



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

sid.inpe.br/mtc-m21d/2024/09.18.21.46-TDI

**UMA ABORDAGEM USANDO REDES NEURAIIS  
ARTIFICIAIS PARA A PREVISÃO DE CURTO PRAZO  
DE ALTURA DE ONDAS MARÍTIMAS EM REGIÃO  
PORTUÁRIA**

Vanessa Colla Paspaltzis

Dissertação de Mestrado do Curso  
de Pós-Graduação em Computação  
Aplicada, orientada pelo Dr. Alan  
James Peixoto Calheiros, aprovada  
em 05 de setembro de 2024.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34T/4C45E38>>

INPE  
São José dos Campos  
2024

**PUBLICADO POR:**

Instituto Nacional de Pesquisas Espaciais - INPE  
Coordenação de Ensino, Pesquisa e Extensão (COEPE)  
Divisão de Biblioteca (DIBIB)  
CEP 12.227-010  
São José dos Campos - SP - Brasil  
Tel.:(012) 3208-6923/7348  
E-mail: pubtc@inpe.br

**CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO  
INTELECTUAL DO INPE - CEPPII (PORTARIA Nº 1216/2024/SEI-  
INPE):**

**Presidente:**

Dr. Thales Sehn Korting - Coordenação-Geral de Ciências da Terra (CGCT)

**Membros:**

Dr. Antonio Fernando Bertachini de Almeida Prado - Conselho de Pós-Graduação (CPG)

Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia, Tecnologia e Ciência Espaciais (CGCE)

Dr. Heyder Hey - Coordenação-Geral de Infraestrutura e Pesquisas Aplicadas (CGIP)

Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)

**BIBLIOTECA DIGITAL:**

Dr. Gerald Jean Francis Banon

Clayton Martins Pereira - Divisão de Biblioteca (DIBIB)

**REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:**

Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)

André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)

**EDITORAÇÃO ELETRÔNICA:**

Ivone Martins - Divisão de Biblioteca (DIBIB)

André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

sid.inpe.br/mtc-m21d/2024/09.18.21.46-TDI

**UMA ABORDAGEM USANDO REDES NEURAIIS  
ARTIFICIAIS PARA A PREVISÃO DE CURTO PRAZO  
DE ALTURA DE ONDAS MARÍTIMAS EM REGIÃO  
PORTUÁRIA**

Vanessa Colla Paspaltzis

Dissertação de Mestrado do Curso  
de Pós-Graduação em Computação  
Aplicada, orientada pelo Dr. Alan  
James Peixoto Calheiros, aprovada  
em 05 de setembro de 2024.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34T/4C45E38>>

INPE  
São José dos Campos  
2024

Dados Internacionais de Catalogação na Publicação (CIP)

---

Paspaltzis, Vanessa Colla.

P264a Uma abordagem usando redes neurais artificiais para a previsão de curto prazo de altura de ondas marítimas em região portuária / Vanessa Colla Paspaltzis. – São José dos Campos : INPE, 2024.

xviii + 64 p. ; (sid.inpe.br/mtc-m21d/2024/09.18.21.46-TDI)

Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2024.

Orientador : Dr. Alan James Peixoto Calheiros.

1. Redes neurais artificiais. 2. LSTM (Long Short Term Memory). 3. Previsão de altura de ondas. 4. Dados meteo-oceanográficos. 5. Previsão de curto prazo. I.Título.

CDU 004.822:551.466

---



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).





MINISTÉRIO DA  
CIÊNCIA, TECNOLOGIA  
E INOVAÇÃO



## INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

### DEFESA FINAL DE DISSERTAÇÃO DE VANESSA COLLA PASPALTZIS BANCA Nº 213/2024, REG. 327807/2021

No dia 05 de setembro de 2024, às 08:30, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Mestre em Computação Aplicada, com a exigência de que o trabalho final a ser publicado deverá incorporar as correções sugeridas pela Banca Examinadora, com revisão pelo(s) orientador(es).

**Título: "UMA ABORDAGEM USANDO REDES NEURAIS ARTIFICIAIS PARA A PREVISÃO DE CURTO PRAZO DE ALTURA DE ONDAS MARÍTIMAS EM REGIÃO PORTUÁRIA"**

#### Membros da Banca:

Dr. Rafael Duarte Coelho dos Santos - presidente - INPE  
Dr. Alan James Peixoto Calheiros - Orientador - INPE  
Dr. Rosio Del Pilar Camayo Maita – Membro Interno – INPE  
Dr. Joseph Harari – Membro Externo – USP



Documento assinado eletronicamente por **Rafael Duarte Coelho dos Santos, Pesquisadora**, em 09/09/2024, às 09:12 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Alan James Peixoto Calheiros, Tecnologista**, em 09/09/2024, às 09:43 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **JOSEPH HARARI (E), Usuário Externo**, em 09/09/2024, às 13:34 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rosio Del Pilar Camayo Maita, Pesquisador**, em 10/09/2024, às 12:10 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site <https://sei.mcti.gov.br/verifica.html>, informando o código verificador **12221612** e o código CRC **E67226CA**.

---

**Referência:** Processo nº 01340.007341/2024-01

SEI nº 12221612

## RESUMO

Esta pesquisa foca na aplicação de redes neurais artificiais, especificamente a LSTM (*Long Short Term Memory*) na previsão de altura de ondas no Terminal Portuário de Barcarena. Neste contexto, foram analisados dados meteo-oceanográficos coletados *in situ* e imagens do satélite GOES16 para desenvolver um modelo de *nowcasting*. A análise exploratória dos dados revelou alta correlação assíncrona entre a altura das ondas, a intensidade dos ventos e a temperatura de brilho do topo das nuvens, refletindo padrões sazonais e ciclos diurnos na área de estudo. Observou-se que os ventos mais intensos e as maiores alturas de onda ocorrem principalmente nos meses de setembro e outubro, com os picos de altura das ondas geralmente no final da tarde e à noite, após a intensificação dos ventos locais. Foram realizados cinco experimentos distintos com a LSTM, variando as combinações das variáveis de entrada, como altura de ondas, ventos, nível do mar e estatísticas de temperatura de brilho das nuvens. Os dados dos anos de 2020 e 2021 foram utilizados para treinamento, enquanto o ano de 2022 foi reservado para teste e avaliação do desempenho dos modelos. Os resultados indicaram que a inclusão de variáveis adicionais, como vento e nível do mar, melhora significativamente a precisão das previsões em comparação com o uso exclusivo de dados de altura de ondas. O modelo que utilizou dados de altura de ondas combinados com dados de ventos e de nível do mar apresentou o melhor desempenho entre todos os modelos testados, obtendo RMSE de 0,0738 m e *Skill* de 0,87 para as previsões para 1 hora. A análise dos erros do modelo mostrou que a precisão das previsões tende a ser menor em condições de ondas mais altas e com aumentos abruptos na altura das ondas. Em relação ao limite operacional portuário, o modelo obteve probabilidade de detecção de 87% e taxa de falsos alarmes de 3%. Portanto, os resultados indicam que a LSTM é uma ferramenta promissora para *nowcasting* de ondas na região, com potencial para aplicação em operações portuárias.

Palavras-chave: Redes Neurais Artificiais. LSTM (Long Short Term Memory). Previsão de Altura de Ondas. Dados Meteo-oceanográficos. Previsão de curto prazo.



# AN APPROACH USING ARTIFICIAL NEURAL NETWORKS FOR SEA WAVE HEIGHT SHORT-TERM PREDICTION IN PORT REGION

## ABSTRACT

This research focus on the application of artificial neural networks, specifically LSTM (Long Short Term Memory), for wave height forecasting at the Barcarena Port Terminal. In situ meteo-oceanographic data and images from the GOES16 satellite were analyzed to develop a wave height nowcasting model. The exploratory data analysis revealed a strong asynchronous correlation between wave height, wind speed, and cloud-top brightness temperature, reflecting seasonal patterns and diurnal cycles in the study area. It was observed that the most intense winds and highest wave heights occur primarily in September and October, with wave height peaks generally occurring in the late afternoon and evening, following local wind intensification. Five different experiments were conducted with LSTM, varying the combinations of input variables, such as wave height, wind speed, sea level, and cloud-top brightness temperature statistics. Data from 2020 and 2021 were used for training, while 2022 data were reserved for testing and model performance evaluation. The results indicated that the inclusion of additional variables, such as wind and sea level, significantly improves forecast accuracy compared to using wave height data alone. The model that combined wave height, wind speed, and sea level data achieved the best performance among all tested models, with an RMSE of 0.0738 m and a Skill of 0.87 for 1-hour forecasts. The error analysis showed that prediction accuracy tends to decrease under higher wave conditions and with abrupt increases in wave height. Regarding the port's operational threshold, the model achieved a detection probability of 87% and a false alarm rate of 3%. Therefore, the results suggest that LSTM is a promising tool for wave nowcasting in the region, with potential applications in port operations.

Keywords: Artificial Neural Networks (ANN). LSTM (Long Short Term Memory). Wave Height Forecasting. Meteo-oceanographic Data. Wave Height Prediction.



## LISTA DE FIGURAS

	<u>Pág.</u>
3.1 Parâmetros de uma onda no oceano. . . . .	7
3.2 Espectro de energia das ondas no mar. Tipos de ondas de superfície e suas características, como energia, comprimento, período e forçantes atuantes. . . . .	8
3.3 Escala Beaufort com a relação entre a velocidade do vento, estado do mar e altura significativa das ondas. . . . .	10
3.4 Esquema ilustrativo de operação de ondógrafos: A) medição pelo orbital e pressão; e B) boia com acelerômetros. . . . .	11
4.1 Localização da estação de monitoramento meteo-oceanográfico. . . . .	17
4.2 Apresentação da estação de monitoramento meteo-oceanográfico em tempo-real, instalada e mantida pela HidroMares no Terminal Portuário de Barcarena. . . . .	18
4.3 Ilustração do perfilador acústico de correntes com sensor de pressão embutido - ADCP ( <i>Acoustic Doppler current profiler</i> , modelo SL 500, que compõe a estação. . . . .	18
4.4 Área de estudo e ilustração do recorte realizado nas imagens do GOES16. . . . .	19
4.5 Ilustração dos dados de $Tb_{IR}$ obtidos a partir de imagens do GOES-16 antes (esquerda) e depois (direita) do recorte da área de interesse. . . . .	20
4.6 Esquema ilustrativo da separação dos conjuntos de entrada e de saída da RNA, elaborado de maneira a permitir previsões operacionais no instante de tempo $t$ , para o instante de tempo " $t+1$ ", com base nos dados de " $t-4h$ até $t$ ". . . . .	22
5.1 Séries temporais dos dados de temperatura de brilho mínima, média, máxima e o desvio padrão obtidos a partir das imagens do GOES16 para a área de estudo. . . . .	28
5.2 Rosa dos ventos para os dados coletados <i>in situ</i> na estação de monitoramento, durante o período de estudo. . . . .	28
5.3 Série temporal dos dados de altura de onda (WH10/10 - alturas 10% mais altas) coletados <i>in situ</i> na estação de monitoramento, durante o período de estudo. . . . .	29
5.4 Gráfico <i>boxplot</i> dos dados de período de pico coletados <i>in situ</i> na estação de monitoramento, durante o período de estudo. . . . .	30
5.5 Série temporal dos dados de variação do nível do mar coletados <i>in situ</i> na estação de monitoramento, durante o período de estudo. . . . .	31

5.6	Matriz de correlação entre séries multivariadas estudadas: temperatura de brilho média ( $Tb_{IRMean}$ ), desvio padrão da temperatura de brilho ( $Tb_{IRStd}$ ), temperatura de brilho mínima ( $Tb_{IRMin}$ ), temperatura de brilho máxima ( $Tb_{IRMax}$ ), intensidade do vento médio ( $WindSpeedAvg$ ), intensidade da componente zonal do vento ( $Wind_U$ ), intensidade da componente meridional do vento ( $Wind_V$ ), nível do mar ( $SeaLevel$ ), período de pico ( $WTPeak$ ) e altura de onda ( $WH10/100$ ). . . . .	32
5.7	Boxplots mensais de altura de onda ( $WH10/100$ , em metros) e de intensidade do vento médio (em m/s). . . . .	33
5.8	Análise de autocorrelação para as séries de altura de onda, intensidade do vento médio, desvio padrão da temperatura de brilho e temperatura de brilho mínima. Os passos de tempo são equivalentes a 20 minutos. . . . .	34
5.9	Análise de autocorrelação para os dados de nível do mar. Passos de tempo são equivalentes a 20 minutos. . . . .	34
5.10	Séries temporais de valores médios e desvio padrão dos dados de altura de onda e de intensidade do vento médio para cada hora do dia. . . . .	36
5.11	Análise de correlação cruzada para as séries de médias diárias de altura de onda e de intensidade do vento médio. . . . .	36
5.12	Análise de correlação cruzada para as séries de médias diárias de altura de onda e de temperatura de brilho média. . . . .	37
5.13	Valores de média (linha contínua) e desvio padrão (área colorida) para o período de 5 horas antes até 5 horas depois de eventos de altura de ondas mais extremos, para os parâmetros estatísticos de temperatura de brilho do topo das nuvens (média, mínimo, máximo e desvio padrão). . . . .	38
5.14	Comparação entre os valores médios dos parâmetros estatísticos de temperatura de brilho (média, mínimo, máximo e desvio padrão) e os valores médios de altura de onda, para o período de 5 horas antes até 5 horas depois de eventos de altura de ondas mais extremas. . . . .	39
5.15	Comparação entre os valores médios dos parâmetros estatísticos de temperatura de brilho (média, mínimo, máximo e desvio padrão) e os valores médios de altura de onda, para o período de 5 horas antes até 5 horas depois de eventos de altura de ondas mais baixas. . . . .	39
5.16	Comparação entre os valores observados e previstos pelo modelo 3 (A) e pelo modelo 5 (B) para o ano de 2022. . . . .	41
5.17	Histograma dos erros das previsões de altura de ondas realizadas para o ano de 2022. Valores de erros positivos indicam superestimação do modelo, enquanto valores de erros negativos indicam subestimação do modelo. . . . .	43



5.18	Matriz de confusão das previsões de altura de ondas realizadas para o ano de 2022, considerando o limite operacional de 0,5 m. . . . .	44
5.19	Série temporal dos erros absolutos obtidos para as previsões geradas para o ano de 2022, com destaque para erros abaixo do percentil 10 e acima do percentil 90. . . . .	45
5.20	Comparação entre os momentos de "erros pequenos"(erro absoluto < percentil 10) e de "erros grandes"(erro absoluto > percentil 10) para cada passo de tempo anterior ao instante de rodada do modelo (t-4h até t) e para o passo de tempo de previsão (t+1h), para o ano de 2022. . . . .	46
5.21	Quantidade de ocorrências de "erros pequenos"(erro absoluto < percentil 10) e de "erros grandes"(erro absoluto > percentil 10) do modelo ao longo de cada mês do ano de 2022. . . . .	47
5.22	Distribuição dos valores de Altura das Ondas e de Intensidade dos Ventos em momentos de "erros pequenos"(erro absoluto < percentil 10) e de "erros grandes"(erro absoluto > percentil 10) do modelo para o ano de 2022. . . . .	47
5.23	Gráficos <i>boxplot</i> representando as alturas das ondas medidas e os erros absolutos das previsões para cada hora do dia. A primeira série de gráficos ilustra as distribuições das alturas das ondas ao longo do dia, enquanto a segunda série apresenta os valores de erro absoluto das previsões correspondentes. . . . .	48
5.24	Gráficos <i>boxplot</i> representando as alturas das ondas e as intensidades dos ventos medidos, bem como os erros absolutos das previsões de altura das ondas para cada hora do dia, para os meses de Março e Setembro de 2022. . . . .	48
5.25	Séries temporais de altura das ondas observada no terminal portuário e prevista pelo modelo, para os meses de março e de setembro de 2022. . . . .	50
5.26	Séries temporais de altura das ondas observada no terminal portuário e prevista pelo modelo, para os dias 13 e 18 de março de 2022, e para os dias 13 a 19 de setembro de 2022, com destaque para os momentos de maiores desvios das previsões em relação aos dados observados. . . . .	51
5.27	Séries temporais de altura das ondas observadas no terminal portuário e previstas pelo modelo, bem como da intensidade e da direção dos ventos observados no local, para o dia 15 de março de 2022, com destaque para o evento das 14:00 onde ocorreu maior divergência entre as alturas observada e prevista. . . . .	52

5.28	Séries temporais de altura das ondas observadas no terminal portuário e previstas pelo modelo, bem como da intensidade e da direção dos ventos observados no local, para os dias 19 e 20 de março de 2022, com destaque para o evento das 20:00 do dia 19, onde ocorreu maior divergência entre as alturas observada e prevista. . . . .	53
5.29	Gráficos do tipo <i>boxplot</i> das diferenças absolutas, em metros, dos dados de altura de onda observada no horário da previsão (t+1) e no horário imediatamente anterior à previsão (t) nos cenários de: maiores erros do modelo (erros > percentil 90); menores erros do modelo (erros < percentil 10); e erros críticos (previsão < 0,5 m e observação > 0,5 m). . . . .	54
5.30	Comparação entre as séries temporais de altura das ondas observadas no terminal portuário, previstas pelo modelo de LSTM e previstas pelo modelo numérico global MERCATOR para o mês de setembro de 2022. . . . .	55

## LISTA DE TABELAS

	<u>Pág.</u>
4.1 Variáveis do Conjunto de Entrada. . . . .	23
5.1 Estatística descritiva dos dados coletados <i>in situ</i> na estação de monitoramento ao longo do período de estudo. . . . .	31
5.2 Resultados dos experimentos com Optuna, variando o conjunto de entrada e os hiper-parâmetros. . . . .	40
5.3 Métricas de desempenho dos modelos 3 e 5 para o ano de 2022. . . . .	41
5.4 Métricas dos maiores e menores erros do modelo para o ano de 2022. . .	45
5.5 Métricas das previsões para 1h, 2h e 3h no futuro para o ano de 2022. . .	55



## LISTA DE ABREVIATURAS E SIGLAS

ABI	–	<i>Advanced Baseline Imager</i>
ADCP	–	<i>Acoustic Doppler Current Profiler</i>
CPTEC	–	Centro de Previsão de Tempo e Estudos Climáticos
FAR	–	<i>False Alarm Rate</i>
IA	–	Inteligência Artificial
LSTM	–	<i>Long short-term memory</i>
MSE	–	<i>Mean Square Error</i>
NOAA	–	<i>National Oceanic and Atmospheric Administration</i>
POD	–	<i>Probability of Detection</i>
P10	–	Percentil de 10%
P66	–	Percentil de 66%
P90	–	Percentil de 90%
P99	–	Percentil de 99%
RNA	–	Redes Neurais Artificiais
RNN	–	<i>Recurrent Nerual Network</i>
RMSE	–	<i>Root Mean Square Error</i>
SISMO	–	Sistema de Informações Meteo-Oceanográficas em Tempo Real
$Tb_{IR}$	–	Temperatura de Brilho
WH10/100	–	Altura das ondas 10% mais altas
ZCIT	–	Zona de Convergência Intertropical



## SUMÁRIO

	<u>Pág.</u>
<b>1 INTRODUÇÃO</b> . . . . .	<b>1</b>
<b>2 OBJETIVOS</b> . . . . .	<b>5</b>
<b>3 REFERENCIAL TEÓRICO</b> . . . . .	<b>7</b>
3.1 Ondas no mar . . . . .	7
3.2 Métodos de medição de ondas . . . . .	9
3.3 Métodos de previsão de ondas . . . . .	11
3.4 Aplicações de algoritmos de aprendizado de máquina na previsão de ondas	13
3.5 Long Short-Term Memory (LSTM) . . . . .	14
<b>4 DADOS E MÉTODOS</b> . . . . .	<b>17</b>
4.1 Dados observados <i>in situ</i> . . . . .	17
4.2 Dados obtidos por imagens de satélite . . . . .	18
4.3 Pré-processamento dos dados . . . . .	19
4.4 Seleção de variáveis, construção dos modelos e otimização de hiperparâ- metros . . . . .	21
4.5 Métricas de avaliação dos resultados . . . . .	23
4.6 Análise do erro do melhor modelo . . . . .	24
<b>5 RESULTADOS E DISCUSSÕES</b> . . . . .	<b>27</b>
5.1 Análise exploratória do conjunto de dados . . . . .	27
5.2 Resultados dos experimentos com a LSTM . . . . .	40
5.3 Resultados do modelo preditivo de altura de ondas . . . . .	42
5.4 Teste de sensibilidade para <i>leadtime</i> superiores a 1 hora . . . . .	55
<b>6 CONCLUSÕES</b> . . . . .	<b>57</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> . . . . .	<b>61</b>





# 1 INTRODUÇÃO

As regiões costeiras possuem importância singular, tanto em aspectos econômicos como em aspectos ecológicos e ambientais. Nessas regiões são geralmente realizadas atividades de pesca, navegação, aquicultura e turismo. Além disso, as zonas costeiras podem apresentar atividades portuárias e industriais. Essas regiões são importantes áreas de ocupação e de uso humano, o que pode ser verificado pelo fato de cerca de dois terços da população humana habitá-las. Portanto, é evidente a necessidade de monitoramento ambiental e gerenciamento dos seus recursos naturais, para que a ocupação e a exploração econômica sejam efetuadas de forma sustentável (MORAES, 2007).

Ao longo dos anos o INPE vem desenvolvendo pesquisas e produtos meteorológicos e oceanográficos que são de suma importância para a tomada de decisões por diversos setores da sociedade. Em 2016, por exemplo, o INPE realizou, sob a coordenação do CPTEC, previsões e serviços meteorológicos para os Jogos Olímpicos e Paraolímpicos com êxito. Atualmente, diversos terminais portuários mantêm monitoramento contínuo de parâmetros ambientais, com coleta e armazenamento de dados meteorológicos e oceanográficos. Algumas empresas do setor privado são contratadas pelos terminais portuários para realização destes serviços. A HidroMares, por exemplo, desenvolveu o Sistema de Informações Meteo-Oceanográficas em Tempo Real (SISMO®) para fornecer informações meteo-oceanográficas para agentes e organizações envolvidas em manobras portuárias, subsidiando tomadas de decisão relacionadas à operação dos portos brasileiros. Logo, como susodito, observa-se que tanto o setor privado como o público necessitam de maiores estudos para melhor atender as demandas da sociedade nesse setor tão produtivo e importante para o desenvolvimento socioeconômico brasileiro.

