PREVISÃO DE ALTURA DE ONDA NA BAÍA DE SANTOS/SP: UMA ABORDAGEM HÍBRIDA DE MODELOS NUMÉRICOS GLOBAIS E INTELIGÊNCIA ARTIFICIAL

WAVE HEIGHT FORECAST IN SANTOS BAY/SP: A HYBRID APPROACH OF GLOBAL NUMERICAL MODELS AND ARTIFICIAL INTELLIGENCE

Alec Gomez Yagi Rodrigues, Alexandra Franciscatto Penteado Sampaio, Matheus Souza Ruiz, Regina de Souza Ferreira, Renan Braga Ribeiro

Universidade Santa Cecília, Faculdade de Engenharia, Curso de Engenharia da Computação
E-mail para contato: alecrodrigues55@gmail.com

RESUMO – A cidade de Santos, no litoral de São Paulo, Brasil, enfrenta com frequência eventos meteoceanográficos extremos, como tempestades e ressacas, que vem se agravando com as alterações climáticas e afetando a segurança portuária e costeira. Este estudo visa avaliar a melhora na previsão destes eventos utilizando técnicas de aprendizado de máquina aplicadas a modelos numéricos globais. Para isto, foram usados algoritmos como Random Forest Regressor, Extra Tree Regressor e Gradient Boosting Regressor, que demostraram alta correlação (R > 0,95) e baixos erros (RMSE < 20cm). No entanto, verifica-se que o modelo apresentou dificuldade em prever os eventos mais intensos, com ondas superiores a 2m de altura. Como trabalho futuro, portanto, recomenda-se incluir um período mais longo e com mais ocorrências de eventos intensos para treinamento dos modelos.

Palavras-chave: Altura de Onda; Previsão Oceanográfica; Aprendizado de Máquina.

ABSTRACT – The city of Santos, located on the coast of São Paulo, Brazil, frequently experiences extreme metoceanographic events, such as storms and swells, exacerbated by climate change, which impact port and coastal safety. This study aims to enhance the forecasting of these events using machine learning techniques applied to global numerical models. Algorithms such as Random Forest Regressor, Extra Tree Regressor, and Gradient Boosting Regressor were employed, demonstrating high correlation (R > 0.95) and low errors (RMSE < 20 cm). However, the models struggled to predict more intense events, particularly waves exceeding 2 meters in height. Future work should focus on incorporating a longer historical period with more occurrences of intense events to improve model training and performance.

Keywords: Wave Height; Oceanographic Forecasting; Machine Learning.

Tipo do trabalho: (X) Iniciação científica ( ) Iniciação tecnológica ( ) Extensão

1 INTRODUÇÃO

A cidade litorânea de Santos, localizada no estado de São Paulo, Brasil, é alvo de frequentes eventos meteoceanográficos extremos, como tempestades, ressacas e elevação do nível do mar (Campos *et al*., 2010). O impacto causado por movimentos de maré depende da época do ano, contudo foi estudado um maior número de eventos extremos em regiões de águas rasas e durante o inverno (Camargo & Harari, 1994). O contínuo aprimoramento da previsão desses eventos através de modelos numéricos de alta resolução (Ribeiro *et al*., 2019) e da implementação de técnicas de aprendizado de máquina (Londhe *et al*., 2016; Callens *et al*., 2020) tem se mostrado importante para mitigar os impactos e garantir a segurança das operações portuárias e das comunidades costeiras.

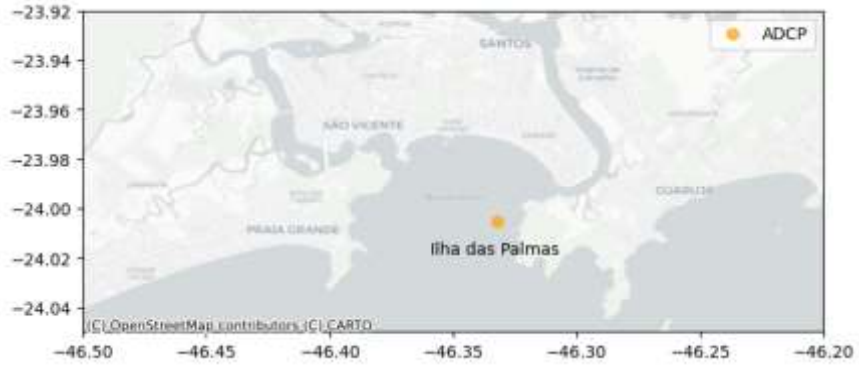
O objetivo deste estudo é o desenvolvimento de um modelo híbrido que ofereça resultados mais assertivos das previsões de altura significativa das ondas, integrando técnicas de aprendizado de máquina a modelos numéricos globais, contribuindo para a gestão de riscos climáticos e operacionais na região de Santos.

2 MATERIAIS E MÉTODOS

2.1 Dados Requeridos

Neste estudo, foram obtidos dados de altura significativa das ondas (Hs, em metros) de um ADCP (A*coustic Doppler Current Profiler*) localizado na estação Ilha das Palmas (Latitude: 24° 0' 18.42" S; Longitude: 46° 19' 55.74" W). Os dados são pertinentes ao sistema de monitoramento meteoceanográfico em tempo real da Praticagem de São Paulo. Também foram utilizados resultados dos modelos de ondas *CMS-Wave* (Ardhuin *et al*., 2010), administrado pela *Copernicus Marine Service* (CMS) e *WAVEWATCH III* (Cheung, 2010), desenvolvido pela *Marine Modeling and Analysis Branch* (MMAB) do *National Centers for Environmental Prediction* (NCEP).

