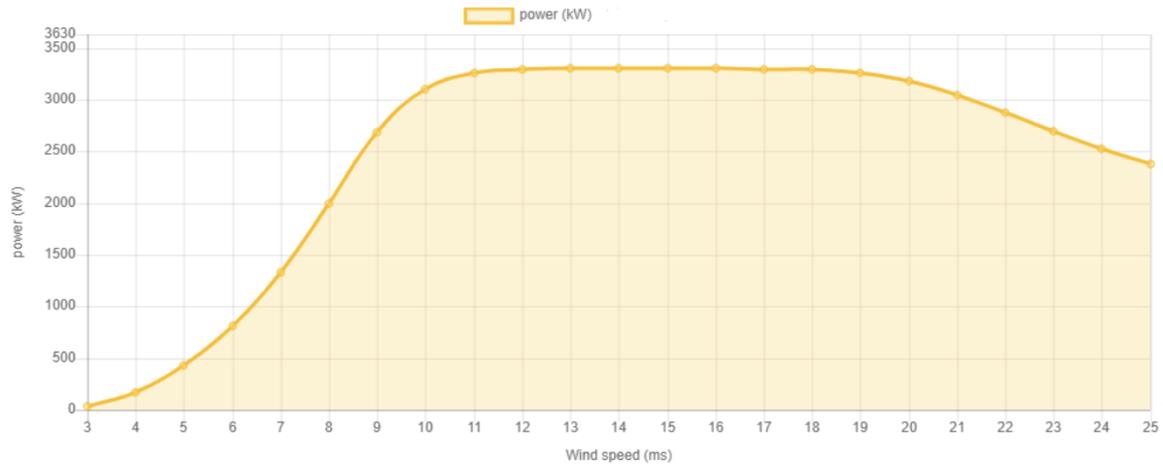


Figura 5 – Curva de potência de um aerogerador [5].

maior desempenho e a velocidade de *cut-off* representa a velocidade máxima na qual o aerogerador pode efetivamente atuar sem danificar seus componentes de forma que, caso seja superada, o rotor é desconectado do gerador, fazendo que as pás girem livremente sem gerar energia [45,47].

2.4 Perfil logarítmico do vento

O perfil logarítmico do vento é um modelo matemático utilizado para descrever como a velocidade do vento varia com a mudança de altitude. Esse perfil é comumente empregado em estudos de energia eólica, engenharia civil e meteorologia para estimar a velocidade do vento em diferentes alturas. A equação do perfil logarítmico do vento é dada pela Equação 2.4, onde h representa a altura em que a velocidade do vento é estimada, V_h é a velocidade do vento na altura h , z é a altura de referência, V_z é a velocidade do vento na altura z , e z_0 é a constante de rugosidade do terreno.

$$V_h = V_z \cdot \left(\frac{\ln\left(\frac{h}{z_0}\right)}{\ln\left(\frac{z}{z_0}\right)} \right) \quad (2.4)$$

Tabela 1 – Valores típicos do coeficiente de rugosidade z_0

Tipo de Superfície	z_0 (m)
Grandes cidades	2,0 - 3,0
Pastagem curta	0,1
Floresta rala, campo com arbustos	0,25
Área urbana ou floresta densa	1,0 - 2,0

Para a utilização desse modelo, a estimativa do coeficiente de rugosidade para um local é um objetivo particularmente difícil. Por conta disso, é comum utilizar valores

tabelados e pequenas correções que correspondam melhor à situação estudada. A Tabela 1 mostra os valores típicos de coeficientes de rugosidade para diferentes tipos de terrenos.

2.5 Interpolação por Splines de ordem superior

A interpolação por splines de ordem superior é um método que utiliza de um spline de ordem k . O spline é uma função formada por vários polinômios de grau k que são ajustados em segmentos consecutivos dentro de um intervalo de dados. Esses polinômios se conectam de maneira que a função spline resultante seja contínua e suas derivadas também permaneçam contínuas até a ordem $k - 1$ nos pontos de junção, que são conhecidos como "knots" ou "nós".

Splines de ordem superior são funções polinomiais que são divididas em partes ao longo de um intervalo e são utilizadas para criar uma interpolação suave dos dados. Elas são particularmente úteis quando é necessário obter uma curva que passe por todos os pontos de forma contínua e bem ajustada. Ao contrário dos polinômios de grau elevado, que frequentemente podem apresentar oscilações indesejadas, os splines oferecem uma abordagem mais equilibrada e eficiente para conectar os pontos, resultando em curvas suaves e precisas.

Para que os polinômios se encaixem de maneira suave entre os diferentes intervalos, eles precisam atender a algumas condições específicas, como:

1. A função $S(x)$ deve passar por todos os pontos de dados dados (x_i, y_i) .
2. As primeiras, segundas, terceiras, quartas e quintas derivadas do spline de ordem 6 devem ser contínuas em cada nó x_i .
3. As condições de extremidade, como valores de derivada nos pontos extremos da função, devem ser mantidas.

Para splines de grau 5 (ordem 6) cada segmento é um polinômio de grau 5, sendo a equação final da interpolação a somatória de todos os segmentos:

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 + e_i(x - x_i)^4 + f_i(x - x_i)^5 \quad (2.5)$$

$$S(x) = \sum_i S_i(x) \quad (2.6)$$

O spline de grau 5 possui maior flexibilidade e suavidade do que o cúbico, grau comumente utilizado para a interpolação de Splines. As condições de continuidade garantem que a função e suas quatro primeiras derivadas sejam contínuas nos nós.

2.6 Imputação Múltipla por Equações Encadeadas

Imputação Múltipla por Equações Encadeadas (MICE) é uma técnica estatística robusta que é amplamente empregada para tratar dados ausentes em conjuntos de dados multivariados. Diferente das abordagens tradicionais, como a imputação simples, o MICE gera múltiplos conjuntos de dados imputados, permitindo capturar a incerteza associada ao processo de imputação.

O método MICE opera por meio de um processo iterativo, no qual as variáveis com dados ausentes são imputadas uma a uma, com base nas informações das outras variáveis. Inicialmente, valores são atribuídos aos dados faltantes usando uma imputação simples, como médias ou medianas. Depois, cada variável com dados ausentes é modelada condicionalmente às demais variáveis, utilizando modelos preditivos apropriados, como regressão linear, regressão logística ou outros métodos adequados para o tipo de dado.

Esse processo é repetido em várias iterações, ajustando continuamente os valores imputados até que os resultados se estabilizem. O MICE geralmente gera entre 5 e 10 conjuntos de dados imputados, o que permite refletir a variabilidade na imputação. Após a imputação, cada conjunto de dados é analisado individualmente, e os resultados são combinados utilizando métodos como as regras de Rubin, que consideram a variabilidade entre as imputações.

O MICE é amplamente utilizado em campos como ciências sociais, medicina e economia, onde a presença de dados ausentes é comum e pode afetar significativamente os resultados se não for adequadamente tratado. Em resumo, o MICE é uma ferramenta poderosa e flexível para imputação de dados ausentes, ajudando a melhorar a precisão das inferências estatísticas em estudos com dados incompletos.

2.7 Otimização Bayesiana

A otimização bayesiana é um método eficaz para otimizar globalmente funções objetivas desconhecidas [48–50]. O problema de otimização é expresso como

$$\mathbf{x}^* = \operatorname{argmax}_{\mathbf{x} \in \mathcal{D}} \mathcal{F}(\mathbf{x}) \quad (2.7)$$

onde $\mathcal{D} \subset \mathbb{R}^d$. As avaliações de funções podem ser ruidosas, representadas como $y = \mathcal{F}(\mathbf{x}) + \varepsilon$ com $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

2.7.1 Processo Gaussiano

O processo gaussiano (GP) [51, 52] é comumente usado como o modelo probabilístico substituto na otimização bayesiana devido à sua flexibilidade na definição de priores para

funções que variam suavemente. Ele é caracterizado por uma função média, $\mu(\mathbf{x})$, e uma função de covariância, $\kappa(\mathbf{x}, \tilde{\mathbf{x}})$. Uma função amostrada de um processo gaussiano é dada por

$$\mathcal{F}(\mathbf{x}) \sim \text{GP}(\mu(\mathbf{x}), \kappa(\mathbf{x}, \tilde{\mathbf{x}})), \quad (2.8)$$

onde $\mathcal{F}(\mathbf{x})$ em qualquer ponto \mathbf{x} é uma variável aleatória distribuída gaussiana com uma média e variância especificadas. A função média anterior em um Processo Gaussiano pode ser assumida como zero, tornando o GP totalmente definido por sua função de covariância. Uma função de covariância comumente usada é o kernel exponencial quadrado, definido como:

$$\kappa(\mathbf{x}, \tilde{\mathbf{x}}) = \exp\left(-\frac{1}{2C}\|\mathbf{x} - \tilde{\mathbf{x}}\|^2\right), \quad (2.9)$$

aqui, C é o parâmetro de escala de comprimento, crucial para determinar a suavidade da função subjacente. Embora existam diversas funções de covariância, neste trabalho, o foco está na função de covariância exponencial quadrática.

Sejam nossas observações da função desconhecida representadas como $\{\mathbf{x}_{1:p}, \mathbf{y}_{1:p}\}$, onde $y = \mathcal{F}(\mathbf{x}) + \varepsilon$ e $f \sim \text{GP}(\cdot)$. O Processo Gaussiano é definido por uma função de covariância $\kappa(\mathbf{x}, \tilde{\mathbf{x}})$. Os valores observados da função $\mathbf{y}_{1:p}$ seguem uma distribuição gaussiana multivariada $\mathbf{y}_{1:p} \sim \text{N}(0, \mathcal{K})$, conforme mostrado na Equação 2.10.

$$\mathcal{K} = \begin{pmatrix} \kappa(\mathbf{x}, \mathbf{x}_1) & \cdots & \kappa(\mathbf{x}, \mathbf{x}_p) \\ \vdots & \ddots & \vdots \\ \kappa(\mathbf{x}, \mathbf{x}_1) & \cdots & \kappa(\mathbf{x}, \mathbf{x}_p) \end{pmatrix} \quad (2.10)$$

onde \mathcal{K} é conhecido como a matriz kernel. Para um novo ponto de dados \mathbf{x}_{p+1} , deixe o valor da função ser \mathbf{y}_{p+1} . Os valores $\mathbf{y}_{1:p}$ e \mathbf{y}_{p+1} juntos seguem uma distribuição gaussiana conjunta, que pode ser expressa como

$$\begin{pmatrix} \mathbf{y}_{1:p} \\ \mathbf{y}_{p+1} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathcal{K} & \boldsymbol{\kappa} \\ \boldsymbol{\kappa}^T & \kappa(\mathbf{x}_{p+1}, \mathbf{x}_{p+1}) \end{pmatrix}\right), \quad (2.11)$$

com $\boldsymbol{\kappa} = (\kappa(\mathbf{x}_1, \mathbf{x}_{p+1}), \kappa(\mathbf{x}_2, \mathbf{x}_{p+1}), \dots, \kappa(\mathbf{x}_p, \mathbf{x}_{p+1}))$. A distribuição preditiva posterior em um novo ponto \mathbf{x}_{p+1} pode ser formulada como $P(\mathbf{y}_{p+1} | \mathbf{x}_{p+1}, \mathbf{x}_{1:p}, \mathbf{y}_{1:p}) \sim \text{N}(\mu(\mathbf{x}_{p+1}), \sigma^2(\mathbf{x}_{p+1}))$ onde a média e a variância posteriores podem ser determinadas como

$$\mu(\mathbf{x}_{p+1}) = \boldsymbol{\kappa}^T (\mathcal{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y}_{1:p} \quad (2.12)$$

$$\sigma^2(\mathbf{x}_{p+1}) = \kappa(\mathbf{x}_{p+1}, \mathbf{x}_{p+1}) - \boldsymbol{\kappa}^T (\mathcal{K} + \sigma^2 \mathbf{I})^{-1} \boldsymbol{\kappa}. \quad (2.13)$$

Devido às suas propriedades, as derivadas parciais de uma GP permanecem gaussianas, pois a diferenciação é uma operação linear. Isso permite a incorporação da derivada da função objetivo desconhecida aplicando os seguintes critérios para funções de covariância, como: $\kappa \left(\partial_{x_d^{(i)}} f^{(i)}, f^{(j)} \right) = \partial_{x_d^{(i)}} \kappa(f^{(i)}, f^{(j)})$.

A otimização bayesiana utiliza uma abordagem eficiente usando uma função de utilidade substituta chamada função de aquisição. Essa função é simples de avaliar e auxilia na identificação do ótimo da função alvo. Um valor de função de aquisição mais alto sugere um valor potencial maior da função alvo.

Neste estudo, utilizou-se a função de aquisição EI [53] devido à sua praticidade e simplicidade, embora a metodologia seja aplicável a qualquer função de aquisição. A função de aquisição EI identifica pontos que devem melhorar a melhor observação atual. Supondo que nosso problema de otimização seja $\operatorname{argmax}_{\mathcal{D}} \mathcal{F}(\mathbf{x})$ e o melhor atual esteja em $\mathbf{x}^+ = \operatorname{argmax}_{\mathbf{x}_i \in \mathcal{D}} \mathcal{F}(\mathbf{x}_i)$, a função de melhoria é definida como:

$$I(\mathbf{x}) = \max\{0, \mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{x}^+)\}, \quad D \equiv \{\mathbf{x}_{1:p}, \mathbf{y}_{1:p}\}. \quad (2.14)$$

$E(I(\mathbf{x}))$ é dado por

$$\alpha(\mathbf{x}) = \begin{cases} (\mu(\mathbf{x}) - \mathcal{F}(\mathbf{x}^+))\Phi(z) + \sigma(\mathbf{x})\phi(z) & \text{if } \sigma(\mathbf{x}) > 0 \\ 0 & \text{if } \sigma(\mathbf{x}) = 0 \end{cases} \quad (2.15)$$

com $z = (\mu(\mathbf{x}) - \mathcal{F}(\mathbf{x}^+))/\sigma(\mathbf{x})$ e $\Phi(z), \phi(z)$ são a CDF e a PDF para a distribuição normal padrão.

2.8 Métricas de desempenho e Validação Cruzada

Neste trabalho, foram utilizadas as métricas amplamente reconhecidas MAE, MAPE, MSE, RMSE e R^2 para avaliar o desempenho preditivo dos modelos no conjunto de validação durante a hiperparametrização, além de apresentar os resultados de precisão no conjunto de teste.

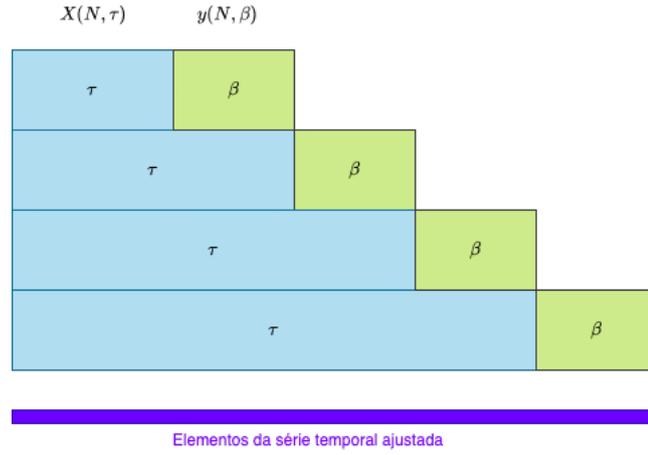


Figura 6 – Validação cruzada para séries temporais de velocidade e direção do vento. Os parâmetros τ e β são selecionados de forma otimizada durante o processo de validação cruzada.

$$\text{MAE} = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (2.16)$$

$$\text{MSE} = \frac{\sum_{i=1}^N [y_i - \hat{y}_i]^2}{N} \quad (2.17)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N [y_i - \hat{y}_i]^2}{N}} \quad (2.18)$$

$$\text{MAPE} = \frac{\sum_{i=1}^N |(y_i - \hat{y}_i)/y_i|}{N} \times 100\% \quad (2.19)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.20)$$

O conjunto de teste gerado pelo modelo de previsão selecionado tem y_i representando os valores observados e \hat{y}_i denotando os valores previstos. Este estudo realiza validação cruzada para as séries temporais de velocidade e direção do vento. A construção de conjuntos de treinamento e validação envolve levar em consideração uma matriz de entrada X com colunas τ e uma matriz de saída Y com colunas β , conforme mostrado na Figura 6. Em cada dobra, o conjunto de validação é usado para avaliar cada modelo usando a métrica MSE. O MSE médio obtido nas respectivas dobras de validação é usado como base para selecionar o melhor modelo.

2.9 Modelos Preditivos

Nesta seção, serão apresentados os modelos preditivos utilizados neste estudo. Todos os modelos tiveram os hiperparâmetros selecionados por meio da Otimização Bayesiana de Hiperparâmetros (BHO), associada à validação cruzada descrita na seção 2.8.

2.9.1 ARIMA

O modelo Autorregressivo Integrado de Médias Móveis (ARIMA) usa os períodos defasados como variáveis independentes para construir uma equação de regressão. Sua expressão geral é [54]:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \quad (2.21)$$

Em que L é o operador de defasagem, $d \in \mathbb{Z}, d > 0$, p e q representam, respectivamente, o período do modelo autorregressivo e o número de termos do modelo de médias móveis, e d representa a ordem da diferenciação. Cada um desses parâmetros representa o hiperparâmetro principal de cada um dos componentes do modelo ARIMA sendo a parte AR associada a p , I associada a d e MA associada a q .

Para a seleção desses hiperparâmetros, utilizou-se o critério de informação de Akaike (AIC), que é amplamente empregado devido à sua capacidade de balancear o ajuste do modelo aos dados e a sua complexidade. A minimização do AIC permite identificar o modelo com o melhor equilíbrio entre precisão e simplicidade, reduzindo o risco de superajuste. Dessa forma, a combinação de parâmetros p , d e q que apresenta o menor valor de AIC indica a configuração ideal de hiperparâmetros para o modelo ARIMA. O AIC é definido pela seguinte expressão:

$$\text{AIC} = 2k - 2 \ln(L), \quad (2.22)$$

onde k é o número de parâmetros estimados no modelo, que inclui os coeficientes p , d e q do modelo ARIMA. L é o valor da função de verossimilhança máxima do modelo ajustado aos dados.

Na Equação 2.22, o termo $-2 \ln(L)$ mede o ajuste do modelo aos dados, com valores menores indicando um melhor ajuste, enquanto o termo $2k$ atua como uma penalidade pela complexidade do modelo, favorecendo configurações mais simples quando apresentam ajustes comparáveis. A minimização do AIC identifica a combinação de parâmetros p , d e q ideal para o modelo, garantindo o melhor equilíbrio entre precisão e simplicidade.

2.9.2 CNN

As redes neurais convolucionais (CNNs) processam dados de imagem com eficiência usando uma estrutura de feed-forward profunda que inclui blocos de convolução. As CNNs podem extrair automaticamente recursos de dados brutos, tornando-os aplicáveis à previsão de séries temporais, tratando sequências como imagens unidimensionais. Eles oferecem as vantagens dos perceptrons multicamadas, como suporte a entradas e saídas multivariadas e aprendizagem de relações complexas, sem exigir aprendizagem direta de observações defasadas, mas em vez disso concentrando-se em representações relevantes de grandes sequências de entrada [55–58].

Uma CNN 1D típica, como demonstrado na Figura 7, consiste em três camadas principais: camadas convolucionais 1D, camadas de pooling e camadas totalmente conectadas. Além disso, dois parâmetros cruciais são a camada de abandono e a função de ativação.

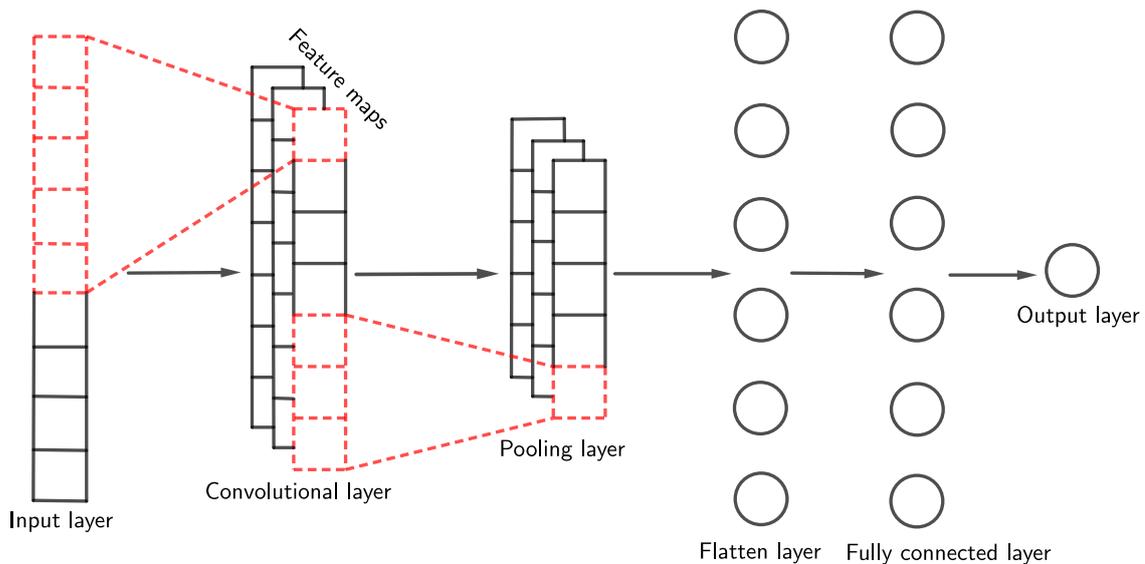


Figura 7 – Arquitetura da rede neural CNN para previsão de velocidade e direção do vento.

Uma camada convolucional unidimensional é usada para identificar características em um vetor. O vetor de entrada $x[p]$, $i = 0, 1, \dots, N - 1$, é alimentado na camada inicial da CNN. Esta camada emprega vários parâmetros:

- **Filters e Kernels:** Eles produzem mapas de recursos aplicando convoluções aos dados de entrada. A eficácia da detecção de recursos depende do número e tamanho dos kernels. Denotando o kernel de convolução de tamanho s como $\kappa[p]$, a saída de convolução resultante $\mathcal{C}[p]$ pode ser calculada como

$$\mathcal{C}[p] = x[p] \otimes \kappa[p] = \sum_{j=0}^{s-1} \kappa[j]x[p - j] \quad (2.23)$$

onde \otimes representa a operação de convolução e $p = 0, 1, 2, \dots, N - 1$. Geralmente, os recursos envolvidos na saída da l ésima camada podem ser expressos como $\mathcal{C}_k^l = \sigma \left(b_k^l + \sum_m \mathcal{C}_m^{l-1} \kappa_{km}^l \right)$. \mathcal{C}_k^l representa o k ésimo recurso na l ésima camada, \mathcal{C}_m^{l-1} denota o m ésimo recurso na $(l - 1)$ ésima camada, κ_{km}^l é o kernel que conecta o k ésimo recurso ao m ésimo recurso, b_m^l é o viés para esses recursos, e σ é a função de ativação [59].

- **Função de ativação:** um parâmetro crucial no modelo CNN, a função de ativação ajuda a aprender e aproximar relacionamentos contínuos e complexos entre variáveis de rede. Funções de ativação comuns incluem RELU, softmax e sigmoid.
- **Stride:** O valor stride determina o movimento do kernel através dos dados de entrada. O passo mais típico é 1, indicando que o kernel muda uma coluna de dados de entrada a cada etapa.
- **Pooling Layer:** Normalmente posicionada após uma camada convolucional, a camada de pooling visa reduzir o tamanho do mapa de características, diminuindo assim os custos computacionais. Existem várias operações de pooling, como pool máximo, pool médio e pool de soma. Este trabalho utiliza pooling máximo 1D, que envolve deslizar um tamanho de pool especificado e passar sobre a entrada e selecionar o valor máximo de cada região, $\mathcal{C}_h^l = \max_{e \in \mathcal{P}_h} \mathcal{C}_e^{l-1}$ onde \mathcal{P}_h representa a região de pooling com índice h .
- **Camada nivelada e camada eliminada:** A camada nivelada converte os dados de entrada em um vetor unidimensional para a camada totalmente conectada (densa). Para evitar o overfitting, que pode ocorrer quando todos os recursos se conectam à camada de nivelamento, uma camada de eliminação é adicionada. Essa camada elimina aleatoriamente alguns neurônios durante o treinamento, reduzindo assim o tamanho e a complexidade do modelo.
- **Camada densa totalmente conectada:** A saída achatada é alimentada na camada densa totalmente conectada, que gera a saída final. Uma função de ativação, como a função RELU descrita anteriormente, é um de seus parâmetros principais.

2.9.3 LSTM

A Long Short-Term Memory (LSTM) é um tipo de rede neural recorrente que permite à rede manter dependências de dados de longo prazo em vários intervalos de tempo. Um esquema de sua arquitetura pode ser visto na Figura 8. Uma células LSTM consiste em uma cadeia de módulos de rede neural repetidos, cada um contendo três portas de controle: a porta de esquecimento, a porta de entrada e a porta de saída. Cada porta compreende uma camada de rede neural sigmóide e uma operação de multiplicação

pontual. As camadas sigmóides produzem números entre 0 e 1, indicando a proporção de informações de entrada que devem ser transmitidas. Para dados de série temporal, o LSTM processa uma sequência de vetores de entrada $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_j, \dots\}$, onde $\mathbf{x}_t \in \mathbb{R}^m$ representa um vetor m -dimensional de leituras para variáveis m em tempo t . Várias dessas séries temporais podem ser geradas abrindo uma janela para uma série temporal maior. Embora o LSTM possa lidar com qualquer dado de série temporal, seu desempenho pode variar dependendo da entrada.

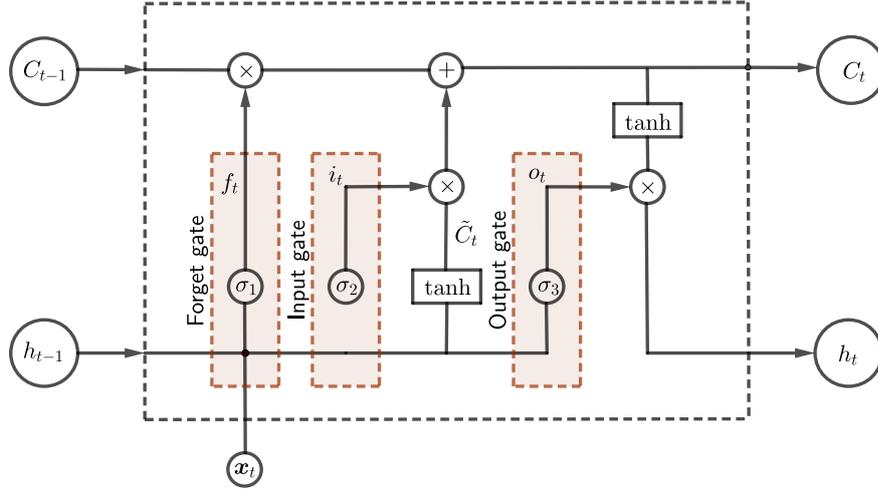


Figura 8 – Arquitetura da rede neural LSTM para previsão de velocidade e direção do vento.

Com as novas informações \mathbf{x}_t no estado t , o módulo LSTM opera da seguinte forma: Primeiro, ele determina quais informações antigas serão descartadas, gerando um número dentro do intervalo $[0,1]$, denotado como

$$f_t = \sigma_1(W_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + b_f), \quad (2.24)$$

onde \mathbf{h}_t é a saída no estado $t - 1$ e W_f, b_f são as matrizes de peso e viés da porta de esquecimento. Então, \mathbf{x}_t é processado antes de ser armazenado no estado da célula. O valor i_t é determinado na porta de entrada, junto com um vetor de valores candidatos \tilde{C}_t gerado por uma camada tanh, para atualizar o novo estado da célula C_t

$$i_t = \sigma_2(W_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + b_i) \quad (2.25)$$

$$\tilde{C}_t = \tanh(W_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + b_c) \quad (2.26)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t, \quad (2.27)$$

aqui $(i, i), (b, c)$ são as matrizes de peso e vieses da porta de entrada e do estado da célula de memória, respectivamente. Finalmente, a porta de saída, que é definida por

$$o_t = \sigma_3(W_o[h_{t-1}, x_t] + b_o) \quad (2.28)$$

$$h_t = o_t \otimes \tanh(C_t), \quad (2.29)$$

onde W_o e b_o são a matriz de peso e polarização da porta de saída, determina uma parte do estado da célula que é gerada. O estado da célula viaja verticalmente por toda a cadeia, preservando internamente as informações sequenciais e permitindo que o LSTM detenha o conhecimento das etapas de tempo subsequentes. [60–64].

2.9.4 Autocodificador LSTM

O padrão ótimo de codificação-decodificação dos dados é aprendido por um auto-codificador, que é uma rede neural não supervisionada. Em geral, ele tem uma camada de entrada, uma camada de saída, uma rede neural codificadora, uma rede neural decodificadora e um espaço latente. Os dados de entrada são compactados pelo codificador no espaço latente, e o decodificador os reconstrói depois. O peso da rede é atualizado comparando a saída reconstruída com os dados iniciais e retropropagando quaisquer erros. O codificador compacta x em uma representação codificada quando recebe uma entrada $\mathbf{y} = \mathcal{E}(\mathbf{x}) \in \mathbb{R}^n$. O decodificador então reconstrói essa representação para produzir a saída $\tilde{\mathbf{x}} = \mathcal{D}(\mathbf{y}) \in \mathbb{R}^m$. A figura 9 ilustra uma rede de autocodificadores LSTM. O autocodificador é treinado para minimizar o erro de reconstrução

$$\mathcal{L} = \frac{1}{2} \|\mathbf{x} - \tilde{\mathbf{x}}\|^2 \quad (2.30)$$

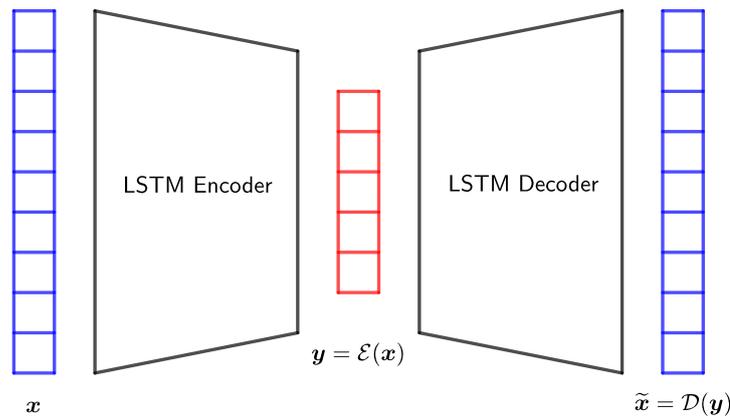


Figura 9 – Ilustração da arquitetura do Autoencoder LSTM usada para previsão de velocidade e direção do vento.

O autocodificador é feito para mais do que apenas replicar entrada para saída. Ao restringir o espaço latente a um tamanho menor que a entrada $n < m$, o autocodificador é forçado a captar os recursos de dados mais cruciais. A dimensão de dados é reduzida

enquanto as informações essenciais da estrutura de dados são mantidas. Houve diferentes propostas para autocodificadores, incluindo autocodificadores vanilla, convolucionais, regularizados e LSTM. No autocodificador LSTM, as redes LSTM são empregadas tanto no codificador quanto no decodificador. O potencial do LSTM para aprender padrões em sequências longas o torna adequado para tarefas como previsão de séries temporais e detecção de anomalias [65–67].

2.9.5 CNN-ALSTM

Um modelo Convolutional Autoencoder LSTM (CNN-ALSTM) é um modelo híbrido que mistura uma Rede Neural Convolucional com um autocodificador LSTM para combinar as vantagens do CNN e do autocodificador LSTM na extração de recursos e capacidade de aprendizado de padrões em dados de sequência longa. Nesses modelos, as camadas de rede CNN extraem recursos espaciais importantes da janela de série temporal e, em seguida, as camadas LSTM aprendem o gradiente da série temporal e as dependências de longo alcance. A estrutura do modelo de autocodificador CNN-LSTM proposto consiste em uma camada de entrada, camadas CNN, camada de agrupamento máximo, camada de codificador LSTM, camada de decodificador LSTM e camada de saída. A figura 10 mostra o esquema do autocodificador convolucional LSTM [32, 68–71].

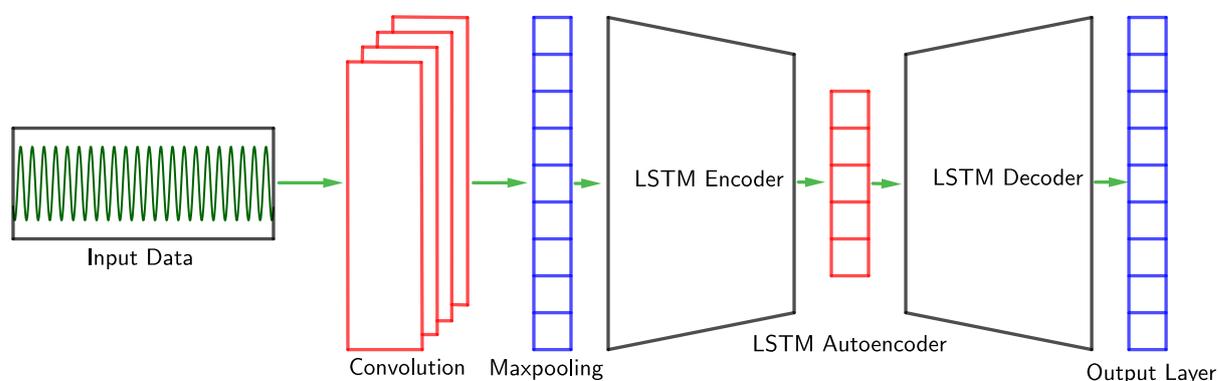


Figura 10 – Arquitetura proposta do CNN-ALSTM para previsão de velocidade e direção do vento.

2.9.6 Modelos Acoplados por Regressão Linear Multivariada

Os modelos acoplados por regressão linear multivariada, assim como outros modelos categorizados como modelos em paralelo, acoplam modelos preditivos diferentes por meio de pesos dados aos mesmos. Tais modelos paralelos geralmente usam o método da média ponderada igual, o método da média ponderada da variância do erro e o método recíproco do erro relativo para paralelizar os modelos [29, 30]. Neste trabalho, optou-se por utilizar os coeficientes de regressão como pesos para acoplar os modelos preditivos, uma vez que essa abordagem se mostrou satisfatória para a previsão de séries temporais complexas, como

a curva de contaminação da COVID-19 [29]. As equações que descrevem nossos modelos acoplados ficam sendo então:

$$y = \beta_0 + \beta_1 x_{modelo_1} + \beta_2 x_{modelo_2} \quad (2.31)$$

onde y representa o valor real da curva, β_0 é uma constante de correção, β_1 é o peso associado aos valores previstos pelo primeiro modelo, chamados de x_{modelo_1} , e β_2 é o peso associado aos valores previstos segundo modelo, chamados de x_{modelo_2} .