Para que o gerenciamento das zonas costeiras ocorra em bases científicas e técnicas adequadas, são necessários ainda estudos que permitam a compreensão e a previsão dos fenômenos naturais atuantes nesses ambientes. Dentre as ferramentas que auxiliam na compreensão e na previsão dos processos meteorológicos e hidrodinâmicos costeiros, podemos citar a modelagem matemática (RAMMING; KOWALIK, 1980) e algoritmos de Inteligência Artificial (IA) (AHMAD, 2019), validados por comparações com dados obtidos em observações na região de estudo.

Os métodos tradicionais de previsão de ondas, principalmente os modelos numéricos globais, podem não apresentar os processos físicos com resolução espacial e temporal necessária para às tomadas de decisões em portos, uma vez que estes pre-

cisam de previsões não só precisas, mas rápidas e de curto prazo, para pontos/locais específicos de interesse para realização de manobras com embarcações e demais atividades portuárias. As técnicas de IA são robustas e, geralmente, precisas o suficiente, apresentando vantagens em relação aos métodos tradicionais de modelagem física, principalmente, no quesito de menor custo computacional de operação ((JAMES et al., 2018),(MAXWELL et al., 2018)). Além disso, as técnicas de IA possibilitam a integração de recursos de aprendizado de máquina em sistemas de modelagem orientados por dados e têm se mostrado uma ferramenta poderosa para analisar dados oceanográficos e de tempo com alta precisão de forma eficiente ((AHMAD, 2019)).

Algoritmos de IA têm sido aplicados em diferentes estudos nas áreas de oceanografia e meteorologia, como na previsão de variações do nível do mar (HSIEH, 2009), previsão de altura de ondas ((ZAMANI et al., 2008); (JAMES et al., 2018); (WEI; CHANG, 2021); (MANDAL; PRABAHARAN, 2006)), previsão de ventos (CHANG et al., 2014), reconstrução de dados faltantes sem séries temporais de altura de onda (ALEXANDRE et al., 2015), detecção de derramamentos de óleos (FRATE et al., 2000) e entre outras aplicações.

A previsão da altura de ondas desempenha um papel crucial na segurança da navegação e nas operações portuárias. As condições das ondas impactam diversas atividades marinhas e podem ser um fator limitante para a execução de determinadas operações. Portanto, estudos voltados para a previsão da altura de ondas são de suma importância para assegurar a eficiência e a segurança dessas atividades. (FAN et al., 2020). Neste contexto, este trabalho tem como objetivo explorar a aplicação de técnicas de IA, mais especificamente de Redes Neurais Artificiais (RNAs), para a previsão de altura das ondas no Terminal Portuário de Barcarena, localizado no estado do Pará, a partir do uso de dados coletados *in situ* através do SISMO® e de dados extraídos das imagens do satélite (e.g. GOES16). A principal finalidade é desenvolver um modelo capaz de compreender e reproduzir os principais padrões e tendências da série temporal de altura de ondas observadas no terminal e gerar previsões de curto prazo (1 hora) com assertividade suficiente para subsidiar operações portuárias.

Uma ferramenta de previsão como esta pode trazer inúmeros benefícios, tanto para o planejamento das atividades portuárias, com para promover maior segurança durante na execução dos trabalhos, evitando tanto acidentes de trabalho como ambientais. Dentre os principais desafios para o estudo proposto neste trabalho, podemos destacar:

- a) Processamento, tratamento e manipulação de séries temporais de dados brutos, com eventuais dados faltantes, bem como informações provenientes de satélites e de modelagem numérica, visto que o sucesso dos modelos de IA dependem fortemente da adequação do conjunto de dados usados para o treinamento do modelo;
- b) Aplicação destes dados aos diferentes algoritmos e metodologias de IA existentes, em busca daqueles que consigam melhor compreender e reproduzir os principais padrões e tendências das séries temporais e, conseqüentemente, os principais fenômenos locais da área de estudo;
- c) Encontrar as ferramentas corretas para obter modelos preditivos de alta resolução temporal e que gerem previsões de alta precisão, com valores próximos aos reais.



## 2 OBJETIVOS

A hipótese científica adotada para este trabalho é que as técnicas de IA já utilizadas em outros estudos em diferentes regiões do globo podem ser utilizadas em regiões costeiras brasileiras, como uma ferramenta para a previsão de curto prazo (*nowcasting*) de altura de onda, representando os fenômenos locais com grau de confiabilidade suficiente para operações portuárias. Para comprovar esta hipótese, o objetivo principal deste estudo foi explorar a aplicação de técnicas de aprendizado de máquina, mais especificamente as RNAs, para a previsão de altura das ondas no Terminal Portuário de Barcarena, localizado no estado do Pará. Os dados meteo-oceanográficos coletados *in situ* através do SISMO® (Sistema de Informações Meteo-Oceanográficas em Tempo Real, desenvolvido pela HidroMares) e dados extraídos de imagens do satélite GOES16 foram utilizados a fim de desenvolver um modelo capaz de compreender e reproduzir os principais padrões e tendências da série temporal de altura de ondas observadas no terminal, e, por conseguinte, realizar *nowcasting*.

Os objetivos específicos para alcançar os desafios inerentes a pesquisa foram:

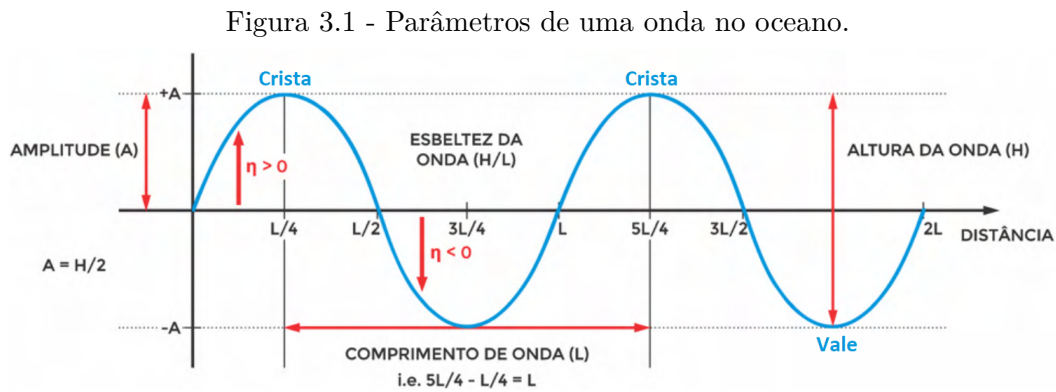
- a) Realizar uma análise exploratória de todo o conjunto de dados a fim de compreender os principais fenômenos meteo-oceanográficos atuantes na área de estudo, bem como realizar o pré-processamento e tratamento necessários destes dados;
- b) Aplicar estes dados a uma RNA bem sucedida na área, realizando diferentes experimentos em busca do modelo que consiga melhor compreender e reproduzir os principais padrões e tendências das séries temporais de altura de ondas;
- c) Avaliar o desempenho do modelo e sua capacidade preditiva em diferentes cenários, simulando a operacionalização desta ferramenta em um sistema em tempo real e comparando com ferramentas já existente e vastamente utilizadas.



### 3 REFERENCIAL TEÓRICO

#### 3.1 Ondas no mar

A superfície do mar apresenta variações contínuas de nível, sendo alternadamente erguida e rebaixada, em relação a uma posição média (HARARI, 2021). Este fenômeno das ondas é geralmente descrito usando os parâmetros apresentados na Figura 3.1.



Fonte: Adaptado de Harari (2021).

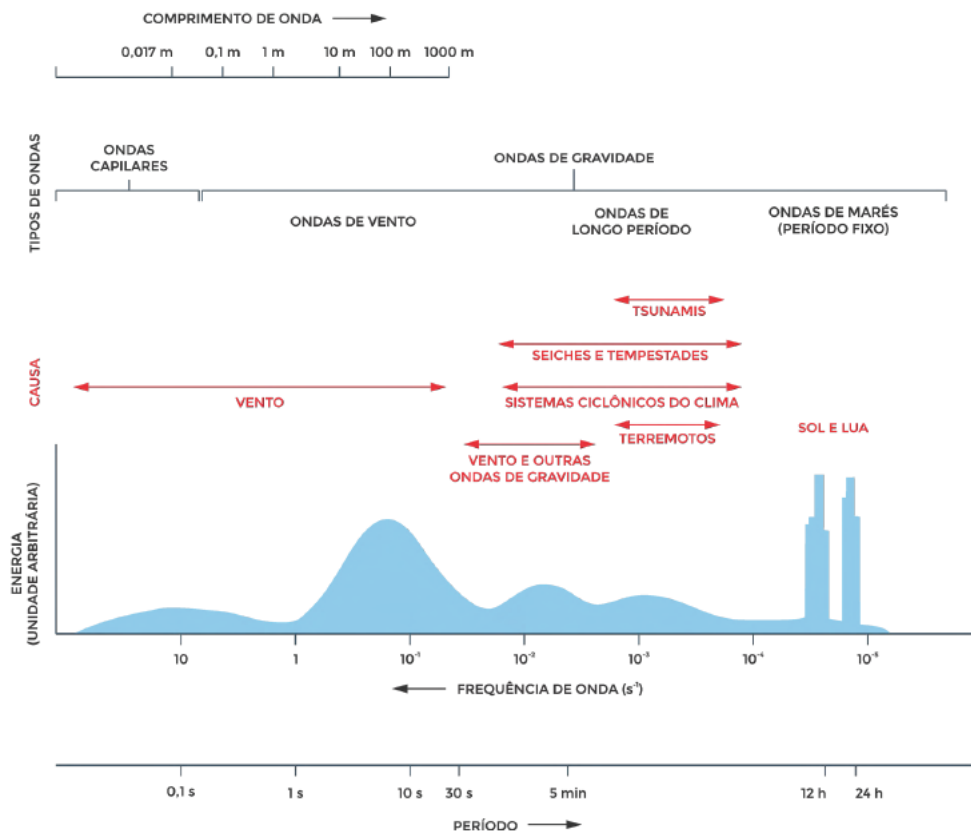
A crista da onda é a parte mais alta da onda sobre o nível médio da água, e o vale da onda é a depressão entre duas cristas consecutivas. A altura da onda é definida como a distância vertical entre a crista da onda e seu vale adjacente. Já o comprimento da onda é a distância horizontal entre duas cristas ou entre dois vales consecutivos. O tempo que uma onda leva para percorrer uma distância igual ao seu comprimento de onda é conhecido como o período da onda. A frequência da onda, por sua vez, é o número de ondas que passam, por segundo, por um ponto fixo (GARRISON, 2002).

Uma onda pode ser definida como um sinal que se propaga, tipicamente se movendo em uma taxa distinta ao do movimento do meio (PEDLOSKY, 2003). As ondas são resultantes da combinação entre forças restauradoras que tendem a trazer o sistema ao estado de repouso inicial e da inércia que faz o sistema ultrapassar a posição que se encontrava quando em repouso. No mar, os principais efeitos geradores de ondas - também chamados de forças perturbadoras - são os ventos, mudanças na pressão atmosférica, erupções vulcânicas, falhas no leito marinho, deslizamentos de terra e a atração gravitacional do Sol e da Lua. A atuação destas últimas forças gera um tipo especial de ondas, as ondas de maré. Já as principais forças que controlam as características das ondas no mar, isto é, as forças restauradoras, são:

tensão da superfície da água, força da gravidade e força de Coriolis (uma força aparente associada à rotação da Terra; importante somente para ondas muito longas) (GARRISON, 2002).

As ondas no mar são classificadas de acordo com suas características física, como seu comprimento e suas forças perturbadoras e restauradoras. A altura da onda não é usada para classificação pois esta pode variar de acordo com diversos fatores, como a profundidade da coluna d'água, a interferência entre outras ondas, a velocidade e a duração do vento, entre outros fatores. A Figura 3.2 mostra a relação entre estas forçantes, o período e a quantidade relativa de energia presente na superfície do oceano para cada tipo de onda. Podemos observar que a maior parte da energia das ondas está concentrada nas ondas de gravidade (isto é, ondas que possuem a gravidade como força restauradora) geradas pelo vento, ou seja, ondas formadas pela transferência da energia do vento para a água, e nas ondas de marés (GARRISON, 2002).

Figura 3.2 - Espectro de energia das ondas no mar. Tipos de ondas de superfície e suas características, como energia, comprimento, período e forçantes atuantes.



Fonte: Harari (2021).



Com relação à altura das ondas, um parâmetro estatístico comumente utilizado em estudos oceanográficos é a altura significativa da onda. Este parâmetro corresponde à média das maiores ondas registradas em uma série de observações, sendo estas um terço do total de ondas observadas. Outro parâmetro estatístico importante é a altura das ondas 10% mais altas, a qual é calculada pela média dos 10% mais altos das ondas observadas. De modo geral, quanto mais forte sopram os ventos, maior a altura das ondas geradas. A relação entre velocidade do vento, altura das ondas e estado do mar pode ser descrita pela Escala de Beaufort (Figura 3.3).

As regiões sobre as quais os ventos sopram são denominadas de pistas de ventos. Nessas regiões, ondas de diferentes amplitudes, comprimentos e períodos são formadas, dependendo da intensidade e duração dos ventos, bem como do tamanho da pista e interações com outras ondas. Ondas de maiores comprimentos se propagam mais rapidamente, se afastando da região de formação. Estas são as ondas conhecidas como *Swell*, e possuem períodos de 10 a 30 segundos. As ondas de picos irregulares existentes no local de atuação dos ventos são chamadas de ondas geradas pelo vento local ou de Vagas, e possuem períodos menores de 10 segundos (HARARI, 2021). Nas proximidades do Terminal Portuário de Barcarena, área de estudo deste trabalho, localizada na Baía de Marajó, estado do Pará, na costa Norte do Brasil, predominam a ação de ondas geradas pelo vento local (COTRIM et al., 2022).

### 3.2 Métodos de medição de ondas

As medições de ondas no mar são essenciais para a compreensão dos processos oceânicos e para a segurança das atividades marítimas. O método mais tradicional é através de equipamentos de medição de variações da altura da superfície do mar, como comentado no Capítulo anterior. Estas medições podem ser feitas por marégrafos ou por sensores de pressão instalados no fundo do mar, os quais registram as variações de pressão causadas pelas ondas que passam acima. Entretanto, equipamentos desenvolvidos especificamente para a medição de ondas possuem características otimizadas para esta aplicação, permitindo a medição não somente da altura e período das ondas, mas também de espectros direcionais e parâmetros estatísticos de ondas já processados (GARRISON, 2002).

Entre os métodos de medição mais comuns, destacam-se as boias de superfície (*Wave Riders*), presas ao fundo por cabos elásticos, que utilizam acelerômetros para registrar o movimento vertical e horizontal da superfície do mar, fornecendo dados contínuos sobre a altura significativa das ondas, seu período e a direção de propagação. Outra técnica amplamente utilizada é através de medidores de velocidade orbital,

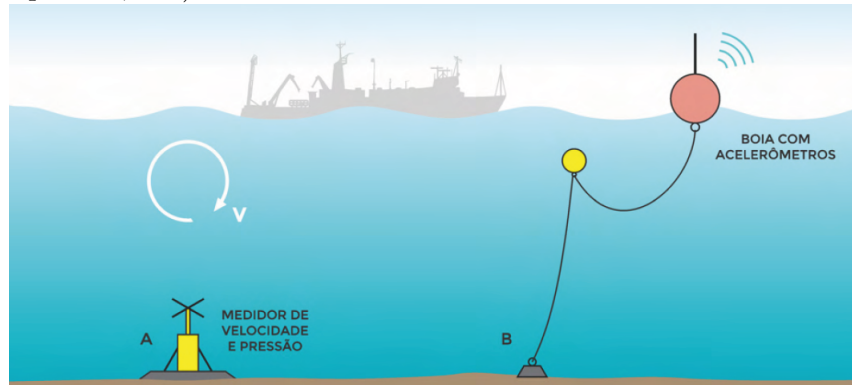
Figura 3.3 - Escala Beaufort com a relação entre a velocidade do vento, estado do mar e altura significativa das ondas.

Beaufort N°	Nome	Velocidade do vento (m/s)	Estado da superfície do mar	Altura significativa das ondas (m)
0	Calmaria	0,0 - 0,2	Mar como um espelho	0
1	Vento muito brando	0,3 - 0,5	Marolas na água em forma de escamas e sem cristas com espuma	0,1 - 0,2
2	Brisa leve	1,6 - 3,3	Pequenas ondulações, cristas translúcidas, mas não quebram	0,3 - 0,5
3	Brisa suave	3,4 - 5,4	Ondulações maiores, as cristas começam a quebrar, espuma espalhada	0,6 - 1
4	Brisa moderada	5,5 - 7,9	Pequenas ondas se tornando mais longas, espuma razoavelmente frequente	1,5
5	Brisa fresca	8,0 - 10,7	Ondas moderadas, tomando formas cada vez mais longas, muita espuma e chances de borrião	2
6	Brisa forte	10,8 - 13,8	Ondas grandes se formando, cristas de espuma branca se espalham por todos os lados e grande probabilidade de borriões	3,5
7	Vento moderado	13,9 - 17,1	Mar revolto, espuma branca de arrebentação, o vento "arranca" camadas de espuma	5
8	Vento fresco	17,2 - 20,7	Ondas razoavelmente altas e com grande comprimento, faixas com espuma branca, fraca arrebentação	7,5
9	Vento forte	20,8 - 24,4	Ondas altas, faixas de espuma densa, borriões começam a atrapalhar a visibilidade	9,5
10	Vento muito forte	24,5 - 28,4	Ondas muito altas, o vento "arranca" faixas de espuma, a superfície do mar fica branca; a visibilidade é afetada	12
11	Tempestade	28,5 - 32,7	Ondas excepcionalmente altas, o mar está coberto por longas faixas de espuma, navios de pequeno e médio porte somem no cavado das ondas, visibilidade bastante afetada	15
12	Furacão	≥ 32,7	Ar repleto de espuma e borriões, mar completamente branco e borriões, visibilidade bastante reduzida	≥ 15

Fonte: Harari (2021).

instrumentos fixos sobre o fundo marinho que medem continuamente a velocidade da água e a pressão hidrostática naquele ponto (Figura 3.4). Além desses métodos, altímetros e radares de sensoriamento remoto também são capazes de realizar medições de ondas. (HARARI, 2021).

Figura 3.4 - Esquema ilustrativo de operação de ondógrafos: A) medição pelo orbital e pressão; e B) boia com acelerômetros.



Fonte: Harari (2021).

### 3.3 Métodos de previsão de ondas

Ondas geradas pelo vento tem sido estudadas por décadas. Os primeiros modelos de previsão de ondas foram equações empíricas que relacionavam a altura e o período de ondas com a velocidade do vento, a duração do vento e o comprimento da pista. Com os avanços tecnológicos no meio computacional e com o desenvolvimento dos métodos numéricos, as equações empíricas puderam ser implementadas em modelos numéricos, e com elas o conhecimento de fenômenos físicos, para estimativa de ondas geradas pelo vento tanto em região oceânica como em região costeira (WEI, 2021), os quais incluem até mesmo informações completas do espectro de onda, como por exemplo os modelos Wavewatch (TOLMAN, 1991) e SWAN (BOOIJ et al., 1996). No entanto, algoritmos de IA estão ganhando cada vez mais visibilidade com o aumento das bibliotecas computacionais públicas disponíveis, tornando possível a análise de grandes quantidades de dados para diversas aplicações baseadas em informações históricas (MINUZZI; FARINA, 2023).

Desta forma, os métodos de previsão de ondas podem ser divididos em duas diferentes abordagens: modelos numéricos, baseados em equações matemáticas e princípios físicos, os quais utilizam a técnica de modelagem numérica; e modelos orientados a dados (*data-driven models*), os quais utilizam dados empíricos e técnicas estatísticas, como as RNAs (WANG; YING, 2023).

A maior limitação da aplicação dos modelos numéricos para previsão de ondas, principalmente para regiões costeiras, está associada ao custo computacional. Os modelos numéricos exigem instalações computacionais de alto desempenho e elevado custo,

os quais não são acessíveis para uma ampla comunidade de pesquisa e operação. Além disso, a acurácia dos modelos numéricos está diretamente associada a resolução espacial e temporal utilizada. Uma resolução espacial ou temporal inadequada pode levar a uma modelagem imprecisa de fenômenos locais, particularmente em áreas costeiras e durante eventos extremos. A simplificação da linha de costa e da batimetria, por exemplo, comumente adotada em modelos de escala global, pode reduzir a acurácia das previsões para locais específicos da zona costeira (KOMEN, 1994);(TOLMAN et al., 2009).

Com os avanços da tecnologia de IA na última década, modelos de aprendizado de máquina tem sido aplicados em estudos de processos costeiros, incluindo a previsão de ondas. As técnicas de IA são robustas e apresentam vantagens em relação aos métodos tradicionais de modelagem física, principalmente no quesito de menor custo computacional ((JAMES et al., 2018),(MAXWELL et al., 2018)). Ainda, os avanços nas tecnologias de IA estão possibilitando a integração de recursos de aprendizado de máquina em sistema de modelagem numérica para preencher lacunas e diminuir as demandas de especialistas humanos (CHAU, 2006).

O estudo realizado por Boukabara et al. (2019) mostrou que as abordagens de IA têm o potencial de atender aos requisitos crescentes para produtos de previsão de curto e médio prazo, incluindo previsões de tempo e projeções climáticas. Na oceanografia, também temos estudos que evidenciam o uso de IA. Peres et al. (2015), por exemplo, desenvolveram um método para aumento de séries temporais de dados de altura significativa de ondas utilizando RNAs, a partir de dados de ventos de reanálises, que foi capaz de gerar reconstruções climatológicas mesma qualidade, e em alguns casos com qualidade superior, àquelas baseadas em modelos físicos (ECMWF ERA-1 e CFSR). Outro exemplo é o estudo realizado por Minuzzi e Farina (2023) para previsão de altura significativa de ondas. Os autores obtiveram acurácia próxima de 87%, mostrando que a técnica baseada em dados pode ser usada como alternativa para os modelos numéricos de alto custo.

Na busca por maior precisão preditiva, modelos híbridos baseados em aprendizado de máquina juntamente com modelos numéricos também estão sendo amplamente aplicados no campo da previsão de altura de onda (WANG; YING, 2023). Dentre os estudos que utilizaram tal abordagem, podemos citar o de Gracia et al. (2021), no qual os autores aplicaram modelos de aprendizado de máquina para melhorar a acurácia do atual sistema de previsão da *Spanish agency Puertos del Estado (PdE)*, e conseguiram reduzir os erros de previsão em média em 36%, demonstrando o

potencial ganho em combinar modelos numéricos e aprendizado de máquina. Outro exemplo é o estudo de [Costa e Soares \(2023\)](#), no qual RNAs foram utilizadas com sucesso para redução de *bias* de reanálise climática.

No Capítulo seguinte, são apresentados mais exemplos de estudos que aplicam técnicas RNAs na previsão de ondas.