Os modelos que serão acoplados por esse método e suas equações gerais ficam descritos nas Equações 2.32, 2.33 e 2.34.

1. ARIMA-LSTM:

$$y_{arima-lstm} = \beta_0 + \beta_1 x_{arima} + \beta_2 x_{lstm} \quad (2.32)$$

2. ARIMA-CNN:

$$y_{arima-cnn} = \beta_0 + \beta_1 x_{arima} + \beta_2 x_{cnn} \quad (2.33)$$

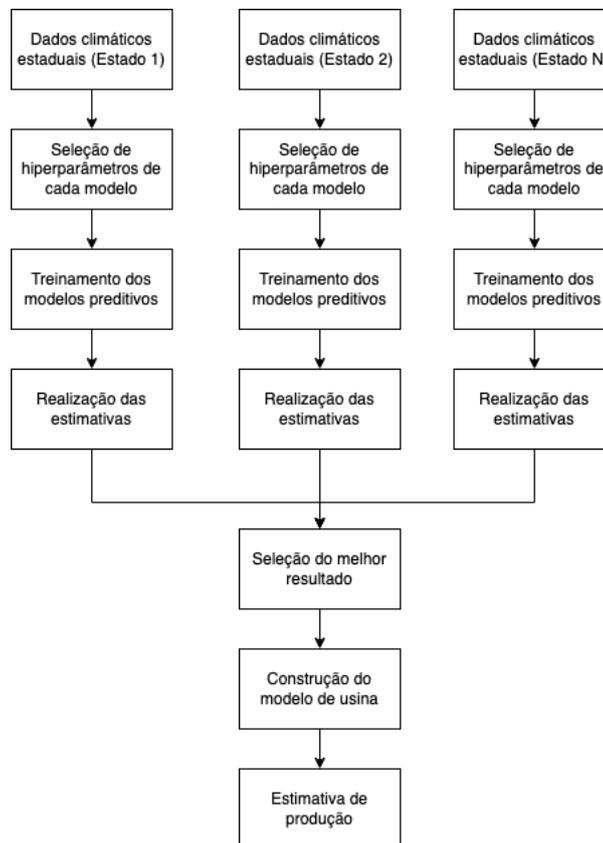
3. CNN-LSTM:

$$y_{cnn-lstm} = \beta_0 + \beta_1 x_{cnn} + \beta_2 x_{lstm} \quad (2.34)$$

3 Metodologia

Neste capítulo é apresentada a metodologia utilizada para estimar a produção de energia eólica em uma usina real. A metodologia consistiu em seis passos principais: o estudo dos dados de cada localidade, a hiperparametrização e o treinamento dos modelos de Redes Neurais Artificiais (RNAs), a modelagem da curva de potência do aerogerador, a estimativa dos dados climáticos, a estimativa da curva de potência dos aerogeradores e a construção do modelo de usina. O fluxograma da Figura 11 ilustra o processo de estimativa de energia eólica realizado neste trabalho.

Figura 11 – Fluxograma das etapas para estimação da produção energética de uma usina.



Na realização deste trabalho, foram utilizados os dados climáticos para os estados da Bahia, Ceará, Piauí, Rio Grande do Norte, Santa Catarina e Rio Grande do Sul. Para cada estado, foi realizada uma análise dos dados fornecidos, com o objetivo de elucidar a dinâmica temporal desses conjuntos e a interdependência entre as variáveis. Após essa análise, o modelo ARIMA e os modelos de redes neurais artificiais, incluindo as arquiteturas CNN, LSTM e CNN-ALSTM, foram formulados e treinados utilizando os dados completos da série temporal.

Com os modelos treinados, seus resultados foram aplicados na base de dados

de treinamento para gerar modelos baseados em regressão linear multivariada, como os modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM. Todos os modelos de redes neurais artificiais tiveram os hiperparâmetros selecionados pelo método BHO. Após a fase de treinamento, previsões para o ano subsequente ao período de treinamento foram geradas, e esses resultados foram comparados com a base de dados de teste, composta pelos dados reais de 2022, utilizando as métricas de avaliação MAPE, MAE, MSE, RMSE e R^2 .

Após a avaliação desses modelos, o melhor entre eles foi selecionado para ser aplicado ao modelo de usina. Os dados gerados para a usina foram comparados com os dados reais de uma planta instalada nas proximidades do ponto de captação dos dados climáticos locais.

3.1 Dados Climáticos

Os dados meteorológicos utilizados neste estudo foram fornecidos pelo INMET. Para selecionar os pontos de captação utilizados neste trabalho, foi realizada uma busca pelos pontos mais próximos a usinas reais e que possuísem uma menor quantidade de dados faltantes. Os dados de treinamento selecionados foram coletados entre os anos de 2011 e 2021, com intervalo horário de dados históricos. Dentre os dados selecionados para compor os conjuntos de treinamento para cada estado o percentual de dados faltantes foi próximo a 10%. Dentre todos os dados fornecidos, doze tipos de dados foram selecionados para serem parâmetros utilizados na análise exploratória inicial. Ao analisar a correlação entre as variáveis, observou-se que a direção do vento possuía uma relação baixa com as outras variáveis utilizadas. Para aumentar o suporte dessa variável, foram introduzidos outros cinco parâmetros que por definição estariam relacionados diretamente com a direção do vento: a projeção da direção do vento nos dois primeiros quadrantes, seu seno, o cosseno, o seno da direção horária do vento e o cosseno da direção horária do vento. Esse passo foi feito para assegurar que as variáveis de interesse tivessem suporte o suficiente para justificar a abordagem multivariada adotada pelos modelos de RNAs. Todos os tipos de dados utilizados para a construção dos modelos de redes neurais, suas unidades de medida e nomenclatura estão descritos na Tabela 2.

Dos dados utilizados, apenas dois sofreram alterações de formato: a *data e hora* foi transformada do formato "dd-mm-aa hh:mm:ss" para *segundos*, e a *direção horária dos ventos* foi transformada de *graus* para *radianos*.

Com o conjunto de dados devidamente estruturado, foi aplicado o método de Imputação Múltipla por Equações Encadeadas (MICE) para preencher os valores ausentes, e em seguida, realizou-se a análise exploratória da base de dados selecionada. Para cada estado, foi construída uma matriz que associa cada linha dos dados de teste às 23 linhas anteriores, resultando em matrizes de input com dimensões de (96310, 24, 18). Nessas

Tabela 2 – Nomenclatura associada a cada variável física explicando a velocidade e direção do vento como alvos.

DADOS E NOMECLATURAS	
Prâmetros	Nomeclaturas
Vento, direção horária (rad)	WIND_DIR_HOUR
Vento, velocidade (m/s)	WIND_HOUR_SPD
Vento, velocidade máxima na hora anterior (m/s)	WIND_MAX_SPD
Humidade relativa horária (%)	UMT_HOUR
Humidade relativa máxima na hora anterior (%)	UMT_MAX
Humidade relativa mínima na hora anterior (%)	UMT_MIN
Temperatura máxima na hora anterior (°C)	TEMP_MAX
Temperatura horária (°C)	TEMP_HOUR
Temperatura mínima na hora anterior (°C)	TEMP_MIN_PREV
Pressão atmosférica máxima na hora anterior (mB)	PRESS_MAX
Pressão atmosférica horária (mB)	PRESS_HOUR
Pressão atmosférica mínima na hora anterior (mB)	PRESS_MIN
Data e Hora (seg)	DATETIME
Seno da direção horária do vento ($\sin(x)$)	WIND_DIR_SIN
Cosseno da direção horária do vento ($\cos(x)$)	WIND_DIR_COS
Seno da projeção da direção horária do vento ($\cos(x)$)	WIND_DIR_PROJ_COS
Cosseno da projeção da direção horária do vento ($\cos(x)$)	WIND_DIR_PROJ_COS
Projeção da direção horária do vento (rad)	WIND_DIR_PROJ_COS

matrizes, 96310 representa o número de observações usadas para o treinamento dos modelos, 24 é o número de horas anteriores consideradas para cada entrada, e 18 corresponde ao número de variáveis climáticas distintas utilizadas. Esses dados foram utilizados para treinar os modelos CNN, LSTM e CNN-ALSTM.

3.2 Métodos de análise do conjunto de dados

Para entender melhor os conjuntos de dados utilizados neste trabalho, foram realizados uma série de testes, medidas e aplicações de modelos estatísticos univariados e multivariados com o objetivo de compreender o comportamento das variáveis de interesse e suas relações com os outros dados utilizados no treinamento dos modelos de redes neurais artificiais (RNAs). Nesta seção, serão descritos todos os procedimentos e ferramentas utilizados para o entendimento do conjunto de dados empregado no treinamento e teste dos modelos apresentados neste estudo.

3.2.1 Matriz de correlações amostrais

A matriz de correlações amostrais é o nome dado à matriz de covariâncias amostrais quando se utilizam variáveis padronizadas. Essa medida descritiva tem como objetivo

principal explicar a estrutura de associação linear entre as variáveis de um conjunto de dados. Para construir a matriz de correlações amostrais, é necessário primeiramente padronizar as variáveis do conjunto de dados, subtraindo a média de cada observação e dividindo o resultado pelo desvio padrão. A Equação 3.1 representa o vetor de variáveis padronizadas relativo a um conjunto de amostras X

$$X_p = \frac{X - \mu}{\sigma} \quad (3.1)$$

onde X_p , μ e σ são respectivamente a padronização, a média e o desvio padrão do conjunto de variáveis X . A matriz de covariâncias amostrais de um conjunto de variáveis X pode ser descrita genericamente pela equação:

$$S = \sum_{i=1}^n \frac{(X_i - \mu)(X_i - \mu)^\top}{n} \quad (3.2)$$

onde X_i é um conjunto de amostras, μ é o vetor de médias $(X_i - \mu)^\top$ é a transposta de $(X_i - \mu)$ e n é o número de observações. A partir dessas duas equações pode-se definir a matriz de correlações amostrais R de um conjunto de variáveis X em função da matriz de covariâncias amostrais como sendo:

$$V = \text{diag} \{S_1^2, \dots, S_n^2\} \quad (3.3)$$

$$R = (V^{\frac{1}{2}})^{-1} S (V^{\frac{1}{2}})^{-1} \quad (3.4)$$

Os valores apresentados na matriz de correlações amostrais estão contidos no intervalo de $[-1, 1]$ onde 1 implica que duas variáveis possuem mesma direção e sentido e -1 implica que duas variáveis possuem mesma direção e sentidos inversos. Para o entendimento dos valores mostrados por essa medida, entende-se que correlações de módulo inferior a 0,5 como sendo correlações fracas, indicando que para valores acima disso consideramos correlações relevantes. Assim, correlações com módulo entre 0,5 e 0,7 podem ser classificadas como moderadas, enquanto correlações acima de 0,7 sugerem uma forte associação entre as variáveis. É importante ressaltar que uma correlação forte não implica necessariamente causalidade.

3.2.2 Análise de Fatores

A análise de fatores é um método estatístico que tem como objetivo descrever a estrutura de dependência de um conjunto de variáveis por meio da identificação de variáveis latentes, chamadas de fatores, que carregam aspectos comuns das variáveis originais do

conjunto de dados. Nesse contexto, para cada variável linearmente independente, é necessário introduzir uma variável latente. Essa variável latente "carrega" todas as informações da variável original, uma vez que ela é a única que contém essas informações específicas. Assim, a variável latente serve para representar a parte única e independente da variável original no modelo.

Neste trabalho a análise de fatores foi utilizada de forma exploratória, propondo sete variáveis latentes para verificar a redundância de informações carregadas pelas variáveis do conjunto de dados. Esses resultados ficarão dispostos em tabelas ilustrativas que mostrarão, para cada fator, a **variância proporcional**, ou variância do conjunto de dados descrita por esse fator, e a **variância acumulada**, ou variância descrita por esse fator somado aos fatores anteriores. A **variância acumulada** do sétimo fator mostrará então a razão da variância do conjunto de dados original que é descrita com os sete fatores.

Para confirmar que a análise de fatores pode ser bem aplicada a um conjunto de dados é necessário efetuar testes que validem essa possibilidade. Para confirmar a aplicabilidade da análise de fatores foram utilizados o método de **Kaiser-Meyer-Olkin (KMO)** e o **teste de esfericidade de Bartlett**.

3.2.2.1 Kaiser-Meyer-Olkin (KMO)

A medida de adequação de Kaiser-Meyer-Olkin é apresentada por meio do cálculo do coeficiente KMO. A Equação 3.5 que descreve esse coeficiente como:

$$KMO = \frac{\sum \sum r_{ij}^2}{\sum \sum r_{ij}^2 + \sum \sum p_{ij}^2} \quad (3.5)$$

onde r_{ij} e p_{ij} são respectivamente o coeficiente de correlação e o coeficiente de correlação parcial entre as variáveis i e j . O elemento p_{ij} é descrito por:

$$p_{ij} = \frac{-r_{ij|rest}}{\sqrt{r_{ii|rest} \cdot r_{jj|rest}}} \quad (3.6)$$

onde $r_{ij|rest}$, $r_{ii|rest}$ e $r_{jj|rest}$ são respectivamente o coeficiente de correlação ajustado pelo efeito das outras variáveis e os coeficientes de autocorrelação entre as variáveis i e j .

Para a interpretação dos valores encontrados para o coeficiente KMO segue a tabela 3.

3.2.2.2 Teste de esfericidade de Bartlett

A condição necessária para que uma análise de fatores produza resultados aceitáveis é a existência de correlações não nulas entre as variáveis analisadas. O teste de esfericidade de Bartlett, então, verifica a existência dessas correlações avaliando se a matriz

Tabela 3 – Valores de referência para o coeficiente KMO.

KMO	Interpretação
0,90 - 1,00	Excelente
0,80 - 0,90	Ótimo
0,70 - 0,80	Bom
0,60 - 0,70	Regular
0,50 - 0,60	Ruim
0,00 - 0,50	Insuficiente

de correlações de um conjunto de dados é próxima o suficiente de uma matriz identidade. Nesse teste, iremos avaliar as seguintes hipóteses:

- **Hipótese Nula (H_0):** A matriz de correlações é próxima o suficiente da matriz identidade. As variáveis são esféricas, ou seja, não relacionadas.
- **Hipótese Alternativa (H_a):** As variáveis não são esféricas ou seja, relacionadas.

Para rejeitar H_0 é necessário um valor de p_{value} menor ou igual a 0,05. Esse valor de p_{value} sugere que há correlações significativas entre as variáveis analisadas. Portanto, as variáveis não são esféricas, e a análise fatorial é apropriada.

3.2.3 Kurtosis

A **curtose** (ou *kurtosis*) é uma medida estatística que descreve as características das caudas de uma distribuição de probabilidade, em comparação a uma distribuição normal. Ela indica o quão "pontuda" ou "achatada" é a distribuição, bem como o grau de concentração de valores extremos nas caudas. Normalmente utiliza-se os valores da chamada "curtose excessiva" (K_{excess}) para categorizar os três diferentes tipos de curtose que existem. Os tipos de curtose relacionados aos possíveis valores utilizados neste trabalho são:

- **Leptocúrtica** $K_{excess} > 0$: São as curtoses que possuem caudas mais longas e pontas mais acentuadas do que uma distribuição normal. Isso indica uma maior concentração de dados em torno da média e uma maior incidência de valores extremos (outliers) nas caudas. Esses dados tendem a ter uma distribuição mais "pontuda".
- **Mesocúrtica** $K_{excess} = 0$: Representa a curtose de uma distribuição normal. Nessa distribuição, as caudas e a pontuação são moderadas, sem uma quantidade excessiva de valores extremos. Esse tipo de distribuição é o ponto de referência para a análise de curtose.

- **Platicúrtica** $K_{\text{excess}} < 0$: Essas curtoses têm caudas mais curtas e uma distribuição mais achatada em relação à normal. Há menos concentração de valores em torno da média, e a distribuição tende a ser mais uniforme, com menor incidência de outliers.

Além do valor da curtose, também é discutido o teste de curtose (kurtosistest). Este teste estatístico é utilizado para determinar se a curtose de uma distribuição de dados é significativamente diferente da curtose de uma distribuição normal. Em outras palavras, ele auxilia na compreensão sobre se a forma das caudas da distribuição analisada é semelhante ou não à de uma distribuição normal. Como resultado deste teste, caso o $pvalue$ seja menor ou igual a 0,05, rejeita-se a hipótese nula, indicando que a distribuição possui caudas não normais.

3.2.4 Skewness

A medida da assimetria (ou skewness) é uma medida estatística que descreve o grau de assimetria de uma distribuição de dados em relação à sua média. Em termos simples, ela indica se os dados estão distribuídos de forma equilibrada em torno da média ou se há uma tendência de concentração em uma das extremidades. A interpretação dos valores de skewness é a seguinte:

- **Skewness** < 0 : A maioria dos dados está concentrada à direita da média, com uma cauda mais longa à esquerda, ou seja, há alguns valores extremos abaixo da média.
- **Skewness** $= 0$: Os dados estão igualmente distribuídos em ambos os lados da média, indicando uma distribuição simétrica.
- **Skewness** > 0 : A maior parte dos dados está concentrada à esquerda da média, com uma cauda mais longa à direita, ou seja, alguns valores extremos estão acima da média.

O teste de assimetria é utilizado para avaliar se a distribuição dos dados é significativamente simétrica ou assimétrica. Ele testa se o valor de skewness é significativamente diferente de zero. A hipótese nula (H_0) do teste afirma que a distribuição dos dados é simétrica ($Skewness = 0$). A interpretação dos resultados para diferentes valores do $pvalue$ é:

- $pvalue > 0,05$: Não há evidências suficientes para rejeitar a hipótese nula, o que significa que a distribuição pode ser considerada simétrica em relação à média.
- $pvalue \leq 0,05$: Há evidências suficientes para rejeitar a hipótese nula, indicando que a distribuição dos dados é assimétrica e a assimetria é estatisticamente significativa.

3.2.5 Jarque-Bera

O teste de Jarque-Bera é um teste estatístico utilizado para verificar se uma amostra de dados segue uma distribuição normal, com base nas medidas de assimetria (skewness) e curtose (kurtosis). A hipótese nula H_0 desse teste supõe a normalidade da distribuição dos dados. Sendo assim, a interpretação do resultado em relação ao valor de p_{value} é:

- $p - value > 0,05$: Indica que os dados não apresentam desvio significativo da normalidade.
- $p - value \leq 0,05$: Os dados são significativamente assimétricos e/ou possuem uma curtose diferente da esperada em uma distribuição normal.

3.2.6 Dickey-Fuller

O teste de Dickey-Fuller (adfuller) é um teste estatístico usado para verificar se uma série temporal contém uma raiz unitária, ou seja, se a série é não estacionária. Em uma série estacionária, a média, variância e autocorrelação dessa série temporal são constantes ao longo do tempo. Isso facilita a identificação de padrões e a criação de previsões confiáveis, já que o comportamento da série no futuro pode ser inferido com base em seu comportamento no passado. Esse teste verifica se a média, a variância e a autocovariância da série mudam ao longo do tempo e sua hipótese nula H_0 é que a série temporal tem uma raiz unitária, sendo essa a condição de estacionaridade. A interpretação do teste de Dickey-Fuller com relação ao p_{value} é dada por:

- $p - value > 0,05$: A série temporal contém uma raiz unitária e **não é estacionária**.
- $p - value \leq 0,05$: A série temporal não contém uma raiz unitária e **é estacionária**.

3.2.7 Outras Métricas

Além dos métodos e métricas introduzidos nessa seção também foram utilizadas as seguintes métricas para entender o comportamento dos conjuntos de dados:

- μ : Média das ocorrências de uma dada variável.
- σ : Desvio padrão de uma dada variável.
- N : Número de observações de uma dada variável.
- $Q1$: Primeiro quartil, também chamado de quartil inferior, é o valor que separa os 25% menores dados do restante.
- $Q2$: Segundo quartil, é o valor que separa os 50% menores dados do restante.

- Q_3 : Terceiro quartil, é o valor que separa os 75% menores dados do restante.

Embora os valores apresentados nesta subseção não sejam analisados em profundidade posteriormente, eles são fundamentais para uma compreensão mais completa dos dados e serão incluídos nas mesmas tabelas que os demais testes e métricas.

3.3 Previsão dos dados climáticos e seleção do modelo preditivo

Para a previsão dos dados de direção e velocidade do vento, foram utilizados os modelos ARIMA, LSTM, CNN e CNN-ALTSM previamente treinados. Esses modelos geraram dados horários para o intervalo de 1 ano além dos dados de treinamento. Para os modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM, a equação gerada pela regressão linear multivariada de cada um de seus componentes foi empregada para combinar os valores das previsões realizadas pelos modelos ARIMA, CNN e LSTM.

Os dados gerados por cada modelo preditivo foram comparados com os valores reais para o ano de 2022 utilizando as métricas R^2 , $MAPE$, MAE , MSE e $RMSE$. O modelo selecionado foi aquele que apresentou o melhor desempenho geral nas métricas de avaliação.

3.4 Modelo de usina

Nesta seção é detalhado o processo de criação do modelo de usina. Os dados utilizados para a construção do modelo incluem informações dos aerogeradores da usina, a altitude da instalação e a quantidade de aerogeradores de cada modelo.

3.4.1 Estimativa da Curva de Potência para cada aerogerador

A curva de potência foi estimada utilizando os dados reais de cada modelo de aerogerador utilizado na usina de referência. Para isso, realizamos a interpolação de splines de ordem superior de grau 5 nos dados da curva de potência dos aerogeradores providos pelo fabricante. Além disso, para cada aerogerador, utilizamos os valores das velocidades de *cut-in*, *ideal*, e *cut-off* para limitar a aproximação às condições de contorno dos aerogeradores. A equação para cada aerogerador é descrita na Equação 3.7.

$$E(x) = \begin{cases} 0, & \text{se } x \leq v_{cut-in} \\ S(x), & \text{se } v_{cut-in} < x < v_{ideal} \\ E_{nominal}, & \text{se } v_{ideal} \leq x < v_{cut-off} \\ 0, & \text{se } x \geq v_{cut-off} \end{cases} \quad (3.7)$$

Onde $E(x)$ é a potência estimada do aerogerador, $S(x)$ é a equação de interpolação de splines de ordem superior, $E_{nominal}$ é a potência nominal do aerogerador e v_{cut-in} , v_{ideal} e $v_{cut-off}$ são respectivamente as velocidades de *cut-in*, ideal, e *cut-off* para o aerogerador.

3.4.2 Potência estimada da usina

Para estimar a potência total da usina, multiplicaram-se as equações de cada tipo de aerogerador pelo número de aerogeradores de cada modelo na usina de referência. Dessa forma, a equação da energia estimada para a usina é descrita pela Equação 3.8.

$$E_{total}(x) = \sum_i N_i \cdot E_i(x) \quad (3.8)$$

Onde x é a velocidade do vento, E_{total} é a estimativa de geração da usina e N_i e E_i são, respectivamente, a quantidade e a energia estimada para o aerogerador de tipo i .

Como o ponto de captação dos dados climáticos não coincide com a localização da usina, é necessário estimar o valor da velocidade do vento. Para isso, utilizou-se a equação do perfil logarítmico da velocidade do vento, levando em consideração os dados de altitude do ponto de captação, a altitude da usina e a altura da torre do aerogerador. Dessa forma, a equação de potência da usina se torna:

$$E_{total}(x) = \sum_i N_i \cdot E_i(V(x, h_i)) \quad (3.9)$$

$$V(x, h_i) = V_z \cdot \left(\frac{\ln\left(\frac{h_i+h}{z_0}\right)}{\ln\left(\frac{z}{z_0}\right)} \right) \quad (3.10)$$

onde h_i é a altura da torre do aerogerador i , h é a altitude da usina, z é a altitude do ponto de captação, V_z é a velocidade do vento prevista para o ponto de captação e z_0 é a constante de rugosidade do terreno.

Como a curva de potência é altamente sensível à velocidade efetiva do vento, foram adotadas duas abordagens distintas para a estimativa. Na primeira abordagem, utilizou-se a equação descrita anteriormente para calcular um intervalo de produção possível, limitado superiormente pela curva de potência $E_u(x)$ e inferiormente por $E_l(x)$, onde essas curvas são descritas pelas Equações 3.11 e 3.12.

$$E_u(x) = k_u \cdot \sum_i N_i \cdot E_i(V(x, h_i)) \quad (3.11)$$

$$E_l(x) = k_l \cdot \sum_i N_i \cdot E_i(V(x, h_i)) \quad (3.12)$$

Nessa Equações 3.11 e 3.12 k_u e k_l são constantes contidas no intervalo $[0, 1]$ com $k_u > k_l$. Os resultados dessa abordagem serão apresentados por meio da contagem da potência real da usina não contidos nesse intervalo, assim como sua representatividade na amostra total.

Na segunda abordagem, utilizou-se a direção do vento estimada pelo modelo preditivo para calcular a velocidade efetiva normal da superfície das pás eólicas. Foi realizada uma regressão com os valores estimados por $E_{total}(x)$, juntamente com os dados reais de velocidade e direção do vento, para ajustar os valores reais de produção da usina no ano de 2021. A equação utilizada para a regressão foi:

$$E_{total}(x, \theta) = \sum_i N_i \cdot E_i(V(x, h_i, \theta)) \quad (3.13)$$

$$V(x, h_i, \theta) = V_z \cdot \cos(\theta - \alpha) \cdot \left(\frac{\ln\left(\frac{h_i+h}{z_0}\right)}{\ln\left(\frac{z}{z_0}\right)} \right) \quad (3.14)$$

onde θ foi a direção do vento estimada para o modelo preditivo e $alpha$ o ângulo da superfície dos aerogeradores.

Após isso, a equação mencionada foi utilizada juntamente com o valor de α calculado para estimar a potência da usina com os dados de velocidade e direção do vento previstos pelo modelo de RNA para o ano de 2022. Para avaliar essa estimativa, foram empregadas como principais métricas de erro o MAPE e o MAE.

No próximo capítulo, serão apresentados todos os resultados obtidos para a análise dos dados climáticos de cada estado, de todos os modelos preditivos utilizados e para as duas abordagens da estimativa de produção pelo modelo de usina. Para esta última, será analisada a efetividade da aproximação para os intervalos de 1 dia, 7 dias, 14 dias e 30 dias.

3.5 Aplicação computacional

Nessa seção serão apresentadas as ferramentas computacionais utilizadas para a aplicação de todos os métodos descritos na metodologia. Todas as implementações realizadas neste trabalho utilizaram como base a linguagem de programação Python em sua versão 3.10.

3.5.1 Análise exploratória dos dados

Na análise exploratória foram utilizadas as bibliotecas **statsmodels**, **sklearn**, **factor_analyzer** e **scipy**. Com essas ferramentas foi possível realizar a montagem da

matriz de correlações, análise fatorial, o cálculo do KMO e todos os testes e métricas estatísticos apresentados nesta seção. Além disso, também foi utilizada a biblioteca **matplotlib** para a construção do gráfico da matriz de correlações e da separação de componentes das séries temporais de cada estado.

3.5.2 Tratamento de dados

Na etapa de tratamento de dados foi onde ocorreu a formatação dos conjuntos de dados de treino e teste, assim como a conversão de unidades de medida, criação de novas colunas e imputação de variáveis faltantes. Para realizar esses processos foram utilizadas as bibliotecas **numpy**, **pandas** e **statsmodels** a qual possui o módulo **MICEData** com o qual foi realizada a imputação.

3.5.3 Otimização de hiperparâmetros e treinamento de modelos preditivos

A otimização dos hiperparâmetros foi realizada de maneira distinta para as RNAs e o modelo ARIMA. Para os modelos baseados em RNAs, utilizou-se o método **BayesianOptimization** da biblioteca **keras_tuner**. Esse método requer a configuração de um **hypermodel**, que deve herdar da classe **HyperModel**. Nesse **hypermodel**, são definidos os hiperparâmetros a serem otimizados durante o processo de otimização bayesiana, assim como a rotina de validação cruzada, que serve de base para a otimização. Para o treinamento das RNAs as bibliotecas utilizadas foram **keras** e **tensorflow**.

Para a realização da otimização e treinamento dos modelos ARIMA, foi utilizada apenas a biblioteca **statsmodels**.

4 Resultados

Os resultados serão apresentados em duas partes diferentes. Na primeira, são exibidos os resultados das análises exploratórias relativas à correlação dos dados, das séries temporais dos dados de interesse e os resultados das previsões realizadas com cada um dos modelos preditivos para cada estado, de forma separada. Em seguida, é apresentado o modelo preditivo selecionado para alimentar o modelo de previsão de geração de energia elétrica, bem como os resultados desse modelo.

4.1 Dados climáticos e modelos

Nesta seção são apresentadas as análises exploratórias realizadas com os dados de cada um dos estados, assim como os resultados de cada modelo preditivo proposto para esse trabalho. Os resultados serão apresentados por meio de gráficos, de correlação e comparativos entre as previsões e os dados reais do conjunto de teste, e Tabelas que trarão as métricas de avaliação, quando se tratando dos modelos preditivos, e correlações e testes estatísticos quando se tratando das análises exploratórias.

4.1.1 Bahia

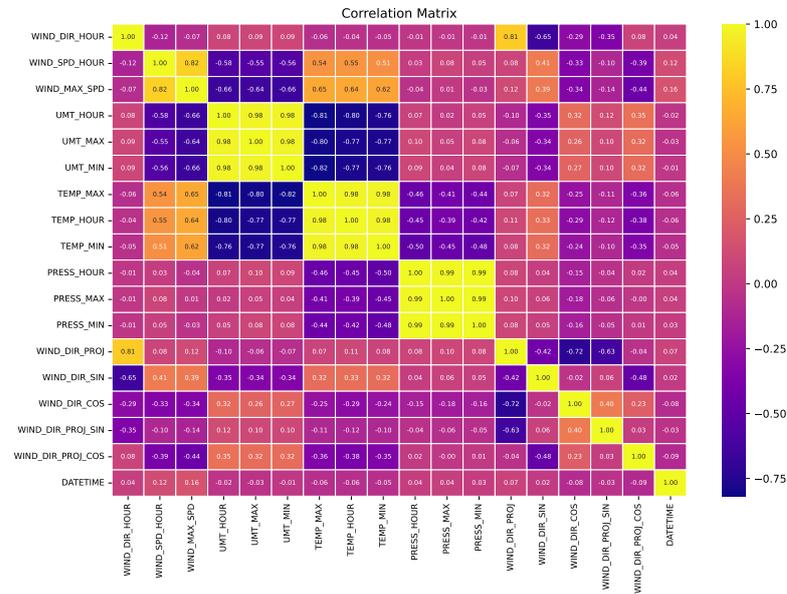
A análise exploratória foi iniciada identificando as correlações entre as variáveis por meio da criação da matriz de correlação dos dados utilizados (Tabela 2). Essa matriz mostra a relação linear entre duas variáveis, com 1 indicando que as variáveis possuem a mesma direção e sentido, -1 indicando que possuem a mesma direção e sentido oposto, e 0 indicando que as variáveis são tangentes. A matriz de correlação está representada no gráfico de calor apresentado na Figura 12.

Para a velocidade dos ventos, os resultados nos mostraram a existência de correlações positivas acima de 0,5 com velocidade máxima do vento na hora anterior, temperatura máxima na hora anterior, temperatura horária e temperatura mínima na hora anterior. Além disso, observou-se a existência de correlações negativas abaixo de $-0,5$ com a umidade horária, umidade máxima na hora anterior e umidade mínima na hora anterior.

Para a direção do vento, observou-se que as únicas correlações consideráveis são a correlação positiva forte com a projeção da direção do vento e a correlação negativa moderada com o seno da direção do vento.

Após a construção da matriz de correlação, foi realizada uma análise de fatores para identificar variáveis latentes, com o objetivo de compreender melhor a correlação entre as variáveis utilizadas na construção do modelo. Antes de proceder com a análise de fatores,

Figura 12 – Bahia - Gráfico de correlação dos dados climáticos referentes ao estado da Bahia.



foi calculado o coeficiente Kaiser-Meyer-Olkin (KMO) e realizado o teste de esfericidade de Bartlett para verificar a viabilidade de aplicação da análise. Os resultados, apresentados na Tabela 4, indicaram uma alta correlação entre os conjuntos de dados, com um KMO de 0,74, o que sugere uma forte correlação entre as variáveis. Isso foi corroborado pelo valor de p_{value} igual a 0,0 para o teste de esfericidade de Bartlett, que rejeita a hipótese nula do teste.

Tabela 4 – Bahia - Resultados de KMO e Teste de Esfericidade de Bartlett

KMO	Teste de Esfericidade de Bartlett p-value
0,74	0,00

Os resultados da análise de fatores ilustrados na Tabela 5, nos mostram a existência de sete variáveis latentes que juntas descrevem 85% das variâncias desse conjunto de dados. Estes resultados, somados com a matriz de correlação nos indicam a capacidade das variáveis utilizadas para o treinamento dos modelos preditivos de contribuir para o aumento de precisão dos ajustes sobre cada uma das variáveis.

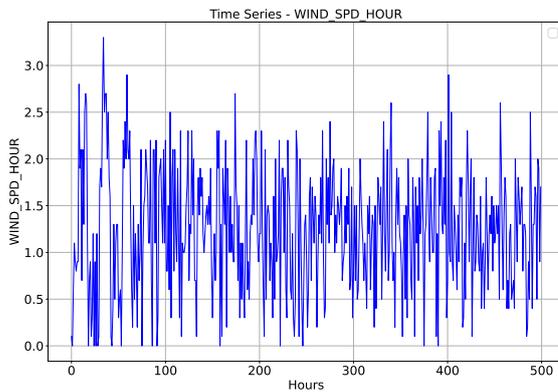
A Figura 13 apresenta as observações de velocidade e direção do vento nas últimas 500 horas e seus respectivos histogramas para todo o conjunto de dados, que inclui 96.432 observações utilizadas do estado da Bahia, conforme ilustrado na Tabela 6.