### 3.4 Aplicações de algoritmos de aprendizado de máquina na previsão de ondas

Conforme mencionado nos Capítulos anteriores, as técnicas de aprendizado de máquina têm sido aplicadas em diferentes estudos nas áreas de oceanografia. Nesta seção, alguns estudos e aplicações relacionados à previsão de ondas serão apresentados.

Um dos estudos pioneiros no assunto foi desenvolvido por [Deo e Naidu \(1998\)](#), que usaram redes neurais do tipo *feed-forward* para prever alturas de ondas com base em dados observados em uma estação de medição de ondas, a partir do princípio de que as RNAs têm a capacidade de reconhecer padrões ocultos nos dados e gerar previsões de valores futuros, e mostraram que as redes neurais geraram resultados de qualidade superior ao de modelos autorregressivos.

O estudo de [Zamani et al. \(2008\)](#) mostrou que modelos orientados a dados, especialmente as RNAs, também do tipo *feed-forward*, podem ser efetivos para previsões de curto prazo de altura significativa de ondas geradas pelo vento, a partir de dados coletados *in situ*, gerando previsões mais assertivas do que modelos de aprendizagem baseada em instâncias (*Instance-Based Learning*). Já [Mandal e Prabakaran \(2006\)](#) utilizaram redes neurais recorrentes (RNN, do inglês *Recurrent Neural Network*) - isto é, um tipo de RNA otimizado para sequências de dados, permitindo que informações anteriores influenciem as saídas atuais por meio de conexões recorrentes - para previsão de altura de ondas, e encontraram resultados superiores aos obtidos com a rede *feed-forward*.

Nos últimos anos, tem-se observado um crescente interesse no uso de um tipo específico de RNN nos estudos voltados para previsões de altura das ondas: a rede LSTM (*Long Short Term Memory*; ([HOCHREITER; SCHMIDHUBER, 1997](#))), como descrito por [Wang e Ying \(2023\)](#). Pesquisadores como [Wei \(2021\)](#), [Abdullah et al. \(2022\)](#), [Fan et al. \(2020\)](#), e entre outros, adotaram redes LSTM em seus estudos, alcançando previsões acuradas.

O modelo desenvolvido por [Wei \(2021\)](#) para a região costeira dos EUA, utilizando dados coletados em quatro boias da *National Data Buoy Center (NDBC)*, da NOAA (*National Oceanic and Atmospheric Administration*), e a rede LSTM, obteve resultados precisos para previsões de curto prazo (1 a 6 horas), tanto para a altura significativa de onda, quanto para o período médio de onda associado a eventos de tempestade. [Fan et al. \(2020\)](#) testaram diferentes técnicas para a previsão de altura de ondas, utilizando dados provenientes de 10 estações de monitoramento instaladas em diferentes regiões do globo terrestre, situadas em ambientes com condições meteo-oceanográficas distintas. De maneira semelhante, o estudo revelou que a rede LSTM pode gerar previsões estáveis e acuradas para até 6 horas, sendo os resultados otimizados na primeira hora de previsão. Os resultados mostraram, ainda, que a assertividades das previsões com a LSTM foram superiores às das outras técnicas testadas, como *Random Forest Algorithm*, *Back Propagation Neural Network*, *Extreme Learning Machine*, *Support Vector Machine* e *Residual Network*.

O estudo de [Abdullah et al. \(2022\)](#), aplicado em diferentes localidades da Indonésia, mostrou que os resultados das previsões de altura significativa de onda são mais acurados quando a previsão é feita para apenas um passo de tempo do futuro, quando em comparação a previsões mais longas e de mais passos de tempo (como 12h, 24h e 48h).

Os estudos mencionados acima apontam que a LSTM é o método mais recomendado para previsões de ondas utilizando dados coletados *in situ*. Desta forma, esta também foi a técnica escolhida para ser explorada no presente trabalho. A seguir uma descrição mais detalhada é apresentada para essa RNA em específico.

### 3.5 Long Short-Term Memory (LSTM)

As RNAs são modelos de aprendizado de máquina inspirados em redes neurais biológicas – o cérebro humano - o qual possui complexas teias de neurônios interconectados. Da mesma forma que nosso cérebro, as RNAs são compostas por neurônios que se ligam uns aos outros formando, de fato, uma rede. O tipo de RNA mais comum são as redes neurais feed-forward, onde os conjuntos de neurônios são organizados em camadas: uma camada de entrada, uma camada de saída e pelo menos uma camada intermediária oculta. Contudo, existem redes que permitem criar uma certa "memória", por exemplo, as RNNs. Neste tipo de rede, as conexões dos neurônios formam uma sequência temporal, agindo como uma memória interna que permite considerar tendências temporais em suas previsões. A principal limitação das RNNs é que são restritas a “olhar para trás no tempo” para poucos passos de tempo. Esse

problema foi resolvido com a LSTM, a qual é um tipo de RNN que possui células de memória de longo prazo (STAUEMEYER; MORRIS, 2019).

A estrutura da LSTM é formada por 3 portas (*gates*), onde cada uma recebe informações de diferentes prazos no tempo, selecionando o que deve ou não ser “esquecido” pela rede:

- *forget gate f*: controla quais partes do estado de longo prazo devem ser deletados. Recebe como entrada a saída do estado anterior, que é somado a um bias e aplicado à uma função sigmoide ( $\sigma$ ) para decidir sobre o que deve ser eliminado (valores próximos de 0) ou mantido (valores próximos de 1), com isso a rede simula uma memória de longo prazo que mantém somente as informações relevantes. O forget gate é definido pela Equação 3.1, onde  $W_i$  são os pesos atribuídos aos seus neurônios,  $h_{t-1}$  é a saída do estado anterior,  $x_t$  é a entrada do estado atual e  $b_i$  é um bias somado à matriz e pesos.

$$f = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (3.1)$$

- *input gate i*: controla quais partes da saída devem ser adicionadas ao estado de longo prazo. Recebe como entrada a saída do estado recém calculado pelo forget gate e submete de forma individual para as funções de ativação sigmoide ( $\sigma$ ) e tangente hiperbólica ( $\tanh$ ), o resultado obtido é multiplicado e enviado para o próximo estado. O input gate pode ser definido pela Equação 3.2.

$$i = \sigma(W_i[h_{t-1}, x_t] + b_i) * \tanh(W_i[h_{t-1}, x_t] + b_i) \quad (3.2)$$

- *cell state c*: Recebe a soma das informações geradas pelo *forget gate* e *input gate*, sendo  $c = f + i$ .

- *output gate o*: controla quais partes do estado de longo prazo devem ser lidas e geradas neste intervalo de tempo. Recebe como entrada a saída gerada pelo cell state aplicada à uma função de ativação tangente hiperbólica ( $\tanh$ ) multiplicado pela saída gerada pelo input gate aplicado à função de ativação sigmoide ( $\sigma$ ). É aplicada uma função de sigmoide no resultado que é enviado para a próxima camada da rede. O output gate pode ser definido pela 3.3.

$$o = \sigma(\tanh(c) * \sigma(i)) \quad (3.3)$$

Assim, a cada intervalo de tempo, algumas memórias são deletadas e algumas são adicionadas. Após a operação de adição (*input gate*), o estado de longo prazo é copiado e passado pela função  $\tanh$ . Em seguida, o resultado é filtrado pelo *output gate*. Desta forma, uma rede LSTM pode aprender a reconhecer uma entrada importante, armazená-la no estado de longo prazo, preservá-la e extraí-la quando necessário. Por este motivo a rede neural LSTM tem sido bem-sucedida na identificação de padrões de longo prazo em séries temporais e previsões de altura de ondas (GÉRON, 2019),(FAN et al., 2020).

Desta forma, neste trabalho também foi decidido utilizar a rede LSTM. Como citado anteriormente, diversos trabalhos utilizaram com sucesso tal técnica para construção de modelos de previsão de onda em locais específicos do globo. Logo, espera-se que, ao aplicar uma técnica de sucesso, seja possível representar com boa qualidade as condições observadas para a região de estudo a partir dos dados disponíveis, trazendo, assim, um estudo que mostra a sua aplicação para uma região que ainda carece de análises mais profundas sobre sua aplicação. Ainda, foi esperado entender com mais detalhes as limitações desta técnica, o que é de suma importância quando para sua utilização em um ambiente operacional.



## 4 DADOS E MÉTODOS

Este capítulo aborda os dados utilizados na pesquisa e os métodos aplicados para alcançar os objetivos supracitados. Assim como a definição da área de estudos e pré-processamento dos dados.

### 4.1 Dados observados *in situ*

O conjunto de dados observados foi coletado *in situ* na estação de monitoramento meteo-oceanográfico em tempo-real instalada no Terminal Portuário de Barcarena (Figura 4.2 e 4.1), localizado no estado do Pará, e mantida pela empresa HidroMares. A estação é composta por um anemômetro, o qual realiza a medição de intensidade e direção de ventos com resolução temporal de 5 minutos, e por um ADCP (sigla em inglês para *Acoustic Doppler Current Profiler*) horizontal, com sensor de pressão embutido (Figura 4.3), o qual realiza a medição de intensidade e direção das correntes, temperatura da água no fundo, pressão (oscilações do nível do mar) e período de pico e altura das ondas, com resolução temporal de 20 minutos.

Figura 4.1 - Localização da estação de monitoramento meteo-oceanográfico.



Foram utilizados neste trabalho os dados coletados no período de 01/01/2020 até 31/12/2022, totalizando 3 anos de dados.

Figura 4.2 - Apresentação da estação de monitoramento meteo-oceanográfico em tempo-real, instalada e mantida pela HidroMares no Terminal Portuário de Barcaarena.

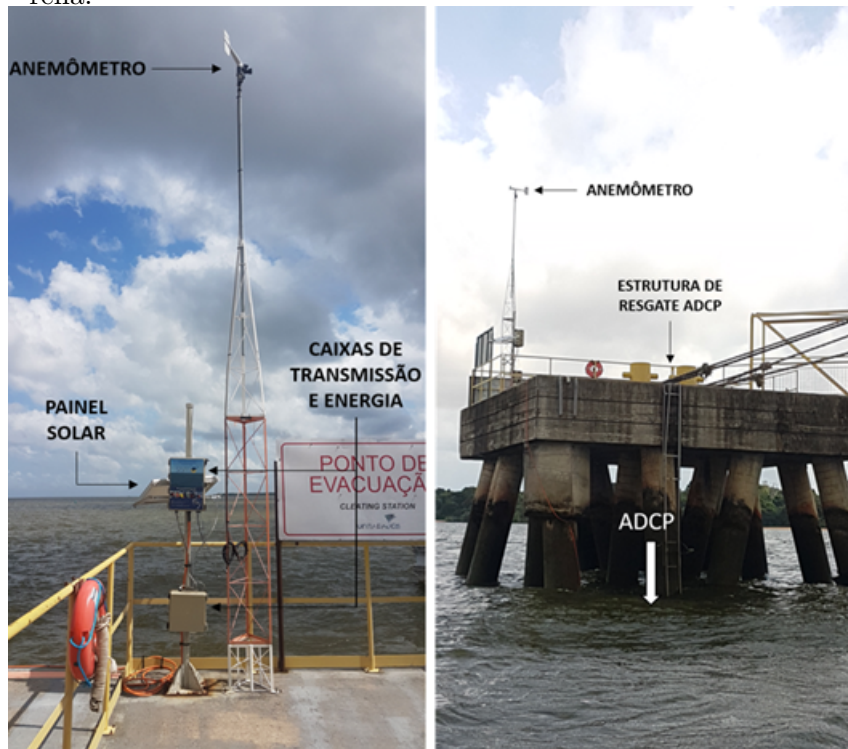


Figura 4.3 - Ilustração do perfilador acústico de correntes com sensor de pressão embutido - ADCP (*Acoustic Doppler current profiler*, modelo SL 500, que compõe a estação.



## 4.2 Dados obtidos por imagens de satélite

Além dos dados observados coletados *in situ*, também foram utilizados dados extraídos de imagens do sensor ABI (*Advanced Baseline Imager*), canal 13, do Satélite Geoestacionário GOES-16 (SCHMIT *et al.*, 2017), operado pela NASA e pela NOAA dos EUA. O GOES-16 orbita a 75,2°W e fornece imagens de alta resolução espacial (2 km no espectro do infravermelho janela) e temporal (10 min) através de 16 bandas espectrais em comprimentos de onda visíveis e infravermelhos através do sensor ABI. O canal 13 corresponde ao comprimento de onda central de 10,3  $\mu\text{m}$  (infravermelho

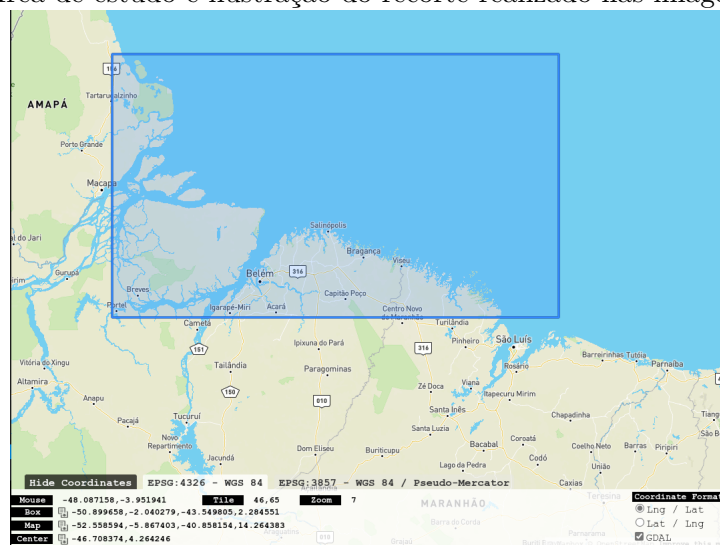
janela), a partir do qual é possível obter dados de temperatura de brilho (K), aqui definido como  $Tb_{IR}$ . Assim como para os dados coletados *in situ*, o período de dados utilizados neste trabalho foi de 01/01/2020 até 31/12/2022.

O objetivo de utilizar os dados de  $Tb_{IR}$ , obtidos pelas imagens do Satélite GOES16, foi inserir indicações de sistemas meteorológicos (e.g. nuvens precipitantes ou não e possíveis tempestades) e ambiente (e.g. gradiente de temperatura espacial e temporal) que podem estar relacionados com o processo de geração de altura de ondas no local.

### 4.3 Pré-processamento dos dados

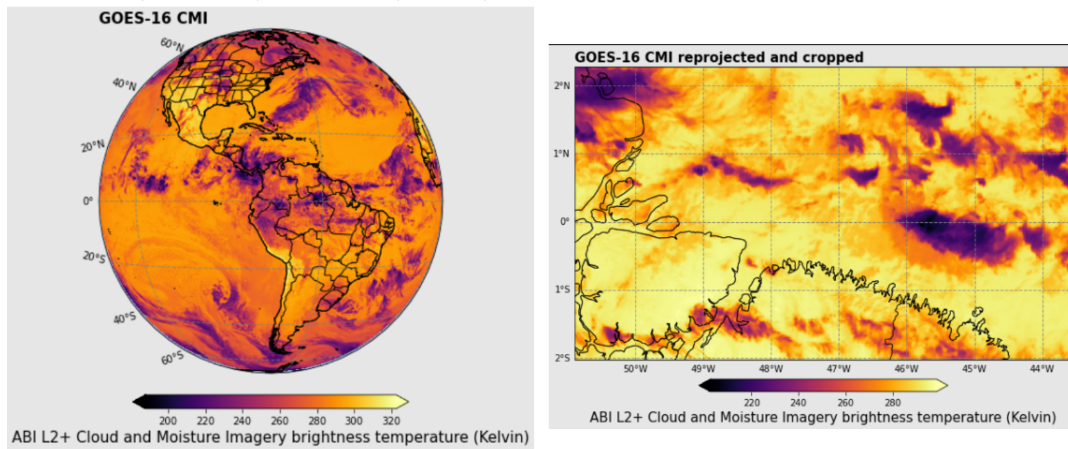
As imagens obtidas pelo GOES16 foram inicialmente processadas visando delimitar a região de interesse e gerar séries temporais para o período de estudo. Assim, foi feito um recorte de cada imagem dentro da área de estudo utilizando os limites de  $50,899658^\circ\text{W}$  a  $43,549805^\circ\text{W}$  e  $2,040279^\circ\text{S}$  a  $2,284551^\circ\text{N}$  (Figura 4.4 e Figura 4.5), sendo estes os valores de longitude máxima, longitude mínima, latitude mínima e latitude máxima, respectivamente. Em seguida, foram calculados parâmetros estatísticos da temperatura de brilho (e.g.  $Tb_{IR}$  mínima, média, máxima e desvio padrão) de cada recorte, com o objetivo de gerar as séries temporais desses parâmetros estatísticos.

Figura 4.4 - Área de estudo e ilustração do recorte realizado nas imagens do GOES16.



Neste trabalho, focamos na previsão de "t+1" (i.e. 1 hora) para explorar a capacidade preditiva da RNA no passo de tempo de previsão mais próximo, onde, geralmente,

Figura 4.5 - Ilustração dos dados de  $Tb_{IR}$  obtidos a partir de imagens do GOES-16 antes (esquerda) e depois (direita) do recorte da área de interesse.



se obtêm os melhores resultados. Para isso, subamostramos as séries temporais do conjunto de dados (observados *in situ* e obtidos por imagem de satélite) a cada 1 hora, de forma que a previsão de "t+1"corresponda a uma previsão para 1 hora no futuro. Com essa abordagem, buscamos garantir uma análise precisa e eficiente, aproveitando a granularidade horária dos dados para avaliar a eficácia do modelo preditivo no curtíssimo prazo.

O tratamento e pré-processamento dos dados consistiu basicamente nas seguintes etapas:

- a) Subamostragem de todas as séries temporais do conjunto de dados a cada 1 hora (como mencionado anteriormente);
- b) Verificação e retirada de dados espúrios, a fim de evitar a inclusão de mais incertezas a RNA;
- c) Decomposição vetorial dos dados de intensidade e direção dos ventos nas componentes zonal (u) e meridional (v);
- d) Análise exploratória do conjunto de dados;
- e) Normalização dos dados;
- f) Separação dos conjuntos de treinamento, validação e teste;
- g) Separação dos conjuntos de entrada e saída da RNA.

Para esta última etapa, diversos experimentos foram realizados visando tanto a otimização dos hiperparâmetros como a definição da melhor combinação de conjuntos de entrada da RNA, como descrito em maior detalhe no Capítulo 4.4.

Todo o processamento dos dados foi realizado utilizando a linguagem de programação Python. A rede neural foi construída utilizando as bibliotecas Keras e Tensorflow.

#### **4.4 Seleção de variáveis, construção dos modelos e otimização de hiperparâmetros**

Foram utilizados os dados dos anos de 2020 e 2021 (total de 2 anos de dados) para treinamento da RNA, sendo 20% deste conjunto de dados utilizado para validação durante o treinamento. Os dados do ano de 2022 foram utilizados para teste e avaliação final do desempenho do melhor modelo (Capítulo 5).

Tendo em vista que um dos objetivos do modelo é permitir sua operacionalização no futuro, visando a geração de previsões operacionais com alta frequência de atualização, os modelos foram construídos para prever a altura das ondas para a próxima hora (1 hora no futuro), com base nos dados das últimas 5 horas. A definição da defasagem de tempo de 5 horas foi escolhida de acordo com os resultados obtidos na análise exploratória de dados, no qual foi encontrado correlação de 0,97 entre altura das ondas e a intensidade dos ventos médios para a defasagem de tempo de 4,66 horas (ver Capítulo 5.1). Assim, os dados foram organizados da seguinte forma:

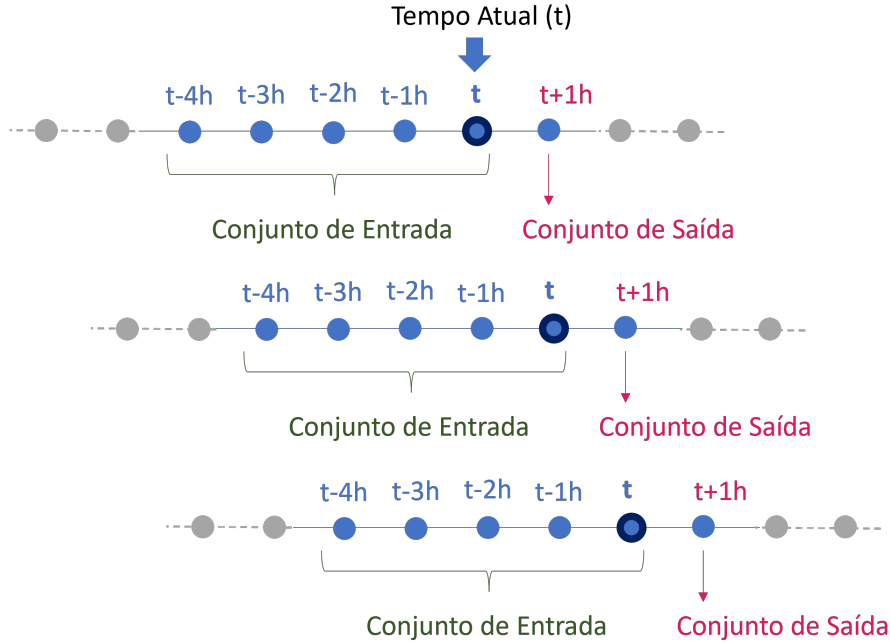
- Conjunto de entrada: últimas 5 horas de dados ( $t-4h$  até  $t$ ).
- Conjunto de saída: altura de onda da próxima hora ( $t+1h$ ).

A Figura 4.6 apresenta um esquema ilustrativo desta separação dos conjuntos de entrada e saída da RNA.

Durante a preparação dos dados para aplicação na RNA, uma atenção especial foi dada aos eventuais dados faltantes das séries temporais (como dados espúrios removidos ou dados inexistentes, oriundos de eventuais falhas de medição e/ou de interrupção temporária de medição). Os conjuntos de dados de entrada e saída selecionados para os experimentos com as RNAs não contém dados faltantes. Isto é, foram utilizados apenas "blocos" de dados sem falhas de nenhum parâmetro.

Com o objetivo de encontrar a melhor combinação de variáveis no conjunto de entrada e de hiper-parâmetros da RNA, 5 diferentes experimentos foram conduzidos com auxílio da biblioteca Optuna (Akiba et al., 2019), variando as variá-

Figura 4.6 - Esquema ilustrativo da separação dos conjuntos de entrada e de saída da RNA, elaborado de maneira a permitir previsões operacionais no instante de tempo  $t$ , para o instante de tempo " $t+1$ ", com base nos dados de " $t-4h$  até  $t$ ".



veis do conjunto de entrada: altura de onda  $WH_{10/100}$ , intensidade do vento médio ( $WindSpeedAvg$ ), intensidade da componente zonal do vento ( $Wind_U$ ), intensidade da componente meridional do vento ( $Wind_V$ ), nível do mar ( $SeaLevel$ ), e as estatísticas de  $Tb_{IR}$  ( $Tb_{IR}stats$ ). A Tabela 4.1 apresenta a relação das variáveis do conjunto de entrada utilizadas em cada experimento. Em todos os experimentos foi mantido o mesmo conjunto de dados de saída: altura de onda ( $WH_{10/100}$ ).