Como pode ser visto nas Figuras 13(a) e 13(c), a velocidade e a direção do vento se comportam de forma diferente. No entanto, elas têm o mesmo atributo de estarem centradas em torno da média de 1,51 e 1,87, respectivamente, com uma variação aproximadamente constante ao longo do tempo. O teste Dickey-Fuller apoia essas suposições de estacionariedade, pois rejeita a hipótese nula de uma raiz unitária na série temporal,

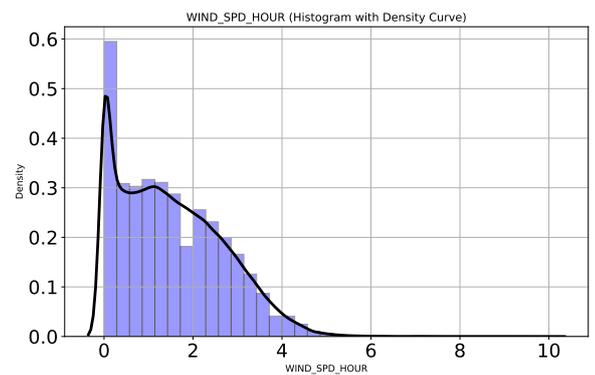
Tabela 5 – Bahia - Variância Explicada por Fator

Fator	Variância Proporcional	Variância Acumulada
Fator 1	0,31	0,31
Fator 2	0,19	0,50
Fator 3	0,12	0,62
Fator 4	0,10	0,70
Fator 5	0,06	0,76
Fator 6	0,06	0,82
Fator 7	0,03	0,85

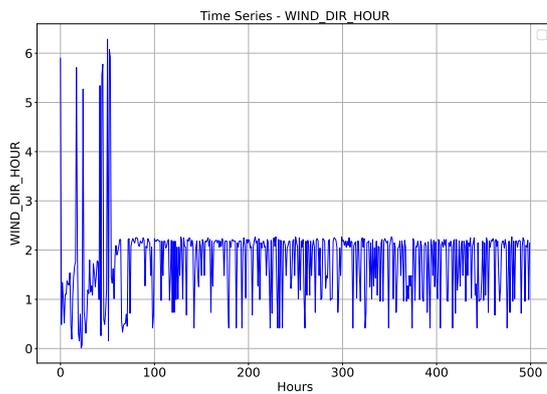
Figura 13 – Bahia - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.



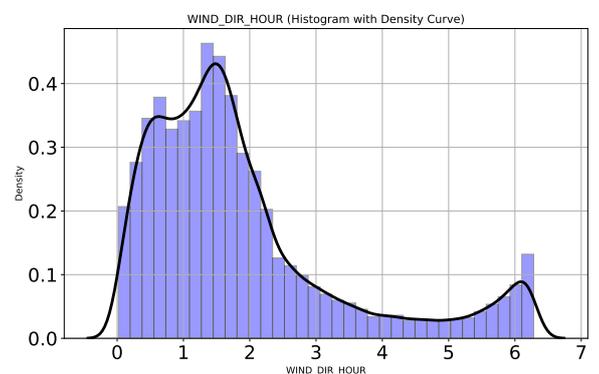
(a)



(b)



(c)



(d)

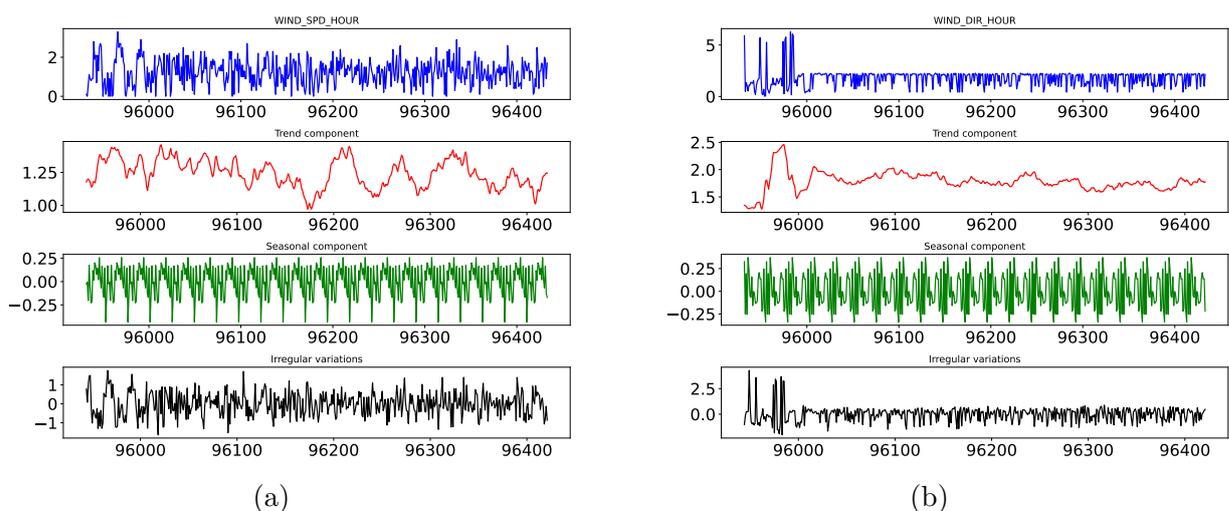
que é o que indica a estacionariedade. Para verificar a não normalidade na série temporal de velocidade e direção do vento, testes de hipóteses como Jarque-Bera, teste de assimetria e teste de curtose foram usados. Com base no valor de p para cada teste, levando em

consideração um nível de significância de 0,05, esses testes contrastam a hipótese nula de que as séries são distribuídas normalmente. A Tabela 6 resume todos os testes usados e as estatísticas computadas. Os histogramas associados à velocidade e direção do vento são mostrados nas Figuras 13(b) e 13(d) com base em todos os registros.

Tabela 6 – Bahia - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionaridade.

	WIND_HOUR_SPD	WIND_DIR_HOUR
N	96432	96432
μ	1,514	1,87
σ	1,16	1,514
x_{\min}	0,00	0,02
x_{\max}	10,00	6,28
Q_1	0,50	0,84
Q_2	1,40	1,48
Q_3	2,30	2,25
kurtosis	-0,07	1,48
kurtosistest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
skewness	0,60	1,45
skewnesstest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
jarque_bera	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
adfuller	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$

Figura 14 – Bahia - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.



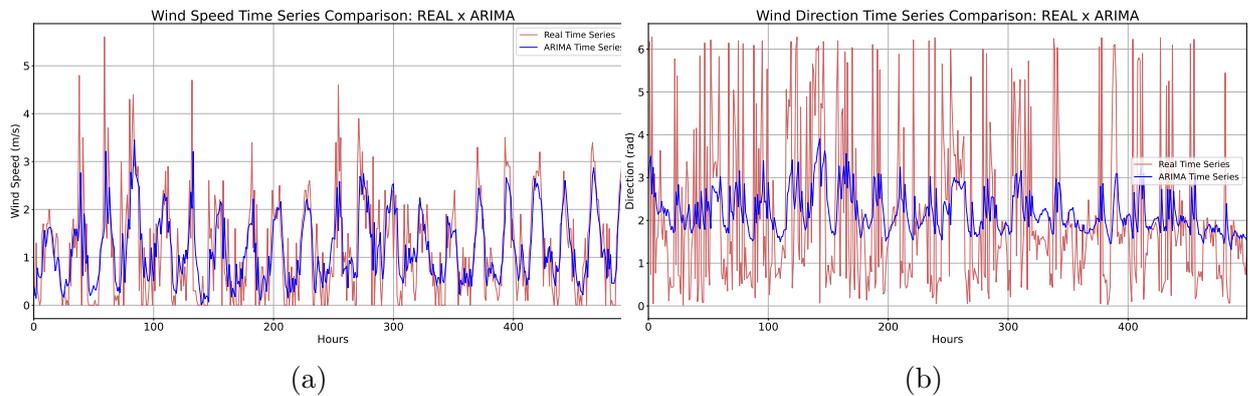
Na Tabela 6, um valor de assimetria positivo de 0,60 confirma que o histograma da velocidade do vento mostra uma assimetria clara à direita que indica que velocidades mais altas do vento são menos frequentes. Isso também ocorre para a direção do vento e pode ser confirmada por meio de um valor de assimetria de 1,45. Uma distribuição de dados

ligeiramente mais plana do que uma distribuição normal pode ser observada com uma curtose de -0,07 para velocidade do vento. As caudas mais leves da distribuição e a menor concentração de valores em torno da média em comparação com uma distribuição normal são o resultado disso. Para a direção do vento, a distribuição de dados é marginalmente mais pontiaguda do que uma distribuição normal, com uma curtose de 0, conforme sugerido pelo valor de curtose de 1,48. Esse resultado implica que há caudas mais pesadas e uma densidade maior de valores próximos à média na distribuição do que em uma distribuição normal.

Na Figura 14, está representada a decomposição das séries temporais de velocidade e direção do vento, com referência a um período de 24 horas. A partir dessa imagem, observa-se uma forte tendência sazonal em ambas as curvas. Além disso, é evidente que a magnitude da componente de erro das séries temporais de ambas as variáveis é bastante alta, sendo aproximadamente metade para a série de velocidade do vento e mais que o dobro para a série de direção do vento.

As Figuras 15, 16, 17 mostram, respectivamente, a comparação das previsões geradas pelos modelos ARIMA, CNN e LSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidas na Tabela 7.

Figura 15 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.



Os resultados apresentados na Tabela 7 indicam que, para a direção do vento, todos os modelos apresentaram uma baixa aderência, com o valor máximo de R^2 sendo de 0,11 para o modelo LSTM. Isso significa que apenas 11% da variância da série temporal foi efetivamente capturada por esse modelo. Além disso, observa-se que, como padrão, todos os modelos apresentaram valores baixos de MAE, mas com valores elevados de MAPE, todos superiores a 190%.

Para a velocidade do vento, os resultados nos mostram uma capacidade de ajuste muito melhor por parte de todos os modelos. O melhor dentre eles, novamente foi o LSTM

Figura 16 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.

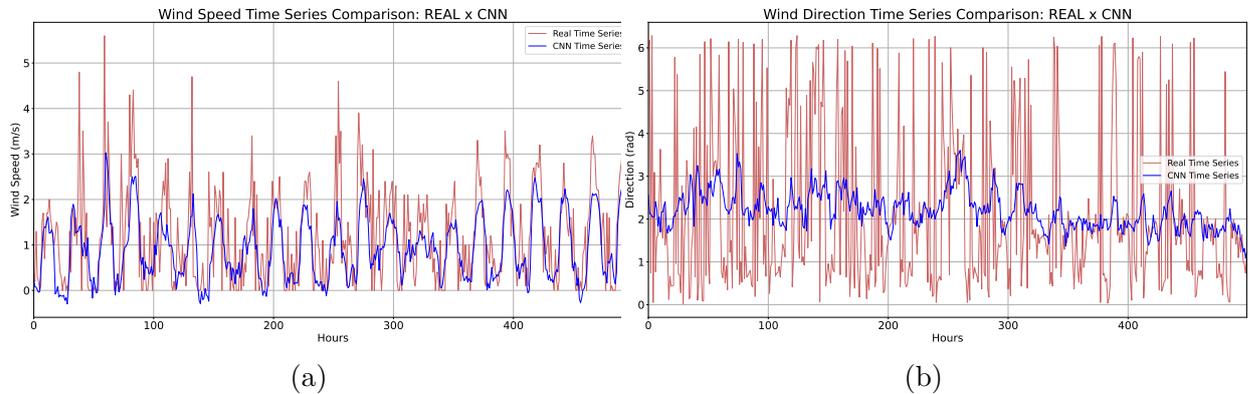
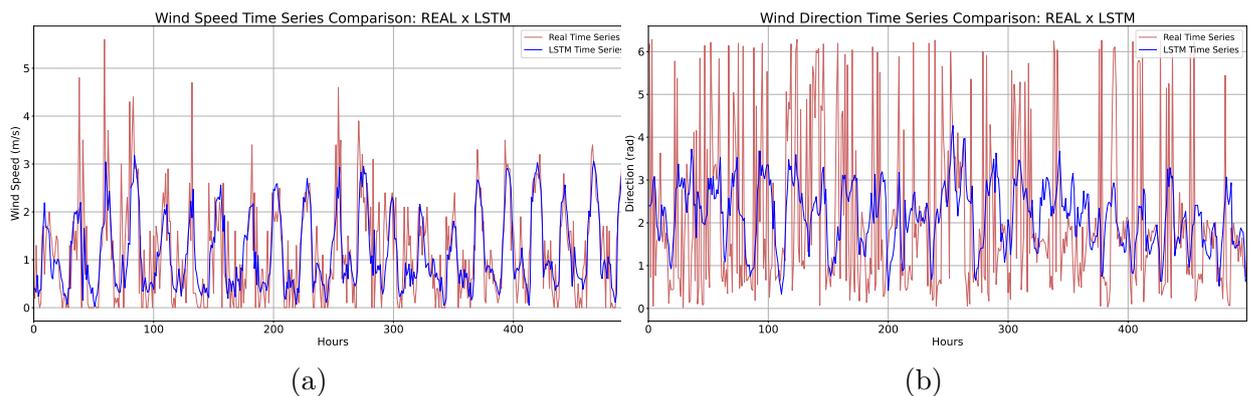


Figura 17 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.



que, com R^2 de 0,57, foi capaz de descrever 57% da variância dos dados. Com exceção do MAPE, melhor performado pelo modelo CNN com valor de 74,16%, o modelo LSTM apresentou o melhor desempenho em todas as outras métricas de avaliação.

As Figuras 18, 19, 20 e 21 representam, respectivamente, a comparação das previsões geradas pelos modelos híbridos ARIMA-CNN, ARIMA-LSTM, CNN-LSTM e CNN-ALSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidos na Tabela 8.

Os resultados apresentados na Tabela 8 indicam uma melhora significativa nos modelos híbridos paralelos que combinam ARIMA com RNAs, em comparação aos modelos individuais, com desempenho superior em quase todas as métricas avaliadas. Em contrapartida, os modelos CNN-ALSTM e CNN-LSTM apresentaram, de forma geral, resultados semelhantes aos modelos individuais, com valores próximos em todas as métricas.

Tabela 7 – Bahia - Métricas de desempenho para CNN, LSTM e ARIMA (Velocidade e Direção do Vento).

Métrica	CNN VEL	LSTM VEL	ARIMA VEL
MAPE	74,16	76,77	80,28
RMSE	0,76	0,71	0,71
MSE	0,58	0,50	0,51
MAE	0,26	0,07	0,07
R^2	0,50	0,57	0,56
Métrica	CNN DIR	LSTM DIR	ARIMA DIR
MAPE	216,06	191,22	195,29
RMSE	1,51	1,49	1,52
MSE	2,29	2,22	2,30
MAE	0,10	0,01	0,03
R^2	0,09	0,11	0,08

Figura 18 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.

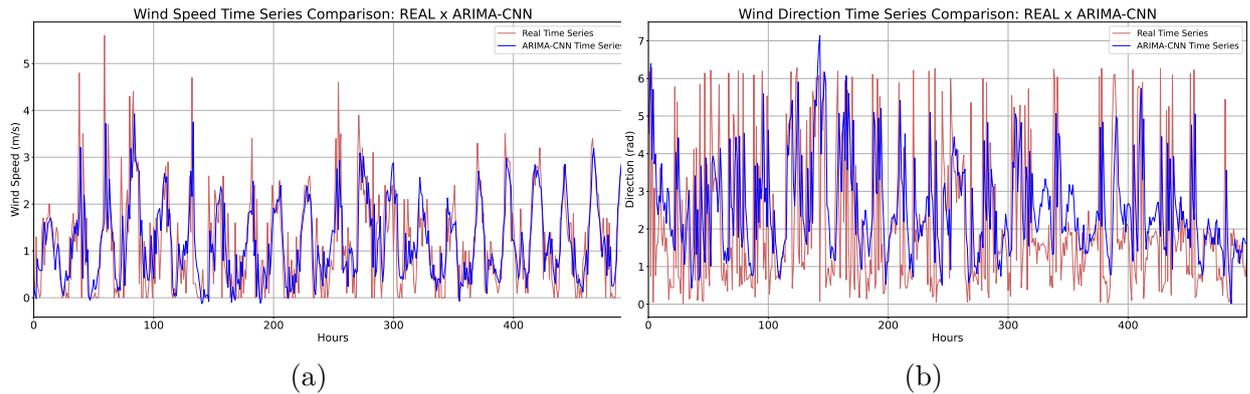


Figura 19 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.

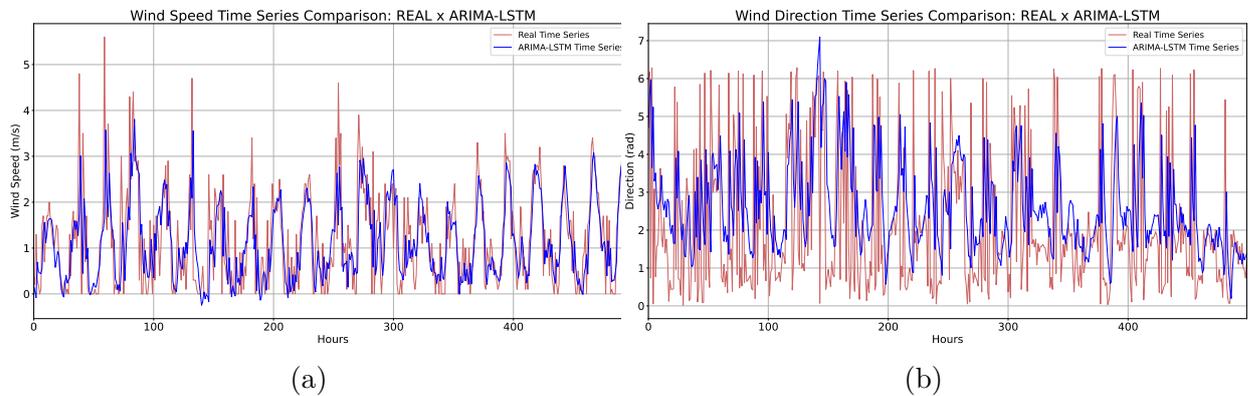


Figura 20 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.

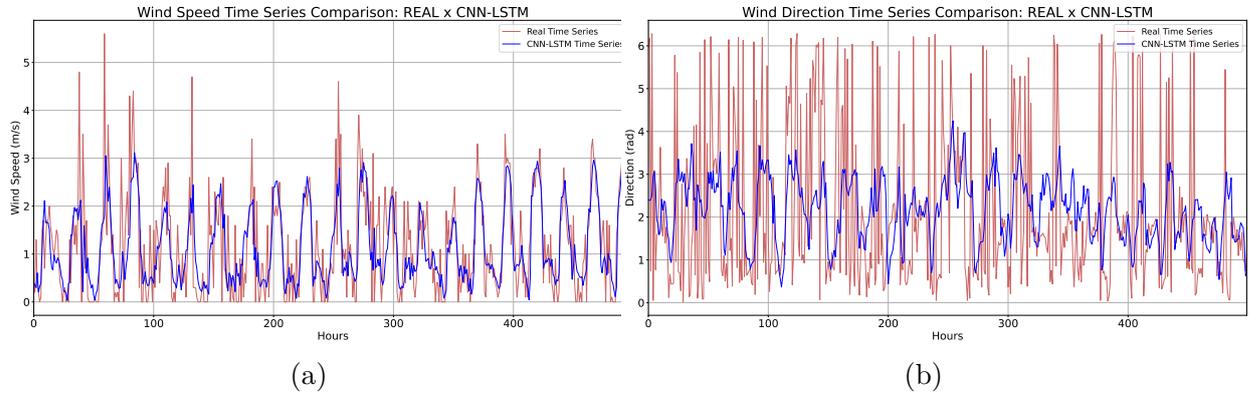
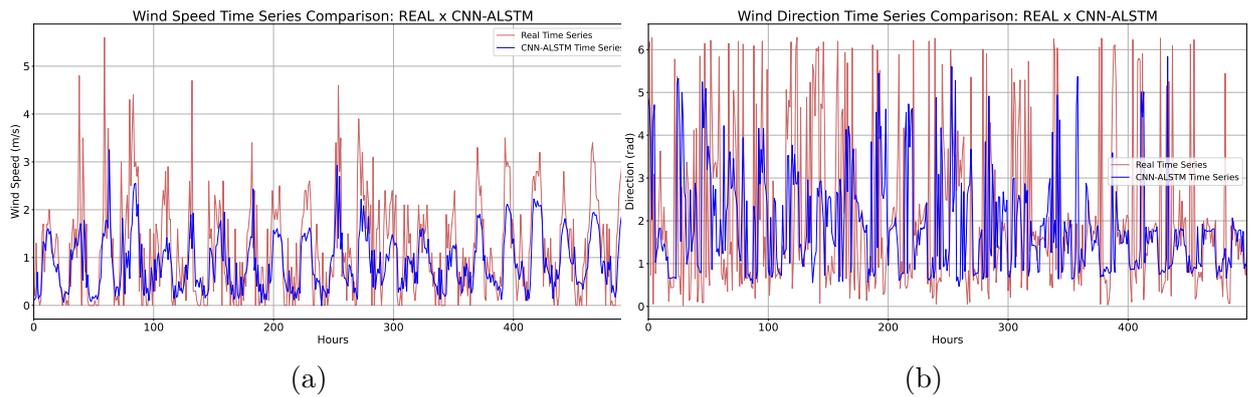


Figura 21 – Bahia - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.



Para a previsão da direção do vento, os modelos ARIMA-LSTM e ARIMA-CNN apresentaram uma melhora expressiva nos resultados de R^2 sendo esses respectivamente de 0,60 e 0,64, onde o melhor desempenho foi obtido pelo modelo ARIMA-CNN. Esses resultados sugerem que o modelo ARIMA-CNN foi capaz de capturar 64% da variância da série temporal enquanto o melhor modelo individual, LSTM, descreveu 11%. Com relação aos parâmetros MAPE, RMSE e MSE, o modelo ARIMA-CNN também apresentou melhoras significativas diminuindo o MAPE de 191,22% para 115,30%, o RMSE de 1,49 para 0,95, o MSE de 2,22 para 0,90 quando comparado com o LSTM. A única métrica que não foi melhorada foi o MAE que teve valores de 0,01 e 0,05 para o LSTM e ARIMA-CNN respectivamente.

Para a velocidade do vento a melhora dos resultados obtidos pelos modelos híbridos se matem. Nesse caso, a performance dos modelos ARIMA-CNN e ARIMA-LSTM mostraram melhorias em todas as métricas de avaliação, com destaque ao parâmetro

Tabela 8 – Bahia - Métricas de desempenho para os modelos combinados (Velocidade e Direção do Vento).

Métrica	CNN- ALSTM VEL	ARIMA- LSTM VEL	ARIMA- CNN VEL	CNN- LSTM VEL
MAPE	71,89	49,26	50,09	75,28
RMSE	0,76	0,47	0,46	0,70
MSE	0,58	0,22	0,22	0,49
MAE	0,29	0,05	0,02	0,08
R^2	0,50	0,81	0,81	0,58
Métrica	CNN- ALSTM DIR	ARIMA- LSTM DIR	ARIMA- CNN DIR	CNN- LSTM DIR
MAPE	166,34	128,61	115,30	191,53
RMSE	1,68	1,00	0,95	1,49
MSE	2,83	1,00	0,90	2,22
MAE	0,18	0,09	0,05	0,01
R^2	-0,13	0,60	0,64	0,11

Tabela 9 – Bahia - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM

Modelo	Tipo	β_2	β_1	β_0
ARIMA-LSTM	Dir	-0,32	2,62	-2,43
ARIMA-LSTM	Vel	-0,06	1,22	-0,24
ARIMA-CNN	Dir	-0,86	2,88	-1,85
ARIMA-CNN	Vel	-0,22	1,34	-0,20
CNN-LSTM	Dir	0,98	0,03	-0,02
CNN-LSTM	Vel	0,90	0,10	-0,01

R^2 de 0,81. Ambos os modelos foram capazes de descrever 81% da variância dos dados, entretanto, por possuir um valor melhor para todas as outras métricas de avaliação, o modelo ARIMA-CNN realizou um melhor ajuste geral para o conjunto de dados do estado da Bahia. Os valores das constantes estimadas para as regressões lineares multivariadas utilizadas para a construção dos modelos híbridos em paralelo estão representados na Tabela 9.

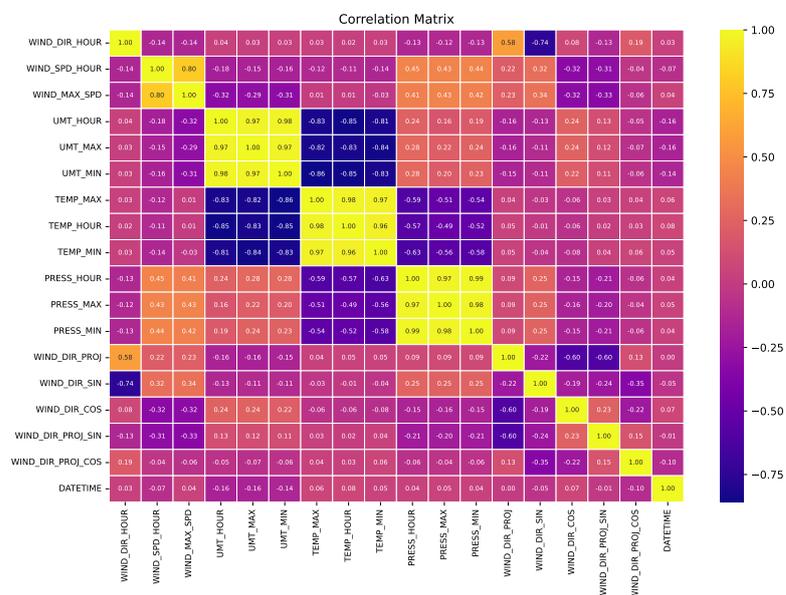
4.1.2 Piauí

Para a velocidade dos ventos, os resultados encontrados na matriz de correlação nos mostram a existência de uma única correlação forte com velocidade máxima do vento na hora anterior. Essa correlação é positiva indicando que as duas possuem direções e sentido muito próximos.

Para a direção do vento, observa-se que suas únicas correlações relevantes são com a projeção da direção do vento e com o seno da direção do vento, sendo essas correlações de

moderada intensidade positiva e forte negativa, respectivamente. A matriz de correlações para o estado do Piauí está representada no gráfico de calor disposto na Figura 22.

Figura 22 – Piauí - Gráfico de correlação dos dados climáticos referentes ao estado do Piauí.



O coeficiente Kaiser-Meyer-Olkin (KMO) calculado para os dados de teste de 0,73 indica que os dados são altamente correlacionados e, portanto, propensos a possuírem variáveis latentes bem descritivas. Essa suposição é confirmada pelo p_{value} de 0,00 no teste de esfericidade de Bartlett que confirma que o conjunto de dados é altamente correlacionado. Esses resultados indicam fortemente a possibilidade de o conjunto possuir variáveis latentes que descrevem grande parte da variância dos dados, o que justifica a realização de uma análise de fatores. Os valores encontrados para o teste e o coeficiente ficam descritos na Tabela 10.

Tabela 10 – Piauí - Resultados de KMO e Teste de Esfericidade de Bartlett

KMO	Teste de Esfericidade de Bartlett p-value
0,73	0,00

Tabela 11 – Piauí - Variância Explicada por Fator

Fator	Variância Proporcional	Variância Acumulada
Fator 1	0,30	0,30
Fator 2	0,17	0,47
Fator 3	0,11	0,58
Fator 4	0,10	0,68
Fator 5	0,09	0,76
Fator 6	0,06	0,83
Fator 7	0,02	0,85

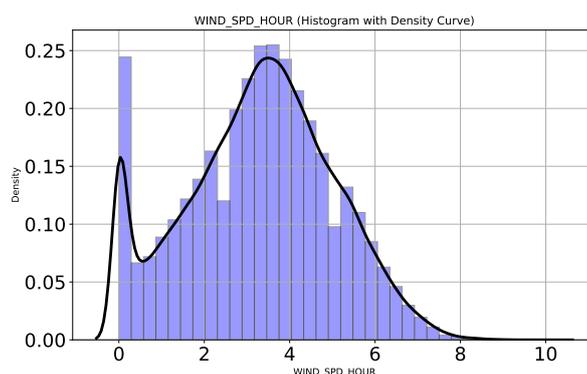
Os resultados da análise de fatores apresentados na Tabela 11 revelam a presença de sete variáveis latentes, que juntas explicam 85% da variância dos dados. Esses achados, em conjunto com a matriz de correlação, sugerem que as variáveis utilizadas no treinamento dos modelos preditivos têm um papel relevante no aumento da precisão dos ajustes para cada uma delas.

A Figura 23 apresenta as observações de velocidade e direção do vento nas últimas 500 horas e seus respectivos histogramas para todo o conjunto de dados, que inclui 96.432 observações utilizadas do estado do Piauí, conforme ilustrado na Tabela 12.

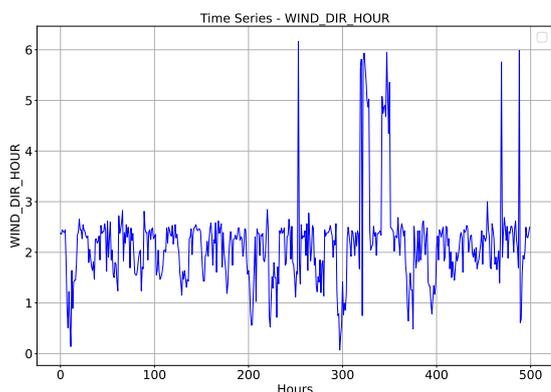
Figura 23 – Piauí - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.



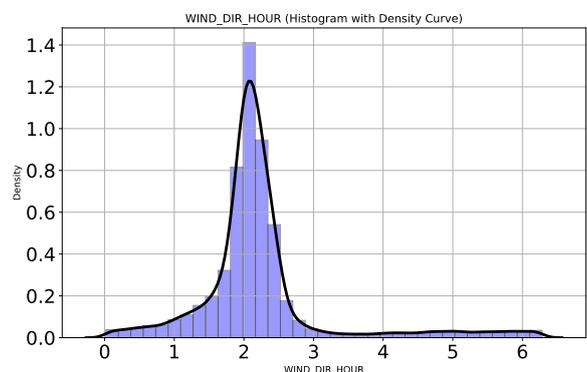
(a)



(b)



(c)



(d)

Como pode ser visto nas Figuras 23(a) e 23(c), a velocidade e a direção do vento se comportam de forma diferente. No entanto, elas têm o mesmo atributo de estarem centradas em torno da média de 3,28 e 2,24, respectivamente, com uma variação

aproximadamente constante ao longo do tempo. O teste Dickey-Fuller apoia essas suposições de estacionariedade, pois rejeita a hipótese nula de uma raiz unitária na série temporal, que é o que indica a estacionariedade. Para verificar a não normalidade na série temporal de velocidade e direção do vento, testes de hipóteses como Jarque-Bera, teste de assimetria e teste de curtose foram usados. Com base no valor de p para cada teste, levando em consideração um nível de significância de 0,05, esses testes contrastam a hipótese nula de que as séries são distribuídas normalmente. A Tabela 12 resume todos os testes usados e as estatísticas computadas. Os histogramas associados à velocidade e direção do vento são mostrados nas Figuras 23(b) e 23(d) com base em todos os registros.

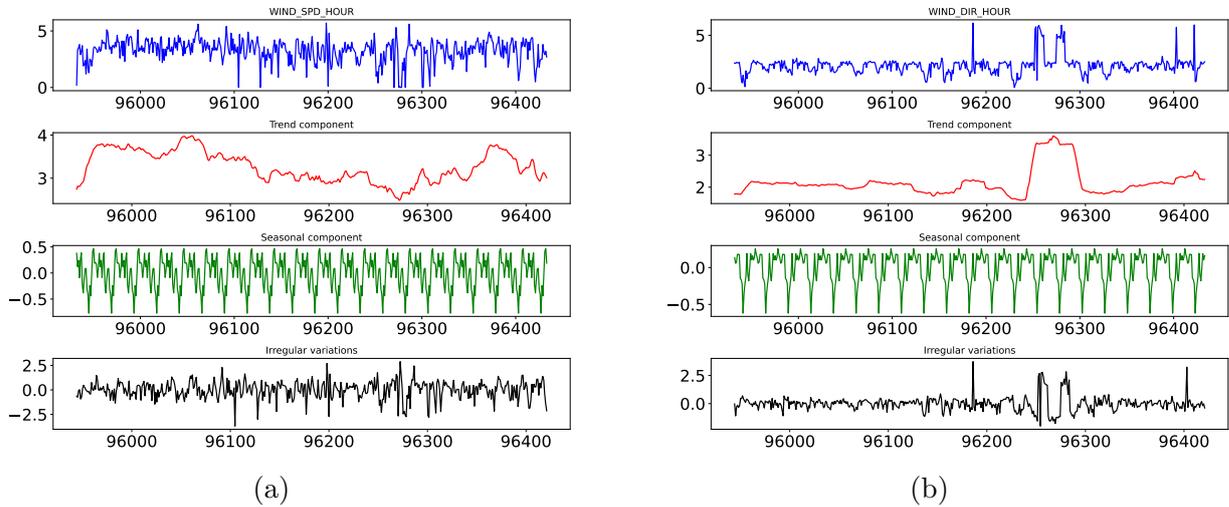
Tabela 12 – Piauí - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionaridade.

	WIND_HOUR_SPD	WIND_DIR_HOUR
N	96432	96432
μ	3,28	2,24
σ	1,74	0,947
\mathbf{x}_{\min}	0,00	0,02
\mathbf{x}_{\max}	10,10	6,28
Q_1	2,10	1,90
Q_2	3,40	2,11
Q_3	4,50	2,39
kurtosis	-0,50	5,67
kurtosistest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
skewness	-0,07	1,91
skewnesstest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
jarque_bera	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
adfuller	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$

Na Tabela 12, um valor de assimetria negativo de -0,07 confirma que o histograma da velocidade do vento apresenta uma leve assimetria à esquerda, indicando que velocidades mais altas que a média são levemente mais frequentes. Para a direção do vento, observa-se uma assimetria à direita, sugerindo que a maioria dos valores de direção do vento estão mais próximos do mínimo observado do que do máximo. Essa assimetria à direita pode ser confirmada por meio de um valor de assimetria de 1,91. Uma distribuição de dados mais plana do que uma distribuição normal pode ser observada com uma curtose de -0,50 para velocidade do vento. As caudas mais leves da distribuição e a menor concentração de valores em torno da média em comparação com uma distribuição normal são o resultado disso. Para a direção do vento, a distribuição de dados é significativamente mais pontiaguda do que uma distribuição normal, como indicado pelo valor de curtose de 5,67. Esse resultado sugere que a distribuição possui caudas mais pesadas e uma maior concentração de valores próximos à média do que em uma distribuição normal (que teria uma curtose de 0).

Na Figura 24 está representada a decomposição das séries temporais de velocidade e

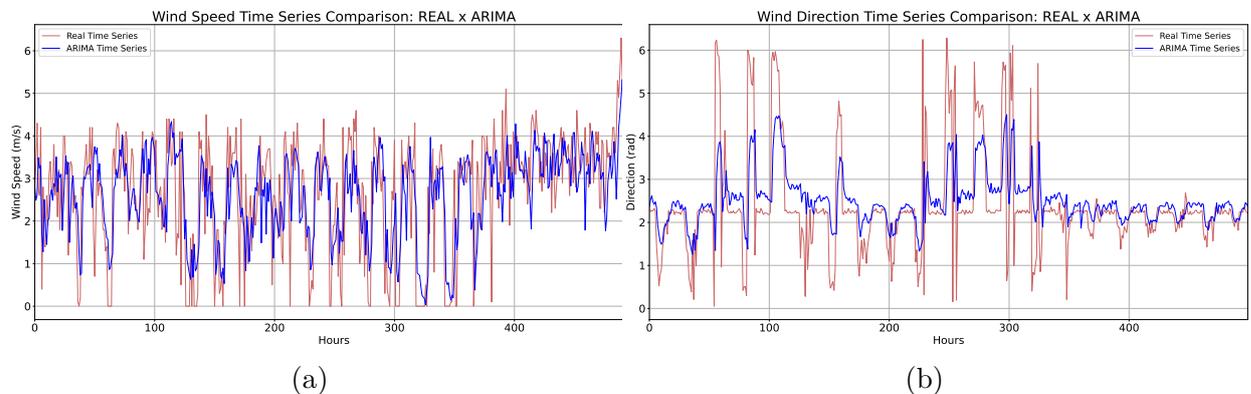
Figura 24 – Piauí - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.



direção do vento com referência a um período de 24 horas. Com base nessa imagem, observa-se uma alta tendência sazonal em ambas as curvas, sendo essa sazonalidade igualmente significativa em ambas as séries. Pode-se observar que a magnitude da componente de erro das séries temporais de ambas as curvas é bastante alta, podendo ser até mais significativa do que a componente de tendência, especialmente no caso da direção do vento.

As Figuras 25, 26, 27 mostram, respectivamente, a comparação das previsões geradas pelos modelos ARIMA, CNN e LSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidos na Tabela 13.

Figura 25 – Piauí - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.



Os resultados apresentados na Tabela 13 nos mostram que, para a direção do vento, todos os modelos tiveram uma baixa adesão com R^2 máximo de 0,26 para o modelo LSTM.

Figura 26 – PiauÍ - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.

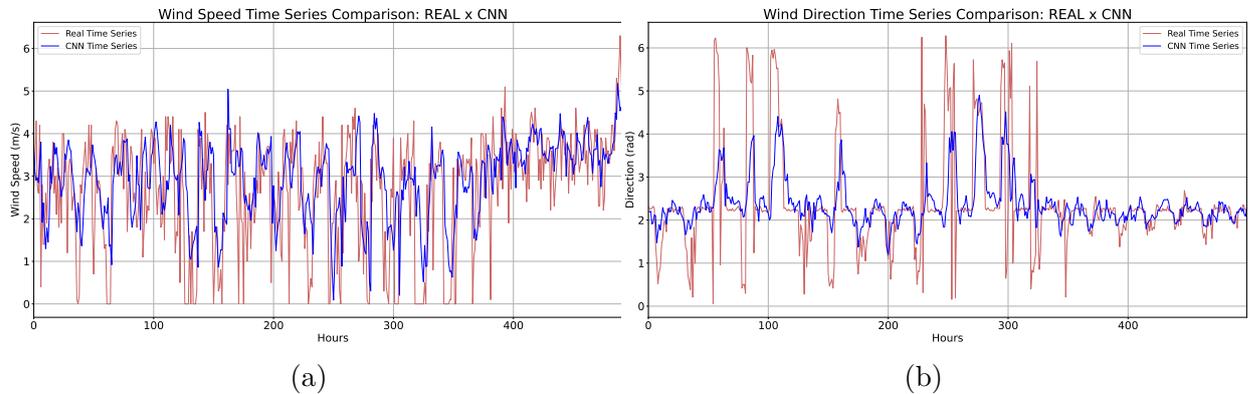
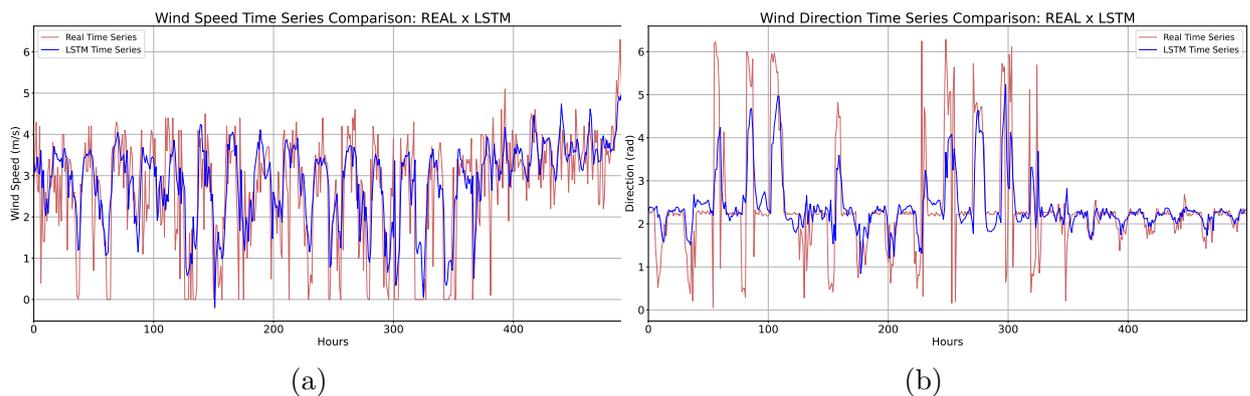


Figura 27 – PiauÍ - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.



Tal resultado implica que apenas 26% da variância da série temporal foi de fato capturada por esse modelo. Além disso, observa-se que, como padrão, todos os modelos apresentaram valores baixos de MAE e MAPE, sendo os melhores resultados, respectivamente, 0,01 e 23,99%, ambos alcançados pelo modelo LSTM. Essa combinação de baixos MAE e MAPE com uma baixa descrição das variâncias pode ser explicada pela alta concentração de valores próximos a um valor central da distribuição de resultados, como visto na Figura 23 (d).

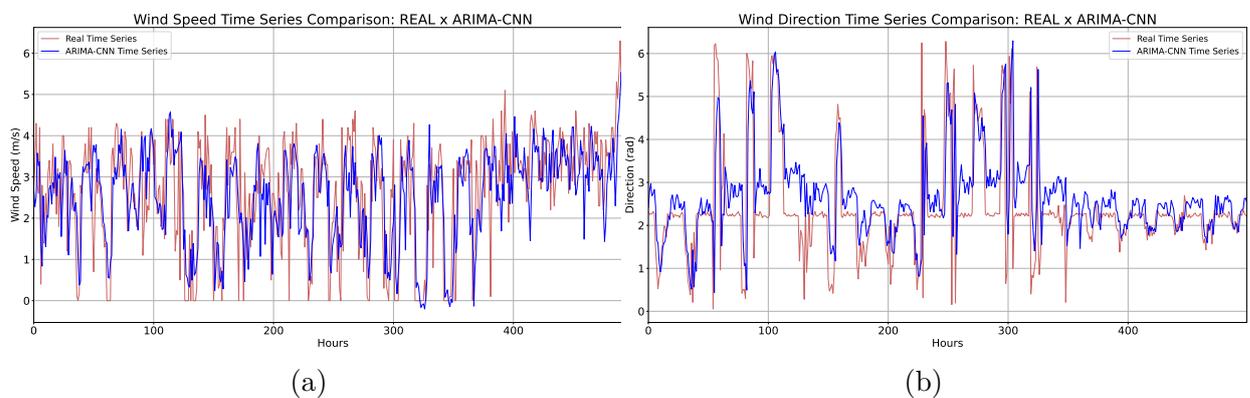
Para a velocidade do vento, os resultados nos mostram uma capacidade de ajuste muito melhor por parte de todos os modelos. O melhor dentre eles, novamente foi o LSTM que, com R^2 de 0,43, foi capaz de descrever 43% da variância dos dados. Com exceção do MAPE, melhor performado pelo modelo ARIMA com valor de 40,19%, o modelo LSTM apresentou o melhor desempenho em todas as outras métricas de avaliação.

Tabela 13 – Piauí - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).

Métrica	CNN VEL	LSTM VEL	ARIMA VEL
MAPE	47,77	41,40	40,19
RMSE	1,08	0,97	0,98
MSE	1,17	0,94	0,97
MAE	0,21	0,01	0,05
R^2	0,29	0,43	0,41
Métrica	CNN DIR	LSTM DIR	ARIMA DIR
MAPE	24,30	23,99	24,27
RMSE	0,60	0,57	0,60
MSE	0,36	0,32	0,36
MAE	0,02	0,01	0,09
R^2	0,15	0,26	0,17

As Figuras 28, 29, 30 e 31 representam, respectivamente, a comparação das previsões geradas pelos modelos híbridos ARIMA-CNN, ARIMA-LSTM, CNN-LSTM e CNN-ALSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidos na Tabela 14.

Figura 28 – Piauí - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.



Os resultados apresentados na Tabela 14 indicam uma melhora significativa nos modelos híbridos paralelos que combinam ARIMA com RNAs, em comparação aos modelos individuais, com desempenho superior em quase todas as métricas avaliadas. Em contrapartida, o modelo CNN-ALSTM e CNN-LSTM apresentaram, de forma geral, resultados semelhantes aos modelos individuais, com valores próximos em todas as métricas.

Para a previsão da direção do vento, os modelos ARIMA-LSTM e ARIMA-CNN apresentaram uma melhora expressiva nos resultados de R^2 sendo esses iguais a 0,69. Esses resultados sugerem que o modelo ARIMA-LSTM foi capaz de capturar 69% da

Figura 29 – PiauÍ - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.

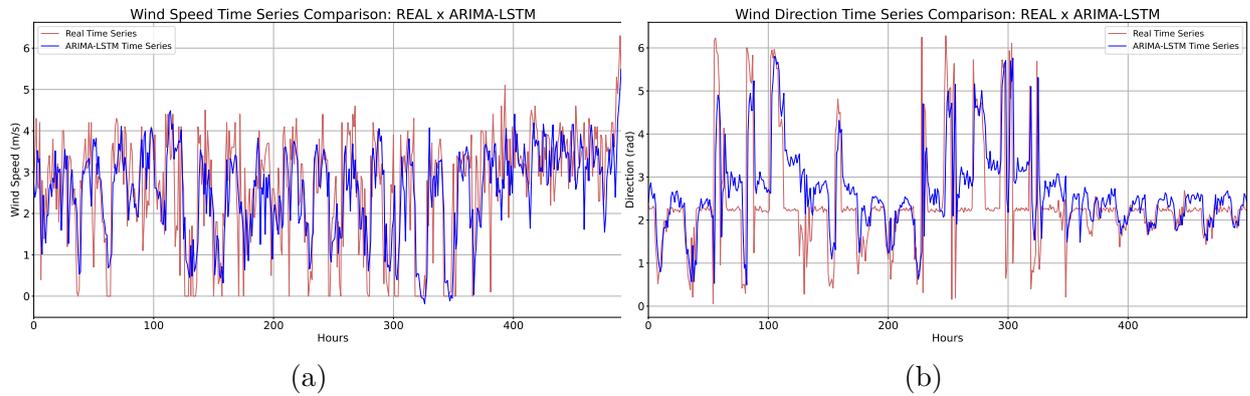


Figura 30 – PiauÍ - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.

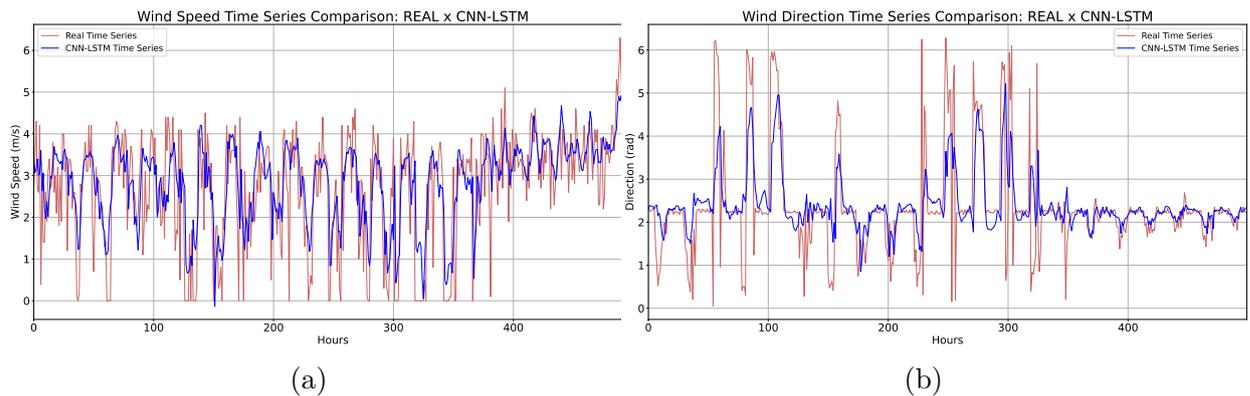
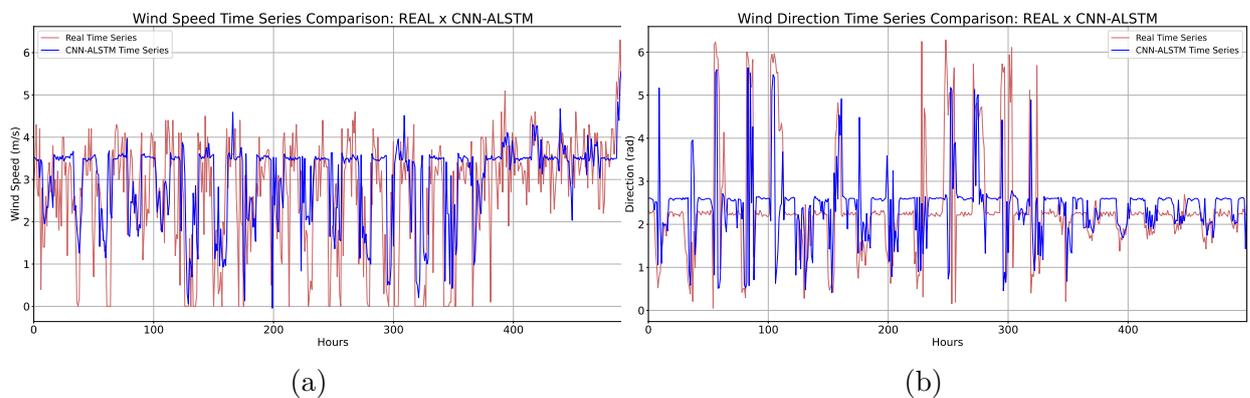


Figura 31 – PiauÍ - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.



variância da série temporal, um resultado melhor que os 26% descritos pelo LSTM, que foi melhor modelo individual. Com relação aos parâmetros MAPE, RMSE e MSE, o modelo ARIMA-CNN também apresentou melhoras significativas diminuindo o MAPE de 41,40% para 15,55%, o RMSE de 0,57 para 0,37 e o MSE de 0,32 para 0,13 quando comparado com o LSTM. O único parâmetro que não foi melhorado pelos modelos híbridos foi o MAE sendo este de 0,15 o ARIMA-LSTM e 0,01 para o LSTM. O melhor desempenho geral foi do modelo ARIMA-LSTM que apresentou métricas de erro mais baixas mesmo com seu R^2 sendo igual ao obtido pelo ARIMA-CNN.

Tabela 14 – Piauí - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).

Métrica	CNN-ALSTM VEL	ARIMA-LSTM VEL	CNN-LSTM VEL	ARIMA-CNN VEL
MAPE	44,19	30,40	41,19	29,83
RMSE	1,04	0,82	0,97	0,83
MSE	1,07	0,67	0,93	0,69
MAE	0,06	0,02	0,02	0,06
R^2	0,35	0,60	0,44	0,59
Métrica	CNN-ALSTM DIR	ARIMA-LSTM DIR	CNN-LSTM DIR	ARIMA-CNN DIR
MAPE	31,32	15,55	23,94	15,89
RMSE	0,73	0,37	0,57	0,36
MSE	0,53	0,13	0,32	0,13
MAE	0,14	0,15	0,01	0,17
R^2	-0,23	0,69	0,26	0,69

Tabela 15 – Piauí - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM

Modelo	Tipo	β_2	β_1	β_0
ARIMA-LSTM	Dir	-0,31	1,90	-1,35
ARIMA-LSTM	Vel	-0,06	1,12	-0,22
ARIMA-CNN	Dir	-0,52	1,96	-1,03
ARIMA-CNN	Vel	-0,14	1,19	-0,17
CNN-LSTM	Dir	0,99	0,01	0,01
CNN-LSTM	Vel	0,95	0,06	-0,04

Para a velocidade do vento, os modelos híbridos mantiveram a melhora nos resultados. Nesse cenário, os modelos ARIMA-CNN e ARIMA-LSTM mostraram avanços em todas as métricas de avaliação, com destaque para o coeficiente de determinação $R^2 = 0,60$ alcançado pelo ARIMA-LSTM, que explicou 60% da variância dos dados. Embora ambos os modelos tenham apresentado bons resultados para as variáveis de interesse, o ARIMA-LSTM se destacou com um desempenho ligeiramente superior em todas as métricas avaliadas. Os valores das constantes estimadas para as regressões lineares multivariadas,

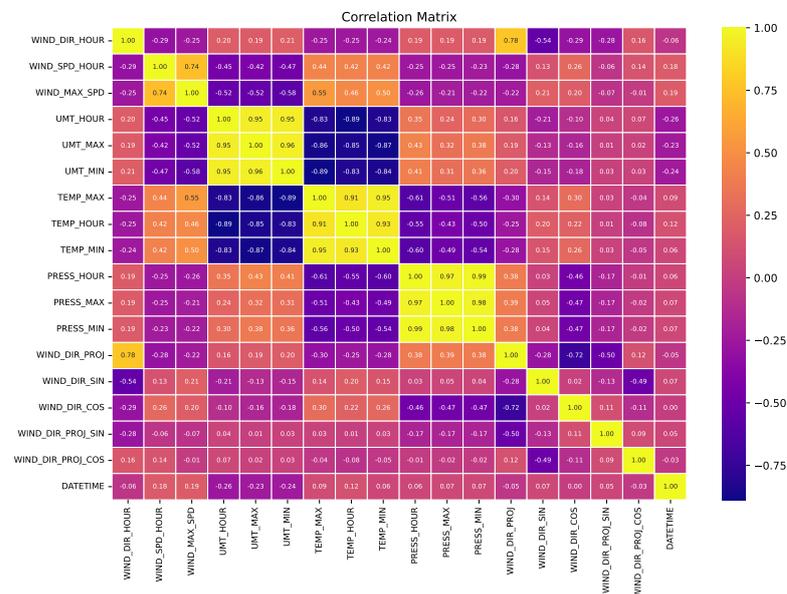
utilizadas na construção dos modelos híbridos em paralelo, estão apresentados na Tabela 15.

4.1.3 Rio Grande do Norte

Para a velocidade dos ventos, os resultados nos mostraram a existência de uma única correlação relevante, sendo esta uma correlação positiva forte com a velocidade máxima do vento na hora anterior. Além disso, foi possível verificar a existência de correlações positivas próximas a 0,50 com a temperatura máxima na hora anterior, temperatura horária e temperatura mínima na hora anterior, assim como correlações negativas moderadas com a umidade horária, umidade máxima na hora anterior e umidade mínima na hora anterior.

Para a direção do vento, observa-se que suas únicas correlações são com a projeção da direção do vento e com o seno da direção do vento, sendo estas correlações uma positiva forte e uma negativa moderada, respectivamente. A matriz de correlação dos dados do estado do Rio Grande do Norte está representada no gráfico de calor disposto na Figura 32.

Figura 32 – Rio Grande do Norte - Gráfico de correlação dos dados climáticos referentes ao estado do Rio Grande do Norte.



O coeficiente Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett evidenciaram uma alta correlação entre as variáveis, o que sugere a existência de variáveis latentes. Os valores apresentados na Tabela 16 mostram um coeficiente KMO de 0,73, indicando uma correlação elevada no conjunto de variáveis. Esse resultado é reforçado pela rejeição da hipótese nula no teste de esfericidade de Bartlett, confirmando a adequação da análise de fatores.

Tabela 16 – Rio Grande do Norte - Resultados de KMO e Teste de Esfericidade de Bartlett

KMO	Teste de Esfericidade de Bartlett p-value
0,73	0,00

Os resultados da análise de fatores ilustrados na Tabela 17, nos mostram a existência de sete variáveis latentes que juntas descrevem 83% das variâncias desse conjunto de dados. Esses resultados, em conjunto com a matriz de correlação, indicam que as variáveis utilizadas no treinamento dos modelos preditivos são altamente correlacionadas sendo capazes de aumentar a precisão dos ajustes para cada uma das variáveis analisadas.

Tabela 17 – Rio Grande do Norte - Variância Explicada por Fator

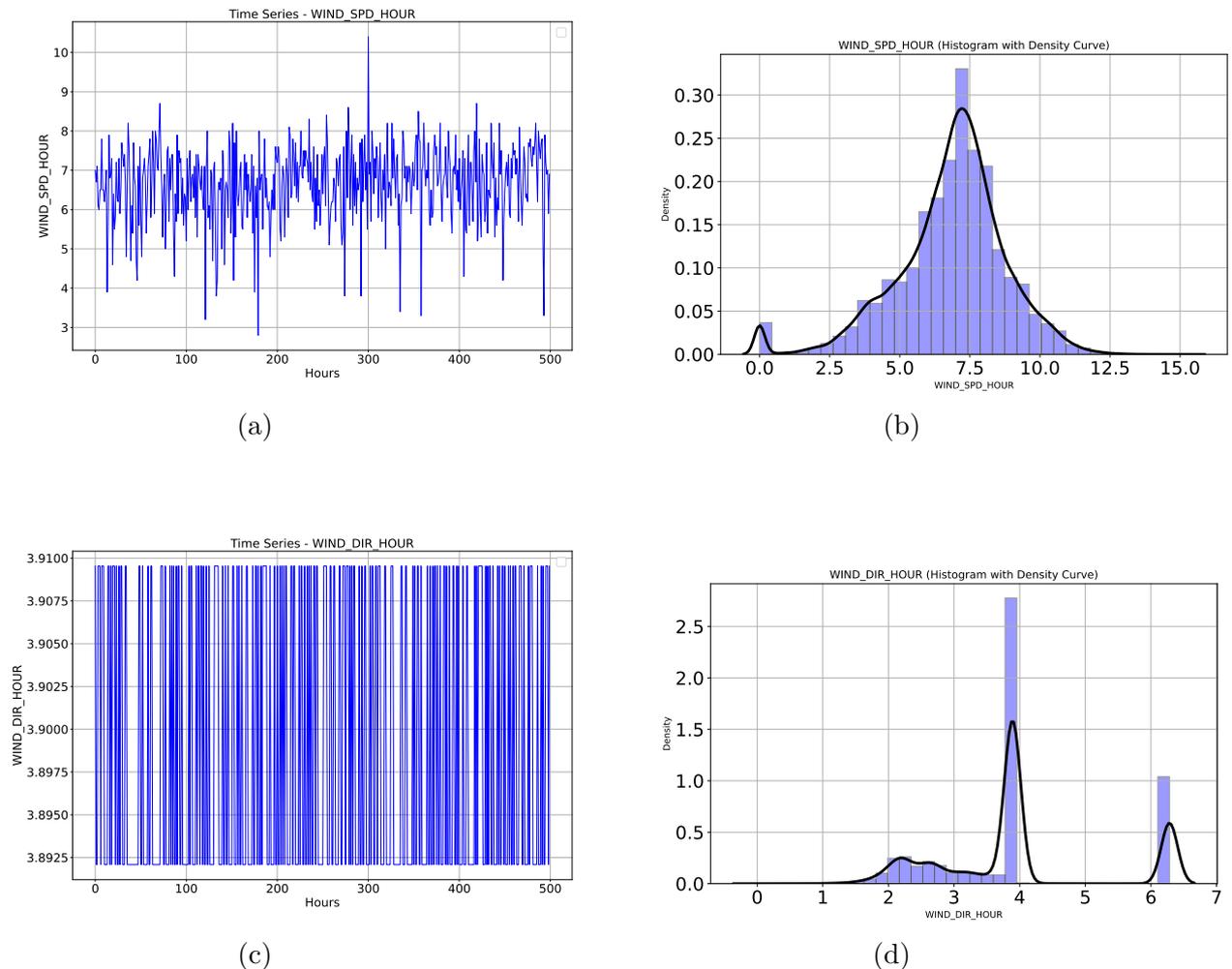
Fator	Variância Proporcional	Variância Acumulada
Fator 1	0,31	0,31
Fator 2	0,17	0,48
Fator 3	0,09	0,57
Fator 4	0,09	0,66
Fator 5	0,08	0,74
Fator 6	0,06	0,81
Fator 7	0,03	0,83

A Figura 33 apresenta as observações de velocidade e direção do vento nas últimas 500 horas e seus respectivos histogramas para todo o conjunto de dados, que inclui 119.664 observações utilizadas do estado do Rio Grande do Norte, conforme ilustrado na Tabela 18. O maior número de observações se deve ao intervalo maior de captação de dados para esse estado, sendo o intervalo entre 2008 e 2021.

Como pode ser visto nas Figuras 33(a) e 33(c), a velocidade e a direção do vento se comportam de forma diferente. No entanto, elas têm o mesmo atributo de estarem centradas em torno da média de 6,81 e 4,17, respectivamente, com uma variação aproximadamente constante ao longo do tempo. O teste Dickey-Fuller apoia essas suposições de estacionariedade, pois rejeita a hipótese nula de uma raiz unitária na série temporal, que indica a estacionariedade. Para verificar a não normalidade na série temporal de velocidade e direção do vento, testes de hipóteses como Jarque-Bera, teste de assimetria e teste de curtose foram usados. Com base no valor de p para cada teste, levando em consideração um nível de significância de 0,05, esses testes contrastam a hipótese nula de que as séries são distribuídas normalmente. A Tabela 18 resume todos os testes usados e as estatísticas computadas. Os histogramas associados à velocidade e direção do vento são mostrados nas Figuras 33(b) e 33(d) com base em todos os registros.

Na Tabela 18, um valor de assimetria negativo de -0,69 confirma que o histograma da velocidade do vento mostra uma assimetria clara à esquerda que indica que velocidades mais baixas do vento são levemente mais frequentes. Para a direção do vento, a assimetria é à direita e pode ser confirmada por meio de um valor de assimetria de 1,45. Uma

Figura 33 – Rio Grande do Norte - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.



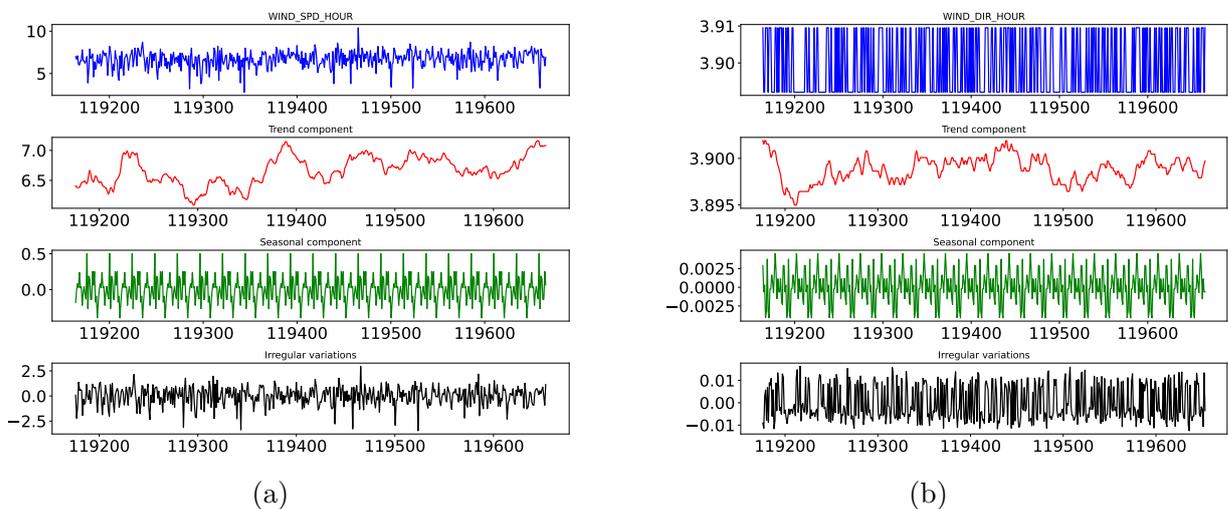
distribuição de dados ligeiramente mais plana do que uma distribuição normal pode ser observada com uma curtose de $-0,91$ para direção do vento. As caudas mais leves da distribuição e a menor concentração de valores em torno da média em comparação com uma distribuição normal são o resultado disso. Para a velocidade do vento, a distribuição de dados é marginalmente mais pontiaguda do que uma distribuição normal conforme sugerido pelo valor de curtose de $1,49$. Esse resultado implica que há caudas mais pesadas e uma densidade maior de valores próximos à média na distribuição do que em uma distribuição normal.

Na Figura 34 está representada a decomposição das séries temporais de velocidade e direção do vento usando de referência um período de 24 horas. Com base nessa imagem, observa-se uma alta tendência sazonal em ambas as curvas, sendo essa sazonalidade de

Tabela 18 – Rio Grande do Norte - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionariedade.

	WIND_HOUR_SPD	WIND_DIR_HOUR
N	119664	119664
μ	6,81	4,17
σ	1,96	1,38
\mathbf{x}_{\min}	0,00	0,02
\mathbf{x}_{\max}	15,30	6,28
Q_1	5,80	3,11
Q_2	7,00	3,94
Q_3	8,00	4,70
kurtosis	1,49	-0,91
kurtosistest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
skewness	-0,69	0,11
skewnesstest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
jarque_bera	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
adfuller	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$

Figura 34 – Rio Grande do Norte - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.



pequena magnitude quando comparada às componentes de tendência e erro. Além disso, a magnitude da componente de erro da série temporal é alta em ambas as curvas, porém se restringe a menos da metade da componente de tendência.

As Figuras 35, 36, 37 mostram, respectivamente, a comparação das previsões geradas pelos modelos ARIMA, CNN e LSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidos na Tabela 19.

Os resultados apresentados na Tabela 19 nos mostram que o único modelo que

Figura 35 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.

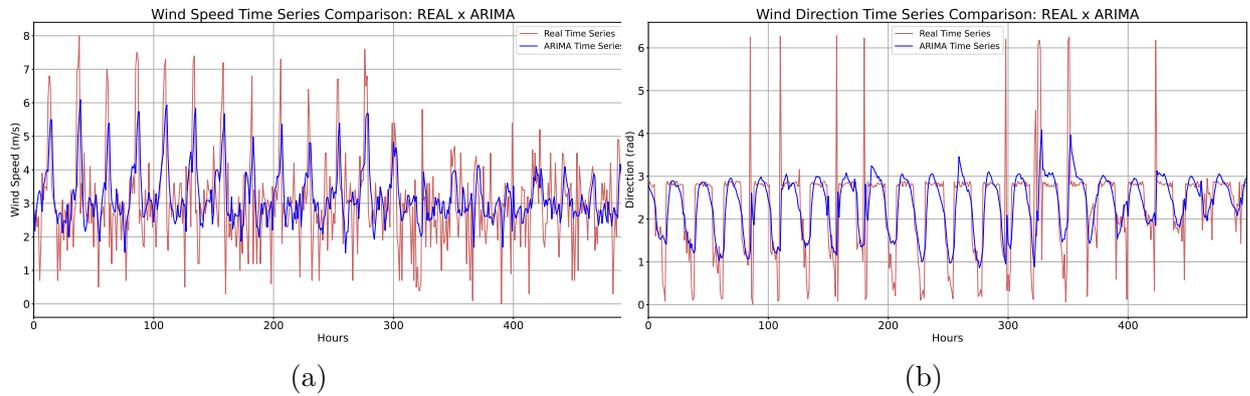


Figura 36 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.

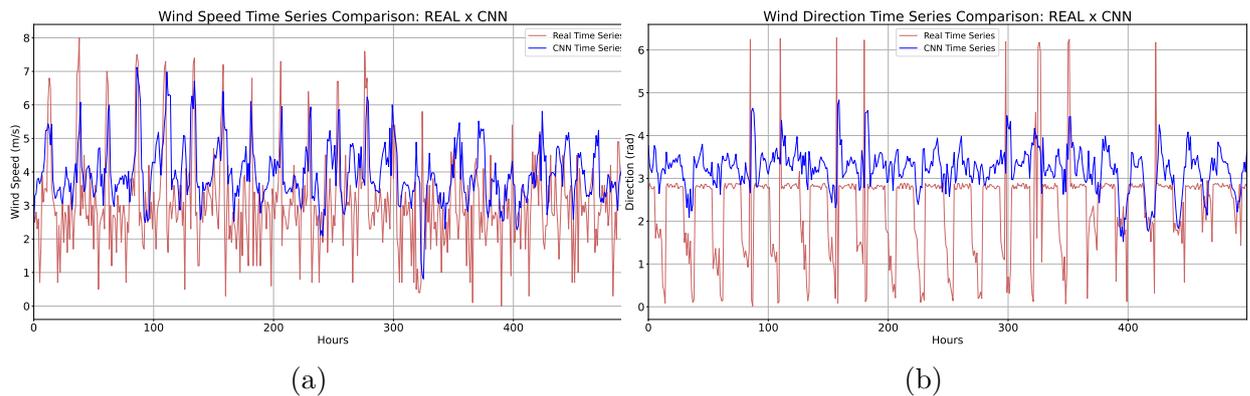
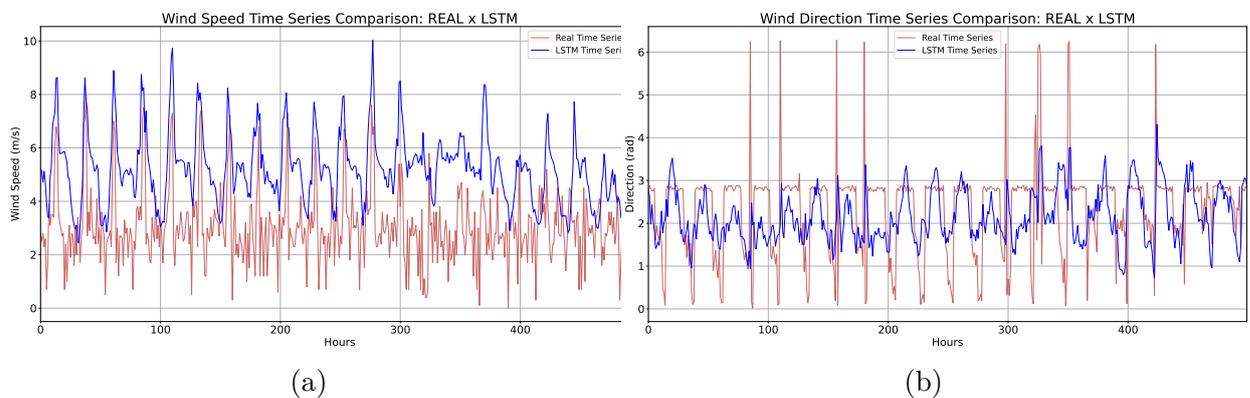


Figura 37 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.