Para otimização dos hiper-parâmetros, foram definidos 50 *trials* por experimento no Optuna. Isto é, 50 tentativas de reduzir a perda (*loss*) variando os hiper-parâmetros em cada um dos 5 experimentos descritos anteriormente. Foram testadas variações do número de camadas LSTM (1 a 3), número de neurônios (1 a 512), taxa de *dropout* (0,0 a 0,7), intensidade do regularizador L2 ( $1e-5$  a  $1e-1$ ), e taxa de aprendizado ( $1e-4$  a  $1e-2$ ). Em todos os experimentos, foi utilizado o otimizador Adam e incluída função de ativação sigmoid na última camada da rede (camada densa).

A métrica utilizada para a definição de melhor modelo durante a realização dos experimentos foi o MSE (*Mean-Square Error*), computado para o conjunto de dados de teste (ano de 2022). O conjunto de teste foi utilizado nesta etapa para avaliar a capacidade dos modelos de generalizar o aprendizado. Isso permitiu comparar o



Tabela 4.1 - Variáveis do Conjunto de Entrada.

Experimento 1	WH <sub>10/100</sub>
Experimento 2	WH <sub>10/100</sub> , $Wind_U$ , $Wind_V$ , $WindSpeedAvg$
Experimento 3	WH <sub>10/100</sub> , $Wind_U$ , $Wind_V$ , $WindSpeedAvg$ , $SeaLevel$
Experimento 4	WH <sub>10/100</sub> , $Wind_U$ , $Wind_V$ , $WindSpeedAvg$ , $Tb_{IR} stats$
Experimento 5	WH <sub>10/100</sub> , $Wind_U$ , $Wind_V$ , $WindSpeedAvg$ , $SeaLevel$ , $Tb_{IR} stats$

desempenho real dos modelos com novos dados, evitando a utilização dos dados do conjunto de treinamento e validação nos quais o modelo já está especializado. Como mencionado em seções anteriores, a RNA escolhida para os experimentos foi a rede LSTM.

#### 4.5 Métricas de avaliação dos resultados

O processo de validação dos resultados de previsão de altura das ondas obtidos neste estudo seguiu três métricas:

i) o método proposto por Willmott (WILLMOTT, 1981), que se baseia na análise do parâmetro adimensional *Skill*,

$$WS = 1 - \frac{\langle (m - o)^2 \rangle}{\langle (|m - \langle o \rangle| + |o - \langle o \rangle|)^2 \rangle}, \quad (4.1)$$

sendo  $m$  e  $o$  as séries temporais de variáveis fornecidas previstas e observadas, respectivamente;  $\langle . \rangle$  representa a média para a série. A métrica *Skill* varia de 0 a 1, onde valores próximos de 1 indicam uma alta concordância entre os valores previstos e observados, e valores próximos de zero indicam pouca concordância, sugerindo que o modelo não está performando bem. Uma característica da métrica *Skill* é que esta é sensível tanto aos erros de magnitude quanto aos erros de fase, proporcionando uma avaliação mais abrangente da performance do modelo. No entanto, esta métrica é menos sensível a valores extremos do que as métricas apresentadas a seguir;

ii) o erro quadrático médio (MSE, em inglês *Mean-Square Error*),

$$MSE = \langle (m - o)^2 \rangle, \quad (4.2)$$

o qual indica a média dos quadrados das diferenças entre os valores previstos e observados, fornecendo uma medida da precisão geral de um modelo. Quanto menor

o MSE, maior a precisão do modelo;

iii) a raiz quadrada do erro médio quadrático (RMSE, em inglês *Root Mean-Square Error*),

$$RMSE = \sqrt{\langle (m - o)^2 \rangle}, \quad (4.3)$$

a qual é a raiz quadrada do MSE. O RMSE também indica o erro médio das previsões de um modelo, sendo mais intuitivo que o MSE por estar na mesma unidade dos dados originais. Assim como para o MSE, quanto menor o RMSE, melhor a assertividade do modelo.

Durante a etapa de análise exploratória dos dados também foi computada a correlação de Pearson entre as séries multivariadas estudadas, a qual é expressa por:

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum x_i^2 - n \bar{x}^2} \sqrt{\sum y_i^2 - n \bar{y}^2}}, \quad (4.4)$$

onde  $n$  é o tamanho das séries, e  $x_i$  e  $y_i$  são as séries em análise.

#### 4.6 Análise do erro do melhor modelo

O desempenho do melhor modelo LSTM escolhido neste estudo para a previsão de ondas pode ser fortemente influenciado pela qualidade e características dos dados utilizados. Fatores como a resolução temporal e espacial dos dados, a presença de ruídos, dados ausentes e a representação adequada de ciclos sazonais e fenômenos climáticos. A distribuição desequilibrada dos dados, especialmente em relação a eventos extremos, pode limitar o modelo, evitando que ele capture alguns padrões nos dados. Além disso, o tamanho do conjunto de dados de treinamento pode ser insuficiente para evitar esses problemas, não permitindo que o modelo generalize bem para novas condições. Este aspecto importante é justamente associado à presença de eventos anômalos ou outliers, como tempestades, que podem introduzir variações inesperadas nos dados em curto prazo e afetar negativamente as previsões. Para entender esses impactos, foi fundamental aplicar uma análise mais detalhada sobre os erros observados por esses modelos com base nas métricas e métodos listados nos capítulos anteriores. Uma análise mais cuidadosa é essencial para entender o skill do modelo LSTM para aplicações operacionais.



Além disso, intercomparações com modelos numéricos como o modelo MERCATOR (COPERNICUS MARINE SERVICE, 2021) e teste de sensibilidade para um *leadtime* superior foram realizados. O objetivo foi verificar se o modelo ajustado de LSTM já obtem resultados satisfatórios ou não para previsões de 1 a 3 horas em comparação a modelo utilizado pela comunidade.



## 5 RESULTADOS E DISCUSSÕES

Neste Capítulo serão apresentados e discutidos os resultados obtidos no desenvolvimento deste trabalho, iniciando pela análise exploratória do conjunto de dados utilizado, seguido dos resultados dos experimentos com a aplicação da LSTM para elaboração do modelo preditivo de altura de ondas. Por fim, foram feitas análises da assertividade do modelo desenvolvido.

### 5.1 Análise exploratória do conjunto de dados

Um passo inicial importante na aplicação de técnicas de IA, assim como em qualquer processo de descoberta de conhecimento, é a análise exploratória do conjunto de dados (MILO; SOMECH, 2020). Assim, nesta seção serão apresentadas e descritas as abordagens utilizadas no tratamento dos dados e as relações encontradas entre eles que antecedem a construção da RNA.

Um dos primeiros dados a serem analisados foram aqueles provenientes de satélites meteorológicos, no caso em questão o GOES16. A Figura 5.1 apresenta as séries temporais dos parâmetros estatísticos (mínimas, máximas, médias e desvio padrão) de temperatura de brilho obtidos através das imagens do GOES16 do canal do infravermelho janela do sensor ABI na área de estudo. Podemos observar que os valores médios apresentam valores majoritariamente maiores entre os meses de julho e outubro, enquanto valores menores são observados principalmente entre os meses de março, abril e maio. Temperaturas mais baixas do topo de nuvem indicam maior ocorrência de cobertura de nuvens neste período, o que pode estar associado a variação da posição da Zona de Convergência Intertropical (ZCIT) ao longo do ano, situando-se na região Norte e Nordeste do Brasil principalmente nos meses de março a maio (Germano, et al. 2017).

Com relação às observações em superfície, os dados de ventos coletados *in situ* na estação de monitoramento meteo-oceanográfico, para o período de estudo, são apresentados na Figura 5.2. Analisando a figura, é possível perceber que os ventos no local são majoritariamente provenientes dos quadrantes E, NE e N, com intensidades predominantes de até 7,5 m/s.

Figura 5.1 - Séries temporais dos dados de temperatura de brilho mínima, média, máxima e o desvio padrão obtidos a partir das imagens do GOES16 para a área de estudo.

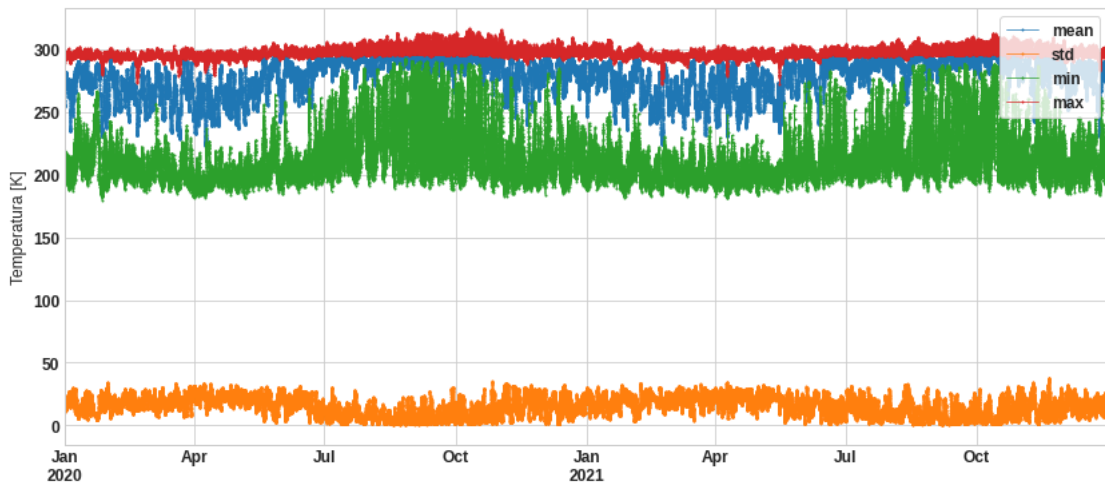
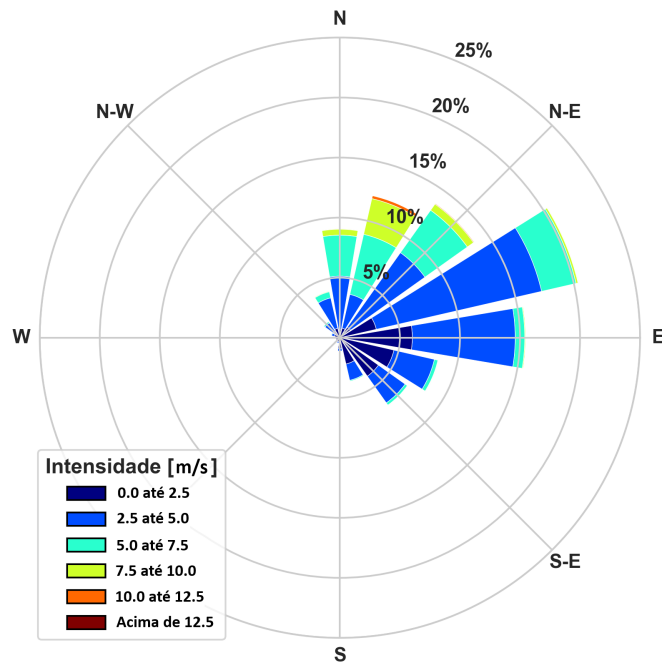


Figura 5.2 - Rosa dos ventos para os dados coletados *in situ* na estação de monitoramento, durante o período de estudo.



Já os dados de altura de ondas coletados *in situ* são apresentados na Figura 5.3. Para as análises, foi utilizado o parâmetro denominado WH10/100, o qual representa as alturas 10% mais altas registradas durante o período de amostragem do equipamento (20 minutos). Este também é o parâmetro utilizado para as tomadas de decisões

do terminal portuário. Para melhor visualização e entendimento da distribuição destes dados, foi feita uma classificação das alturas em três faixas, utilizado como referência os percentis de 66% ( $P66 = 0,29$  m) e de 99% ( $P99 = 1,23$  m): ondas baixas e moderadas ( $< P66$ ); ondas altas ( $\geq P66$  e  $< P99$ ); e ondas muito altas ( $\geq P99$ ). Com base na figura é possível observar que, apesar da ocorrência de ondas com altura maior ou igual a 1,23 m ser baixa para o local, estas ondas podem ocorrer ao longo de todo o ano. Ao longo do período de estudo, as maiores ocorrências destas ondas foram entre os meses de setembro e outubro de 2020. A altura máxima de onda registrada no período de estudo foi de 2,28 m, e a altura média de 0,30 m.

A Figura 5.4 resume a distribuição dos dados de período de pico de onda observados, isto é, os períodos correspondentes as maiores energias do espectro de ondas medidos, no formato de gráfico do tipo *boxplot*. Podemos observar que os valores de período de pico variam em torno de 3 s e 7 s, sendo sempre inferiores a 10 s, o que mostra a predominância de ondas geradas pelo vento local na área de estudo (COTRIM et al., 2022).

Figura 5.3 - Série temporal dos dados de altura de onda (WH10/10 - alturas 10% mais altas) coletados *in situ* na estação de monitoramento, durante o período de estudo.

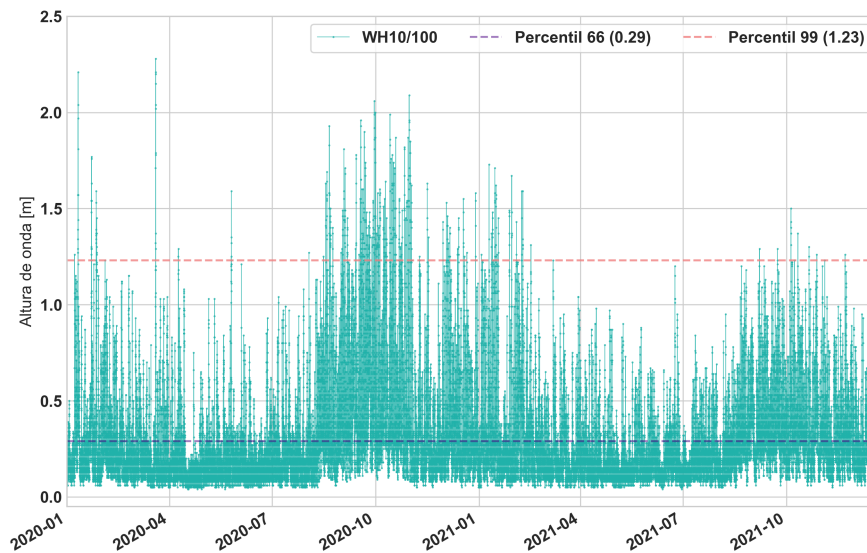
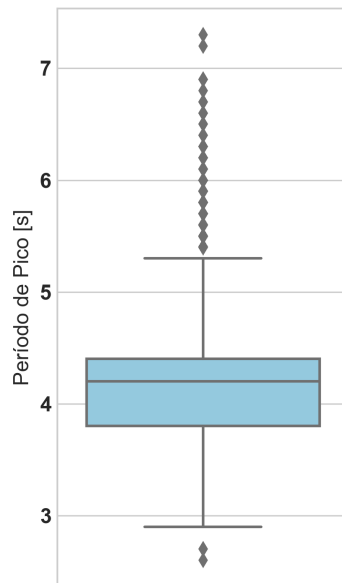


Figura 5.4 - Gráfico *boxplot* dos dados de período de pico coletados *in situ* na estação de monitoramento, durante o período de estudo.



Na Figura 5.5 é apresentada a série temporal dos dados de nível do mar observados na estação de monitoramento, os quais são coletados pelo mesmo equipamento que realiza as medições de ondas (ADCP). Durante o período analisado, o nível médio do mar foi de 1,61 m, com altura mínima de 0,0 m e altura máxima de 3,69 m. Na Tabela 5.1 estão sintetizadas as estatísticas descritivas das variáveis de intensidade dos ventos, altura das ondas 10% mais altas, período de pico e nível do mar coletados *in situ* na estação de monitoramento.

Figura 5.5 - Série temporal dos dados de variação do nível do mar coletados *in situ* na estação de monitoramento, durante o período de estudo.

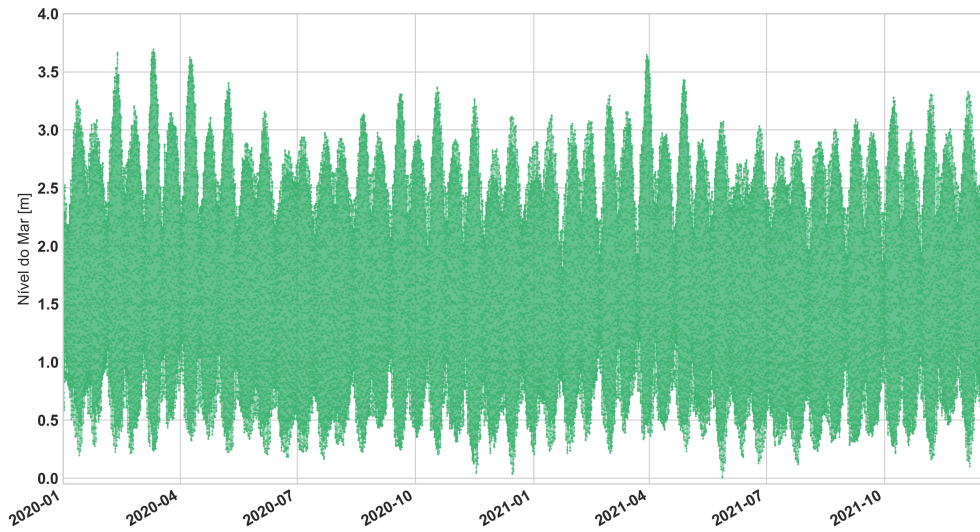
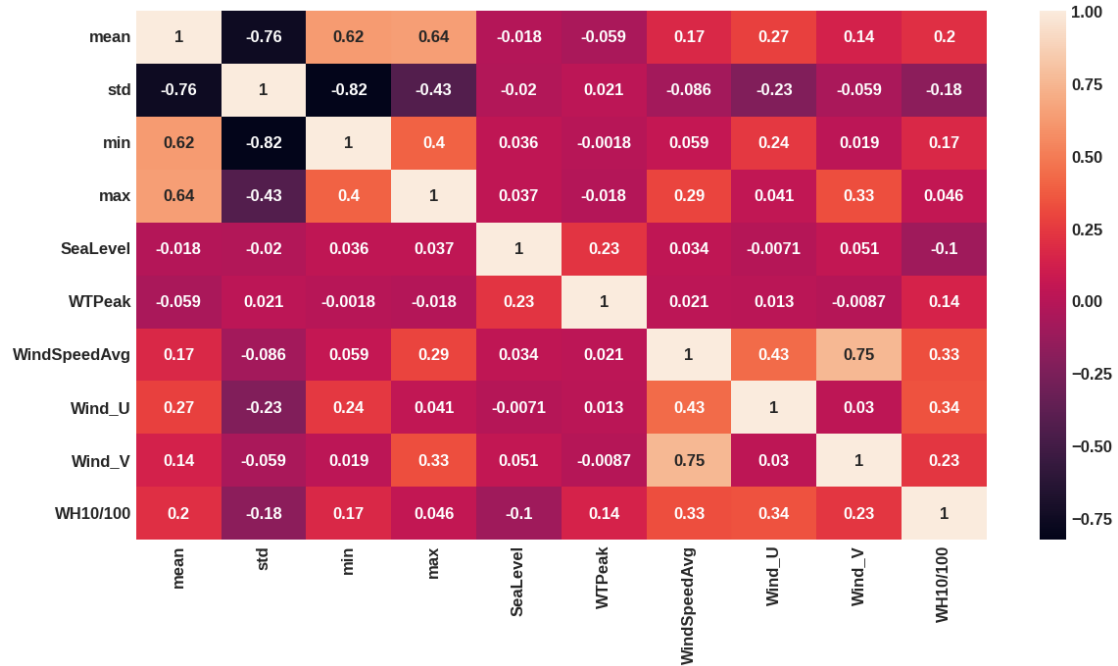


Tabela 5.1 - Estatística descritiva dos dados coletados *in situ* na estação de monitoramento ao longo do período de estudo.

	Nível do Mar [m]	Altura de Onda [m]	Período de Pico [s]	Vento Médio [m/s]
Média	1.61	0.30	4.14	3.9
Desvio Padrão	0.81	0.24	0.50	1.98
Mínimo	0.01	0.04	2.6	0.00
Perecentil 25	0.89	0.14	3.8	2.48
Perecentil 50	1.54	0.22	4.2	3.61
Perecentil 75	2.32	0.37	4.4	5.04
Máximo	3.69	2.28	7.3	15.94

De modo a analisar as relações entre as variáveis, inicialmente foi calculada a correlação de Pearson entre eles. A Figura 5.6 apresenta a matriz de correlação de Pearson entre as séries temporais das diferentes variáveis do conjunto de dados. As variáveis com maiores correlações com altura de onda foram a intensidade do vento médio (correlação de 0.33) e a intensidade das componentes zonal (0.34) e meridional do vento (0.23). A correlação entre nível do mar e altura de onda foi de -0,1, e entre período de pico e altura de onda, de 0,14. Já para a temperatura de brilho, a correlação mais alta encontrada com a altura de onda foi de 0,11, para a temperatura de brilho média. Cabe ressaltar que nenhuma defasem foi aplicada nestas correlações.

Figura 5.6 - Matriz de correlação entre séries multivariadas estudadas: temperatura de brilho média ( $Tb_{IRMean}$ ), desvio padrão da temperatura de brilho ( $Tb_{IRStd}$ ), temperatura de brilho mínima ( $Tb_{IRMin}$ ), temperatura de brilho máxima ( $Tb_{IRMax}$ ), intensidade do vento médio ( $WindSpeedAvg$ ), intensidade da componente zonal do vento ( $Wind_U$ ), intensidade da componente meridional do vento ( $Wind_V$ ), nível do mar ( $SeaLevel$ ), período de pico ( $WTPeak$ ) e altura de onda ( $WH10/100$ ).



Com objetivo de explorar em maior detalhe as relações entre intensidade dos ventos e alturas das onda, as quais apresentaram correlações com valores abaixo do esperado, foi realizada uma análise para verificação de sazonalidade a partir de boxplots mensais (Figura 5.7). Com base nesta figura, podemos observar que há um padrão muito semelhante nas variações mensais de ventos e de altura de onda. Entre os meses de abril e julho, tanto os valores de alturas de ondas como de intensidade dos ventos são mais baixos. Já durante os meses de setembro e outubro, as alturas de ondas e as intensidades dos ventos são predominantemente mais altas. Portanto, esta é uma evidência de que as variações sazonais de altura de ondas estão associadas às variações sazonais de ventos no local de estudo, apesar de as correlações de pearson calculadas diretamente entre as variáveis, para as séries completas de dados a cada 20 min, terem sido muito baixas.

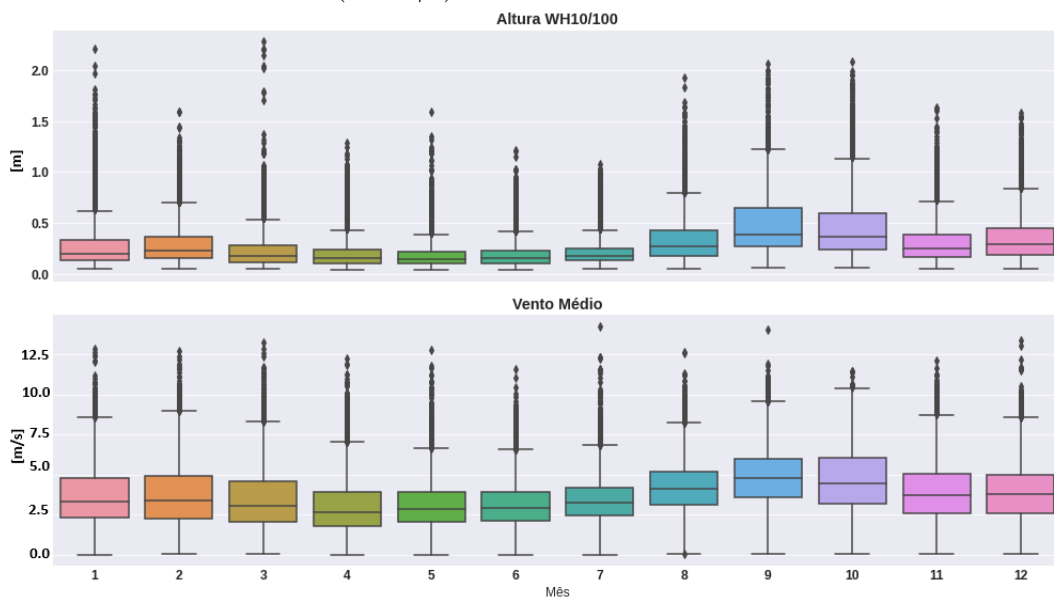
As variações sazonais na velocidade do vento na costa Norte do Brasil, e portanto, na área de estudo, estão associadas à atuação de sistemas meteorológicos de grande



escala, como a intensificação da alta subtropical do Atlântico Sul entre os meses de setembro e novembro e a migração da ZCIT, a qual é mais ativa sobre o Norte e Nordeste do Brasil, principalmente, nos meses de março a maio (Germano, et al. 2017).

Os gráficos também mostram que os maiores valores de altura registrados ocorreram, por sua vez, nos meses de janeiro e março, provavelmente ocasionados por eventos locais mais intensos, visto que estes dados estão apresentados como outliers.

Figura 5.7 - Boxplots mensais de altura de onda (WH10/100, em metros) e de intensidade do vento médio (em m/s).



A fim de avançar na análise exploratória dos dados, foi realizada também a análise de autocorrelação para as variáveis de altura de onda, intensidade do vento e temperatura de brilho (Figura 5.8). Analisando a figura, é possível observar que todas as variáveis apresentam um padrão semelhante: iniciam com autocorreção elevada e com diminuição gradativa conforme os passos de tempo avançam, seguido de um aumento dos valores de autocorreção após aproximadamente 24-25 horas. Este padrão sugere a ocorrência de ciclos diurnos no local tanto para as ondas como para ventos. Cabe ressaltar que a região amazônica é conhecida por ciclos diurnos de convecção bem definidos (GERMANO et al., 2017).

A autocorrelação dos dados de altura de onda mostra também um aumento após 12,3 horas, padrão que também é observado nos dados de nível do mar (Figura

5.9). O nível do mar apresenta um sinal de maré bem definido, com ciclos de 6,33 horas, 12,33 horas e 24,33 horas. O regime de marés na área de estudo é semidiurno, caracterizado por duas marés altas e duas marés baixas ao longo do dia (FRANCO, 2005).

Figura 5.8 - Análise de autocorrelação para as séries de altura de onda, intensidade do vento médio, desvio padrão da temperatura de brilho e temperatura de brilho mínima. Os passos de tempo são equivalentes a 20 minutos.

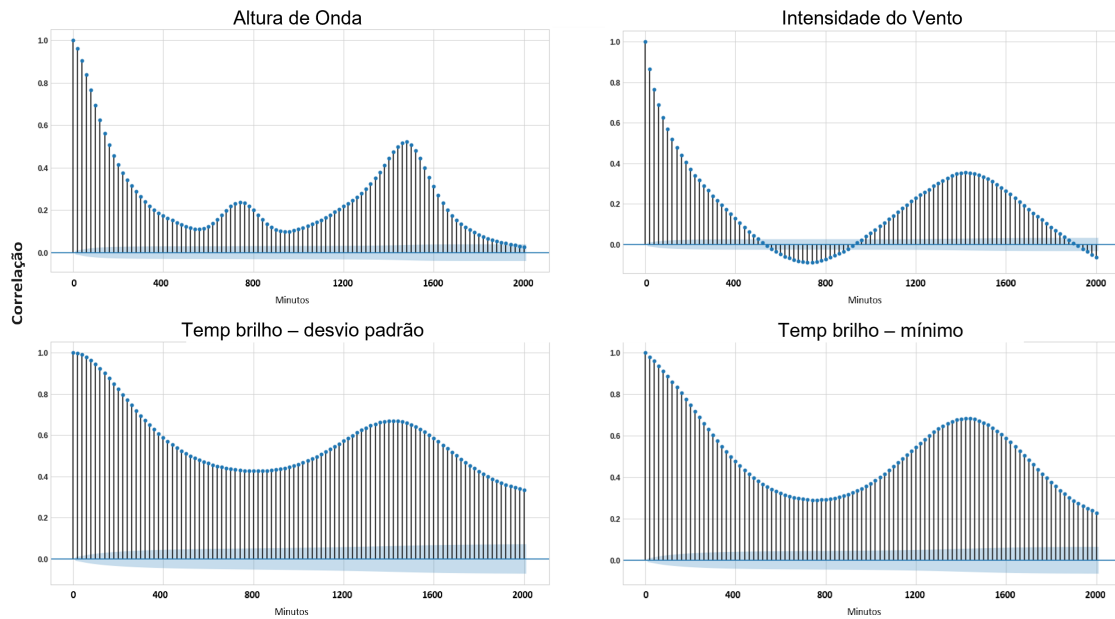
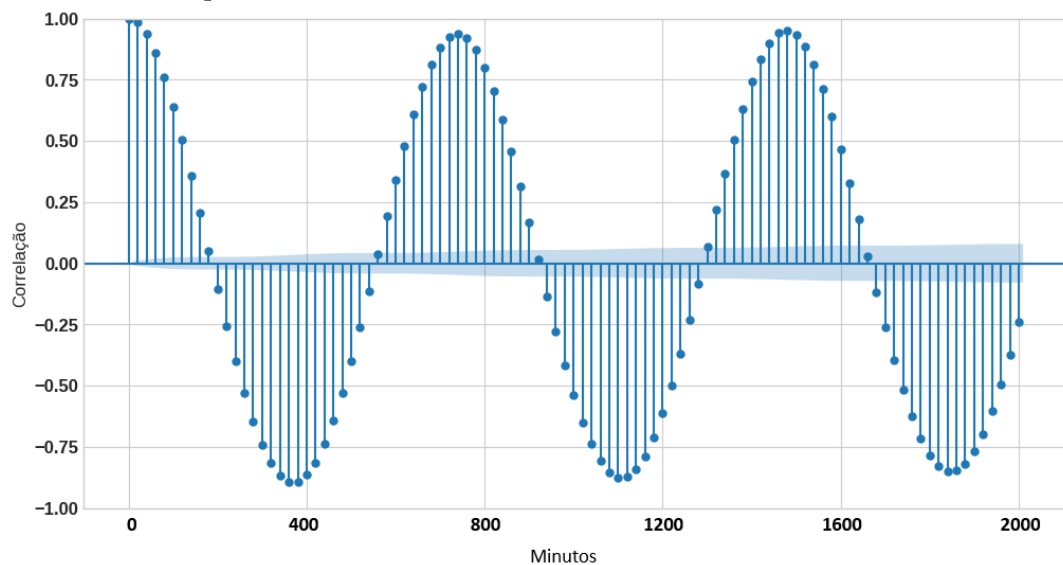


Figura 5.9 - Análise de autocorrelação para os dados de nível do mar. Passos de tempo são equivalentes a 20 minutos.



Tendo em vista o padrão de aumentos nos valores de autocorrelação dos parâmetros após aproximadamente 24-25 horas, e na tentativa de explorar as possíveis relações entre as séries multivariadas estudadas, foi realizada uma análise de ciclo diurno para os dados de altura de onda, de intensidade dos ventos e de temperatura de brilho média. Para esta análise, os dados foram inicialmente agrupados pelo horário do dia. Em seguida, foi feito o cálculo do valor médio para cada horário (Figura 5.10). Por fim, foi definida a correlação cruzada entre as séries médias diurnas (Figuras 5.12 e 5.11).

Os resultados mostram que há uma elevada correlação tanto entre altura de onda e intensidade do vento (correlação de 0,97), como entre altura de onda e temperatura de brilho média (correlação de -0,94), com um *lag* de 280 min (4h40) para a primeira, e de 140 min (2h20) para a segunda. Isto é, em média, durante o dia, 4 horas e 40 minutos após o maior pico médio diário de intensidade do vento, e 2 horas e 20 minutos após o menor pico diário de temperatura de brilho média para a área de estudo, ocorre o pico máximo diário de altura de onda na estação de monitoramento. Tais resultados indicam que estes ciclos devem ser considerados no momento da escolha do número de passos de tempo de dados de entrada para representar os processos locais com maior fidelidade e sensibilidade. Ainda, os resultados mostram que, em média, as ondas tendem a ser mais altas entre às 18:00 e às 02:00, enquanto os ventos mais intensos tendem a ocorrer entre às 14:00 e às 20:00. Com relação à temperatura de brilho média diária, os menores valores tendem a ocorrer entre às 16:00 e às 21:00. Portanto, apesar dos baixos valores encontrados na matriz de correlação, os resultados das análises exploratórias mostram que existe uma associação assíncrona entre as séries de altura de onda e de temperatura de brilho média, e entre as séries de altura de onda e de intensidade do vento médio. Este resultado pode refletir os ciclos diurnos naturais da convecção, maré e brisa da região.

Figura 5.10 - Séries temporais de valores médios e desvio padrão dos dados de altura de onda e de intensidade do vento médio para cada hora do dia.

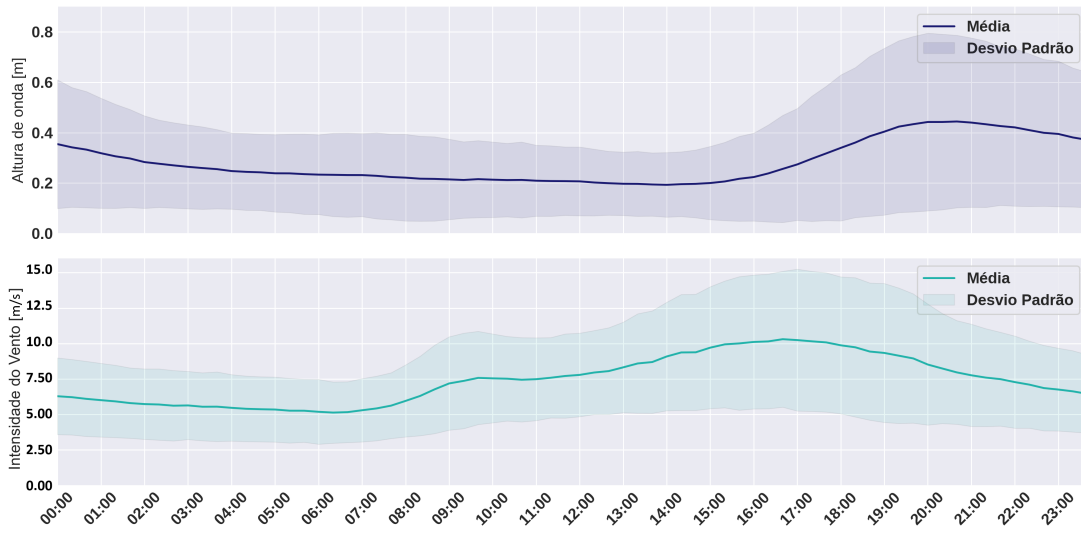


Figura 5.11 - Análise de correlação cruzada para as séries de médias diárias de altura de onda e de intensidade do vento médio.

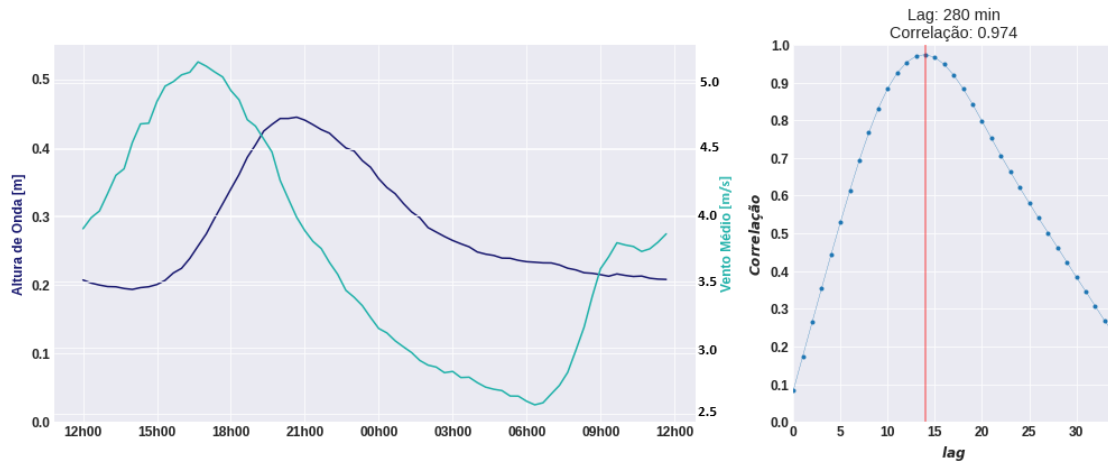
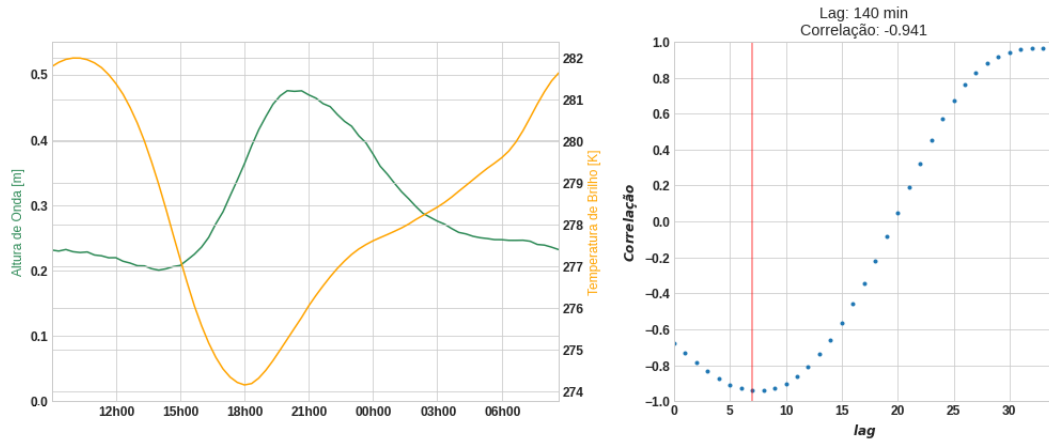


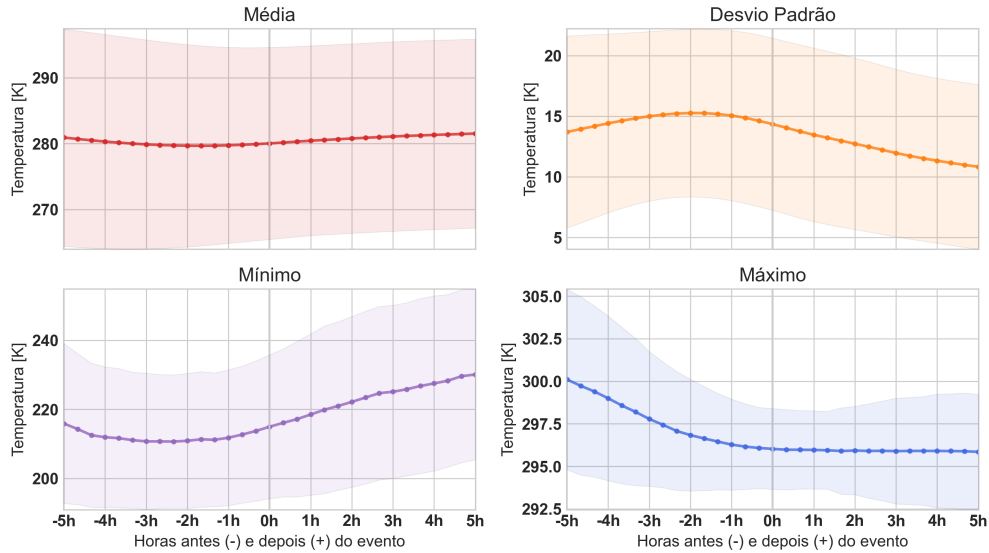
Figura 5.12 - Análise de correlação cruzada para as séries de médias diárias de altura de onda e de temperatura de brilho média.



Para analisar as relações entre os dados de altura de onda e as estatísticas de temperatura de brilho em maior detalhe, foi feito um agrupamento destes dados para análise de 5 horas antes até 5 horas depois de cada evento de onda mais extrema ( $>$  Percentil 99) e de ondas baixas ( $<$  Percentil 66), a fim de avaliar possíveis tendências de crescimento ou decaimento nos dados de temperatura de brilho que possam auxiliar na identificação de precursores de eventos de ondas mais extremas.

A Figura 5.13 apresenta os valores de média e desvio padrão para o período de 5 horas antes até 5 horas depois de eventos de altura de ondas mais extremos observados nas proximidades do terminal portuário para os parâmetros estatísticos de temperatura de brilho. Já as Figuras 5.14 e 5.15 apresentam comparações entre os valores médios de cada parâmetro estatístico da temperatura de brilho e os valores médios de altura de onda, para o período de 5 horas antes até 5 horas depois de eventos de altura de ondas mais extremos e de altura de ondas mais baixas, respectivamente.

Figura 5.13 - Valores de média (linha contínua) e desvio padrão (área colorida) para o período de 5 horas antes até 5 horas depois de eventos de altura de ondas mais extremos, para os parâmetros estatísticos de temperatura de brilho do topo das nuvens (média, mínimo, máximo e desvio padrão).



Os resultados indicam que há uma associação entre os dados de temperatura de brilho de topo das nuvens e os dados de altura de ondas: tendência de decaimento na temperatura de brilho média, mínima e máxima, e de crescimento no desvio padrão da temperatura de brilho na área de estudo, em momentos que antecedem os picos de altura em eventos de alturas de ondas mais extremas. Lembrando que, temperaturas mais baixas do topo de nuvem estão associadas a nuvens com topo mais alto, sugerindo assim a presença de uma cobertura de nuvens, e até mais possíveis tempestades, durante a ocorrência de eventos de ondas maiores. Para eventos de alturas de ondas mais baixas, o padrão oposto foi observado. Portanto, estes parâmetros estatísticos de temperatura de brilho podem possivelmente ser utilizados como precursores de eventos de altura de ondas mais extremas.

Figura 5.14 - Comparação entre os valores médios dos parâmetros estatísticos de temperatura de brilho (média, mínimo, máximo e desvio padrão) e os valores médios de altura de onda, para o período de 5 horas antes até 5 horas depois de eventos de altura de ondas mais extremas.

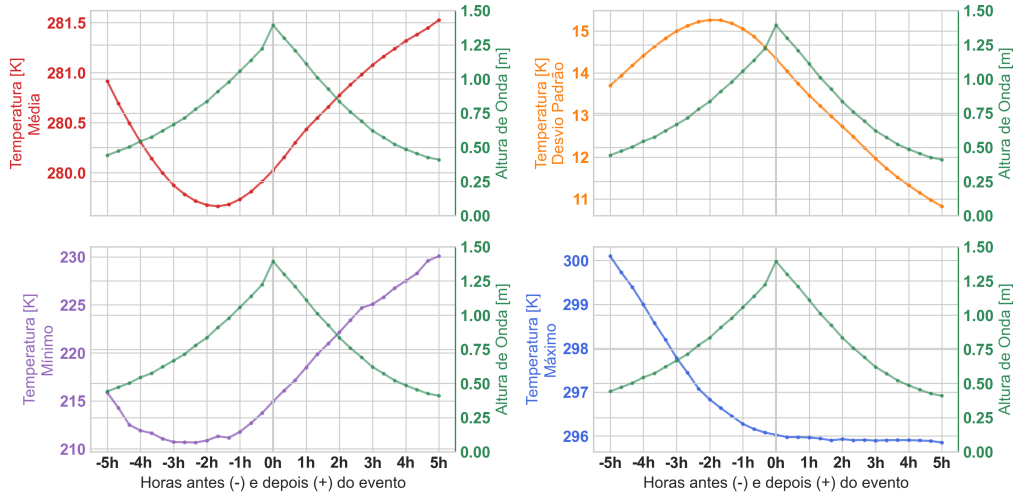
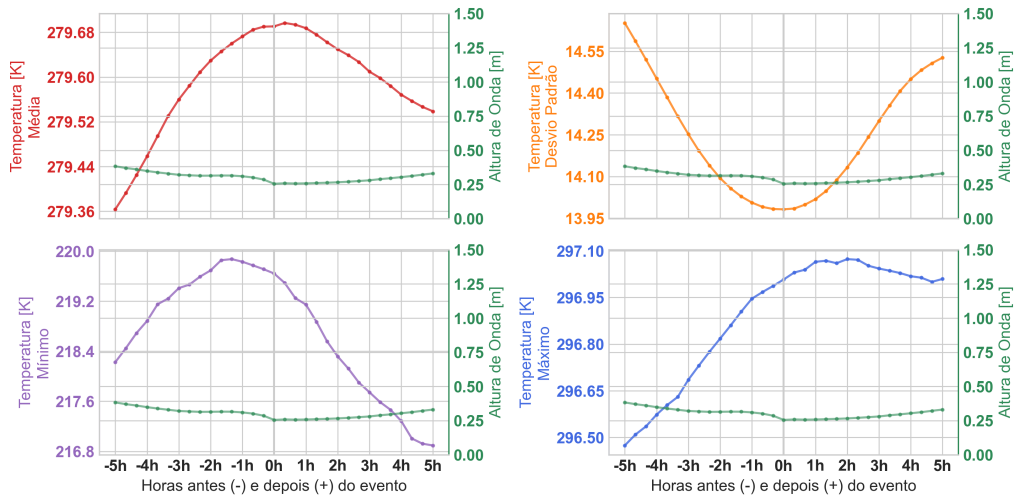


Figura 5.15 - Comparação entre os valores médios dos parâmetros estatísticos de temperatura de brilho (média, mínimo, máximo e desvio padrão) e os valores médios de altura de onda, para o período de 5 horas antes até 5 horas depois de eventos de altura de ondas mais baixas.