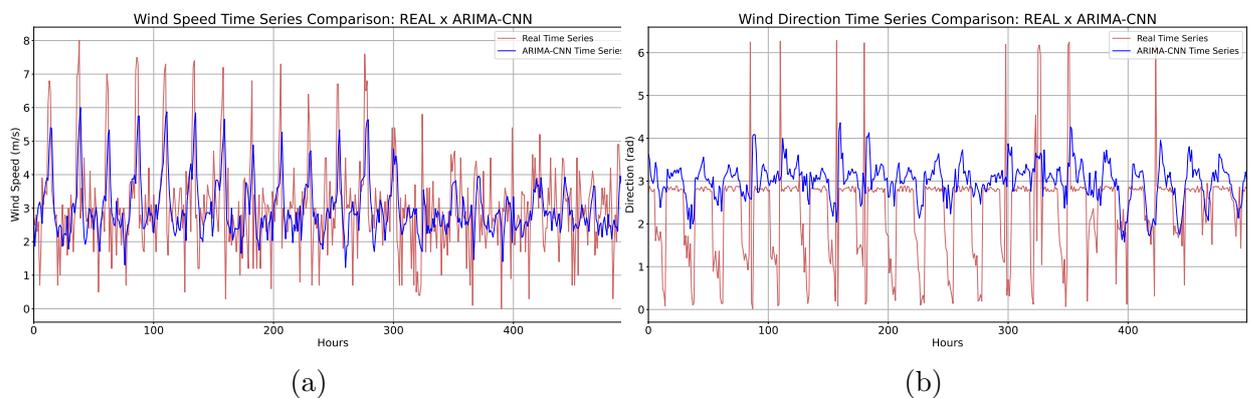
performou melhor que a linha reta da média das observações reais foi o ARIMA com os R^2 de 0,16, explicando apenas 16% da variância da série temporal para a velocidade do vento, e 0,20, explicando apenas 20% da variância da série temporal para a direção do vento. A alta dificuldade de modelos baseados em RNAs para ajustar a curva pode ser devido a alta frequência de oscilações sazonais.

Tabela 19 – Rio Grande do Norte - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).

Métrica	CNN VEL	LSTM VEL	ARIMA VEL
MAPE	78,95	143,94	53,79
RMSE	1,44	2,44	1,09
MSE	2,06	5,93	1,18
MAE	0,58	1,96	0,02
R^2	-0,46	-3,20	0,16
Métrica	CNN DIR	LSTM DIR	ARIMA DIR
MAPE	124,44	86,84	67,66
RMSE	1,09	0,97	0,79
MSE	1,18	0,94	0,63
MAE	0,63	0,18	0,01
R^2	-0,50	-0,19	0,20

As Figuras 38, 39, 40 e 41 representam, respectivamente, a comparação das previsões geradas pelos modelos híbridos ARIMA-CNN, ARIMA-LSTM, CNN-LSTM e CNN-ALSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidos na Tabela 20.

Figura 38 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.



Os resultados apresentados na Tabela 20 indicam uma melhora significativa nos modelos híbridos paralelos que combinam ARIMA com RNAs, em comparação aos modelos individuais, com desempenho superior em quase todas as métricas avaliadas para a

Figura 39 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.

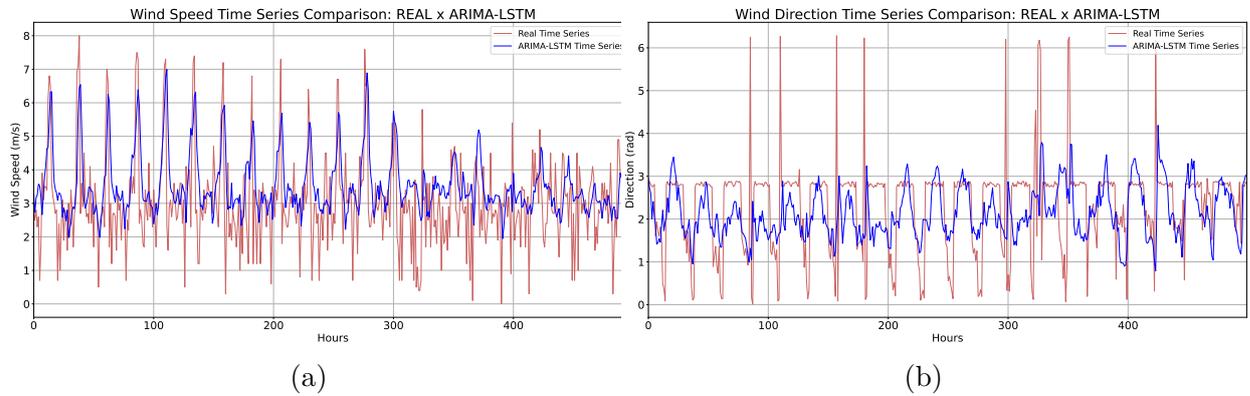


Figura 40 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.

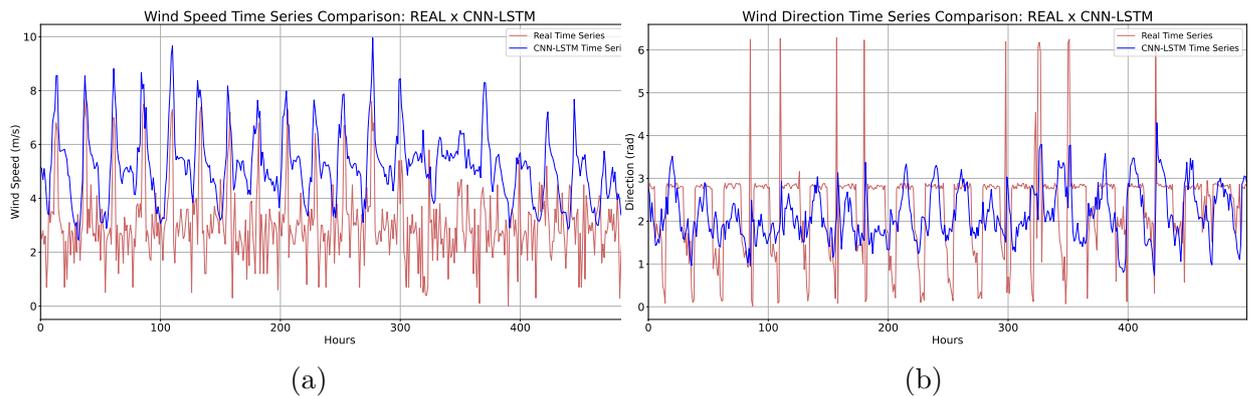
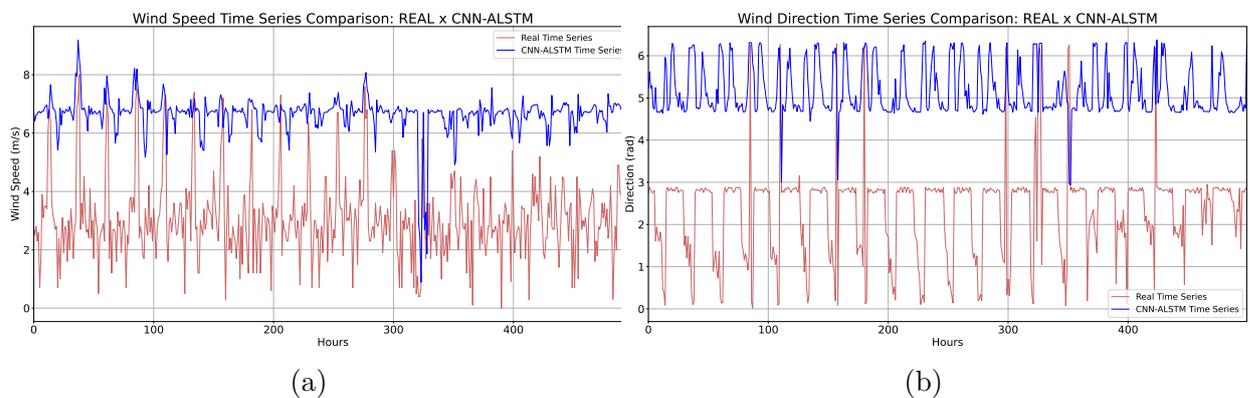


Figura 41 – Rio Grande do Norte - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.



velocidade do vento. Quanto à direção do vento, a utilização de modelos de RNAs para compor os modelos híbridos resultou em uma divergência nos resultados, piorando as métricas quando comparados com o modelo individual ARIMA.

Tabela 20 – Rio Grande do Norte - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).

Métrica	CNN-ALSTM VEL	ARIMA-LSTM VEL	CNN-LSTM VEL	ARIMA-CNN VEL
MAPE	239,77	52,84	142,50	37,60
RMSE	4,03	0,92	2,41	0,83
MSE	16,24	0,85	5,80	0,69
MAE	3,76	0,29	1,94	0,24
R^2	-10,51	0,40	-3,11	0,51
Métrica	CNN-ALSTM DIR	ARIMA-LSTM DIR	CNN-LSTM DIR	ARIMA-CNN DIR
MAPE	273,08	83,47	86,89	107,42
RMSE	2,86	0,93	0,96	0,94
MSE	8,15	0,87	0,93	0,88
MAE	2,62	0,19	0,18	0,44
R^2	-9,37	-0,11	-0,18	-0,12

Tabela 21 – Rio Grande do Norte - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM

Modelo	Tipo	β_2	β_1	β_0
ARIMA-LSTM	Dir	0,92	0,08	-0,03
ARIMA-LSTM	Vel	0,29	0,78	-0,49
ARIMA-CNN	Dir	0,83	0,16	-0,06
ARIMA-CNN	Vel	0,07	1,01	-0,48
CNN-LSTM	Dir	0,99	0,01	-0,01
CNN-LSTM	Vel	0,98	0,02	-0,01

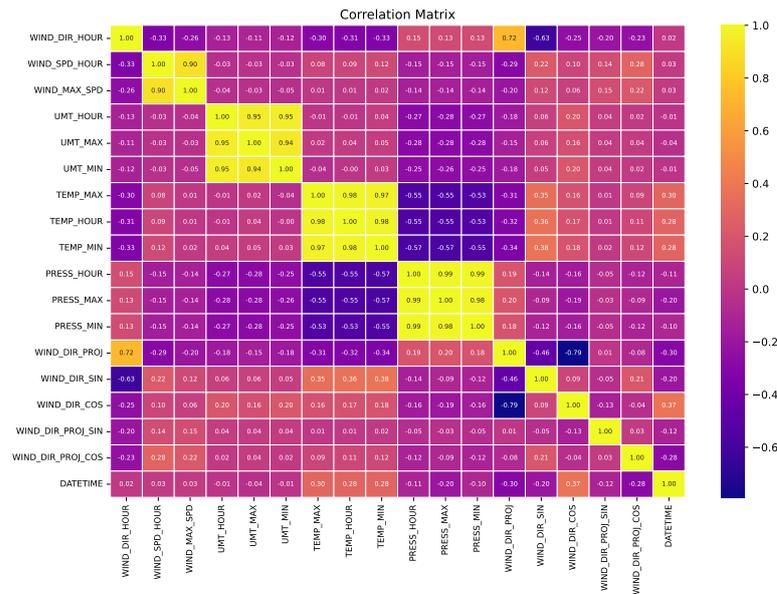
Os resultados dos modelos ARIMA-LSTM e ARIMA-CNN para a previsão da velocidade do vento se mostraram promissores para os outros estados. No entanto, a incapacidade de aplicar esses modelos devido à divergência representa um ponto negativo para sua utilização no contexto deste estado. Para este estado, o modelo que apresentou a melhor performance foi o ARIMA, sendo o único dos modelos que não divergiu. Os valores das constantes estimadas para as regressões lineares multivariadas, utilizadas na construção dos modelos híbridos em paralelo, estão apresentados na Tabela 21.

4.1.4 Santa Catarina

A matriz de correções para o estado do Santa Catarina disposta no gráfico de calor na Figura 42 nos mostra que, para a velocidade dos ventos, existe apenas uma correlação

relevante, sendo esta positiva forte, com velocidade máxima do vento na hora anterior. Para a direção do vento, observa-se que suas únicas correlações com módulos superiores a 0,50 são com a projeção da direção do vento e com o seno da direção do vento, sendo estas correlações positiva e negativa, respectivamente.

Figura 42 – Santa Catarina - Gráfico de correlação dos dados climáticos referentes ao estado de Santa Catarina.



O coeficiente Kaiser-Meyer-Olkin (KMO) calculado para os dados de teste de 0,68 indica que os dados são altamente correlacionados e, portanto, propensos a possuírem variáveis latentes bem descritivas. Essa suposição é confirmada pelo teste de esfericidade de Bartlett que confirma a hipótese de que o grupo dos nossos dados é altamente correlacionado. Esses resultados indicam fortemente a possibilidade de o conjunto possuir variáveis latentes que descrevem grande parte da variância dos dados, o que justifica a realização de uma análise de fatores. Os valores encontrados para o teste e o coeficiente ficam descritos na Tabela 22. Os resultados da análise de fatores ilustrados na Tabela 23, nos mostram a existência de sete variáveis latentes que juntas descrevem 83% das variâncias desse conjunto de dados. Esses resultados, somados com a matriz de correlação nos confirmam a alta correlação entre as variáveis utilizadas e sugerem sua capacidade de contribuir para o aumento de precisão dos ajustes sobre cada uma das variáveis.

Tabela 22 – Resultados de KMO e Teste de Esfericidade de Bartlett

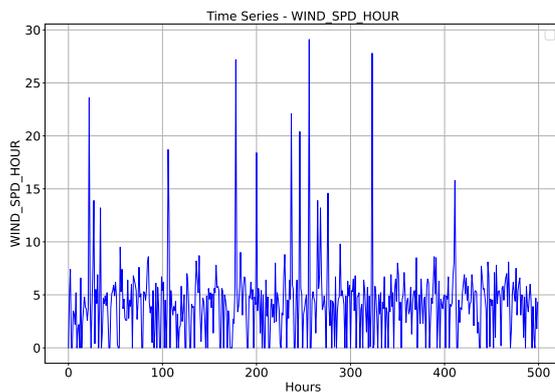
KMO	Teste de Esfericidade de Bartlett p-value
0,68	0,00

A Figura 43 apresenta as observações de velocidade e direção do vento nas últimas 500 horas e seus respectivos histogramas para todo o conjunto de dados, que inclui 96.432 observações utilizadas do estado de Santa Catarina, conforme ilustrado na Tabela 24.

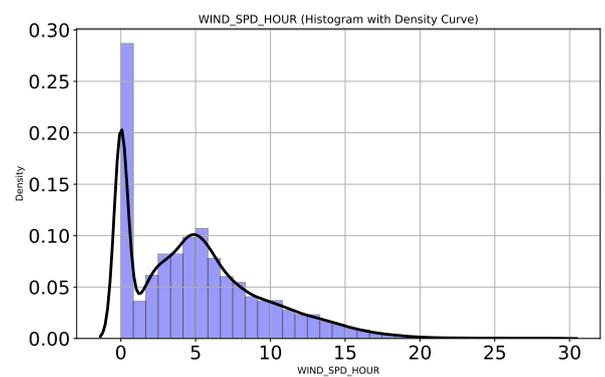
Tabela 23 – Santa Catarina - Variância Explicada por Fator

Fator	Variância Proporcional	Variância Acumulada
Fator 1	0,18	0,18
Fator 2	0,17	0,34
Fator 3	0,16	0,50
Fator 4	0,12	0,62
Fator 5	0,10	0,72
Fator 6	0,06	0,79
Fator 7	0,05	0,83

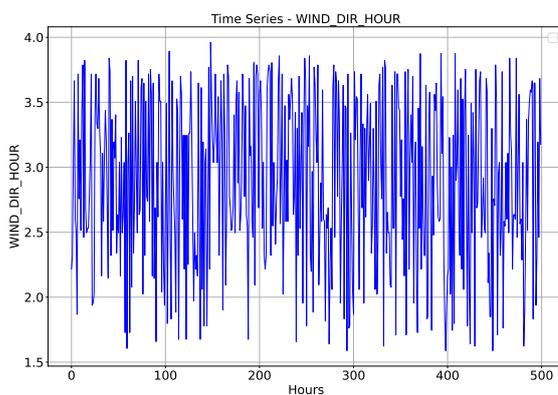
Figura 43 – Santa Catarina - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.



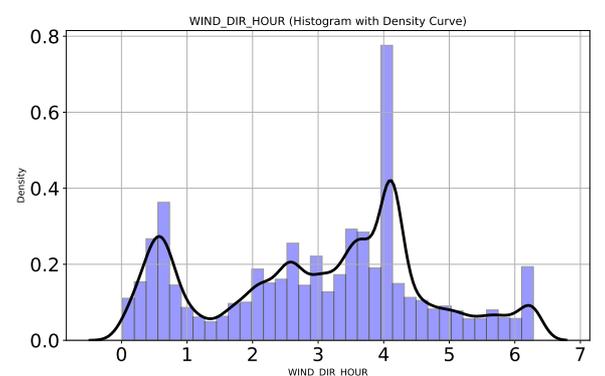
(a)



(b)



(c)



(d)

Como pode ser visto nas Figuras 43(a) e 43(c), a velocidade e a direção do vento se comportam de forma diferente. No entanto, elas têm o mesmo atributo de estarem centradas em torno da média de 5,13 e 3,03, respectivamente, com uma variação

aproximadamente constante ao longo do tempo. O teste Dickey-Fuller apoia essas suposições de estacionariedade, pois rejeita a hipótese nula de uma raiz unitária na série temporal, que é o que indica a estacionariedade. Para verificar a não normalidade na série temporal de velocidade e direção do vento, testes de hipóteses como Jarque-Bera, teste de assimetria e teste de curtose foram usados. Com base no valor de p para cada teste, levando em consideração um nível de significância de 0,05, esses testes contrastam a hipótese nula de que as séries são distribuídas normalmente. A Tabela 24 resume todos os testes usados e as estatísticas computadas. Os histogramas associados à velocidade e direção do vento são mostrados nas Figuras 43(b) e 43(d) com base em todos os registros.

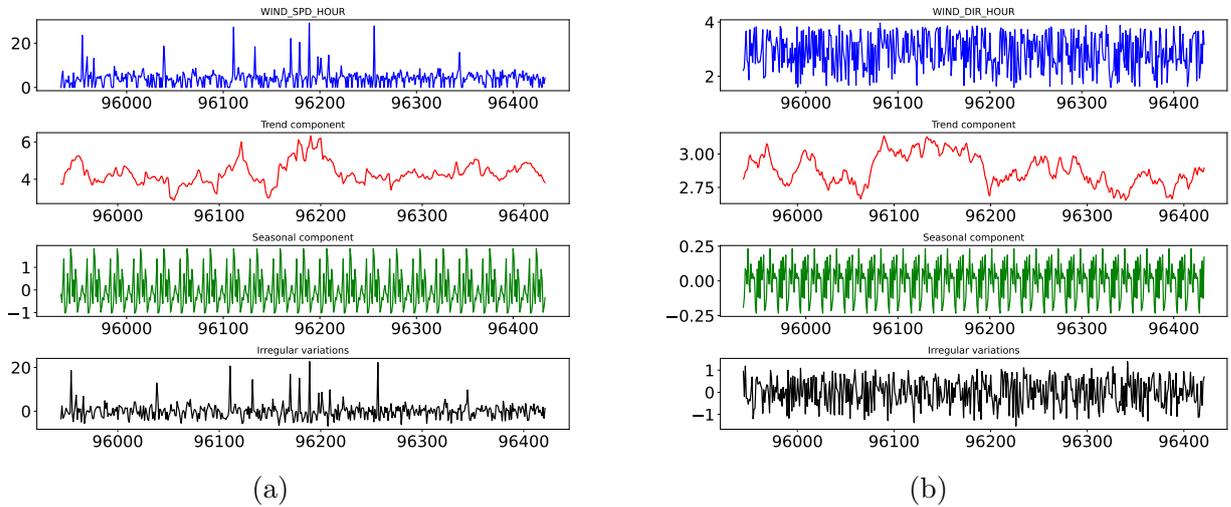
Tabela 24 – Santa Catarina - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionariedade.

	WIND_HOUR_SPD	WIND_DIR_HOUR
N	96432	96432
μ	5,13	3,03
σ	4,52	1,65
\mathbf{x}_{\min}	0,00	0,02
\mathbf{x}_{\max}	29,10	6,28
Q_1	1,20	1,80
Q_2	4,60	3,35
Q_3	7,60	4,12
kurtosis	1,08	-0,812
kurtosistest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
skewness	0,97	-0,12
skewnesstest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
jarque_bera	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
adfuller	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$

Na Tabela 24, um valor de assimetria positivo de 0,97 confirma que o histograma da velocidade do vento mostra uma assimetria clara à direita que indica que velocidades mais altas do vento são menos frequentes. Para a direção do vento, o efeito é o contrário. A assimetria à esquerda pode ser confirmada por meio de um valor de assimetria de -0,12. Uma distribuição de dados ligeiramente mais plana do que uma distribuição normal pode ser observada com uma curtose de -0,82 para a direção do vento. As caudas mais leves da distribuição e a menor concentração de valores em torno da média em comparação com uma distribuição normal são o resultado disso. Para a velocidade do vento, a distribuição de dados é marginalmente mais pontiaguda do que uma distribuição normal conforme sugerido pelo valor de curtose de 1,08. Esse resultado implica que há caudas mais pesadas e uma densidade maior de valores próximos à média na distribuição do que em uma distribuição normal.

Na Figura 44 está representada a decomposição das séries temporais de velocidade e direção do vento usando de referência um período de 24 horas. Com base nessa imagem,

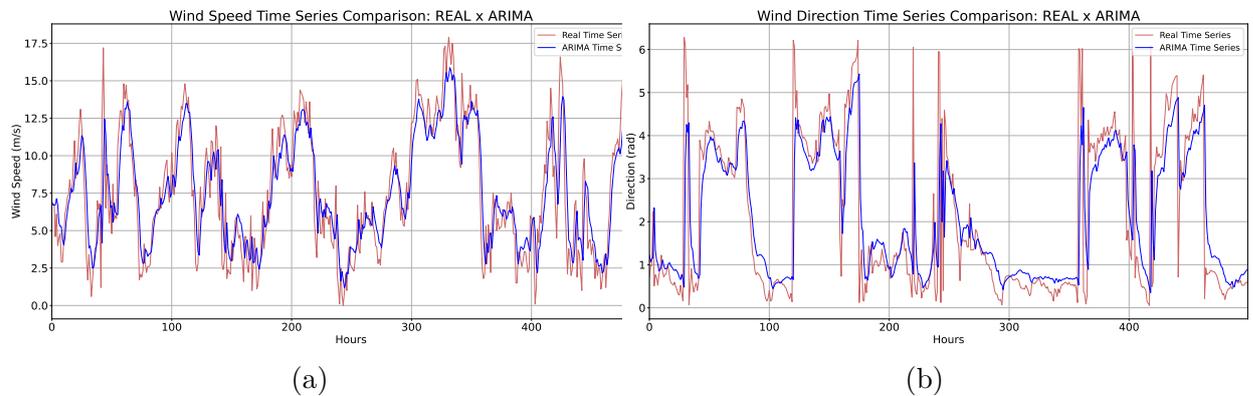
Figura 44 – Santa Catarina - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.



observa-se uma alta tendência sazonal em ambas as curvas. Além disso, verifica-se que a magnitude da componente de erro da série temporal de ambas as curvas é bem alta, quase se igualando à componente de tendência no caso da direção do vento e sendo maior que a tendência no caso da velocidade do vento.

As Figuras 45, 46, 47 mostram, respectivamente, a comparação das previsões geradas pelos modelos ARIMA, CNN e LSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidos na Tabela 25.

Figura 45 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.



Os resultados apresentados na Tabela 25 mostram que, dentre os modelos individuais, o CNN e o ARIMA foram os que realizaram os melhores ajustes para ambos os dados de interesse. Assim como para os outros estados, observa-se que a série da velocidade do

Figura 46 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.

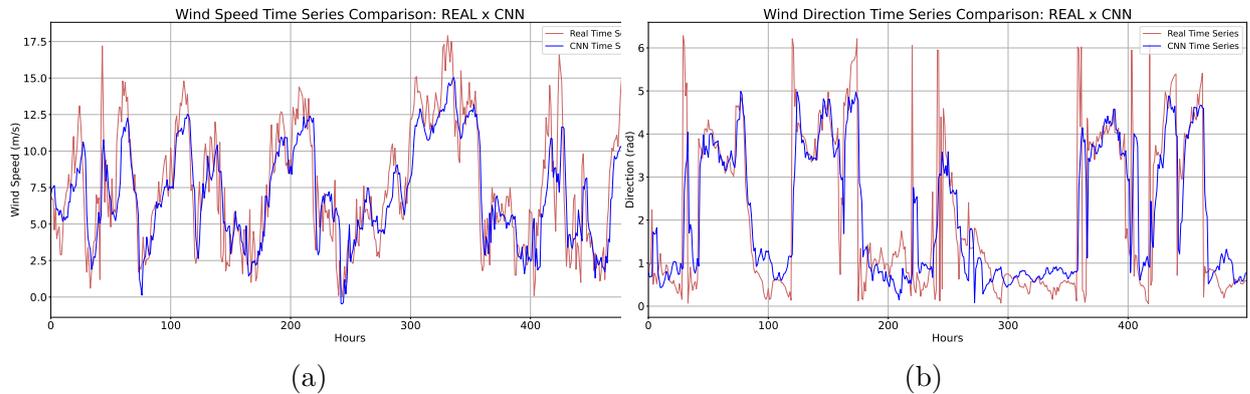
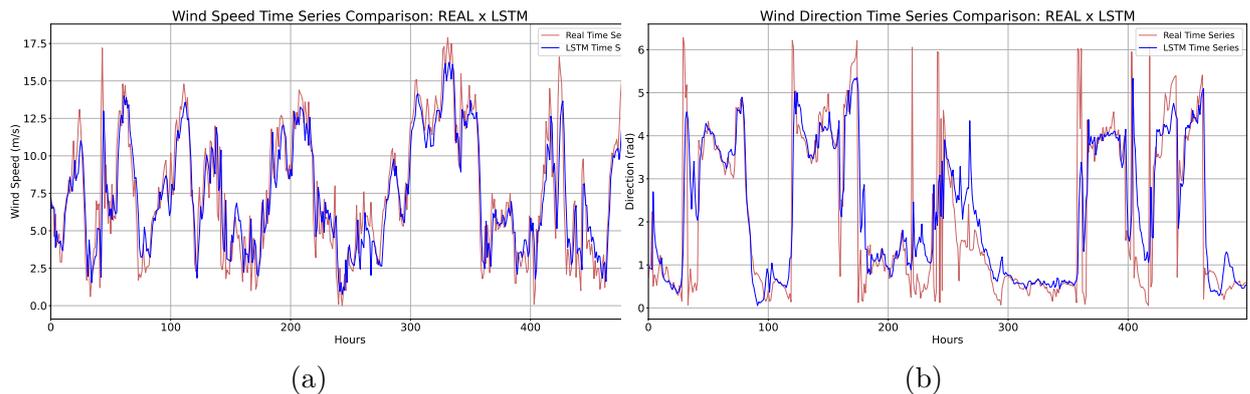


Figura 47 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.



vento foi melhor ajustada do que a direção do vento. Dentre os resultados para a velocidade do vento, destacou-se o modelo ARIMA, que conseguiu alcançar um MAPE de 31,25% e um R^2 de 0,65, descrevendo 65% da variância dessa curva. Seus resultados se mostraram superiores em todas as métricas de avaliação.

Para a direção do vento, os resultados nos mostram que, assim como para a velocidade do vento, as melhores métricas de avaliação foram obtidas pelos modelos CNN e ARIMA. Uma característica em comum dos resultados foi um alto valor de MAPE sendo o menor de 307,22%, associado ao modelo ARIMA. O melhor resultado geral, assim como para a métrica MAPE, foi proporcionado pelo modelo ARIMA com R^2 de 0,41, foi capaz de descrever 41% da variância dos dados.

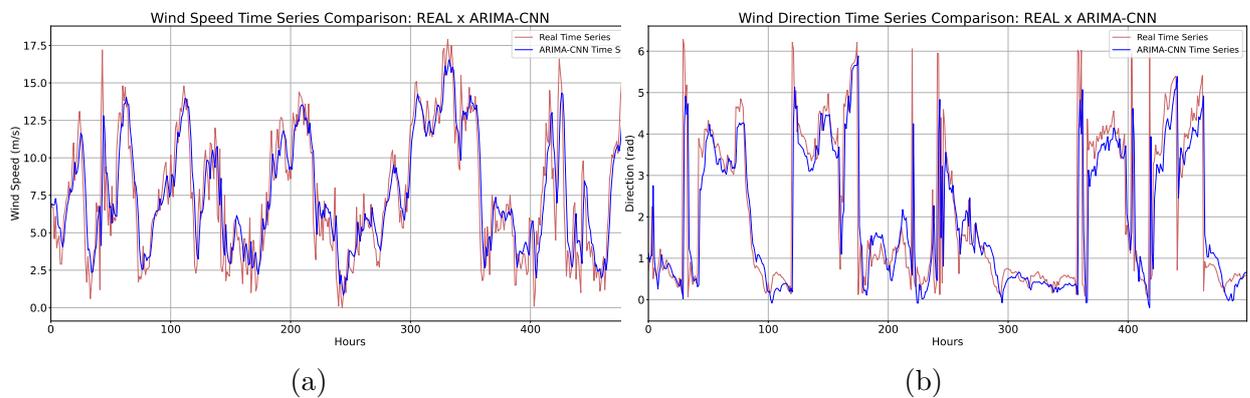
As Figuras 48, 49, 50 e 51 representam, respectivamente, a comparação das previsões geradas pelos modelos híbridos ARIMA-CNN, ARIMA-LSTM, CNN-LSTM e

Tabela 25 – Santa Catarina - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).

Métrica	CNN VEL	LSTM VEL	ARIMA VEL
MAPE	32,97	39,00	31,25
RMSE	2,01	2,64	1,75
MSE	4,02	6,96	3,06
MAE	0,38	1,68	0,14
R^2	0,54	0,20	0,65
Métrica	CNN DIR	LSTM DIR	ARIMA DIR
MAPE	311,99	400,39	307,22
RMSE	1,05	1,17	1,05
MSE	1,11	1,37	1,10
MAE	0,05	0,54	0,10
R^2	0,41	0,27	0,41

CNN-ALSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidos na Tabela 26.

Figura 48 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.



Os resultados apresentados na Tabela 26 mostram uma melhoria significativa nos modelos híbridos paralelos que combinam ARIMA com redes neurais, em comparação aos modelos individuais, demonstrando um desempenho superior em todas as métricas avaliadas. Por outro lado, os modelos CNN-ALSTM e CNN-LSTM exibiram, de maneira geral, resultados semelhantes aos dos modelos individuais, com métricas bastante próximas.

Para a previsão da direção do vento, os modelos ARIMA-LSTM e ARIMA-CNN apresentaram uma melhora expressiva nos resultados de R^2 sendo esses respectivamente de 0,59 e 0,58, onde o melhor desempenho foi obtido pelo modelo ARIMA-LSTM. Esses resultados sugerem que o modelo ARIMA-LSTM foi capaz de capturar 59% da variância da série temporal enquanto o melhor modelo individual, ARIMA, foi capaz de descrever

Figura 49 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.

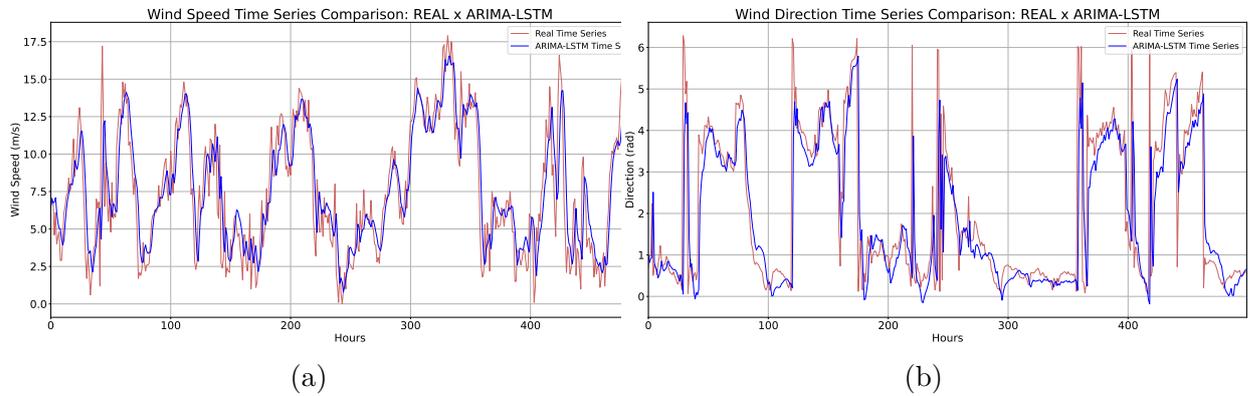


Figura 50 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.

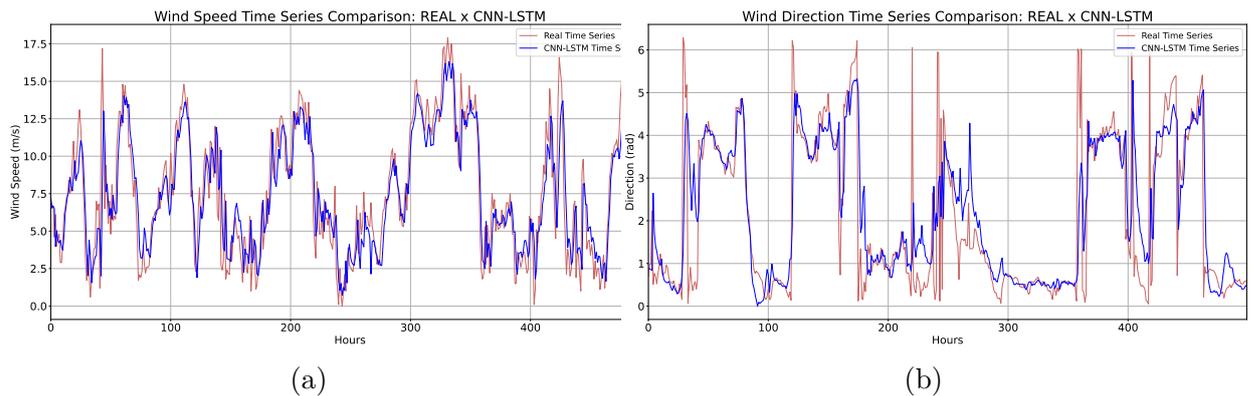


Figura 51 – Santa Catarina - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.