Desta forma, foi decidido utilizar todas as variáveis disponíveis nos experimentos com a RNA, a fim de avaliar o desempenho da rede ao variar as variáveis no conjunto de entrada. Os resultados dos experimentos com a LSTM na previsão de altura de onda são apresentados a seguir.

## 5.2 Resultados dos experimentos com a LSTM

A Tabela 5.2 apresenta os resultados dos experimentos realizados com auxílio do Optuna (AKIBA et al., 2019), sintetizando a combinação dos hiper-parâmetros que geraram os melhores resultados (menor valor de *loss*) em cada experimento, bem como o número da *trial* que gerou tais resultados. A Tabela apresenta também os números de camadas (*layers*) e neurônios (*units*) utilizados, além dos valores das taxas de esquecimento (*dropout*) e de aprendizado (*learning rate*) que geraram os melhores resultados de cada experimento.

Tabela 5.2 - Resultados dos experimentos com Optuna, variando o conjunto de entrada e os hiper-parâmetros.

Experimento	trial	loss	units	layers	dropout	regularizer	learning_rate
1	7	0.00268	46	2	0.09205	0.00029	0.00482
2	41	0.00193	102	2	0.54915	0.00001	0.00018
3	14	0.00123	333	2	0.00151	0.00001	0.00046
4	46	0.00184	66	2	0.08601	0.00001	0.00285
5	34	0.00126	360	2	0.57963	0.00001	0.00027

Analisando a tabela, podemos verificar que os experimentos que geraram os melhores resultados, com base no valor de *loss*, foram o experimento 3 e o experimento 5. Ambos os experimentos tiveram no conjunto de entrada as variáveis de altura das ondas, intensidade dos ventos, componentes zonal e meridional dos ventos e nível do mar. O experimento 5 contou ainda com as estatísticas de temperatura de brilho no conjunto de entrada, sendo esta a diferença entre os dois experimentos. Os experimentos 3 e 5 foram os únicos que consideraram a variável de nível do mar no conjunto de entrada. Por serem regidas por forças previsíveis e cíclicas, as marés não apresentam comportamentos aleatórios ou caóticos, o que torna o sinal de maré determinístico. A característica cíclica deste sinal de maré, em conjunto com o sinal de vento local, mostrou-se a melhor combinação de variáveis para previsão de altura de ondas no local de estudo.

O experimento que obteve o pior desempenho foi o experimento 1, no qual o modelo foi treinado somente com dados de altura das ondas. A inclusão dos dados de vento no experimento 2 reduziu os erros médios deste modelo em 28%, e a inclusão das estatísticas de temperatura de brilho, juntamente com os dados de ventos, no experimento 4, reduziu os erros médios em 31%. Contudo, como mencionado anteriormente, a maior redução nos erros médios ocorreu quando adicionado os dados



de nível do mar, juntamente com os dados de ventos (experimento 3): redução de 54% em comparação ao modelo treinado apenas com dados de altura de ondas. A inclusão dos dados de temperatura de brilho a este modelo treinado com as variáveis de altura de ondas, ventos e nível do mar, no experimento 5, gerou resultados muito semelhantes, porém ligeiramente inferiores.

A Figura 5.16 e a Tabela 5.3 apresentam as validações das previsões geradas pelos modelos construídos no experimento 3 e 5, agora chamados de "modelo 3 e 5", para o conjunto de teste (ano de 2022). O modelo 3 apresentou RMSE de 0,0738 m, enquanto o modelo 5 apresentou RMSE de 0,0742 m. Os valores de  $R^2$  foram de 0,87 e 0,86, respectivamente, para o modelo 3 e o modelo 5.

Figura 5.16 - Comparação entre os valores observados e previstos pelo modelo 3 (A) e pelo modelo 5 (B) para o ano de 2022.

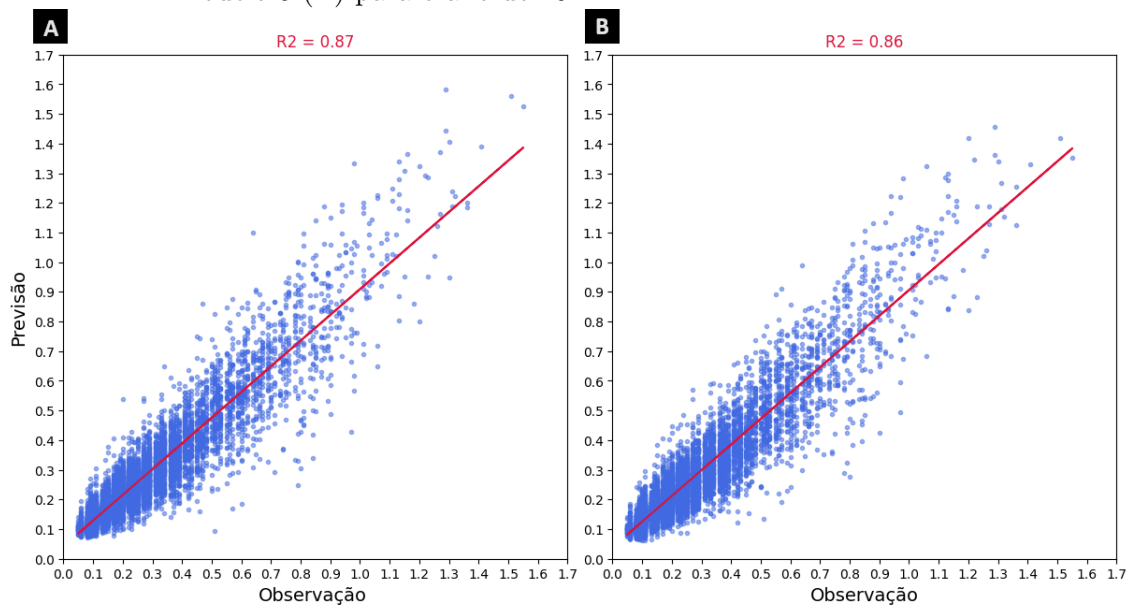


Tabela 5.3 - Métricas de desempenho dos modelos 3 e 5 para o ano de 2022.

Modelo	MSE	RMSE	MAE	Skill	$R^2$
3	0,0054 m <sup>2</sup>	0,0738 m	0,0514 m	0,87	0,87
5	0,0055 m <sup>2</sup>	0,0742 m	0,0523 m	0,86	0,86

Não foram encontrados na literatura estudos sobre o desenvolvimento de modelos

preditivos de altura das ondas específicos para a região de estudo deste trabalho ou suas proximidades, impossibilitando a comparação direta com os resultados aqui obtidos. Assim, este trabalho apresenta pela primeira vez uma previsão de altura das ondas na região costeira de Barcarena, mais especificamente nas proximidades do Terminal Portuário de Barcarena, utilizando técnicas de IA. Desta forma, os resultados foram comparados com estudos realizados em outras localidades, que utilizaram técnicas semelhantes às utilizadas neste trabalho.

Os resultados de ambos os modelos (3 e 5) foram superiores aos encontrados por [Fan et al. \(2020\)](#) em 9 das 10 estações utilizadas pelos autores, os quais obtiveram valores de RMSE variando entre 0,0818 m e 0,3293 m em seus modelos desenvolvidos também com a LSTM, para previsão de 1 hora no futuro, para estações oceanográficas instaladas em locais distintos do globo terrestre (costa leste dos Estados Unidos, Inglaterra, Espanha, costa sudeste do Brasil, Coreia, Havaí e na costa oeste do Canadá). O melhor resultado obtido pelos autores foi de 0,0738 m de RMSE, para a estação situada na costa do estado da Georgia, nos Estados Unidos, o mesmo resultado obtido pelo modelo 3 desenvolvido neste trabalho.

Apesar dos resultados semelhantes entre os modelos 3 e 5, o modelo 3 teve um desempenho superior, apresentando os melhores resultados entre todos os experimentos realizados. Além disso, este modelo utilizou menos variáveis no conjunto de entrada do que o modelo 5, o que facilita o processo de treinamento e operacionalização. Por estes motivos, o modelo 3 foi o modelo escolhido para o restante das análises apresentadas neste trabalho.

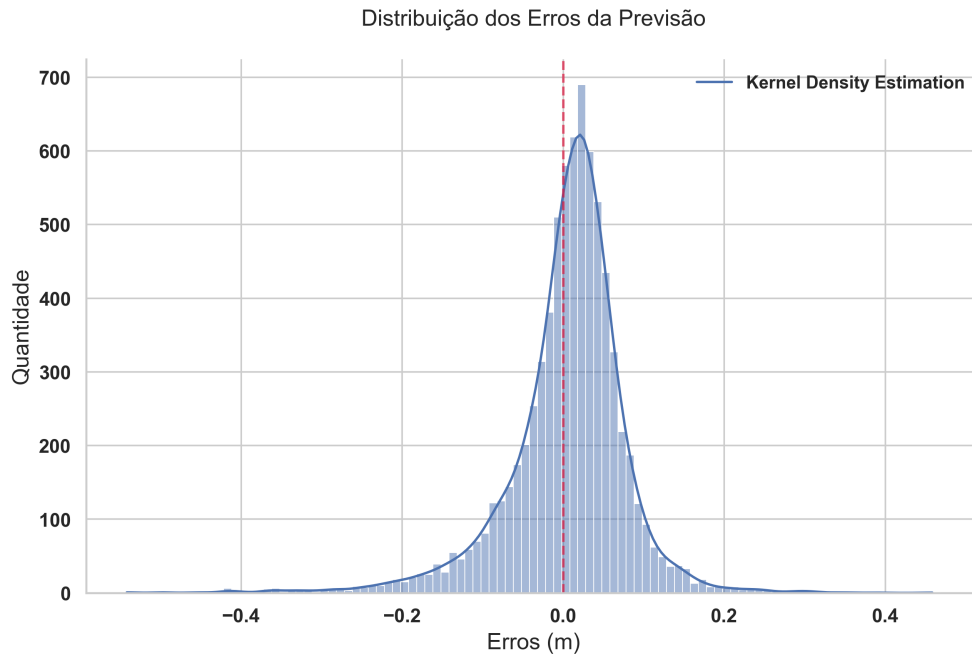
O Capítulo 5.3 apresenta as análises e validações realizadas com o modelo preditivo de altura de ondas (modelo 3), as quais tiveram com objetivo avaliar o desempenho do modelo em diferentes cenários, buscando entender e quantificar o erros do modelo, bem como verificar a assertividade com relação ao limite operacional do terminal portuário.

### **5.3 Resultados do modelo preditivo de altura de ondas**

A Figura 5.17 apresenta a distribuição dos erros das previsões de altura de ondas realizadas para o ano de 2022. Podemos observar que o histograma mostra uma distribuição assimétrica, levemente deslocada para a direita. Isto indica que, de modo geral, o modelo possui uma tendência de superestimar as alturas das ondas. Analisando a figura, também é possível perceber que os erros estão concentrados na faixa de -0,1 m a 0,1 m. A percentagem de ocorrência de erros absolutos acima de

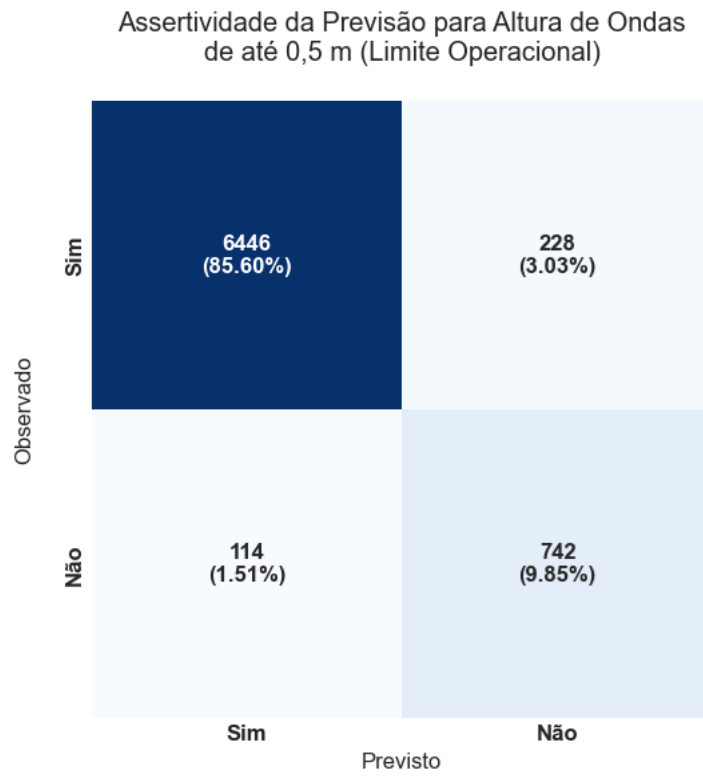
10 cm foi de 12,1 %. Como visto no Capítulo anterior, o erro médio absoluto (MAE) para o ano de 2022 foi de 0,05 m (5 cm). Já o erro máximo para o período foi de 0,54 m (54 cm).

Figura 5.17 - Histograma dos erros das previsões de altura de ondas realizadas para o ano de 2022. Valores de erros positivos indicam superestimação do modelo, enquanto valores de erros negativos indicam subestimação do modelo.



A Figura 5.18 apresenta a matriz de confusão das previsões de altura de ondas realizadas para o ano de 2022, considerando o limite operacional do terminal portuário de 0,5 m. A Probabilidade de Detecção (da sigla em inglês *POD: Probability of Detection*) do modelo foi de 87% e a Taxa de Falsos Alarmes (da sigla em inglês *FAR: False Alarm Rate*) foi de 3%. Este último resultado indica que no ano de 2022 o modelo teria gerado falso alarme, prevendo ondas acima de 0,5 m quando a altura observada foi inferior a esse limite, em apenas 3% das ocorrências. Sendo assim, os resultados mostram que o modelo é capaz de prever com boa assertividade se as alturas das ondas estarão acima ou abaixo do limite operacional de 0,5 m estabelecido pelo terminal portuário.

Figura 5.18 - Matriz de confusão das previsões de altura de ondas realizadas para o ano de 2022, considerando o limite operacional de 0,5 m.



Para obter um melhor entendimento da precisão das previsões e dos erros do modelo, foi feita a análise de dois conjuntos de dados distintos: erros absolutos abaixo do Percentil 10 e acima do Percentil 90. Essa abordagem permite comparar momentos de maiores erros com momentos de erros menores, fornecendo uma visão mais completa do desempenho do modelo. A Figura 5.19 apresenta a série temporal dos erros absolutos obtidos pelo modelo para o ano de 2022, com destaque para os erros absolutos abaixo do Percentil 10 (erros pequenos) e acima do Percentil 90 (erros grandes). A Tabela 5.4 sintetiza esses resultados de modo quantitativo, apresentando as métricas das previsões destes dois conjuntos. Podemos observar que o erro médio absoluto para o conjunto de "erros pequenos" foi de 0,003 m, enquanto para o conjunto de "erros grandes" foi de 0,173 m. Ainda, nota-se na tabela que a performance do modelo cai consideravelmente para os maiores erros, o que mostra grande discrepância entre os erros, onde o Skill apresenta valores de 0,48 em comparação aos 0,99.

Figura 5.19 - Série temporal dos erros absolutos obtidos para as previsões geradas para o ano de 2022, com destaque para erros abaixo do percentil 10 e acima do percentil 90.

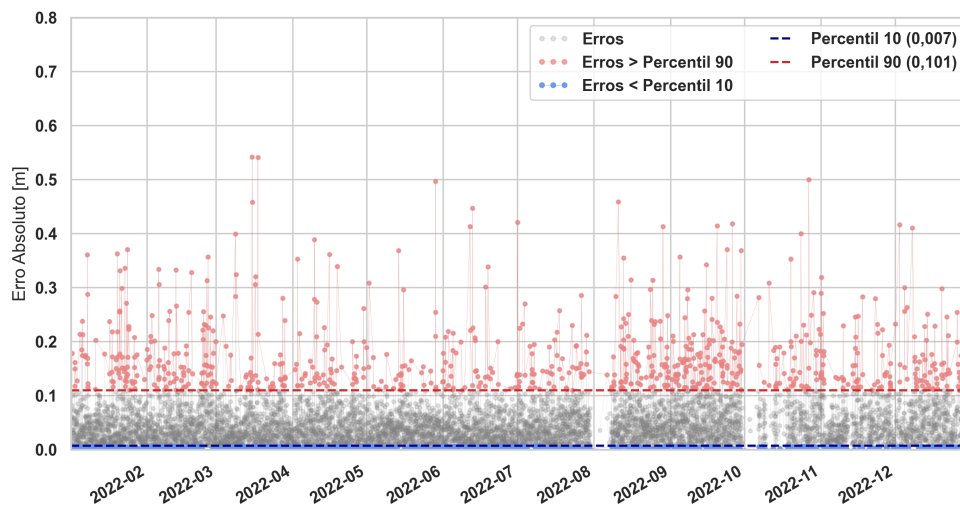
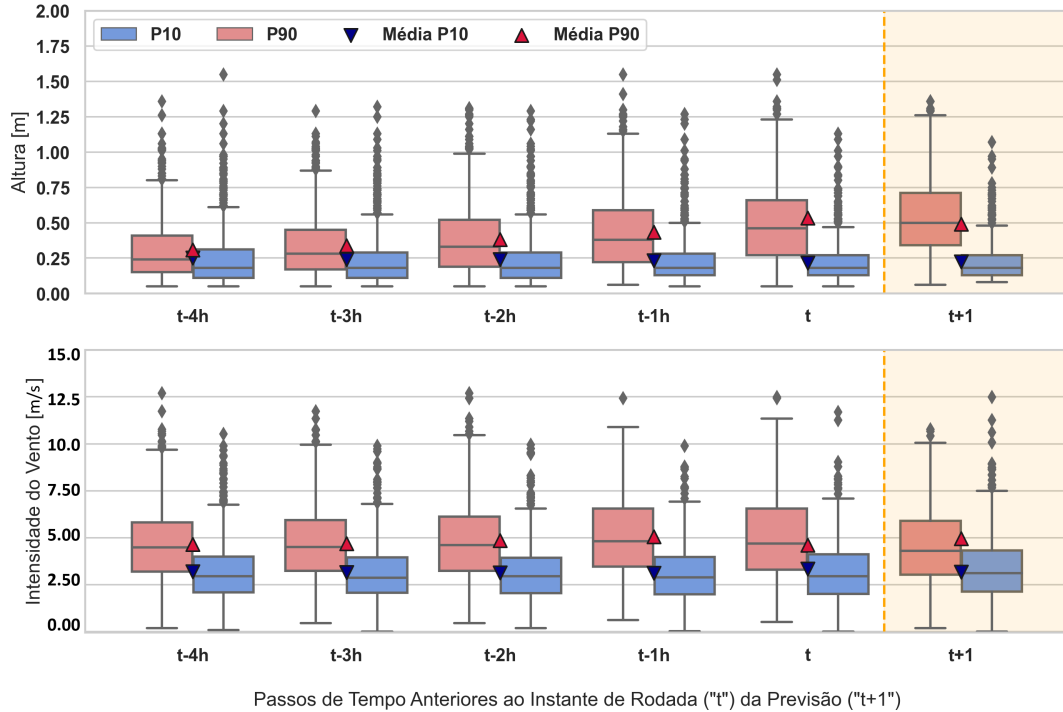


Tabela 5.4 - Métricas dos maiores e menores erros do modelo para o ano de 2022.

Conjunto	RMSE	MAE	Skill
(erros < p10)	0,004 m	0,003 m	0,99
(erros > p90)	0,187 m	0,173 m	0,48

A Figura 5.20 mostra uma comparação dos dois conjuntos de dados ("erros grandes" e "erros pequenos") na forma de gráficos do tipo *boxplot* para cada passo de tempo anterior ao instante de rodada do modelo ( $t-4h$  até  $t$ ), e também para o passo de tempo de previsão ( $t+1$ ), para cada ocorrência observada no ano de 2022. Os resultados mostram que os "erros grandes" de previsão ocorrem, em média, em momentos de alturas de ondas mais altas do que os "erros pequenos". De modo geral, as ocorrências de "erros grandes" mostram uma tendência de aumento gradual das alturas de ondas conforme os passos de tempo se aproximam do horário de rodada do modelo. Já as ocorrências de "erros pequenos" não apresentam tal tendência.

Figura 5.20 - Comparação entre os momentos de "erros pequenos" (erro absoluto < percentil 10) e de "erros grandes" (erro absoluto > percentil 10) para cada passo de tempo anterior ao instante de rodada do modelo (t-4h até t) e para o passo de tempo de previsão (t+1h), para o ano de 2022.



A Figura 5.21 apresenta a quantidade de ocorrências de "erros pequenos" e de "erros grandes" para cada mês ao longo do ano de 2022. Podemos observar que o mês de setembro foi o mês que apresentou as maiores ocorrências de "erros grandes", com 114 ocorrências, seguido dos meses de fevereiro e agosto, com 79 e 78 ocorrências, respectivamente. O mês de setembro também foi o mês com ventos predominantemente mais intensos e alturas de ondas predominantemente mais altas, como visto anteriormente no Capítulo 5.1, Figura 5.7, na análise exploratória dos dados de 2020 e 2021. Já o mês de março de 2022 foi o que apresentou a maior ocorrência de "erros pequenos", com 86 ocorrências, seguido dos meses de janeiro e abril, com 84 e 82 ocorrências, respectivamente. Na Figura 5.22 estão ilustrados os histogramas dos dados de altura de onda e de intensidade do vento para os dois conjuntos analisados, evidenciando a predominância de ventos menos intensos, e alturas mais baixas de ondas, no conjunto de "erros pequenos", quando em comparação com o conjunto de "erros grandes".

Figura 5.21 - Quantidade de ocorrências de "erros pequenos"(erro absoluto < percentil 10) e de "erros grandes"(erro absoluto > percentil 10) do modelo ao longo de cada mês do ano de 2022.

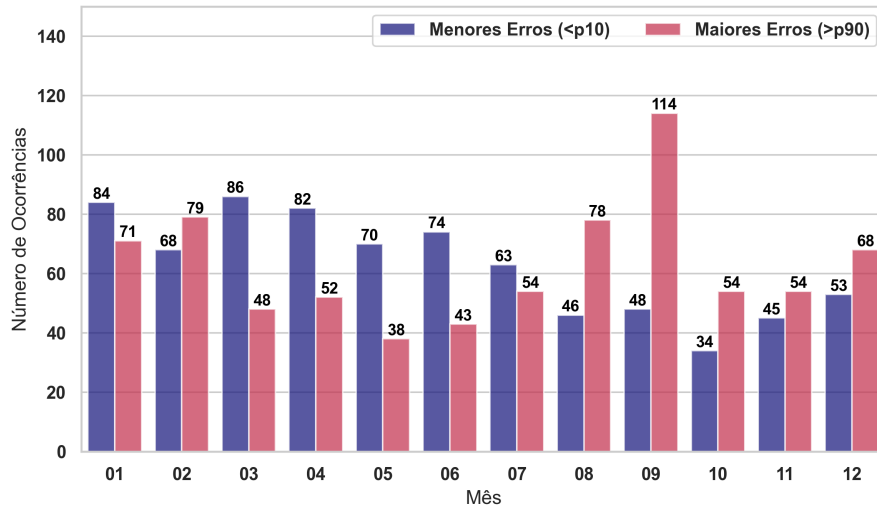
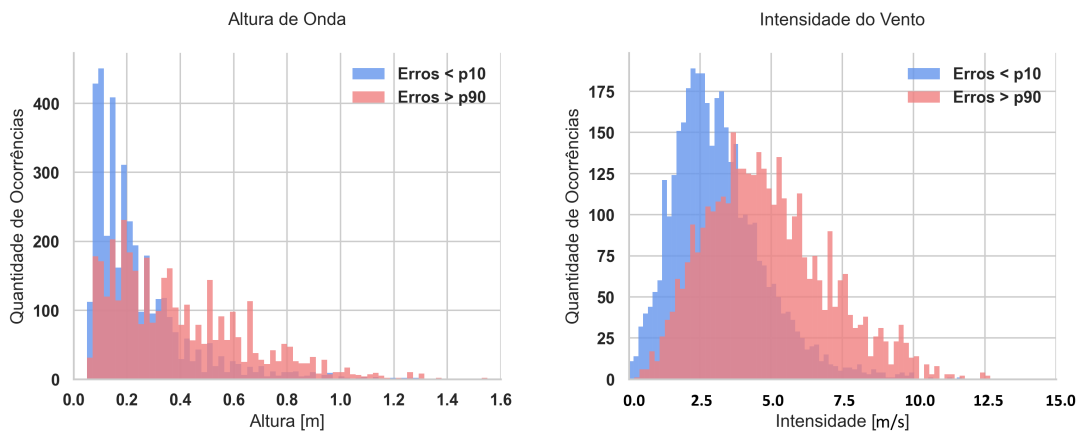


Figura 5.22 - Distribuição dos valores de Altura das Ondas e de Intensidade dos Ventos em momentos de "erros pequenos"(erro absoluto < percentil 10) e de "erros grandes"(erro absoluto > percentil 10) do modelo para o ano de 2022.



As Figuras 5.23 e 5.24 apresentam a distribuição das alturas de ondas medidas ao longo do dia, bem como a distribuição dos erros absolutos das previsões de altura das ondas, também a cada hora do dia. A primeira figura mostra os resultados para o ano completo de 2022 (Figura 5.23), enquanto a segunda figura compara os resultados dos meses de março e setembro de 2022, os quais tiveram as maiores ocorrências de erros pequenos e de erros grandes, respectivamente.

Figura 5.23 - Gráficos *boxplot* representando as alturas das ondas medidas e os erros absolutos das previsões para cada hora do dia. A primeira série de gráficos ilustra as distribuições das alturas das ondas ao longo do dia, enquanto a segunda série apresenta os valores de erro absoluto das previsões correspondentes.

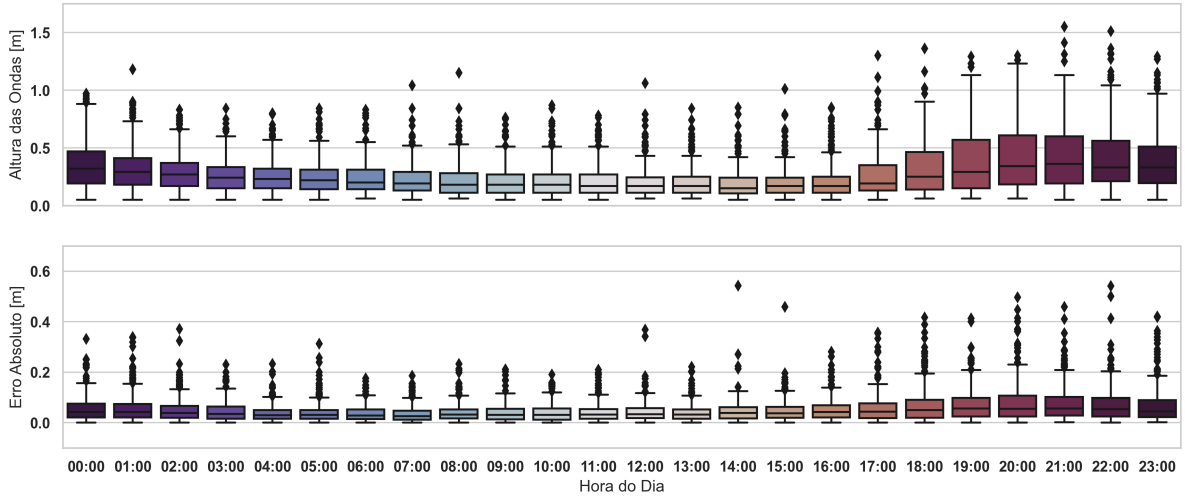
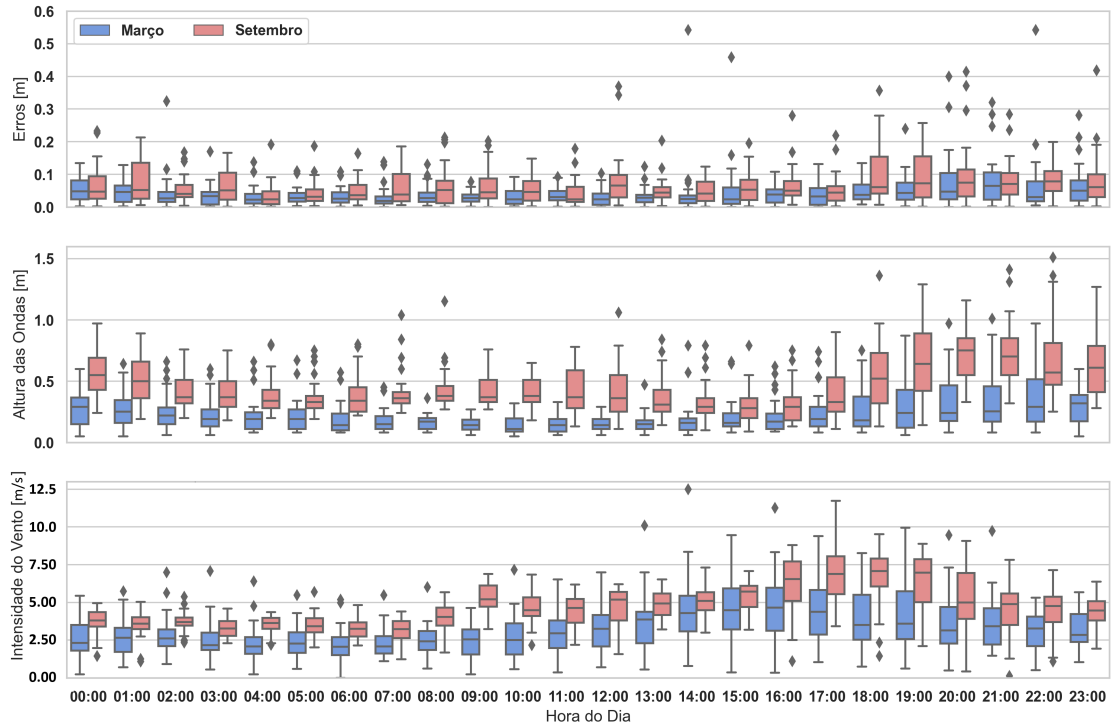


Figura 5.24 - Gráficos *boxplot* representando as alturas das ondas e as intensidades dos ventos medidos, bem como os erros absolutos das previsões de altura das ondas para cada hora do dia, para os meses de Março e Setembro de 2022.





Analisando as Figuras 5.23 e a Figura 5.24 podemos observar que os erros das previsões tendem a ser maiores nos horários do dia em que as alturas das ondas também são predominantemente mais altas, principalmente entre às 18:00 e às 22:00. Este comportamento é ainda mais evidente para o mês de setembro, no qual foram observadas as maiores ocorrências de "erros grandes" do modelo. A Figura 5.24 também mostra a defasagem temporal existente entre os horários de ventos mais intensos e ondas mais altas, como visto em capítulos anteriores, com a intensificação do vento antecedendo o aumento na altura das ondas.

Na Figura 5.25 estão ilustradas as séries temporais de altura de ondas observadas e previstas ao longo dos meses de março e de setembro de 2022. Podemos observar a semelhança entre as curvas de dados observados e previstos para ambos os meses, evidenciando que mesmo no mês com maiores ocorrências de "erros grandes" (setembro), o modelo teve resultados satisfatórios, acompanhando as tendências de aumento e diminuição das alturas das ondas na maioria das ocorrências. Na Figura 5.26, estas séries temporais estão apresentadas em maior detalhe para dois períodos onde ocorreram desvios mais significativos entre as curvas de dados observados e previstos: (i) entre os dias 14 e 16 de março; e (ii) entre os dias 13 e 19 de setembro. Com exceção do dia 15 de março e 19 de setembro, os desvios apresentados pelo modelo foram majoritariamente associados a subestimativa ou superestimativas das previsões, na qual os dados observados apresentaram alturas de ondas mais altas ou mais baixas daquelas previstas pelo modelo, mas sem erros de fase, e também sem ocasionar impactos operacionais no terminal com relação a assertividade do limite operacional.

Já nos dias 15 de março e 19 de setembro, no entanto, ocorreram subestimativas do modelo com maiores impactos operacionais, visto que as alturas previstas pelo modelo foram abaixo do limite operacional de 0,5 m, enquanto as observações foram acima deste limite (Figura 5.26). No dia 15 de março, podemos observar que no horário de 15:00 o modelo previu uma altura de onda acima de 1 m enquanto o dado medido foi inferior a 0,8 m. Este evento sucedeu o erro significativo do modelo mencionado anteriormente, que ocorreu para a previsão das 14:00, onde a altura de onda prevista foi de 0,25 m, abaixo do limite operacional do terminal, e o dado observado foi superior a 0,75 m, estando portanto acima do limite operacional do terminal. Neste caso, o modelo parece ter tido um atraso na "resposta" do aumento das alturas das ondas, o que ocasionou a superestimação da altura das ondas nos horários seguintes, às 15:00. Na sequência, o modelo consegue normalizar a assertividade, voltando a se aproximar dos dados observados a partir da previsão das 16:00. Este

evento específico ocorreu após uma intensificação da velocidade do vento, como mostra a Figura 5.27. No horário de 13:00, quando foi feita a previsão para às 14:00, o vento já havia intensificado, contudo, a resposta do modelo somente aconteceu na previsão das 15:00. Este erro foi muito semelhante ao ocorrido no dia 19 de setembro (Figura 5.28) e pode estar associado a uma limitação da metodologia aplicada, na qual optou-se por utilizar os dados subamostrados a cada 1 hora, diminuindo assim a resolução temporal das séries. Outra hipótese é que o modelo não consegue apresentar bom desempenho quando ocorrem variações abruptas entre dois períodos consecutivos. A Figura 5.29 corrobora com esta hipótese, indicando que os maiores erros do modelo, e também os erros mais críticos (quando a previsão diverge da observação com relação ao limite operacional do terminal), ocorrem majoritariamente em momentos de maiores variações entre dados de altura de ondas consecutivos.

Figura 5.25 - Séries temporais de altura das ondas observada no terminal portuário e prevista pelo modelo, para os meses de março e de setembro de 2022.

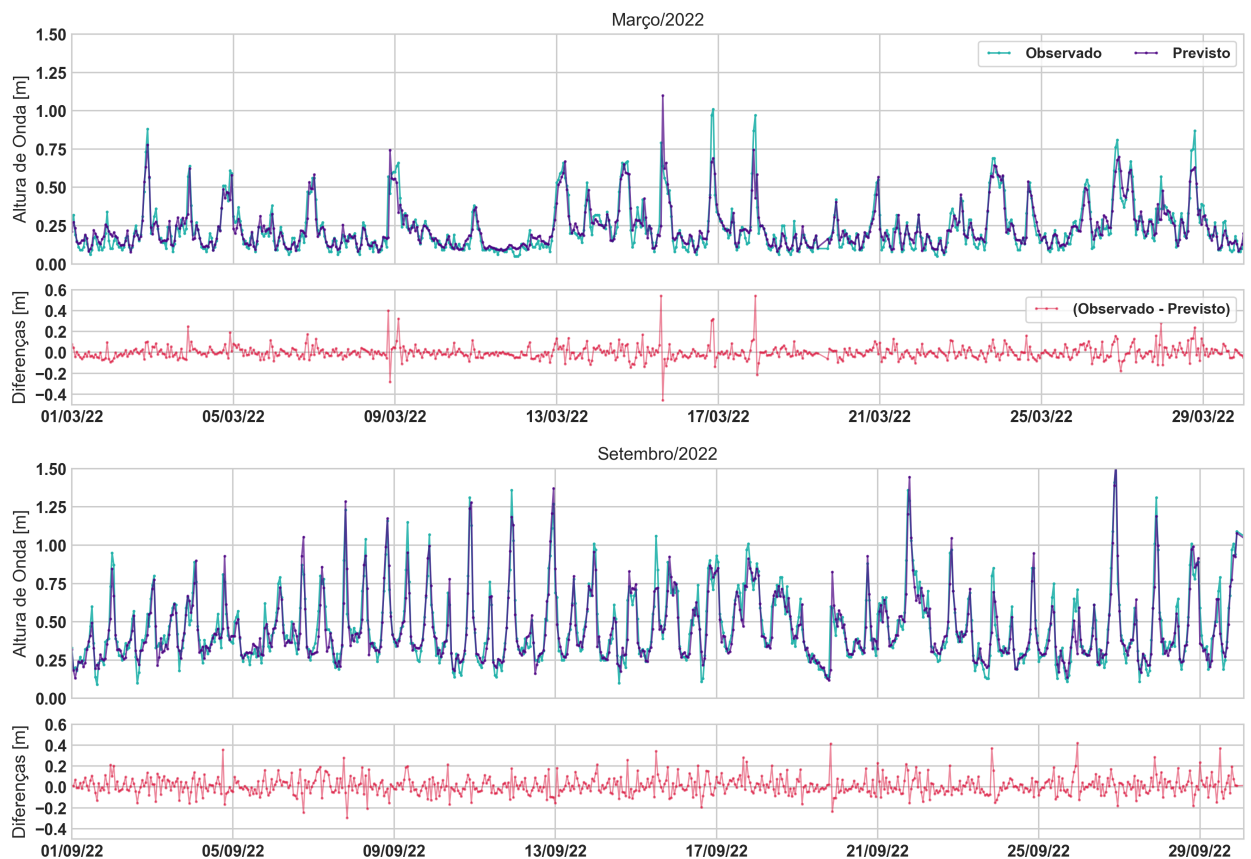


Figura 5.26 - Séries temporais de altura das ondas observada no terminal portuário e prevista pelo modelo, para os dias 13 e 18 de março de 2022, e para os dias 13 a 19 de setembro de 2022, com destaque para os momentos de maiores desvios das previsões em relação aos dados observados.

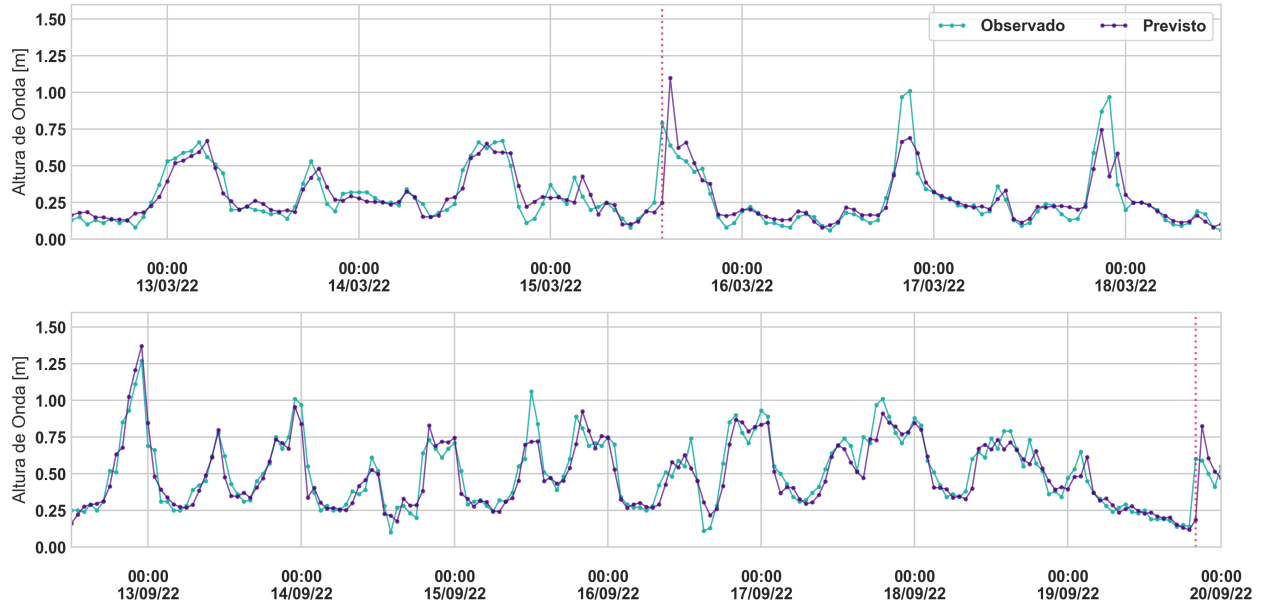


Figura 5.27 - Séries temporais de altura das ondas observadas no terminal portuário e previstas pelo modelo, bem como da intensidade e da direção dos ventos observados no local, para o dia 15 de março de 2022, com destaque para o evento das 14:00 onde ocorreu maior divergência entre as alturas observada e prevista.

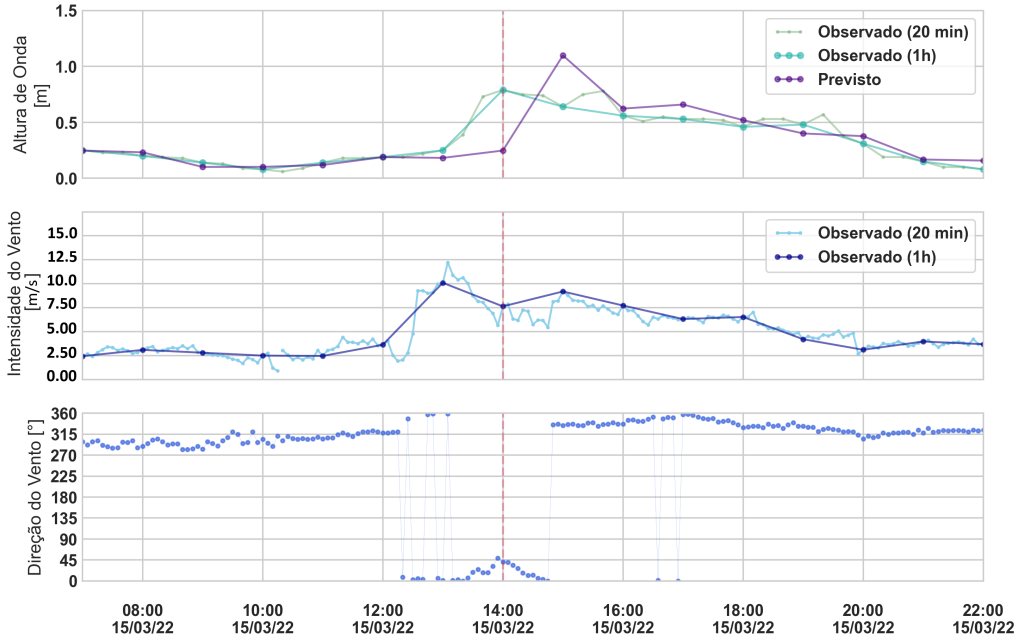


Figura 5.28 - Séries temporais de altura das ondas observadas no terminal portuário e previstas pelo modelo, bem como da intensidade e da direção dos ventos observados no local, para os dias 19 e 20 de março de 2022, com destaque para o evento das 20:00 do dia 19, onde ocorreu maior divergência entre as alturas observada e prevista.