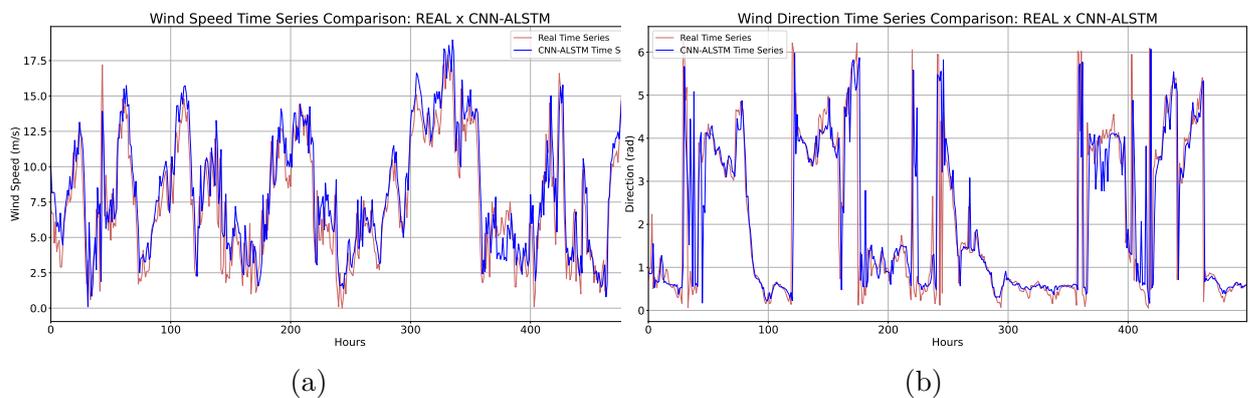


Tabela 26 – Santa Catarina - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).

Métrica	CNN- ALSTM VEL	ARIMA- LSTM VEL	CNN- LSTM VEL	ARIMA- CNN VEL
MAPE	35,49	24,06	38,46	24,56
RMSE	2,18	1,41	2,58	1,42
MSE	4,75	2,00	6,68	2,02
MAE	0,01	0,17	1,61	0,27
R^2	0,46	0,77	0,24	0,767
Métrica	CNN- ALSTM DIR	ARIMA- LSTM DIR	CNN- LSTM DIR	ARIMA- CNN DIR
MAPE	365,97	251,97	392,03	269,74
RMSE	1,28	0,87	1,15	0,89
MSE	1,63	0,76	1,32	0,80
MAE	0,39	0,08	0,49	0,02
R^2	0,13	0,59	0,30	0,58

Tabela 27 – Santa Catarina - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM

Modelo	Tipo	β_2	β_1	β_0
ARIMA-LSTM	Dir	-0,21	1,35	-0,44
ARIMA-LSTM	Vel	0,26	0,81	-0,35
ARIMA-CNN	Dir	-0,33	1,45	-0,38
ARIMA-CNN	Vel	0,05	1,02	-0,35
CNN-LSTM	Dir	0,99	0,01	-0,06
CNN-LSTM	Vel	0,98	0,02	0,01

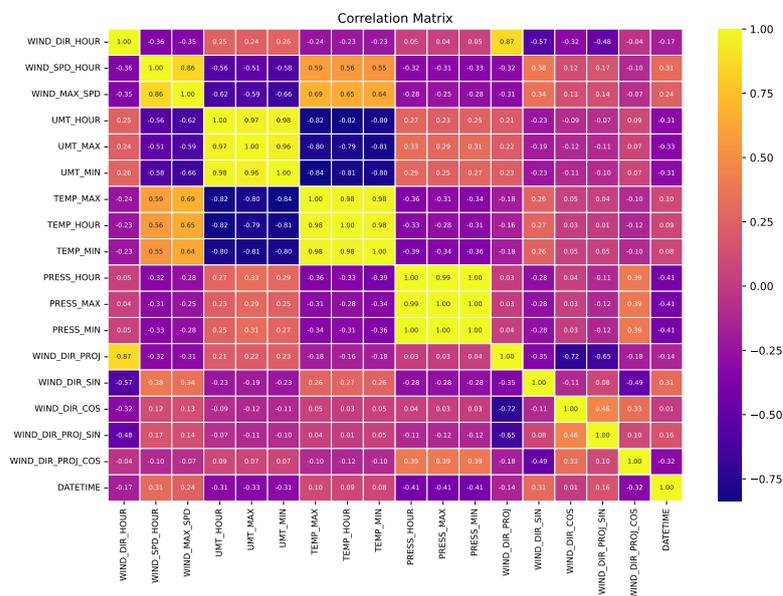
41% . Com relação aos parâmetros MAPE, RMSE e MSE, o modelo ARIMA-LSTM também apresentou melhoras significativas diminuindo o MAPE de 307,22% para 251,97%, o RMSE de 1,05 para 0,87, o MSE de 1,10 para 0,76 e MAE de 0,10 para 0,08 quando comparado com o ARIMA.

Para a velocidade do vento, a melhora dos resultados obtidos pelos modelos híbridos se mantém. Nesse caso, a performance dos modelos ARIMA-CNN e ARIMA-LSTM mostrou melhorias em todas as métricas de avaliação, com destaque para o coeficiente de determinação $R^2 = 0,77$ alcançado pelo ARIMA-LSTM, que foi capaz de explicar 77% da variância dos dados. Para o estado de Santa Catarina, o modelo que apresentou o melhor desempenho geral nas métricas avaliativas foi o ARIMA-LSTM, performando melhor tanto para a direção quanto para a velocidade do vento. Os valores das constantes estimadas para as regressões lineares multivariadas, utilizadas na construção dos modelos híbridos em paralelo, estão apresentados na Tabela 27.

4.1.5 Ceará

A matriz de correções para o estado do Ceará disposta no gráfico de calor na Figura 52 nos mostra que, para a velocidade dos ventos, existem sete correlações relevantes, sendo as quatro positivas com velocidade máxima do vento na hora anterior, temperatura máxima na hora anterior, temperatura mínima na hora anterior e temperatura horária, e as três negativas com a umidade máxima na hora anterior, umidade mínima na hora anterior e com a umidade horária. Para a direção do vento, observa-se que as únicas correlações relevantes são com a projeção da direção do vento e com o seno da direção do vento, apresentando correlações positivas e negativas, respectivamente.

Figura 52 – Ceará - Gráfico de correlação dos dados climáticos referentes ao estado do Ceará.



O coeficiente Kaiser-Meyer-Olkin (KMO) calculado para os dados de teste de 0,75 indica que os dados são altamente correlacionados e, portanto, propensos a possuírem variáveis latentes bem descritivas. Essa suposição é confirmada pelo teste de esfericidade de Bartlett que confirma a hipótese de que o grupo dos nossos dados é altamente correlacionado. Esses resultados indicam fortemente a possibilidade de o conjunto possuir variáveis latentes que descrevem grande parte da variância dos dados, o que justifica a realização de uma análise de fatores. Os valores encontrados para o teste e o coeficiente ficam descritos na Tabela 28.

Tabela 28 – Ceará - Resultados de KMO e Teste de Esfericidade de Bartlett

KMO	Teste de Esfericidade de Bartlett p-value
0,75	0,00

Os resultados da análise de fatores ilustrados na Tabela 29, nos mostram a existência de sete variáveis latentes que juntas descrevem 87% das variâncias desse conjunto de dados.

Esse resultado, somado com a matriz de correlação nos sugerem a capacidade das variáveis utilizadas para o treinamento dos modelos preditivos de contribuir para o aumento de precisão dos ajustes sobre cada uma das variáveis.

Tabela 29 – Ceará - Variância Explicada por Fator

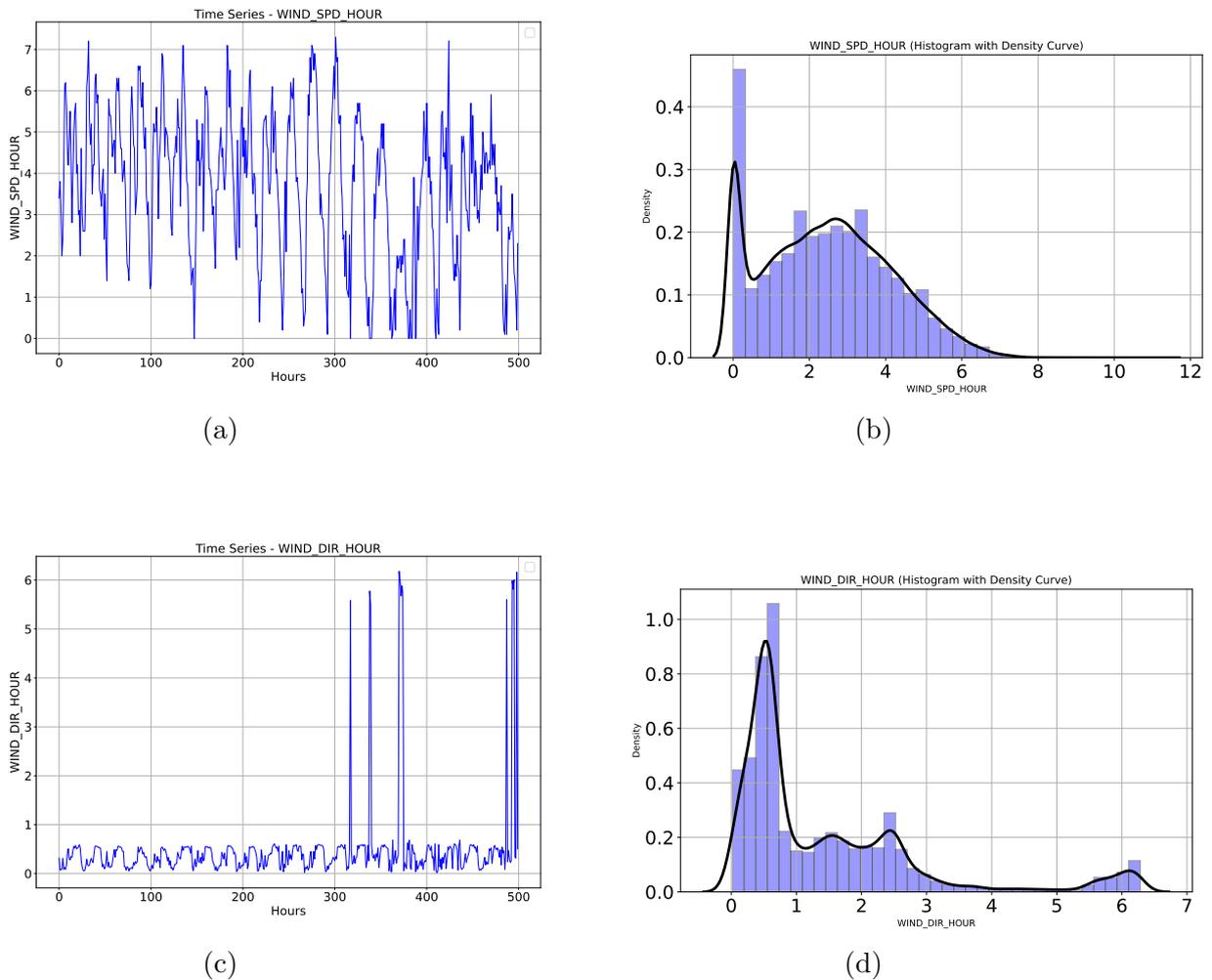
Fator	Variância Proporcional	Variância Acumulada
Fator 1	0,31	0,31
Fator 2	0,18	0,49
Fator 3	0,15	0,64
Fator 4	0,08	0,72
Fator 5	0,08	0,80
Fator 6	0,05	0,84
Fator 7	0,03	0,87

A Figura 53 apresenta as observações de velocidade e direção do vento nas últimas 500 horas e seus respectivos histogramas para todo o conjunto de dados, que inclui 96.432 observações utilizadas do estado do Ceará, conforme ilustrado na Tabela 30.

Como pode ser visto nas Figuras 53(a) e 53(c), a velocidade e a direção do vento se comportam de forma diferente. No entanto, elas têm o mesmo atributo de estarem centradas em torno da média de 2,45 e 1,47, respectivamente, com uma variação aproximadamente constante ao longo do tempo. O teste Dickey-Fuller apoia essas suposições de estacionariedade, pois rejeita a hipótese nula de uma raiz unitária na série temporal, que é o que indica a estacionariedade. Para verificar a não normalidade na série temporal de velocidade e direção do vento, testes de hipóteses como Jarque-Bera, teste de assimetria e teste de curtose foram usados. Com base no valor de p para cada teste, levando em consideração um nível de significância de 0,05, esses testes contrastam a hipótese nula de que as séries são distribuídas normalmente. A Tabela 30 resume todos os testes usados e as estatísticas computadas. Os histogramas associados à velocidade e direção do vento são mostrados nas Figuras 53(b) e 53(d) com base em todos os registros.

Na Tabela 30, um valor de assimetria positivo de 0,27 confirma que o histograma da velocidade do vento mostra uma assimetria clara à direita que indica que velocidades mais altas do vento são menos frequentes. Isso também ocorre para a direção do vento e pode ser confirmada por meio de um valor de assimetria de 1,79. Uma distribuição de dados ligeiramente mais plana do que uma distribuição normal pode ser observada com uma curtose de -0,64 para velocidade do vento. As caudas mais leves da distribuição e a menor concentração de valores em torno da média em comparação com uma distribuição normal são o resultado disso. Para a direção do vento, a distribuição de dados é marginalmente mais pontiaguda do que uma distribuição normal conforme sugerido pelo valor de curtose de 3,01. Esse resultado implica que há caudas mais pesadas e uma densidade maior de valores próximos à média na distribuição do que em uma distribuição normal.

Figura 53 – Ceará - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.



Na Figura 54, é apresentada a decomposição das séries temporais de velocidade e direção do vento, com referência a um período de 24 horas. A partir da imagem, observa-se uma forte tendência sazonal em ambas as curvas, sendo essa sazonalidade mais pronunciada na série temporal da direção do vento. Além disso, a magnitude da componente de erro nas séries temporais de ambas as curvas é bastante alta, quase equivalente à componente de tendência.

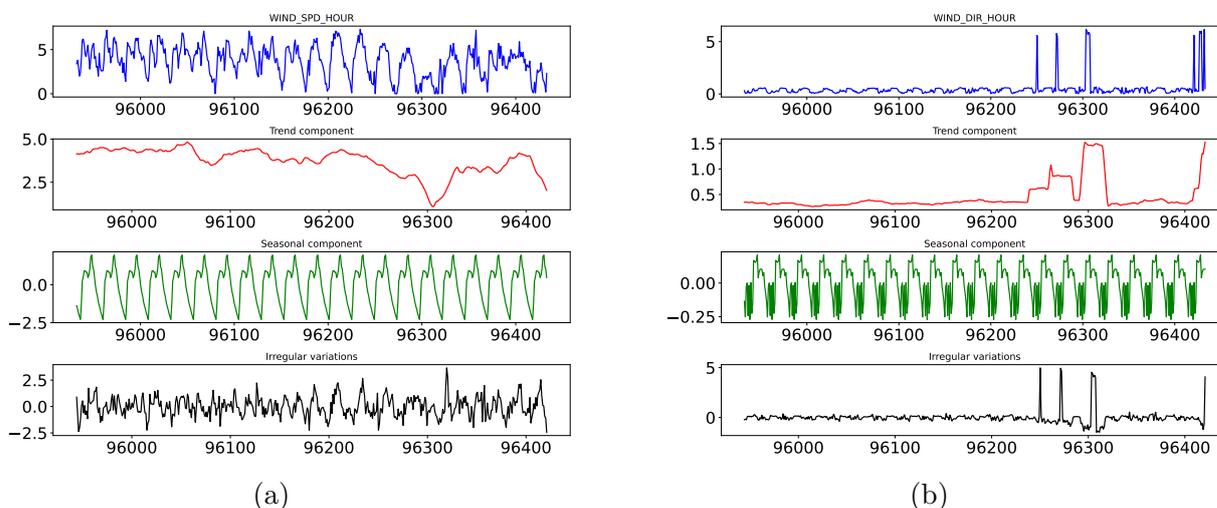
As Figuras 55, 56, 57 mostram, respectivamente, a comparação das previsões geradas pelos modelos ARIMA, CNN e LSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidas na Tabela 31.

Os resultados apresentados na Tabela 7 nos mostram que, para a direção do vento, todos os modelos tiveram uma baixa adesão com R^2 máximo de 0,27 para o modelo

Tabela 30 – Ceará - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionaridade.

	WIND_HOUR_SPD	WIND_DIR_HOUR
N	96432	96432
μ	2,45	1,47
σ	1,67	1,44
\mathbf{x}_{\min}	0,00	0,02
\mathbf{x}_{\max}	11,20	6,28
Q_1	1,10	0,51
Q_2	2,40	1,00
Q_3	3,60	2,06
kurtosis	-0,64	3,01
kurtosistest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
skewness	0,27	1,79
skewnesstest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
jarque_bera	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
adfuller	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$

Figura 54 – Ceará - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.



LSTM. Tal resultado implica que apenas 27% da variância da série temporal foram de fato capturadas por esse modelo. Além disso, observa-se que, como padrão, todos os modelos apresentaram valores baixos de MAE e valores elevados de MAPE, todos acima de 187,08%, resultado alcançado pelo modelo ARIMA.

Para a velocidade do vento, os resultados nos mostram uma capacidade de ajuste muito melhor por parte de todos os modelos. O melhor dentre eles, novamente foi o LSTM que, com R^2 de 0,69, foi capaz de descrever 69% da variância dos dados. Com exceção do MAE, melhor performado pelo modelo CNN com valor de 0,13, o modelo LSTM apresentou

Figura 55 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.

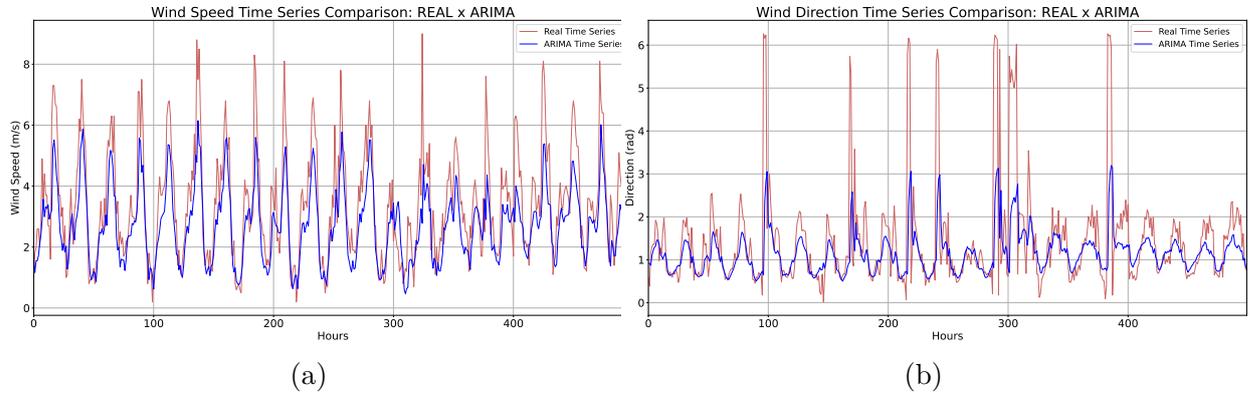


Figura 56 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.

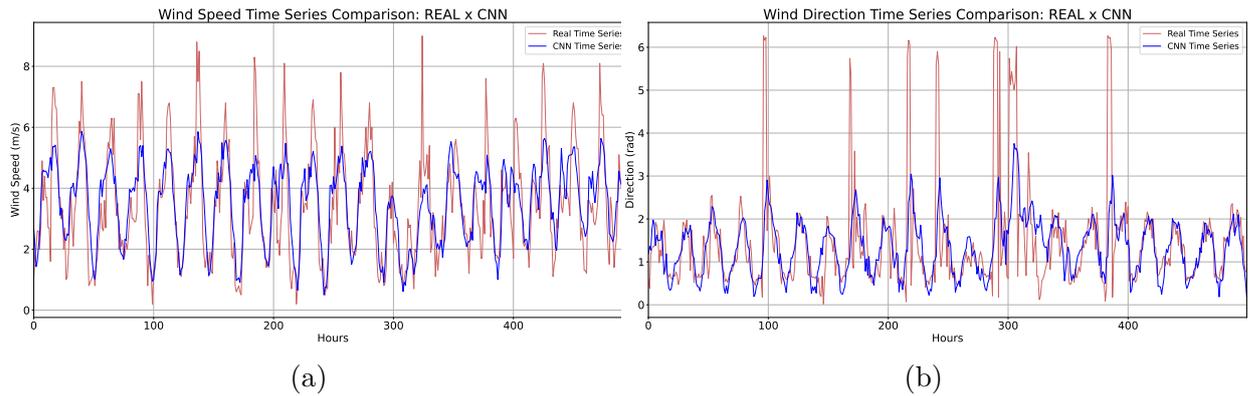
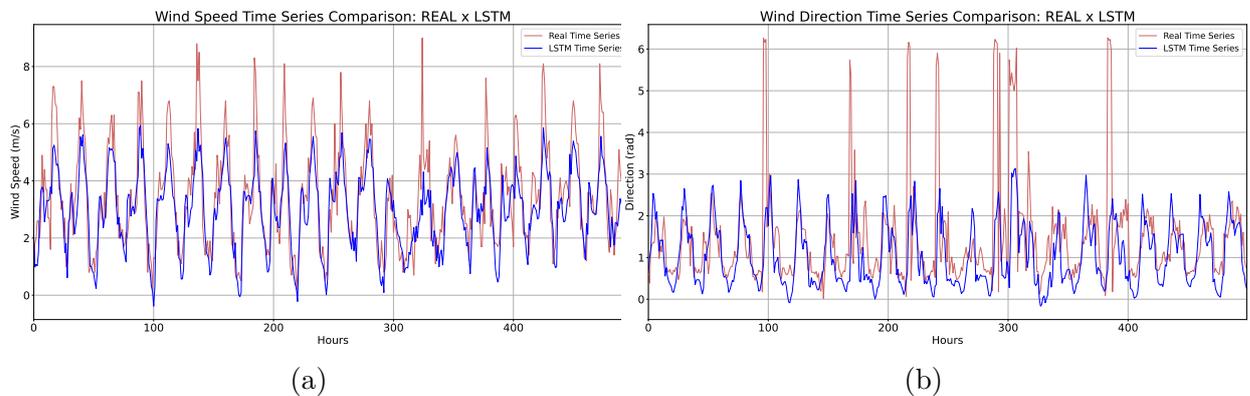


Figura 57 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.



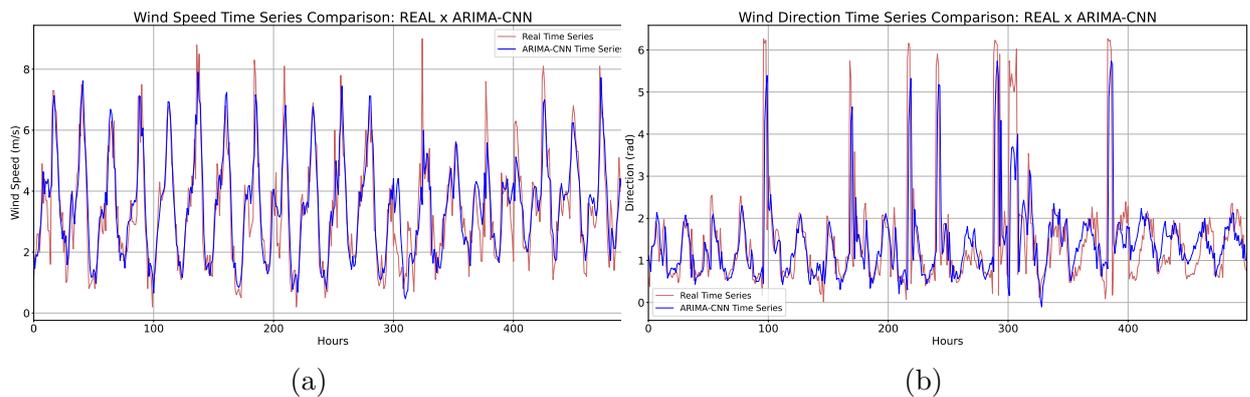
o melhor desempenho em todas as outras métricas de avaliação.

Tabela 31 – Ceará -Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).

Métrica	CNN VEL	LSTM VEL	ARIMA VEL
MAPE	84,64	59,81	52,91
RMSE	1,07	0,95	1,05
MSE	1,14	0,90	1,10
MAE	0,13	0,15	0,54
R^2	0,61	0,69	0,63
Métrica	CNN DIR	LSTM DIR	ARIMA DIR
MAPE	236,334	198,65	187,09
RMSE	1,59	1,58	1,60
MSE	2,53	2,51	2,57
MAE	0,02	0,16	0,34
R^2	0,26	0,27	0,25

As Figuras 58, 59, 60 e 61 representam, respectivamente, a comparação das previsões geradas pelos modelos híbridos ARIMA-CNN, ARIMA-LSTM, CNN-LSTM e CNN-ALSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidos na Tabela 32.

Figura 58 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.



Os resultados apresentados na Tabela 32 indicam uma melhora significativa nos modelos híbridos paralelos que combinam ARIMA com RNAs, em comparação aos modelos individuais, com desempenho superior em quase todas as métricas avaliadas. Em contrapartida, os modelos CNN-ALSTM e CNN-LSTM apresentaram, de forma geral, resultados semelhantes aos modelos individuais, com valores próximos em todas as métricas.

Para a previsão da direção do vento, os modelos ARIMA-LSTM e ARIMA-CNN apresentaram uma melhora expressiva nos resultados de R^2 sendo esses respectivamente

Figura 59 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.

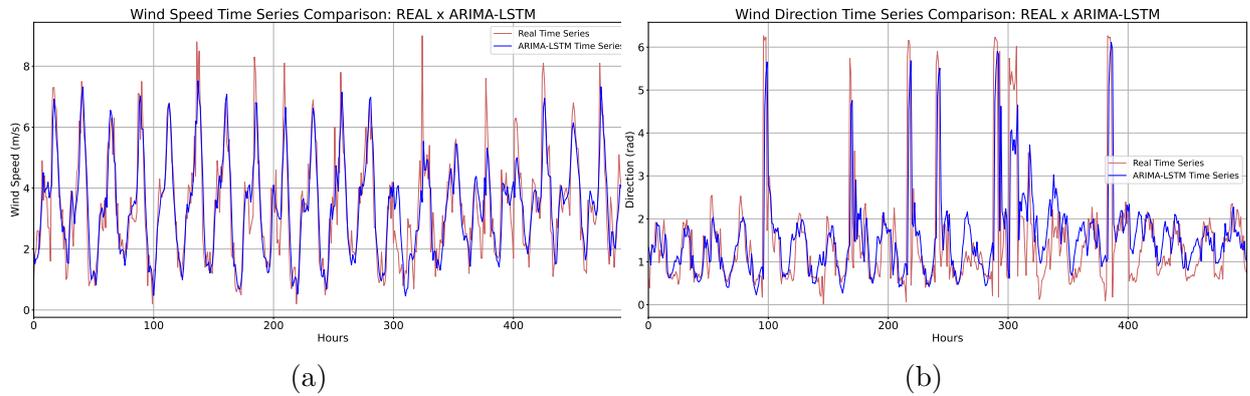


Figura 60 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.

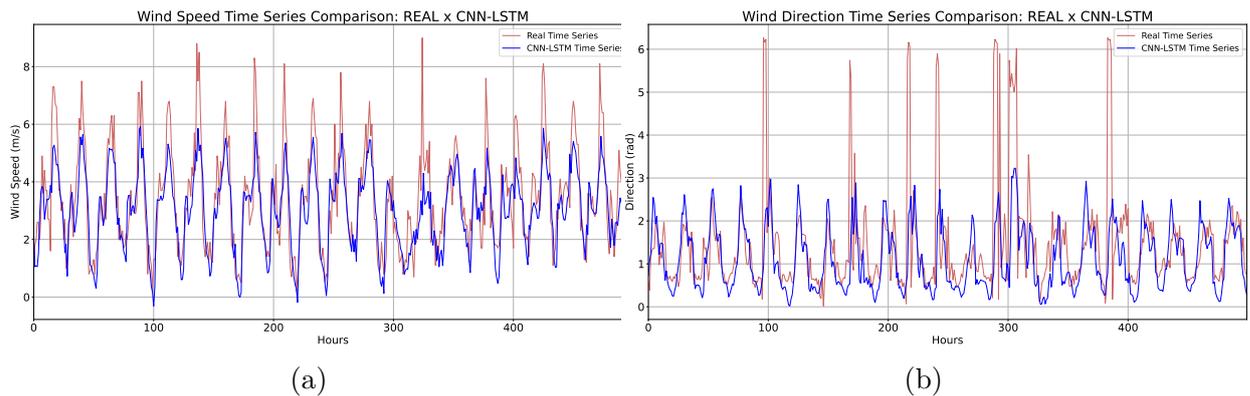
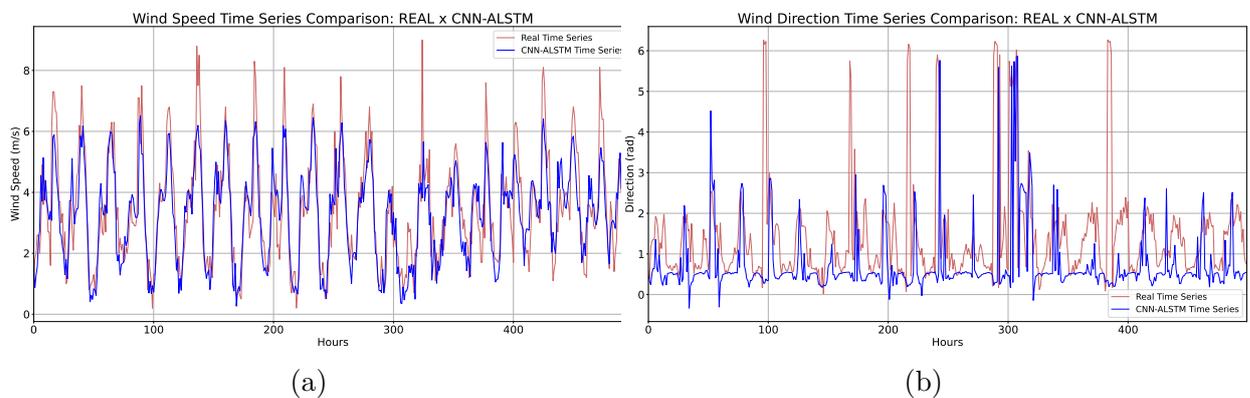


Figura 61 – Ceará - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.



de 0,82 e 0,84, onde o melhor desempenho foi obtido pelo modelo ARIMA-CNN. Esses resultados sugerem que o modelo ARIMA-CNN foi capaz de capturar 84% da variância da série temporal enquanto o melhor modelo individual, LSTM, descreveu 27%. Com relação aos parâmetros MAPE, RMSE e MSE, o modelo ARIMA-CNN também apresentou melhoras significativas diminuindo o MAPE de 198,65% para 132,92%, o RMSE de 1,58 para 0,74, o MSE de 2,51 para 0,55 e o MAE de 0,16 para 0,04 quando comparado com o LSTM.

Tabela 32 – Ceará - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).

Métrica	CNN- ALSTM VEL	ARIMA- LSTM VEL	CNN- LSTM VEL	ARIMA- CNN VEL
MAPE	55,74	40,52	59,79	42,12
RMSE	0,95	0,65	0,94	0,65
MSE	0,90	0,42	0,89	0,42
MAE	0,09	0,08	0,14	0,04
R^2	0,69	0,86	0,70	0,86
Métrica	CNN- ALSTM DIR	ARIMA- LSTM DIR	CNN- LSTM DIR	ARIMA- CNN DIR
MAPE	133,16	144,64	207,63	132,92
RMSE	1,88	0,79	1,57	0,74
MSE	3,53	0,63	2,47	0,55
MAE	0,42	0,16	0,11	0,04
R^2	-0,04	0,81	0,28	0,84

Tabela 33 – Ceará - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM

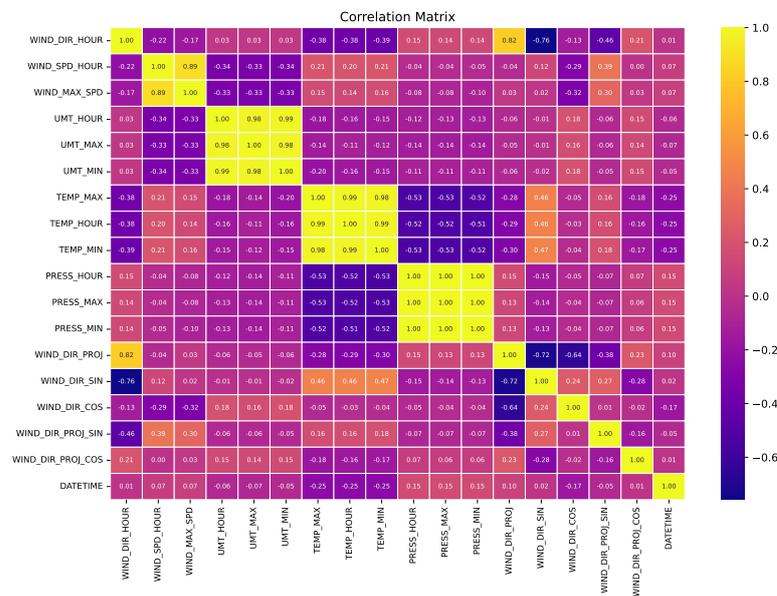
Modelo	Tipo	β_2	β_1	β_0
ARIMA-LSTM	Dir	-0,47	2,43	-0,72
ARIMA-LSTM	Vel	0,22	1,09	-0,23
ARIMA-CNN	Dir	-0,57	2,41	-0,57
ARIMA-CNN	Vel	0,07	1,26	-0,19
CNN-LSTM	Dir	0,93	0,10	0,04
CNN-LSTM	Vel	0,94	0,09	-0,02

Para a velocidade do vento, a melhora dos resultados obtidos pelos modelos híbridos se mantém. Nesse caso, a performance dos modelos ARIMA-CNN e ARIMA-LSTM mostrou melhorias em todas as métricas de avaliação, com destaque para o coeficiente de determinação $R^2 = 0,86$ alcançado pelo ARIMA-LSTM, que foi capaz de explicar 86% da variância dos dados. Para o conjunto de dados do estado do Ceará, o modelo que apresentou o melhor desempenho geral nas métricas de avaliação foi o ARIMA-CNN. Os valores das constantes estimadas para as regressões lineares multivariadas, utilizadas na construção dos modelos híbridos em paralelo, estão apresentados na Tabela 33.