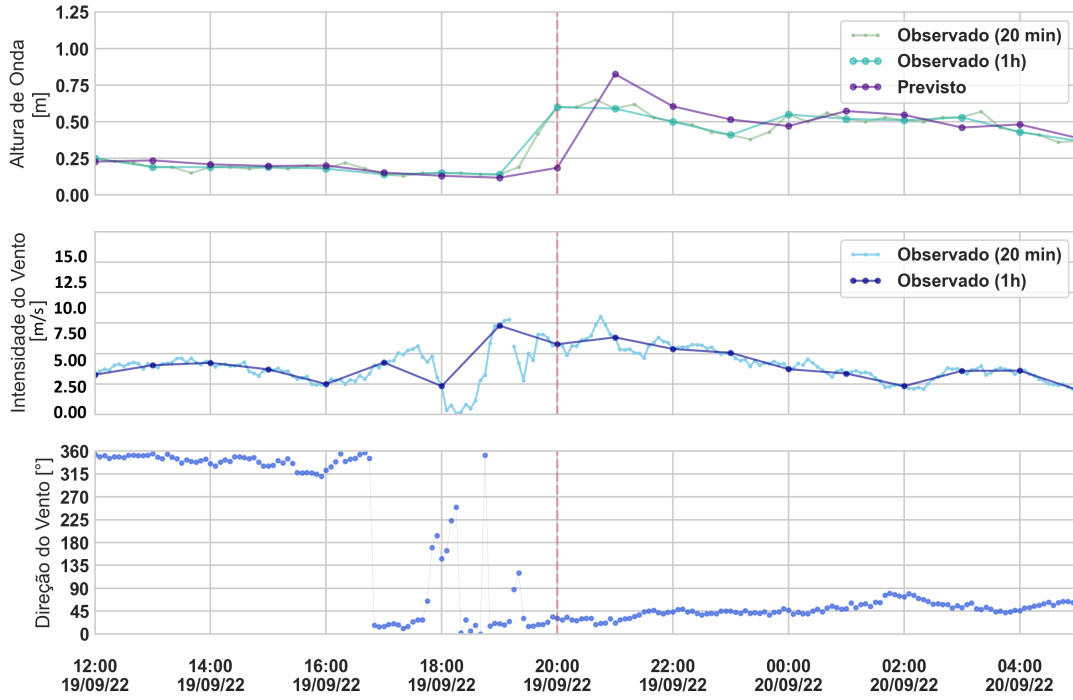
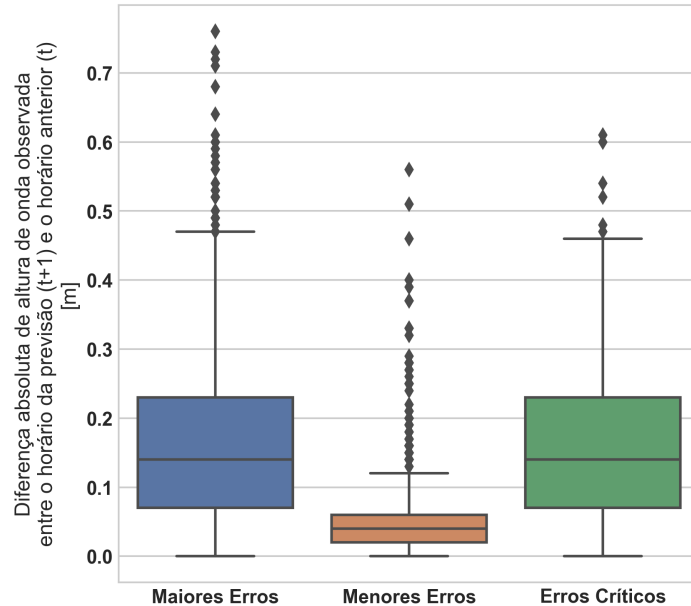
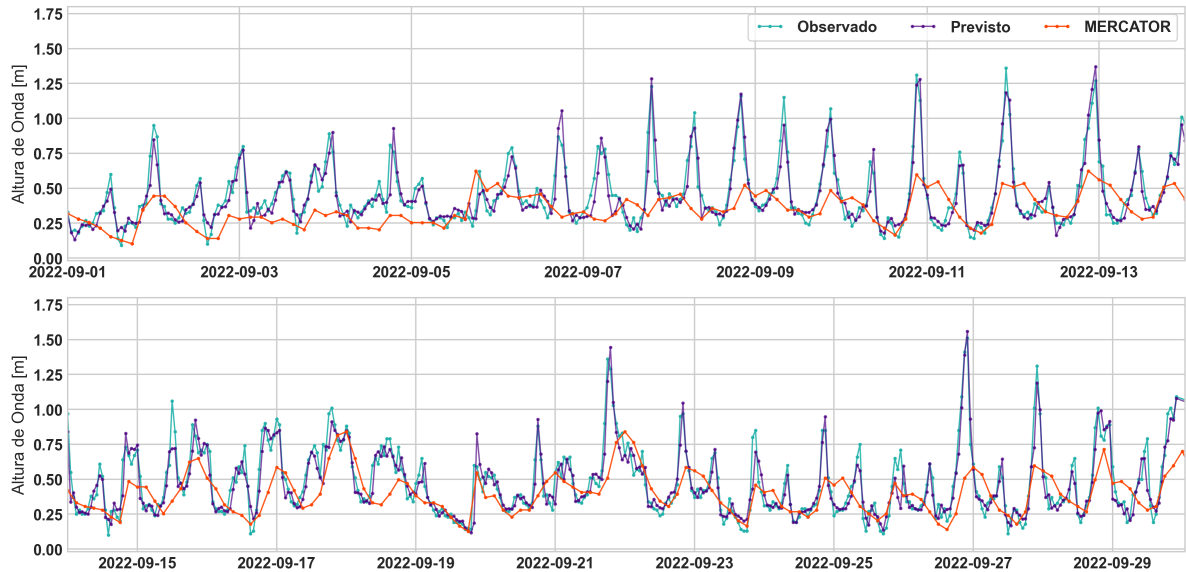


Figura 5.29 - Gráficos do tipo *boxplot* das diferenças absolutas, em metros, dos dados de altura de onda observada no horário da previsão ( $t+1$ ) e no horário imediatamente anterior à previsão ( $t$ ) nos cenários de: maiores erros do modelo (erros  $>$  percentil 90); menores erros do modelo (erros  $<$  percentil 10); e erros críticos (previsão  $<$  0,5 m e observação  $>$  0,5 m).



Por fim, a Figura 5.30 apresenta a comparação das alturas de ondas observadas na estação de monitoramento, previstas pelo modelo de LSTM desenvolvido neste trabalho, e previstas pelo modelo numérico global MERCATOR, da *E.U. Copernicus Marine Service*, para o mês de setembro de 2022, no qual o modelo de LSTM apresentou as maiores ocorrências de erros grandes. Este modelo global possui resolução espacial de  $0.083^\circ \times 0.083^\circ$  e fornece previsões com resolução temporal de 3 horas. Estas são as previsões atualmente utilizadas pelo terminal portuário. Como evidenciado na Figura, o modelo de LSTM obteve resultados superiores ao MERCATOR ao longo de todo o período. O modelo MERCATOR em muitos momentos subestimou a altura das ondas, o que resultou em um RMSE de 0,23 m e Skill de 0,08. O modelo de LSTM, por sua vez, obteve RMSE de 0,09 m (redução de 61%) e Skill de 0,84 para este período.

Figura 5.30 - Comparação entre as séries temporais de altura das ondas observadas no terminal portuário, previstas pelo modelo de LSTM e previstas pelo modelo numérico global MERCATOR para o mês de setembro de 2022.



#### 5.4 Teste de sensibilidade para *leadtime* superiores a 1 hora

Com objetivo de avaliar a capacidade do modelo desenvolvido em gerar previsões para intervalos de tempo mais longos no futuro (até 3 horas) e quantificar a redução na assertividade em cada passo de tempo, foi realizado um experimento modificando o conjunto de saída da RNA (Figura 4.6) de modo que este apresentasse os dados de "t+2h" (*leadtime* de 2 horas) e de "t+3h" (*leadtime* de 3 horas), ao invés de "t+1h" como feito anteriormente. Não foi feita alteração na arquitetura da RNA, a rede apenas foi retreinada com os novos conjuntos de saída para determinar as métricas do modelo nestes cenários. A Tabela 5.5 apresenta estes resultados para esse teste de sensibilidade do modelo.

Tabela 5.5 - Métricas das previsões para 1h, 2h e 3h no futuro para o ano de 2022.

Previsão	RMSE	MAE	Skill	R <sup>2</sup>
+1h	0,074 m	0,051 m	0,87	0,87
+2h	0,098 m	0,068 m	0,76	0,76
+3h	0,116 m	0,079 m	0,67	0,68

Nota-se que há um aumento de 32,4% no valor de RMSE para previsões para 2

horas do futuro, e de 56,8% para previsões para 3 horas no futuro, em comparação com o valor de RMSE para previsões para 1 hora no futuro. Com relação à métrica Skill, houve diminuição de 5% e de 15%, respectivamente, para as previsões de +2h e +3h.

Os resultados indicam, portanto, que há uma diminuição na assertividade das previsões do modelo para intervalos de tempo mais longos, principalmente para previsões de +3h. Contudo, apesar da redução, os valores de RMSE obtidos para as previsões de até 3 horas se mantiveram abaixo do RMSE calculado para o modelo global MERCATOR no mês de setembro de 2022, como visto no capítulo anterior, sugerindo que tais resultados podem estar ainda com qualidade superior à do modelo global.

Para trabalhos futuros, sugerimos que seja feita a reavaliação dos hiper-parâmetros e da arquitetura da RNA, de acordo com os novos conjuntos de dados de saída, a fim de buscar resultados mais otimizados para estes cenários.



## 6 CONCLUSÕES

Neste trabalho foi analisada a aplicação da rede neural artificial LSTM para o desenvolvimento de um modelo de previsão de curto prazo de altura de ondas no Terminal Portuário de Barcarena, a partir de dados observados *in situ* e extraídos de imagens do satélite GOES16. Também foi realizada análise exploratória do conjunto de dados e identificados os principais fenômenos meteo-oceanográficos atuantes no local de estudo, como as variações sazonais e ciclos diurnos.

A análise exploratória dos dados revelou uma elevada correlação assíncrona entre as séries temporais de altura de onda e intensidade do vento médio, bem como entre as séries de altura de onda e temperatura de brilho do topo das nuvens, obtidas a partir das imagens do satélite GOES-16. Esses padrões podem refletir os ciclos diurnos naturais de convecção, marés e brisas da região. Foi observado que os ventos mais intensos e as maiores alturas de onda ocorrem predominantemente nos meses de setembro e outubro na área de estudo. Além disso, foi verificado que as maiores alturas de ondas normalmente ocorrem no final da tarde e durante a noite, após a intensificação dos ventos locais.

Os resultados indicam que o pico máximo diário de altura das ondas na estação de monitoramento ocorre, em média, 4h40 após o maior pico diário de intensidade do vento e 2h20 após o menor pico diário de temperatura de brilho na área de estudo. Além disso, foi identificada tendência de decaimento na temperatura de brilho média, mínima e máxima, acompanhada de um aumento no desvio padrão da temperatura de brilho, em momentos que antecedem os picos de altura de onda em eventos mais extremos (altura superior ao percentil 99), sugerindo a presença de cobertura de nuvens e possíveis tempestades associadas a esses eventos.

Os resultados dos experimentos com a LSTM mostraram que a inclusão de variáveis adicionais, como dados de ventos e nível do mar, melhora significativamente a precisão das previsões em comparação com o uso exclusivo de dados de altura de ondas. O modelo que incorporou as variáveis supracitadas superou os demais, reduzindo os erros médios em 54% em relação ao modelo treinado apenas com dados de altura de ondas. A inclusão dos dados de temperatura de brilho também melhorou as previsões, principalmente quando comparado ao modelo que utilizou apenas dados de ventos e altura de ondas no treinamento, contudo, seu desempenho não superou o modelo treinado com dados de ventos, altura de ondas e nível do mar, o qual obteve o melhor desempenho entre todos os modelos testados. Quando os dados de temperatura de brilho foram adicionados ao modelo de melhor desempenho, os resultados

foram semelhantes, porém com desempenho ligeiramente inferior. Estes resultados ressaltam a importância de selecionar cuidadosamente o conjunto de entrada para aplicação da LSTM para se obter previsões mais assertivas de altura de ondas.

O modelo preditivo de altura de ondas desenvolvido utilizando dados de ventos, nível do mar e altura de ondas obteve resultados satisfatórios, com RMSE de 0,0738 m e *Skill* de 0,87 para previsões para 1 hora no futuro, reproduzindo os principais padrões e tendências dos dados de altura de ondas observadas no terminal para este curtíssimo prazo (*nowcasting* de 1 hora) ao longo do ano de 2022. Os resultados deste modelo foram comparados com as previsões do modelo numérico global MERCATOR para o mês de setembro de 2022, mostrando uma redução de 61% nos erros médios em relação ao modelo global. Com relação ao limite operacional do terminal portuário, o modelo obteve probabilidade de detecção de 87% e Taxa de Falsos Alarmes de 3% para o ano de 2022, o que indica que o modelo foi capaz de prever com boa confiabilidade se as alturas das ondas estavam acima ou abaixo do limite operacional na próxima hora. A análise dos erros do modelo revelou que, apesar dos bons resultados encontrados, as previsões tendem a ser menos precisas em condições de ondas mais altas e quando há um aumento abrupto na altura das ondas entre medidas consecutivas. Esta última condição pode levar a um atraso de 1 hora na resposta do modelo.

O teste de sensibilidade do modelo para previsões superiores a 1 hora no futuro indicaram que há um aumento de 32,4% e de 56,8% no RMSE, para previsões para 2 horas e para 3 horas no futuro, respectivamente, quando comparado à previsão para 1 hora no futuro. Para trabalhos futuros, sugere-se que seja feita reavaliação dos hiper-parâmetros da rede, de modo a otimizar o modelo para as previsões mais longas, visto que maiores *leadtimes* podem ser de grande valia para o planejamento de atividades portuárias.

Este estudo ofereceu, portanto, a oportunidade de pesquisar e implementar uma ferramenta de IA para a previsão de altura de ondas na região costeira de Barcarena, especificamente nas proximidades do Terminal Portuário de Barcarena. Os resultados obtidos para o ano de 2022 indicam que a LSTM é uma ferramenta promissora para *nowcasting* de ondas na região, com potencial para aplicação em operações portuárias.

Por fim, visando aumentar a precisão das previsões, sugere-se para trabalhos futuros a utilização de dados com maior resolução temporal, como por exemplo de 20 minutos, a fim de tornar o modelo mais sensível às variações de pequena escala

temporal observadas tanto nos dados de ventos como nos dados de ondas do local de estudo. Ainda, recomenda-se que sejam realizados estudos com modelos híbridos no local, combinando a LSTM com modelos numéricos numéricos, principalmente para geração de previsões mais longas. A LSTM pode ser utilizada para correção do *bias* de um modelo numérico, por exemplo. Outra abordagem seria utilizar as previsões de modelos numéricos meteorológicos como *input* da LSTM para previsão de altura de ondas.



## REFERÊNCIAS BIBLIOGRÁFICAS

- ABDULLAH, F.; NINGSIH, N.; AL-KHAN, T. Significant wave height forecasting using long short-term memory neural network in Indonesian waters. **Journal of Ocean Engineering and Marine Energy**, v. 8, n. 2, p. 183–192, 2022. 13, 14
- AHMAD, H. Machine learning applications in oceanography. **Aquatic Research**, v. 2, n. 3, p. 161–169, 2019. 1, 2
- AKIBA, T.; SANO, S.; YANASE, T.; OHTA, T.; KOYAMA, M. Optuna: A next-generation hyperparameter optimization framework. In: **Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**. New York, NY, USA: Association for Computing Machinery, 2019. (KDD '19), p. 2623–2631. ISBN 9781450362016. Disponível em: <<<https://doi.org/10.1145/3292500.3330701>>>. 40
- ALEXANDRE, E.; CUADRA, L.; NIETO-BORGE, J.; CANDIL-GARCÍA, G.; PINO, M. D.; SALCEDO-SANZ, S. A hybrid genetic algorithm—extreme learning machine approach for accurate significant wave height reconstruction. **Ocean Modelling**, v. 92, p. 115–123, 2015. 2
- BOUIJ, N.; HOLTHUIJSEN, L.; RIS, R. The "swan" wave model for shallow water. In: **Coastal Engineering 1996: Proceedings of the Twenty-Fifth International Conference**. Orlando, FL, USA: ASCE, 1996. p. 668–676. 11
- BOUKABARA, S.-A.; KRASNOPOLSKY, V.; STEWART, J. Q.; MADDY, E. S.; SHAHROUDI, N.; HOFFMAN, R. N. Leveraging modern artificial intelligence for remote sensing and nwp: benefits and challenges. **Bulletin of the American Meteorological Society**, v. 100, n. 12, p. ES473–ES491, 2019. 12
- CHANG, W.-Y. et al. A literature review of wind forecasting methods. **Journal of Power and Energy Engineering**, v. 2, n. 04, p. 161, 2014. 2
- CHAU, K. A review on the integration of artificial intelligence into coastal modeling. **Journal of environmental management**, v. 80, n. 1, p. 47–57, 2006. 12
- COPERNICUS MARINE SERVICE. **Mediterranean Sea - high resolution diurnal subskin sea surface temperature analysis**. 2021. Disponível em: <<<https://doi.org/10.48670/moi-00171>>>. 25
- COSTA, R. M. C. M. O.; SOARES, C. G. Enhancing the accuracy of metocean hindcasts with machine learning models. **Ocean Engineering**, v. 287, p. 115724, 2023. ISSN 0029-8018. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S002980182302108X>>>. 13
- COTRIM, C. d. S.; SEMEDO, A.; LEMOS, G. Brazil wave climate from a high-resolution wave hindcast. **Climate**, v. 10, n. 4, 2022. ISSN 2225-1154. Disponível em: <<<https://www.mdpi.com/2225-1154/10/4/53>>>. 9, 29

DEO, M.; NAIDU, C. S. Real time wave forecasting using neural networks. **Ocean Engineering**, v. 26, n. 3, p. 191–203, 1998. ISSN 0029-8018. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S0029801897100257>>>. 13

FAN, S.; XIAO, N.; DONG, S. A novel model to predict significant wave height based on long short-term memory network. **Ocean Engineering**, v. 205, p. 107298, 2020. 2, 13, 14, 16, 42

FRANCO, A. S. F. **Marés: fundamentos de oceanografia física**. 2. ed. Rio de Janeiro, Brasil: Interciência, 2005. ISBN 9788571931535. 34

FRATE, F. D.; PETROCCHI, A.; LICHTENEGGER, J.; CALABRESI, G. Neural networks for oil spill detection using ers-sar data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 38, n. 5, p. 2282–2287, 2000. 2

GARRISON, T. **Fundamentos de oceanografia**. 2. ed. São Paulo: Cengage Learning, 2002. 7, 8, 9

GERMANO, M. F.; VITORINO, M. I.; COHEN, J. C. P.; COSTA, G. B.; SOUTO, J. I. d. O.; REBELO, M. T. C.; SOUSA, A. M. L. de. Analysis of the breeze circulations in eastern Amazon: an observational study. **Atmospheric Science Letters**, v. 18, n. 2, p. 67–75, 2017. 33

GÉRON, A. **Mãos à obra: aprendizado de máquina com Scikit-Learn & TensorFlow**. [S.l.]: Alta Books, 2019. 16

GRACIA, S.; OLIVITO, J.; RESANO, J.; BRIO, B. Martin-del; ALFONSO, M. de; ÁLVAREZ, E. Improving accuracy on wave height estimation through machine learning techniques. **Ocean Engineering**, v. 236, p. 108699, 2021. 12

HARARI, J. **Noções de oceanografia**. São Paulo: Instituto Oceanográfico da Universidade de São Paulo, 2021. E-book. ISBN 978-65-995854-0-1. 7, 8, 9, 10, 11

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997. 13

HSIEH, W. W. **Machine learning methods in the environmental sciences: neural networks and kernels**. Cambridge: Cambridge University Press, 2009. 2

JAMES, S. C.; ZHANG, Y.; O'DONNCHA, F. A machine learning framework to forecast wave conditions. **Coastal Engineering**, v. 137, p. 1–10, 2018. 2, 12

KOMEN, G. Dynamics and modelling of ocean waves. **Dynamics of Atmosphere and Oceans**, v. 25, n. 4, p. 276, 1994. 12

MANDAL, S.; PRABAHARAN, N. Ocean wave forecasting using recurrent neural networks. **Ocean Engineering**, v. 33, n. 10, p. 1401–1410, 2006. 2, 13

MAXWELL, A. E.; WARNER, T. A.; FANG, F. Implementation of machine-learning classification in remote sensing: an applied review. **International Journal of Remote Sensing**, v. 39, n. 9, p. 2784–2817, 2018. 2, 12

MILO, T.; SOMECH, A. Automating exploratory data analysis via machine learning: an overview. In: **ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT DATA, 2020. Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data.** San Francisco, CA, USA: ACM, 2020. p. 2617–2622. 27

MINUZZI, F. C.; FARINA, L. A deep learning approach to predict significant wave height using long short-term memory. **Ocean Modelling**, v. 181, p. 102151, 2023. ISSN 1463-5003. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S1463500322001652>>>. 11, 12

MORAES, A. C. R. **Contribuições para a gestão da zona costeira do Brasil: elementos para uma geografia do litoral brasileiro.** [S.l.]: Annablume, 2007. 1

PEDLOSKY, J. **Waves in the ocean and atmosphere: introduction to wave dynamics.** [S.l.]: Springer, 2003. 7

PERES, D.; IUPPA, C.; CAVALLARO, L.; CANCELLIERE, A.; FOTI, E. Significant wave height record extension by neural networks and reanalysis wind data. **Ocean Modelling**, v. 94, p. 128–140, 2015. 12

RAMMING, H.-G.; KOWALIK, Z. **Numerical modelling of marine hydrodynamics.** [S.l.]: Elsevier, 1980. 1

SCHMIT, T. J.; GRIFFITH, P.; GUNSHOR, M. M.; DANIELS, J. M.; GOODMAN, S. J.; LEBAIR, W. J. A closer look at the abi on the goes-r series. **Bulletin of the American Meteorological Society**, v. 98, n. 4, p. 681–692, 2017. 18

STAUDEMAYER, R. C.; MORRIS, E. R. Understanding lstm—a tutorial into long short-term memory recurrent neural networks. **arXiv preprint arXiv:1909.09586**, 2019. 15

TOLMAN, H. L. A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents. **Journal of Physical Oceanography**, v. 21, n. 6, p. 782–797, 1991. 11

TOLMAN, H. L. et al. User manual and system documentation of wavewatch iii tm version 3.14. **Technical note, MMAB contribution**, v. 276, n. 220, 2009. 12

WANG, M.; YING, F. Point and interval prediction for significant wave height based on lstm-gru and kde. **Ocean Engineering**, v. 289, p. 116247, 2023. ISSN 0029-8018. Disponível em: <<<https://www.sciencedirect.com/science/article/pii/S0029801823026318>>>. 11, 12, 13

WEI, C.-C.; CHANG, H.-C. Forecasting of typhoon-induced wind-wave by using convolutional deep learning on fused data of remote sensing and ground measurements. **Sensors**, v. 21, n. 15, p. 5234, 2021. 2

WEI, Z. Forecasting wind waves in the us atlantic coast using an artificial neural network model: towards an ai-based storm forecast system. **Ocean Engineering**, v. 237, p. 109646, 2021. [11](#), [13](#), [14](#)

WILLMOTT, C. J. On the validation of models. **Physical Geography**, v. 2, n. 2, p. 184–194, 1981. [23](#)

ZAMANI, A.; SOLOMATINE, D.; AZIMIAN, A.; HEEMINK, A. Learning from data for wind–wave forecasting. **Ocean Engineering**, v. 35, n. 10, p. 953–962, 2008. [2](#), [13](#)