4.1.6 Rio Grande do Sul

A matriz de correções para o estado do Rio Grande do Sul disposta no gráfico de calor na Figura 62 nos mostra que, para a velocidade dos ventos, existe apenas uma correlação relevante, sendo esta positiva forte, com velocidade máxima do vento na hora anterior. Para a direção do vento, observa-se que suas únicas correlações fortes são com a projeção da direção do vento e com o seno da direção do vento, sendo estas correlações positiva e negativa, respectivamente.

Figura 62 – Rio Grande do Sul - Gráfico de correlação dos dados climáticos referentes ao estado do Rio Grande do Sul.



O coeficiente Kaiser-Meyer-Olkin (KMO) calculado para os dados de teste de 0,71 indica que os dados são altamente correlacionados e, portanto, propensos a possuírem variáveis latentes bem descritivas. Essa suposição é confirmada pelo teste de esfericidade de Bartlett que confirma a hipótese de que o grupo dos nossos dados é altamente correlacionado. Esses resultados indicam fortemente a possibilidade de o conjunto possuir variáveis latentes que descrevem grande parte da variância dos dados, o que justifica a realização de uma análise de fatores. Os valores encontrados para o teste e o coeficiente ficam descritos na Tabela 34.

Tabela 34 – Resultados de KMO e Teste de Esfericidade de Bartlett

KMO	Teste de Esfericidade de Bartlett p-value
0,71	0,00

Os resultados da análise de fatores ilustrados na Tabela 35, nos mostram a existência de sete variáveis latentes que juntas descrevem 84% das variâncias desse conjunto de dados. Esses resultados, em conjunto com a matriz de correlação nos confirmam a alta correlação com os dados e sugerem a capacidade das variáveis utilizadas para o treinamento dos

modelos preditivos de contribuir para o aumento de precisão dos ajustes sobre cada uma das variáveis.

Tabela 35 – Rio Grande do Sul - Variância Explicada por Fator

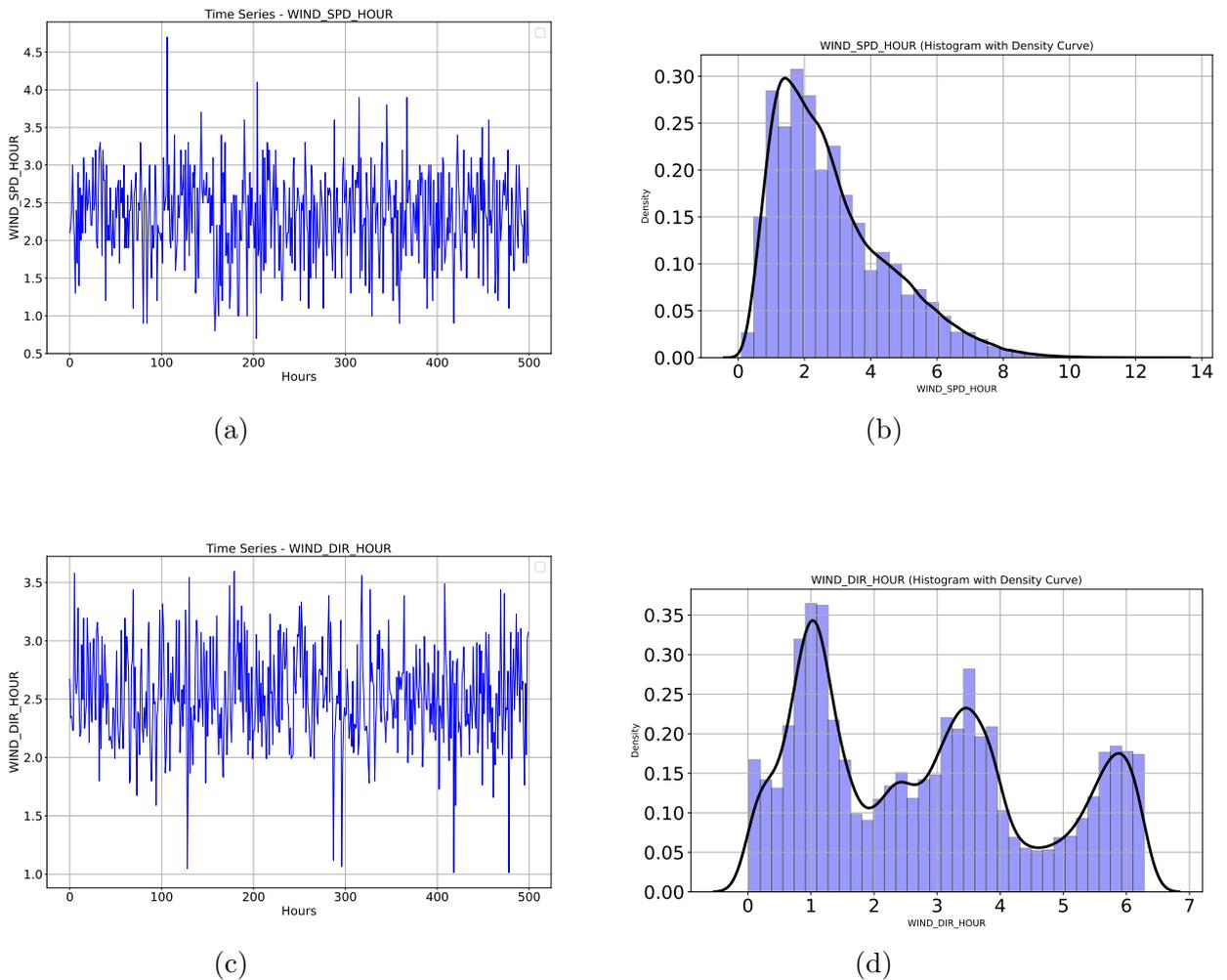
Fator	Variância Proporcional	Variância Acumulada
Fator 1	0,17	0,17
Fator 2	0,17	0,34
Fator 3	0,16	0,50
Fator 4	0,15	0,65
Fator 5	0,11	0,76
Fator 6	0,07	0,82
Fator 7	0,02	0,84

A Figura 63 apresenta as observações de velocidade e direção do vento nas últimas 500 horas e seus respectivos histogramas para todo o conjunto de dados, que inclui 96.432 observações utilizadas do estado do Rio Grande do Sul, conforme ilustrado na Tabela 36.

Como pode ser visto nas Figuras 63(a) e 63(c), a velocidade e a direção do vento se comportam de forma diferente. No entanto, elas têm o mesmo atributo de estarem centradas em torno da média de 2,83 e 2,81, respectivamente, com uma variação aproximadamente constante ao longo do tempo. O teste Dickey-Fuller apoia essas suposições de estacionariedade, pois rejeita a hipótese nula de uma raiz unitária na série temporal, que é o que indica a estacionariedade. Para verificar a não normalidade na série temporal de velocidade e direção do vento, testes de hipóteses como Jarque-Bera, teste de assimetria e teste de curtose foram usados. Com base no valor de p para cada teste, levando em consideração um nível de significância de 0,05, esses testes contrastam a hipótese nula de que as séries são distribuídas normalmente. A Tabela 36 resume todos os testes usados e as estatísticas computadas. Os histogramas associados à velocidade e direção do vento são mostrados nas Figuras 63(b) e 63(d) com base em todos os registros.

Na Tabela 36, um valor de assimetria positivo de 1,03 confirma que o histograma da velocidade do vento mostra uma assimetria clara à direita que indica que velocidades mais altas do vento são menos frequentes. Isso também ocorre para a direção do vento, mostrando que a maioria dos valores de direção do vento se encontram abaixo da média das observações. Essa assimetria à direita pode ser confirmada por meio de um valor de assimetria de 0,36. Uma distribuição de dados ligeiramente mais plana do que uma distribuição normal pode ser observada com uma curtose de -1,08 para direção do vento. As caudas mais leves da distribuição e a menor concentração de valores em torno da média em comparação com uma distribuição normal são o resultado disso. Para a velocidade do vento, a distribuição de dados é marginalmente mais pontiaguda do que uma distribuição normal conforme sugerido pelo valor de curtose de 0,79. Esse resultado implica que há caudas mais pesadas e uma densidade maior de valores próximos à média na distribuição do que em uma distribuição normal.

Figura 63 – Rio Grande do Sul - As Figuras (a) e (c) mostram séries temporais associadas à velocidade e direção do vento, respectivamente, considerando as últimas 500 observações horárias. As Figuras (b) e (d) mostram histogramas para velocidade e direção do vento, respectivamente, considerando o número total de observações horárias.



Na Figura 64 está representada a decomposição das séries temporais de velocidade e direção do vento usando de referência um período de 24 horas. Com base nessa imagem, observa-se uma alta tendência sazonal em ambas as curvas. Além disso, observa-se que a magnitude da componente de erro da série temporal de ambas as curvas é bem alta, se aproximando da metade da dimensão da curva de tendência.

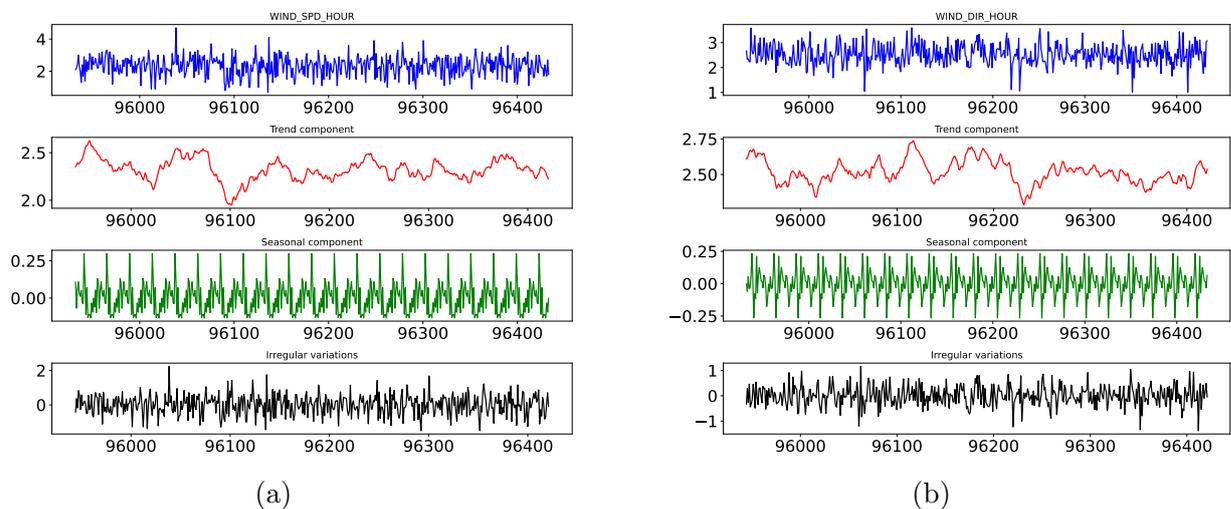
As Figuras 65, 66, 67 mostram, respectivamente, a comparação das previsões geradas pelos modelos ARIMA, CNN e LSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidos na Tabela 37.

Os resultados apresentados na Tabela 37 mostram que, para a direção do vento, todos os modelos conseguiram explicar cerca de 40% da variância da série temporal, com o modelo LSTM apresentando o R^2 máximo de 0,42. Isso implica que o modelo foi capaz

Tabela 36 – Rio Grande do Sul - Estatísticas principais para a velocidade e direção do vento. Testes de hipótese associados à normalidade, assimetria, curtose e estacionariedade.

	WIND_HOUR_SPD	WIND_DIR_HOUR
N	96432	96432
μ	2,83	2,81
σ	1,72	1,83
x_{\min}	0,10	0,02
x_{\max}	13,10	6,28
Q_1	1,50	1,12
Q_2	2,40	2,71
Q_3	3,80	3,94
kurtosis	0,79	-1,08
kurtosistest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
skewness	1,03	0,36
skewnesstest	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
jarque_bera	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$
adfuller	$p\text{-value} = 0,00$	$p\text{-value} = 0,00$

Figura 64 – Rio Grande do Sul - As Figuras (a) e (b) exibem a decomposição aditiva de séries temporais vinculadas à velocidade e direção do vento, respectivamente. Essas Figuras mostram componentes de tendência, sazonais e irregulares, com foco nas 500 observações horárias mais recentes.



de capturar 42% da variância da série temporal. Além disso, observa-se que, como padrão, todos os modelos apresentaram valores baixos de MAE e valores altos de MAPE, todos superiores a 199%.

Para a direção do vento, os resultados nos mostram uma capacidade superior de ajuste por parte de todos os modelos. O melhor dentre eles, novamente foi o LSTM que, com R^2 de 0,70, foi capaz de descrever 70% da variância dos dados. Com exceção do MAPE e do MAE, melhor performado pelo modelo ARIMA com valores de 37,48% e 0,01

Figura 65 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA em comparação ao conjunto de teste, respectivamente.

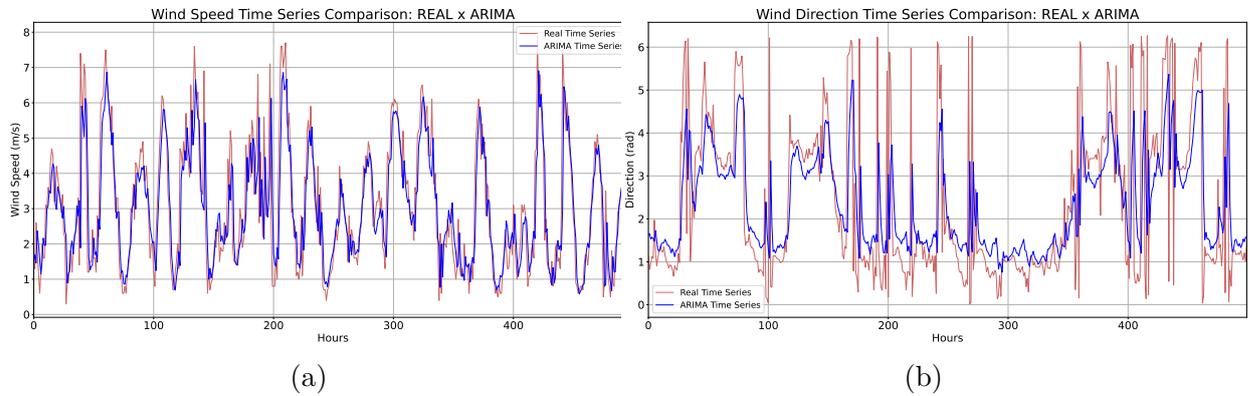


Figura 66 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN em comparação ao conjunto de teste, respectivamente.

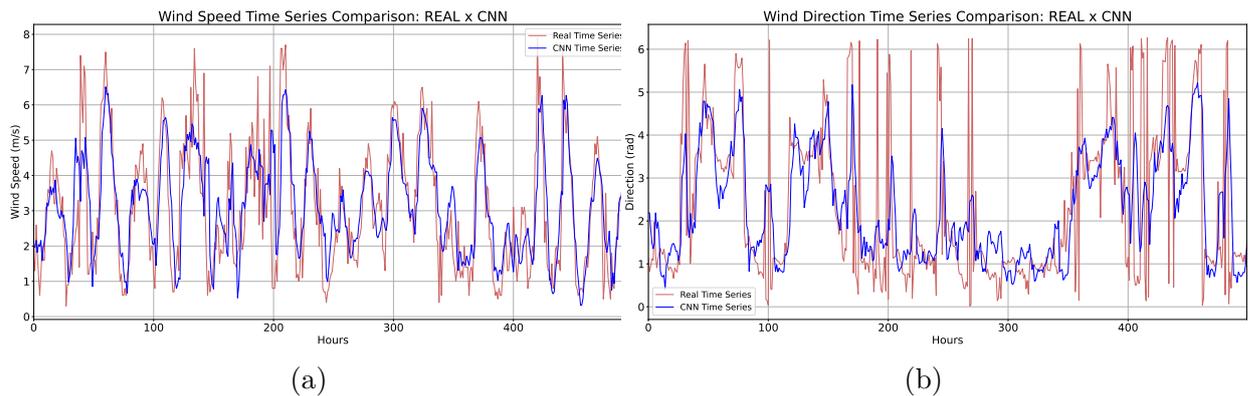
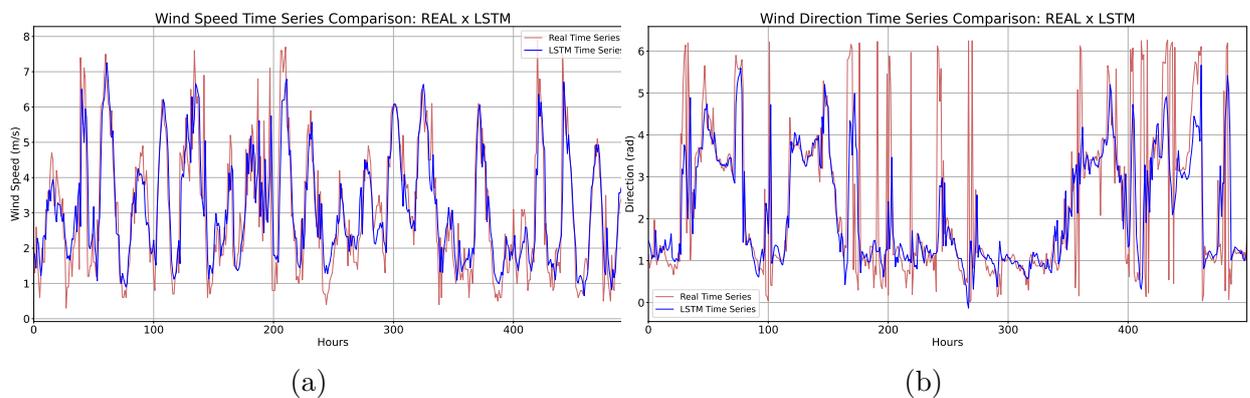


Figura 67 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo LSTM em comparação ao conjunto de teste, respectivamente.



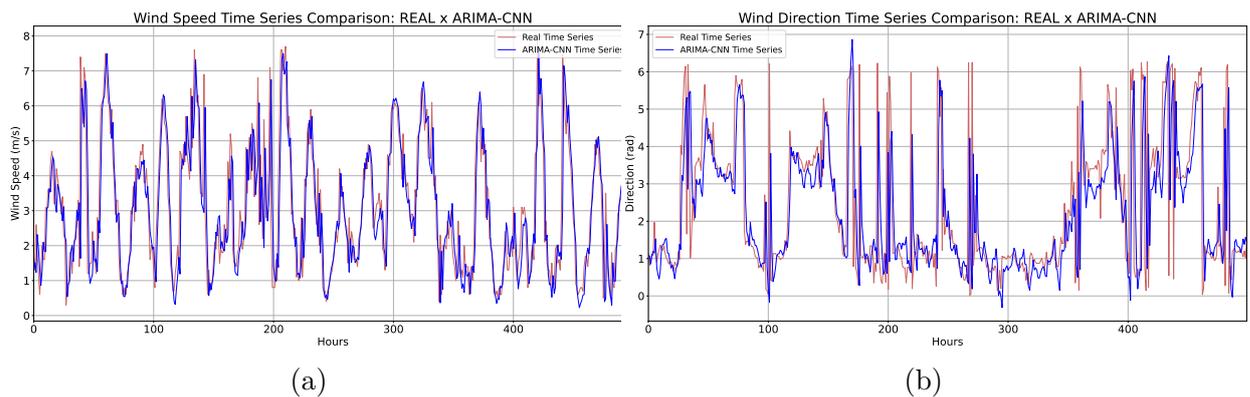
respectivamente, o modelo LSTM apresentou o melhor desempenho em todas as outras métricas de avaliação.

Tabela 37 – Rio Grande do Sul - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).

Métrica	CNN VEL	LSTM VEL	ARIMA VEL
MAPE	49,45	39,01	37,48
RMSE	1,05	0,87	0,89
MSE	1,10	0,76	0,80
MAE	0,13	0,03	0,01
R^2	0,56	0,70	0,68
Métrica	CNN DIR	LSTM DIR	ARIMA DIR
MAPE	219,89	199,39	212,18
RMSE	1,41	1,41	1,43
MSE	2,00	1,98	2,05
MAE	0,02	0,01	0,03
R^2	0,41	0,42	0,40

As Figuras 68, 69, 70 e 71 representam, respectivamente, a comparação das previsões geradas pelos modelos híbridos ARIMA-CNN, ARIMA-LSTM, CNN-LSTM e CNN-ALSTM com os dados de teste. As métricas de avaliação para esses três modelos são exibidos na Tabela 38.

Figura 68 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-CNN em comparação ao conjunto de teste, respectivamente.



Os resultados apresentados na Tabela 38 indicam uma melhora significativa nos modelos híbridos paralelos que combinam ARIMA com RNAs, em comparação aos modelos individuais, com desempenho superior em quase todas as métricas avaliadas. Em contrapartida, os modelos CNN-ALSTM e CNN-LSTM apresentaram, de forma geral, resultados semelhantes aos modelos individuais, com valores próximos em todas as métricas.

Figura 69 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo ARIMA-LSTM em comparação ao conjunto de teste, respectivamente.

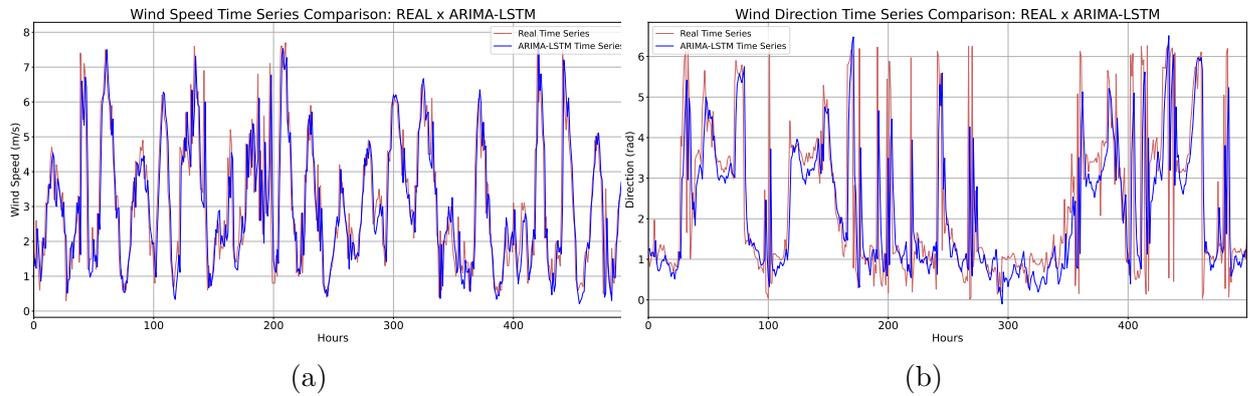


Figura 70 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-LSTM em comparação ao conjunto de teste, respectivamente.

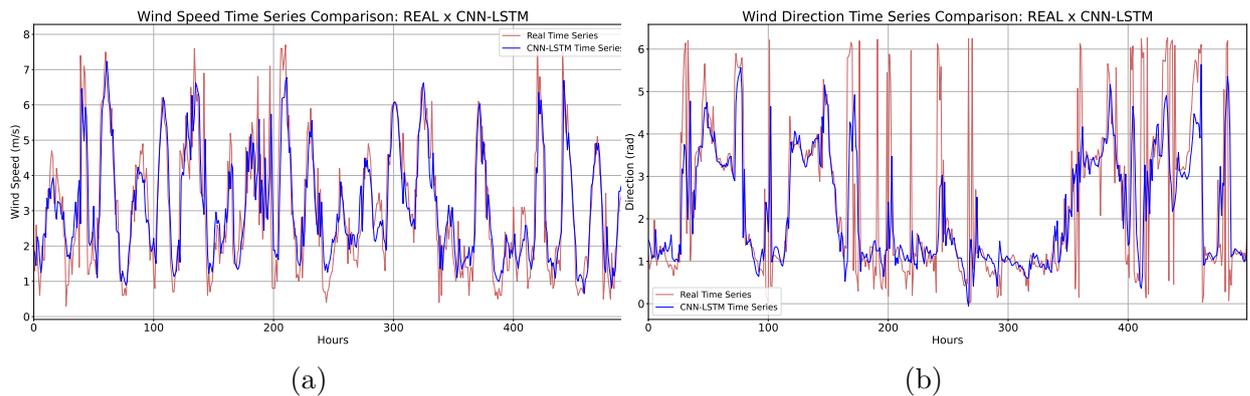


Figura 71 – Rio Grande do Sul - As Figuras (a) e (b) exibem a previsão da velocidade e da direção do vento com o modelo CNN-ALSTM em comparação ao conjunto de teste, respectivamente.

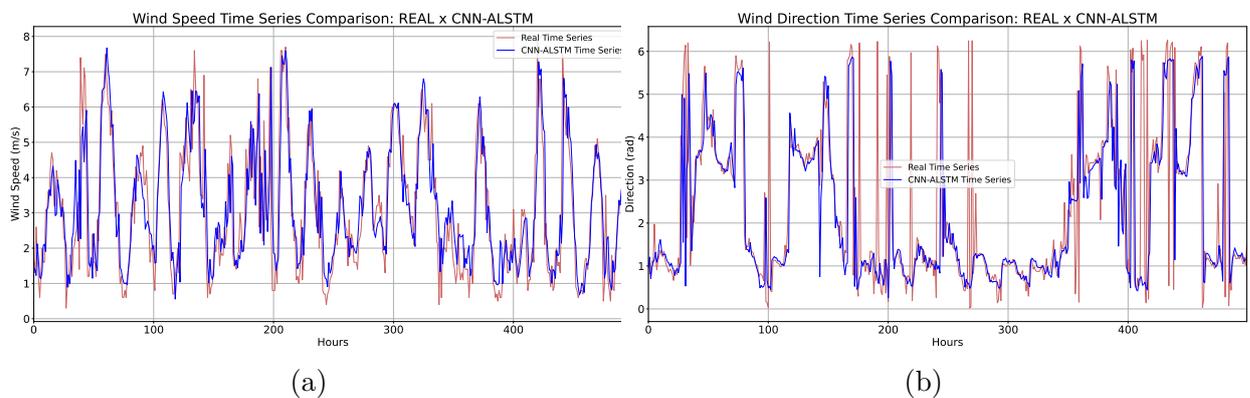


Tabela 38 – Rio Grande do Sul - Métricas de desempenho para Modelos Híbridos (Velocidade e Direção do Vento).

Métrica	CNN-ALSTM VEL	ARIMA-LSTM VEL	CNN-LSTM VEL	ARIMA-CNN VEL
MAPE	38,79	13,57	38,91	13,40
RMSE	0,89	0,33	0,87	0,33
MSE	0,78	0,11	0,76	0,11
MAE	0,07	0,03	0,03	0,04
R^2	0,69	0,96	0,70	0,96
Métrica	CNN-ALSTM DIR	ARIMA-LSTM DIR	CNN-LSTM DIR	ARIMA-CNN DIR
MAPE	173,36	98,79	199,60	85,75
RMSE	1,60	0,75	1,40	0,71
MSE	2,55	0,56	1,96	0,51
MAE	0,07	0,06	0,01	0,07
R^2	0,25	0,83	0,42	0,85

Tabela 39 – Rio Grande do Sul - Parâmetros dos Modelos ARIMA-LSTM, ARIMA-CNN e CNN-LSTM

Modelo	Tipo	β_2	β_1	β_0
ARIMA-LSTM	Dir	-0,21	1,58	-1,05
ARIMA-LSTM	Vel	-0,06	1,22	-0,44
ARIMA-CNN	Dir	-0,46	1,75	-0,82
ARIMA-CNN	Vel	-0,05	1,20	-0,42
CNN-LSTM	Dir	0,96	0,05	-0,01
CNN-LSTM	Vel	0,97	0,03	-0,01

Para a previsão da direção do vento, os modelos ARIMA-LSTM e ARIMA-CNN apresentaram uma melhora expressiva nos resultados de R^2 sendo esses respectivamente de 0,83 e 0,85, onde o melhor desempenho foi obtido pelo modelo ARIMA-CNN. Esses resultados sugerem que o modelo ARIMA-CNN foi capaz de capturar 85% da variância da série temporal enquanto o melhor modelo individual, LSTM, foi capaz de descrever 42%. Com relação aos parâmetros MAPE, RMSE e MSE, o modelo ARIMA-CNN também apresentou melhoras significativas diminuindo o MAPE de 199,39% para 98,79%, o RMSE de 1,41 para 0,75 e o MSE de 1,98 para 0,56 quando comparado com o LSTM. O único parâmetro que não foi melhorado pelos modelos híbridos foi o MAE. O modelo híbrido que obteve o menor valor para essa métrica foi o ARIMA-LSTM, com 0,06, em comparação a 0,01 do LSTM.

Para a velocidade do vento, a melhora dos resultados obtidos pelos modelos híbridos se mantém. Nesse caso, a performance dos modelos ARIMA-CNN e ARIMA-LSTM mostrou melhorias em todas as métricas de avaliação, com destaque para o coeficiente de

determinação $R^2 = 0,96$ alcançado pelo ARIMA-CNN, que foi capaz de explicar 96% da variância dos dados. O modelo que mais se destacou para o conjunto de dados do Rio Grande do Sul foi o modelo híbrido ARIMA-CNN, apresentando os melhores resultados gerais nas métricas de avaliação para ambas as variáveis de interesse. Os valores das constantes estimadas para as regressões lineares multivariadas, utilizadas na construção dos modelos híbridos em paralelo, estão apresentados na Tabela 27.

4.2 Considerações gerais

Os resultados para os estados indicaram uma consistente superioridade dos modelos híbridos em paralelo, especialmente o ARIMA-LSTM e o ARIMA-CNN. Ambos os modelos apresentaram desempenhos muito próximos em todos os casos analisados, sugerindo que o ARIMA é um componente fundamental para a construção de um modelo híbrido com melhor performance.

Dentre os resultados, destacaram-se os obtidos nos estados do Ceará e do Rio Grande do Sul, sendo este último o que apresentou o melhor desempenho dos modelos híbridos, com destaque para o ARIMA-CNN. Esse modelo demonstrou o melhor conjunto de métricas para ambas as variáveis de interesse de forma simultânea e, por essa razão, foi selecionado para alimentar o modelo de previsão de geração de energia da usina.

4.3 Modelo de Usína

Como o estado que apresentou o melhor resultado foi o Rio Grande do Sul, a usina eólica de Xangri-lá foi selecionada para realizar a aproximação, considerando que é a usina mais próxima ao ponto de captação dos dados climáticos utilizados no treinamento do modelo. Os dados de posicionamento, bem como a distância calculada entre o ponto de captação e a usina, estão apresentados na Tabela 40.

Tabela 40 – Dados de localização da usina e do ponto de captação de dados climáticos.

	Latitude	Longitude	Altura	Distância
Ponto de Captação	-29.35027777	-49.73305554	4.66m	-
Usína de Xangri-lá	-29.77794444	-50.07472222	1626m	57.91km

Substituindo os dados de altitude na equação do perfil logarítmico do vento e considerando o valor de z_0 igual a 1 m, dado que o ponto de captação está localizado em um centro urbano, a equação utilizada para estimar a velocidade do vento na altitude da usina é expressa como:

$$V_h = V_z \cdot \left(\frac{\ln \left(\frac{1626+h_t}{1} \right)}{\ln \left(\frac{4.66}{1} \right)} \right) \quad (4.1)$$

onde h_t é a altura da torre de cada aerogerador.

A usina eólica de Xangri-lá, da Honda Energy, conta atualmente com dez aerogeradores, sendo nove do modelo Vestas V112/3000 e um do modelo Vestas V136/3450 [72].

Os aerogeradores Vestas V112/3000 são do tipo HAWT, com potência nominal de 3 MW e equipados com três pás de 55 metros de comprimento, cada uma. Já o modelo Vestas V136/3450 possui potência nominal de 3,45 MW e é equipado com três pás de 67 metros de comprimento [72–74].

Dados detalhados sobre as características de cada um desses aerogeradores e suas torres estão apresentados na Tabela 41.

Tabela 41 – Características dos aerogeradores V112/3000 e V136/3450 utilizados no modelo de usina.

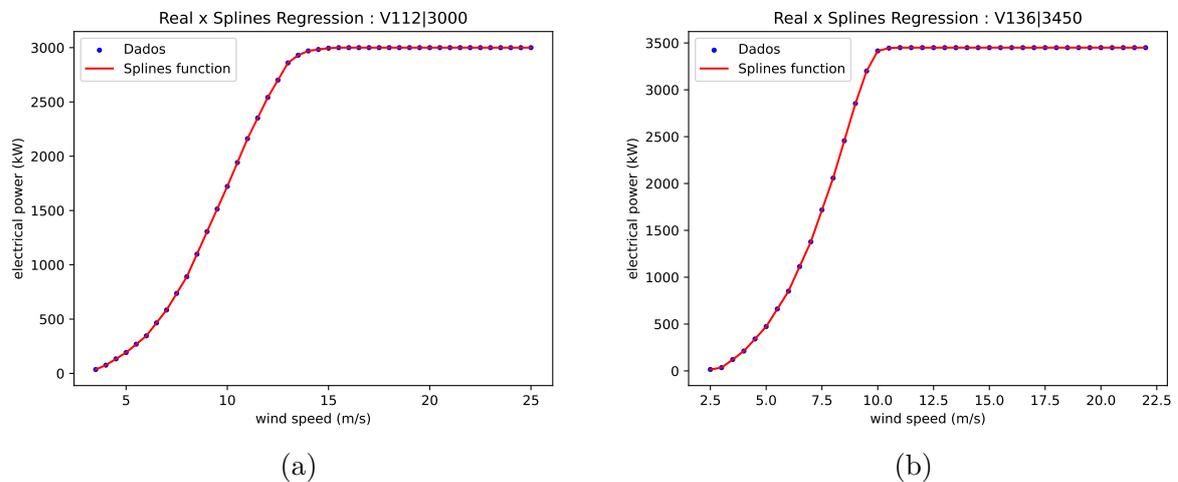
Modelo	Número de pás	Comprimento das pás	Potência nominal	Vel cut-in	Vel cut-off	Vel ideal	Altura Torre
V112/3000	3	55m	3000kW	3.5m/s	25m/s	15.5m/s	94m
V136/3450	3	67m	3450kW	2.5m/s	22m/s	11m/s	112m

Substituindo a altura de cada torre na Equação 4.1, obtêm-se as seguintes equações do perfil logarítmico do vento para cada aerogerador:

$$V_{V/112} = V_z \cdot \left(\frac{\ln \left(\frac{1626+94}{1} \right)}{\ln \left(\frac{4.66}{1} \right)} \right) = V_z \cdot 4.84081 \quad (4.2)$$

$$V_{V/136} = V_z \cdot \left(\frac{\ln \left(\frac{1626+112}{1} \right)}{\ln \left(\frac{4.66}{1} \right)} \right) = V_z \cdot 4.84757 \quad (4.3)$$

Figura 72 – Curvas de potência interpoladas para os diferentes aerogeradores comparados com os dados reais de potência.



As Equações (4.1) e (4.2) mostram, respectivamente, os resultados da interpolação de Splines de ordem 5 para as curvas de potência dos aerogeradores de modelos V112/3000 e V136/3450. As curvas de potência interpoladas são representadas em comparação com os dados reais na Figura 72.

$$\begin{aligned}
y = & 1644.1503 - 1609.5422(x - 3.5)_+^5 - 1548.3700(x - 7)_+^5 \\
& - 1402.5127(x - 11)_+^5 - 1092.2300(x - 13)_+^5 + 40.3060(x - 15)_+^5 \\
& + 2109.2857 \left[(x - 3.5)_+^5 + (x - 7)_+^5 + (x - 11)_+^5 + (x - 13)_+^5 + (x - 15)_+^5 \right] \\
& + 899.1558 \left[(x - 3.5)_+^5 + (x - 7)_+^5 + (x - 11)_+^5 + (x - 13)_+^5 + (x - 15)_+^5 \right] \quad (4.4) \\
& + 1622.7610 \left[(x - 3.5)_+^5 + (x - 7)_+^5 + (x - 11)_+^5 + (x - 13)_+^5 + (x - 15)_+^5 \right] \\
& + 1257.1511 \left[(x - 3.5)_+^5 + (x - 7)_+^5 + (x - 11)_+^5 + (x - 13)_+^5 + (x - 15)_+^5 \right] \\
& + 1368.1456 \left[(x - 3.5)_+^5 + (x - 7)_+^5 + (x - 11)_+^5 + (x - 13)_+^5 + (x - 15)_+^5 \right]
\end{aligned}$$

$$\begin{aligned}
y = & 15.0000 + 104.6650(x - 3)_+^5 - 249.9692(x - 8)_+^5 \\
& + 1030.1106(x - 9.5)_+^5 + 154.1384(x - 10)_+^5 + 4043.0005(x - 15)_+^5 \\
& + 3062.9630 \left[(x - 3)_+^5 + (x - 8)_+^5 + (x - 9.5)_+^5 + (x - 10)_+^5 + (x - 15)_+^5 \right] \\
& + 3702.4636 \left[(x - 3)_+^5 + (x - 8)_+^5 + (x - 9.5)_+^5 + (x - 10)_+^5 + (x - 15)_+^5 \right] \quad (4.5) \\
& + 3293.6536 \left[(x - 3)_+^5 + (x - 8)_+^5 + (x - 9.5)_+^5 + (x - 10)_+^5 + (x - 15)_+^5 \right] \\
& + 3476.1946 \left[(x - 3)_+^5 + (x - 8)_+^5 + (x - 9.5)_+^5 + (x - 10)_+^5 + (x - 15)_+^5 \right] \\
& + 3429.7850 \left[(x - 3)_+^5 + (x - 8)_+^5 + (x - 9.5)_+^5 + (x - 10)_+^5 + (x - 15)_+^5 \right]
\end{aligned}$$

Utilizando os dados de produção horária da usina de Xangri-lá fornecidos pelo Operador Nacional do Sistema Elétrico (ONS) [75] e os dados climáticos reais, foi estimada a direção da superfície dos aerogeradores em 7,588 radianos. As Figuras 73 (a), (b), (c) e (d) comparam a potência média real, a potência média considerando a direção do vento e o intervalo de potências estimado sem a direção do vento, para diferentes períodos de tempo. Os valores utilizados foram obtidos a partir da média da potência horária para os intervalos de 24 horas, 7 dias, 14 dias e 30 dias. Nessas figuras, a linha vermelha representa a potência média real, a linha azul a potência média com direção do vento, e o intervalo é delimitado por 100% e 25% de aproveitamento das potências estimadas sem a direção do vento. A Tabela 42 apresenta os diferentes valores de MAPE e MAE obtidos pela estimativa de potência com direção do vento, a contagem de elementos da potência real fora do intervalo de previsão e seu valor proporcional à quantidade total de dados na amostra.

Os resultados apresentados na Tabela 42 indicam uma melhora consistente em ambos os ajustes com o aumento do intervalo utilizado para calcular a média da potência

Figura 73 – As Figuras (a), (b), (c) e (d) mostram a projeção da potência utilizando a direção do vento e a potência média real sobre o intervalo projetado sem a direção do vento para os intervalos de tempo de 1 dia, 7 dias, 14 dias e 30 dias.

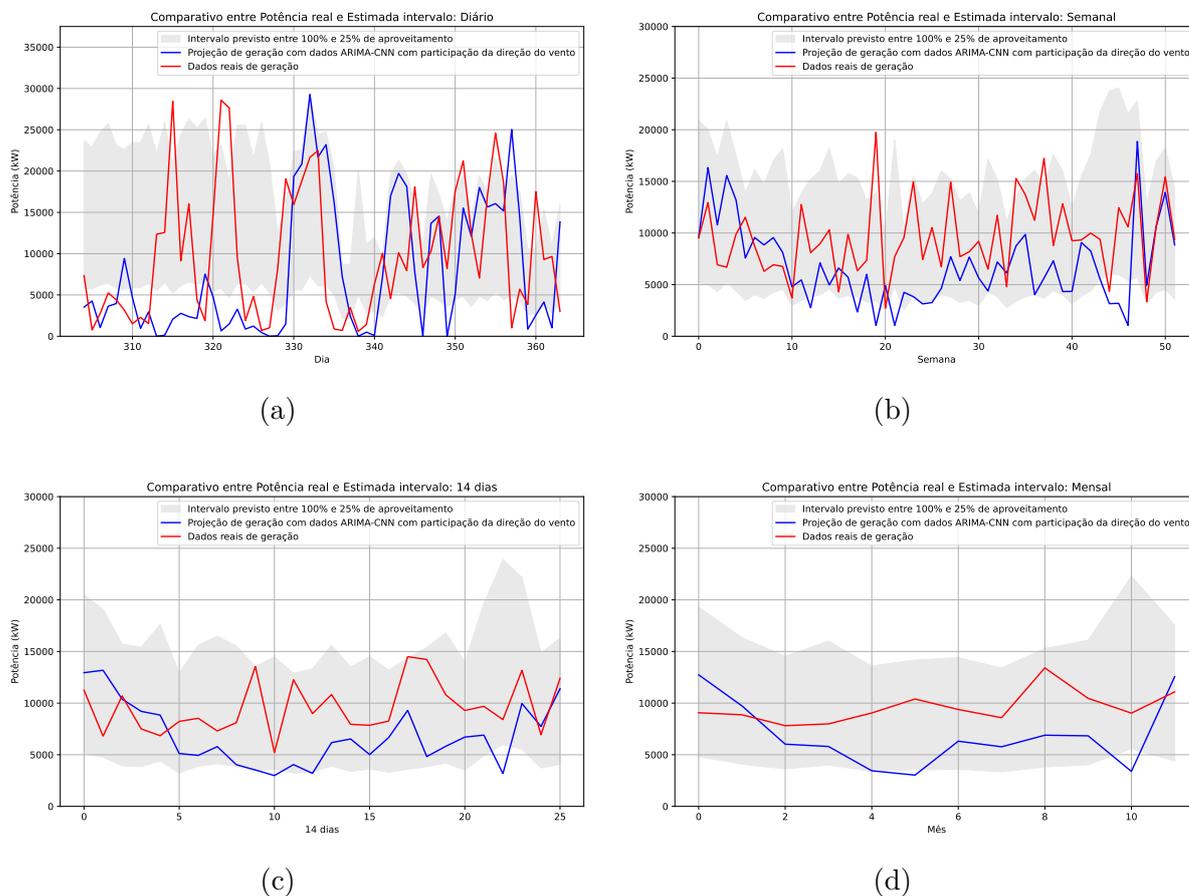


Tabela 42 – Características dos aerogeradores V112/3000 e V136/3450 utilizados no modelo de usina.

Intervalo	MAE	MAPE	Pontos fora do intervalo	Proporção fora do intervalo
1 hora	9846.04 kW	1837.10%	5984	68%
24 horas	7751.55 kW	189.50%	184	50.55%
7 dias	4219.44 kW	43.06%	4	7.69%
14 dias	3705.41 kW	38.06%	2	7.69%
30 dias	3719.07 kW	38.16%	0	0%

da usina. Para a potência estimada com a direção do vento, as métricas MAE e MAPE diminuíram de 9846,04 kW e 1839,1% no regime horário para 3705,31 kW e 38,0% no regime de 14 dias, com resultados próximos aos obtidos no regime de 30 dias. Na abordagem do intervalo de potências sem a direção do vento, a proporção de pontos fora do intervalo reduziu de 68% no regime horário para 7,69% nos regimes de 7 e 14 dias.

A dificuldade em ajustar os resultados para períodos curtos pode ter sido influenciada por diversos fatores não considerados na construção do modelo. No caso da estimativa

de potência utilizando a velocidade do vento, a distorção pode ter sido significativamente afetada pela impossibilidade de projetar a direção observada no ponto de captação para a localização da usina de Xangri-lá, que está a mais de 57 km de distância. Além disso, os aerogeradores da usina de Xangri-lá possuem sistemas que permitem a rotação das pás, tanto para a segurança do equipamento quanto para otimizar a produtividade, fazendo com que a estimativa da direção da superfície dos aerogeradores por um valor fixo, como foi adotado, não consiga representar com precisão a componente de velocidade efetivamente utilizada para alimentar o aerogerador.

Outra questão que pode ter impactado as projeções é a incapacidade da equação do perfil logarítmico do vento, por si só, em descrever as alterações de velocidade causadas por relevos e movimentações complexas de massas de ar, uma vez que essa projeta a velocidade do vento apenas na coordenada z . A dificuldade de projetar os dados climáticos gerados pelo modelo híbrido no espaço se apresenta, também, como o principal fator limitante da abordagem do intervalo de potências sem considerar a direção do vento.

Em ambos os casos apresentados, a melhoria dos resultados com o aumento do período utilizado para o cálculo das médias pode ser entendida como uma suavização das componentes sazonais e de erro da série temporal de potência da usina no intervalo horário, o que aproxima os valores da curva resultante à componente de tendência, possivelmente menos afetada por variações locais dos dados climáticos. Em intervalos mais avançados, a abordagem do intervalo de potências demonstrou ser eficiente em projetar a geração provável da usina analisada, mesmo diante das dificuldades relacionadas à simplicidade da aproximação.

5 Conclusão

Neste trabalho, foram avaliados diferentes modelos preditivos baseados em Redes Neurais Artificiais (RNAs) e ARIMA para a previsão da direção e velocidade do vento em diversos estados brasileiros. Os modelos preditivos analisados foram classificados em três categorias principais: modelos simples, como CNN, LSTM e ARIMA; modelos híbridos em paralelo, incluindo CNN-LSTM, ARIMA-LSTM e ARIMA-CNN; e o modelo híbrido em série, representado pelo CNN-ALSTM. Os resultados, para cada estado, foram avaliados com base nas métricas de desempenho $MAPE$, MAE , $RMSE$, MSE e R^2 , obtidas a partir da comparação entre os dados estimados de velocidade e direção do vento gerados por esses modelos e os dados reais referentes ao ano de 2022.

Com base nos resultados apresentados nas análises do capítulo anterior, observa-se que os modelos híbridos paralelos, especialmente o ARIMA-CNN e o ARIMA-LSTM, demonstraram desempenho superior na estimativa da velocidade e direção do vento em comparação com os outros métodos avaliados. O melhor desempenho foi obtido pelo modelo ARIMA-CNN, aplicado aos dados climáticos do estado do Rio Grande do Sul. Nesse caso, o aumento de precisão em relação ao melhor modelo individual, o LSTM, foi evidenciado pelas métricas de avaliação. Para a velocidade do vento, o ARIMA-CNN apresentou os seguintes valores: $MAPE$ de 13,40, $RMSE$ de 0,33, MSE de 0,11, MAE de 0,04 e R^2 de 0,96. Em contrapartida, o LSTM obteve $MAPE$ de 39,01, $RMSE$ de 0,87, MSE de 0,76, MAE de 0,03 e R^2 de 0,70. Esses resultados reforçam a eficácia do modelo ARIMA-CNN na captura de padrões complexos dos dados climáticos.

Para a direção do vento, o aumento de desempenho foi ainda mais expressivo. O modelo LSTM apresentou os seguintes valores: $MAPE$ de 199,39, $RMSE$ de 1,41, MSE de 1,98, MAE de 0,01 e R^2 de 0,42. Em comparação, o modelo ARIMA-CNN alcançou resultados significativamente melhores, com $MAPE$ de 85,75, $RMSE$ de 0,71, MSE de 0,51, MAE de 0,01 e R^2 de 0,85. Esses resultados evidenciam a capacidade superior do ARIMA-CNN em capturar a variabilidade e os padrões relacionados à direção do vento.

Por outro lado, os modelos híbridos CNN-LSTM e CNN-ALSTM, apesar de suas diferentes estruturas, apresentaram métricas de avaliação próximas ou inferiores às dos modelos simples em todos os estados analisados, indicando limitações na sua eficácia comparativa.

Os resultados também evidenciaram que os dados gerados pelos modelos híbridos podem ser aplicados na estimativa da produção de energia em usinas reais. As análises realizadas demonstraram um desempenho progressivamente melhor para ambas as abordagens de previsão à medida que o intervalo de tempo utilizado para o cálculo das médias foi

ampliado. Na primeira abordagem, com a média horária, verificou-se que 68% dos pontos estavam fora do intervalo de predição. Esse valor caiu para 7,69% com a média de 14 dias e chegou a 0% com a média de 30 dias, quando todos os pontos passaram a estar contidos no intervalo. Na segunda abordagem, observou-se um padrão semelhante: o *MAPE* foi de 1837,10% para médias de 1 hora, reduziu-se para 43,06% para médias de 7 dias e chegou a 38,06% para médias de 14 dias. Esses resultados destacam que a ampliação do intervalo de tempo para o cálculo das médias contribui para maior precisão e confiabilidade nas estimativas de produção energética, atenuando o impacto das variações de curto prazo nos dados climáticos utilizados.

Os resultados menos precisos observados em ambas as abordagens, especialmente ao utilizar períodos mais curtos para o cálculo das médias, como intervalos de 1 hora e 24 horas, ressaltam a necessidade de aprimorar a técnica, principalmente na projeção de dados climáticos para diferentes pontos no espaço. Essa limitação permeia toda a implementação do modelo preditivo, evidenciando a importância de refinar as técnicas empregadas nas projeções verticais e horizontais da velocidade do vento. A abordagem atual, baseada no perfil logarítmico do vento, restringe-se a projeções na dimensão vertical, enquanto as projeções associadas à direção do vento abrangem tanto as dimensões vertical quanto horizontal.

Uma possível abordagem para a solução dos problemas de projeção de dados climáticos e a utilização de Sistemas de Informação Geoespacial (GIS) combinados com redes neurais artificiais ou modelos de ajuste multivariados para efetuar as mesmas. Além disso, o desenvolvimento de um método para simular a mudança de angulação das pás dos aerogeradores em função da velocidade e direção históricas do vento pode representar um avanço significativo na viabilização do método de projeção que utiliza a direção do vento.

Referências

- [1] Kleidon, Axel: *Life, hierarchy, and the thermodynamic machinery of planet Earth*. *Physics of life reviews*, 7(4):424–460, 2010. Citado 2 vezes nas páginas 8 e 24.
- [2] BRASILEIRO, EÓLICO: *Atlas do potencial eólico Brasileiro*, 2001. Citado 2 vezes nas páginas 8 e 25.
- [3] Almeida, Antenor Timo Pinheiro de e Micael Martins da Silva: *Desenvolvimento e montagem de um gerador eólico com pás compósitas*. 2011. Citado 3 vezes nas páginas 8, 25 e 26.
- [4] Manwell, James F, Jon G McGowan e Anthony L Rogers: *Wind energy explained: theory, design and application*. John Wiley & Sons, 2010. Citado 3 vezes nas páginas 8, 26 e 27.
- [5] WindBox: *Curva de Potência de Aerogeradores: O Que É e Como Funciona*. <https://windbox.com.br/blog/curva-de-potencia/>, 2024. Acesso em: 12 set. 2024. Citado 2 vezes nas páginas 8 e 28.
- [6] International Energy Agency: *Wind Energy*, Accessed on: 13 de dezembro de 2024. <https://www.iea.org/energy-system/renewables/wind>. Citado na página 20.
- [7] European Commission: *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Renovation Wave for Europe - greening our buildings, creating jobs, improving lives*, 2023. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52023DC0669>, acesso em 2024-04-22. Citado na página 20.
- [8] U.S. Department of Energy: *Energy Secretary Granholm Announces Ambitious New 30GW Offshore Wind Deployment Target*, Accessed on: 13 de dezembro de 2024. <https://www.energy.gov/articles/energy-secretary-granholm-announces-ambitious-new-30gw-offshore-wind-deployment>. Citado na página 20.
- [9] The White House: *Fact Sheet: Biden Administration Jumpstarts Offshore Wind Energy Projects to Create Jobs*, 2021. <https://www.whitehouse.gov/briefing-room/statements-releases/2021/03/29/fact-sheet-biden-administration-jumpstarts-offshore-wind-energy-projects-to-cr> acesso em Accessed on: 13 de dezembro de 2024. Citado na página 20.

- [10] ÅF Pöyry: *China Expects to Achieve 2030 Wind and Solar Ambitions Ahead in 2025*, Accessed on: 13 de dezembro de 2024. <https://afry.com/en/insight/china-expects-achieve-2030-wind-and-solar-ambitions-ahead-in-2025>. Citado na página 20.
- [11] Mongabay-India: *Webinar: What does India need to meet its 2030 renewable energy targets?*, Accessed on: 13 de dezembro de 2024. <https://india.mongabay.com/2022/07/webinar-what-does-india-need-to-meet-its-2030-renewable-energy-targets/>. Citado na página 20.
- [12] Japan External Trade Organization (JETRO): *Wind Energy Key to Net Zero Emission Targets*, Accessed on: 13 de dezembro de 2024. <https://www.jetro.go.jp/en/invest/insights/japan-insight/wind-energy-key-to-net-zero-emission-targets.html>. Citado na página 20.
- [13] *Power BI Visualization*. <https://app.powerbi.com/view?r=eyJrIjoibjNjc4OGYyYjQtYWM2ZC00YjllLWJlYmEtYzdkNTQ1MTc1NjM2IiwidCI6IjQwZDZmOWI4LWV>. Accessed on: 13 de dezembro de 2024. Citado na página 20.
- [14] Power Technology: *Brazil aims to triple clean energy by 2030*, Accessed on: 13 de dezembro de 2024. <https://www.power-technology.com/news/brazil-triple-clean-energy-2030/?cf-view>. Citado na página 20.
- [15] EPBR: *Triplicar energias renováveis até 2030 é difícil, mas possível, diz BNEF*, Accessed on: 13 de dezembro de 2024. <https://epbr.com.br/triplicar-energias-renovaveis-ate-2030-e-dificil-mas-possivel-diz-bnef/>. Citado na página 20.
- [16] Empresa de Pesquisa Energética: *Balço Energético Nacional - 2022*, 2022. Citado na página 20.
- [17] Associação Brasileira de Energia Eólica: *Boletim Anual de Geração Eólica 2016*, 2016. Citado na página 20.
- [18] Associação Brasileira de Energia Eólica: *Boletim Anual de Geração Eólica 2022*, 2022. Citado na página 20.
- [19] EVWind: *Wind Farms and Grid Integration: Challenges and Solutions*, 2023. <https://www.evwind.es/2023/06/23/wind-farms-and-grid-integration-challenges-and-solutions/92414>. Citado na página 21.

- [20] Chen, Hao, Qixia Zhang e Yngve Birkelund: *Machine learning forecasts of Scandinavian numerical weather prediction wind model residuals with control theory for wind energy*. Energy Reports, 8(Supplement 13):661–668, 2022, ISSN 2352-4847. Citado na página 21.
- [21] Würth, Ines, Lluís Valldecabres, Emil Simon, Corinna Möhrlein, Bahadır Uzunoglu, Christian Gilbert, Gregor Giebel, David Schlipf e Anton Kaifel: *Minute-Scale Forecasting of Wind Power—Results from the Collaborative Workshop of IEA Wind Task 32 and 36*. Energies, 12:712, 2019. <https://www.mdpi.com/1996-1073/12/4/712>. Citado na página 21.
- [22] Manwell, James F., Jon G. McGowan e Anthony L. Rogers: *Wind Energy Explained: Theory, Design and Application*. John Wiley & Sons, 2009. Citado na página 21.
- [23] Tsai, Wei Chun, Chih Ming Hong, Chien Shun Tu, Wen Mao Lin e Chih Hung Chen: *A Review of Modern Wind Power Generation Forecasting Technologies*. Sustainability, 15:10757, 2023. <https://www.mdpi.com/2071-1050/15/14/10757>. Citado na página 21.
- [24] Hive Power: *Grid Stability Issues with Renewable Energy: How They Can Be Solved*, Accessed on: 13 de dezembro de 2024. <https://www.hivepower.tech/blog/grid-stability-issues-with-renewable-energy-how-they-can-be-solved>. Citado na página 21.
- [25] Zucatelli, P. J., E. G. S. Nascimento, A. Á. B. Santos, A. M. G. Arce e D. M. Moreira: *An investigation on deep learning and wavelet transform to nowcast wind power and wind power ramp: A case study in Brazil and Uruguay*. Energy, 230:120842, 2021, ISSN 0360-5442. <https://doi.org/10.1016/j.energy.2021.120842>. Citado na página 21.
- [26] Jacondino, William Duarte, Ana Lucia da Silva Nascimento, Leonardo Calvetti, Gilberto Fisch, Cesar Augustus Assis Beneti e Sheila Radman da Paz: *Hourly day-ahead wind power forecasting at two wind farms in northeast Brazil using WRF model*. Energy, 230:120841, 2021, ISSN 0360-5442. <https://doi.org/10.1016/j.energy.2021.120841>. Citado na página 21.
- [27] Jong, Pieter de, Roger Dargaville, Jeremy Silver, Steven Utembe, Asher Kiperstok e Ednildo Andrade Torres: *Forecasting high proportions of wind energy supplying the Brazilian Northeast electricity grid*. Applied Energy, 195:538–555, 2017, ISSN 0306-2619. <https://doi.org/10.1016/j.apenergy.2017.03.058>. Citado na página 21.
- [28] Paula, M., C. Marilaine, F. José Nuno e C. Wallace: *Predicting Long-Term Wind Speed in Wind Farms of Northeast Brazil: A Comparative Analysis Through Machine*

- Learning Models*. IEEE Latin America Transactions, 18(11):2011–2018, November 2020. Citado na página 21.
- [29] Jin, Yongchao, Renfang Wang, Xiaodie Zhuang, Kenan Wang, Honglian Wang, Chenxi Wang e Xiyin Wang: *Prediction of COVID-19 data using an ARIMA-LSTM hybrid forecast model*. Mathematics, 10(21):4001, 2022. Citado 3 vezes nas páginas 21, 39 e 40.
- [30] Xu, Guoyan, Zixu Jing, Yingchi Mao e Xinyue Su: *A dam deformation prediction model based on ARIMA-LSTM*. Em *2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService)*, páginas 205–211. IEEE, 2020. Citado 2 vezes nas páginas 21 e 39.
- [31] Jana, Debasish, Jayant Patil, Sudheendra Herkal, Satish Nagarajaiah e Leonardo Duenas-Osorio: *CNN and Convolutional Autoencoder (CAE) based real-time sensor fault detection, localization, and correction*. Mechanical Systems and Signal Processing, 169:108723, 2022. Citado na página 21.
- [32] Thakur, Dipanwita, Suparna Biswas, Edmond SL Ho e Samiran Chattopadhyay: *Convae-lstm: Convolutional autoencoder long short-term memory network for smartphone-based human activity recognition*. IEEE Access, 10:4137–4156, 2022. Citado 2 vezes nas páginas 21 e 39.
- [33] Gao, Honghao, Binyang Qiu, Ramon J Duran Barroso, Walayat Hussain, Yueshen Xu e Xinheng Wang: *Tsmae: a novel anomaly detection approach for internet of things time series data using memory-augmented autoencoder*. IEEE Transactions on network science and engineering, 2022. Citado na página 21.
- [34] Nguyen, H Du, Kim Phuc Tran, Sébastien Thomassey e Moez Hamad: *Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management*. International Journal of Information Management, 57:102282, 2021. Citado na página 21.
- [35] Rayi, Vijaya Krishna, SP Mishra, Jyotirmayee Naik e PK Dash: *Adaptive VMD based optimized deep learning mixed kernel ELM autoencoder for single and multistep wind power forecasting*. Energy, 244:122585, 2022. Citado na página 21.
- [36] Du, Liang, Ruobin Gao, Ponnuthurai Nagaratnam Suganthan e David ZW Wang: *Bayesian optimization based dynamic ensemble for time series forecasting*. Information Sciences, 591:155–175, 2022. Citado na página 21.
- [37] Kong, Jian Lei, Xiao Meng Fan, Xue Bo Jin, Ting Li Su, Yu Ting Bai, Hui Jun Ma e Min Zuo: *BMAE-Net: A data-driven weather prediction network for smart agriculture*. Agronomy, 13(3):625, 2023. Citado na página 21.

- [38] Instituto Nacional de Meteorologia: *National Institute of Meteorology*. Online. <https://portal.inmet.gov.br/dadoshistoricos>, Accessed on: April 16, 2024. Citado na página 22.
- [39] Ponkumar, G., S. Jayaprakash e K. Kanagarathinam: *Advanced Machine Learning Techniques for Accurate Very-Short-Term Wind Power Forecasting in Wind Energy Systems Using Historical Data Analysis*. *Energies*, 16:5459, 2023. <https://doi.org/10.3390/en16145459>. Citado na página 22.
- [40] Solas, M., N. M. Cepeda e J. P. Viegas: *Convolutional Neural Network for Short-term Wind Power Forecasting*. Em *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, páginas 1–5, 2019. Citado na página 22.
- [41] Gomes, Samuel Luz: *CIÊNCIA E TECNOLOGIA DO CEARÁ PROGRAMA DE PÓS-GRADUAÇÃO EM ENERGIAS RENOVÁVEIS*. Citado 2 vezes nas páginas 24 e 25.
- [42] Fleming, PD e SD Probert: *The evolution of wind-turbines: An historical review*. *Applied energy*, 18(3):163–177, 1984. Citado na página 25.
- [43] Leung, Dennis YC e Yuan Yang: *Wind energy development and its environmental impact: A review*. *Renewable and sustainable energy reviews*, 16(1):1031–1039, 2012. Citado na página 25.
- [44] Kaldellis, John K e Dimitris Zafirakis: *The wind energy (r) evolution: A short review of a long history*. *Renewable energy*, 36(7):1887–1901, 2011. Citado na página 25.
- [45] Miranda, Amauri Ghellere Garcia *et al.*: *Estudo de ciclo de vida de gerador eólico de pequeno porte utilizado em propriedade rural*. 2018. Citado 3 vezes nas páginas 26, 27 e 28.
- [46] Gontijo, Gustavo Figueiredo: *CONTROLES DE UM SISTEMA EÓLICO NA CONFIGURAÇÃO DFIG COM FUNCIONALIDADES DE CONDICIONAMENTO DE ENERGIA*. Tese de Doutorado, Universidade Federal do Rio de Janeiro, 2016. Citado na página 26.
- [47] Nowaczyk, Eduardo Ribas: *Estudo da otimização do posicionamento de aerogeradores em parques eólicos*. 2013. Citado na página 28.
- [48] Shahriari, Bobak, Kevin Swersky, Ziyu Wang, Ryan P Adams e Nando De Freitas: *Taking the human out of the loop: A review of Bayesian optimization*. *Proceedings of the IEEE*, 104(1):148–175, 2015. Citado na página 30.

- [49] Snoek, Jasper, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Mostofa Patwary, Mr Prabhat e Ryan Adams: *Scalable bayesian optimization using deep neural networks*. Em *International conference on machine learning*, páginas 2171–2180. PMLR, 2015. Citado na página 30.
- [50] Ryan, Elizabeth G, Christopher C Drovandi, James M McGree e Anthony N Pettitt: *A review of modern computational algorithms for Bayesian optimal design*. *International Statistical Review*, 84(1):128–154, 2016. Citado na página 30.
- [51] Williams, Christopher KI e Carl Edward Rasmussen: *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006. Citado na página 30.
- [52] Bartók, Albert P e Gábor Csányi: *G gaussian approximation potentials: A brief tutorial introduction*. *International Journal of Quantum Chemistry*, 115(16):1051–1057, 2015. Citado na página 30.
- [53] Mockus, J, V Tiesis e A Zilinskas: *The application of Bayesian methods for seeking the extremum, vol. 2*. L Dixon and G Szego. *Toward Global Optimization*, 2, 1978. Citado na página 32.
- [54] ArunKumar, KE, Dinesh V Kalaga, Ch Mohan Sai Kumar, Masahiro Kawaji e Timothy M Brenza: *Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends*. *Alexandria engineering journal*, 61(10):7585–7603, 2022. Citado na página 34.
- [55] Fukushima, Kunihiko e Sei Miyake: *Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition*. *Competition and cooperation in neural nets*, 45:267–285, 1982. Citado na página 35.
- [56] Lang, Christian, Florian Steinborn, Oliver Steffens e Elmar W Lang: *Applying a 1D-CNN network to electricity load forecasting*. Em *Theory and Applications of Time Series Analysis: Selected Contributions from ITISE 2019 6*, páginas 205–218. Springer, 2020. Citado na página 35.
- [57] LeCun, Yann, Bernhard Boser, John Denker, Donnie Henderson, Richard Howard, Wayne Hubbard e Lawrence Jackel: *Handwritten digit recognition with a back-propagation network*. *Advances in neural information processing systems*, 2, 1989. Citado na página 35.
- [58] Hussain, Dostdar, Tahir Hussain, Aftab Ahmed Khan, Syed Ali Asad Naqvi e Akhtar Jamil: *A deep learning approach for hydrological time-series prediction: A case study of Gilgit river basin*. *Earth Science Informatics*, 13:915–927, 2020. Citado na página 35.

- [59] Saini, Manali, Udit Satija e Madhur Deo Upadhayay: *Light-weight 1-D convolutional neural network architecture for mental task identification and classification based on single-channel EEG*. arXiv preprint arXiv:2012.06782, 2020. Citado na página 36.
- [60] Sagheer, Alaa e Mostafa Kotb: *Time series forecasting of petroleum production using deep LSTM recurrent networks*. *Neurocomputing*, 323:203–213, 2019. Citado na página 38.
- [61] Shahid, Farah, Aneela Zameer e Muhammad Muneeb: *Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM*. *Chaos, Solitons & Fractals*, 140:110212, 2020. Citado na página 38.
- [62] Xiang, Zhongrun, Jun Yan e Ibrahim Demir: *A rainfall-runoff model with LSTM-based sequence-to-sequence learning*. *Water resources research*, 56(1):e2019WR025326, 2020. Citado na página 38.
- [63] Hu, Caihong, Qiang Wu, Hui Li, Shengqi Jian, Nan Li e Zhengzheng Lou: *Deep learning with a long short-term memory networks approach for rainfall-runoff simulation*. *Water*, 10(11):1543, 2018. Citado na página 38.
- [64] Peng, Lu, Shan Liu, Rui Liu e Lin Wang: *Effective long short-term memory with differential evolution algorithm for electricity price prediction*. *Energy*, 162:1301–1314, 2018. Citado na página 38.
- [65] Malhotra, Pankaj, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal e Gautam Shroff: *LSTM-based encoder-decoder for multi-sensor anomaly detection*. arXiv preprint arXiv:1607.00148, 2016. Citado na página 39.
- [66] Provotar, Oleksandr I, Yaroslav M Linder e Maksym M Veres: *Unsupervised anomaly detection in time series using lstm-based autoencoders*. Em *2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT)*, páginas 513–517. IEEE, 2019. Citado na página 39.
- [67] Principi, Emanuele, Damiano Rossetti, Stefano Squartini e Francesco Piazza: *Unsupervised electric motor fault detection by using deep autoencoders*. *IEEE/CAA Journal of Automatica Sinica*, 6(2):441–451, 2019. Citado na página 39.
- [68] Ibrahim, Mohamed Sayed, Sawsan Morkos Gharghory e Hanan Ahmed Kamal: *A hybrid model of CNN and LSTM autoencoder-based short-term PV power generation forecasting*. *Electrical Engineering*, páginas 1–17, 2024. Citado na página 39.
- [69] Ye, Zhuang e Jianbo Yu: *Health condition monitoring of machines based on long short-term memory convolutional autoencoder*. *Applied Soft Computing*, 107:107379, 2021. Citado na página 39.

- [70] Yu, Jianbo, Xing Liu e Lyujiangnan Ye: *Convolutional long short-term memory autoencoder-based feature learning for fault detection in industrial processes*. IEEE Transactions on Instrumentation and Measurement, 70:1–15, 2020. Citado na página 39.
- [71] Dasan, Evangelin e Ithayarani Panneerselvam: *A novel dimensionality reduction approach for ECG signal via convolutional denoising autoencoder with LSTM*. Biomedical Signal Processing and Control, 63:102225, 2021. Citado na página 39.
- [72] The Wind Power: *Xangri-lá (Brasil) - Parque Eólico*, n.d. https://www.thewindpower.net/windfarm_es_20883_xangri-la.php, acesso em 2024-09-20, Acesso em: 20 set. 2024. Citado na página 103.
- [73] AutoIndústria: *Honda Energy amplia parque eólico em Xangri-lá*, 2020. <https://www.autoindustria.com.br/2020/02/07/honda-energy-amplia-parque-eolico-em-xangri-la/>, acesso em 2024-09-20, Acesso em: 20 set. 2024. Citado na página 103.
- [74] Sala de Imprensa Honda: *Honda comemora oito anos da operação de seu parque eólico*, 2023. <https://saladeimprensa.honda.com.br/releases/honda-comemora-oito-anos-da-operacao-de-seu-parque-eolico>, acesso em 2024-09-20, Acesso em: 20 set. 2024. Citado na página 103.
- [75] Operador Nacional do Sistema Elétrico (ONS): *Geração Média Diária Horária*, 2024. <https://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/geracao-media-diaria-horaria.aspx>, Acessado em: 24 set. 2024. Citado na página 104.