Figura 1. Localização geográfica do ADCP Ilha das Palmas.



2.2 Algoritmos Utilizados

Os algoritmos de aprendizado de máquina aplicados em cada uma das etapas foram *Random Forest Regressor* (RF), *Extra Trees Regressor* (ET) e *Gradient Boosting Regressor* (GB) que apresentaram resultados satisfatórios e melhores, comparados a de redes neurais, um método conhecido por sua eficiência, porém de difícil entendimento e praticidade (Callens *et al*., 2020). Para análise e medição de resultados foram utilizadas técnicas como *feature importance* e plots de resultados, bem como os índices estatísticos viés, RMSE (*Root Mean Square Error*), R (correlação de Pearson) e índice *skill* (Willmott, 1982).

2.3 Tratamento dos dados medidos

Para a realização dos métodos de aprendizado de máquina, os dados foram divididos entre treinamento (início de 2022 até o final de 2023) e teste (início de 2024 até junho). Foi necessário tratar os dados, removendo períodos com dados ausentes ou inconsistentes, e, em seguida, preencher os períodos ausentes com o valor médio registrado de altura significativa da onda.

2.4 Seleção de hiperparâmetros e atributos

Todo algoritmo de aprendizado de máquina pode ser aprimorado e ajustado com hiperparâmetros, que tem função de otimizar o modelo estatístico. Para cada caso, é preciso verificar quais hiperparâmetros melhor se adequam aos algoritmos utilizados; após esta verificação com os dados utilizados, os resultados são apresentados na Tabela 1 a seguir.

Tabela 1. Tabela de hiperparâmetros dos métodos usados.

Hiperparâmetros	***Random Forest Regressor***	***Extra Tree Regressor***	***Gradient Boosting Regressor***
n_estimators	900	900	900
min_sample_split	15	10	5
min_sample_leaf	10	6	10
max_features	0.2	0.4	0.2
max_depth	12	9	6
bootstrap	True	True	-
subsample	-	-	0.1
learning_rate	-	-	0.01
ccp_alpha	0.0	0.0	0.0
criterion	squared_error	squared_error	squared_error

As variáveis escolhidas para todos os três modelos vieram da realização de um processo chamado *Feature Importance*, que analisa os parâmetros utilizados para um aprendizado de máquina e avalia qual deles teve maior peso e relevância em seu resultado (Wojtas & Chen, 2020).

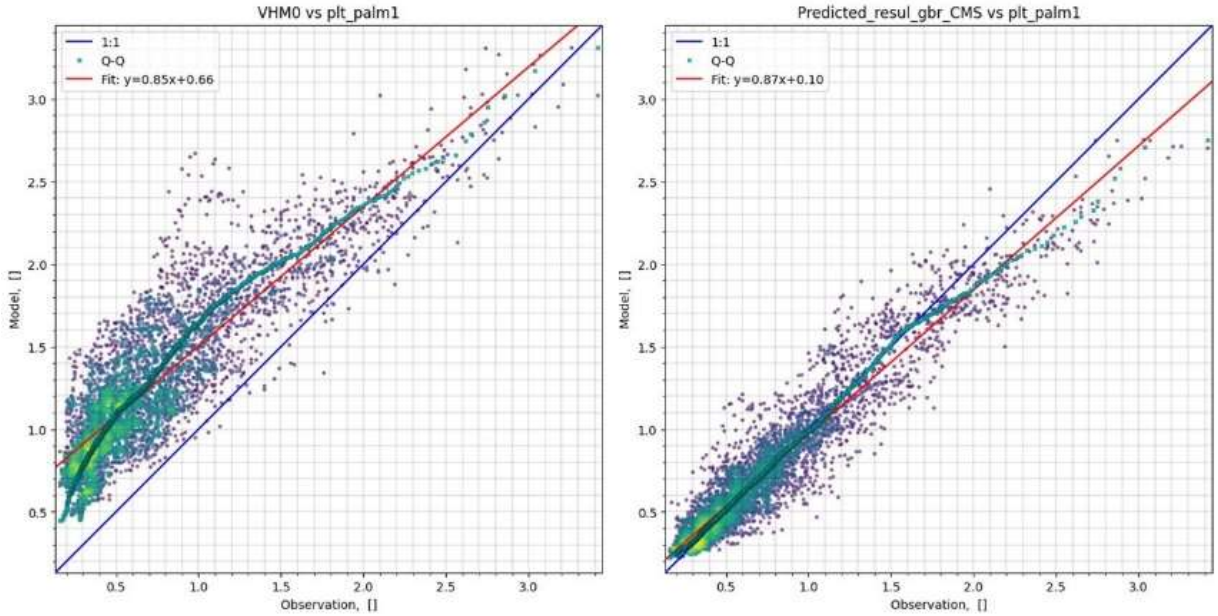
3 RESULTADOS E DISCUSSÃO

A Tabela 2 apresenta os índices estatísticos para os modelos globais e para os três algoritmos de aprendizado de máquina aplicados. Para os modelos globais, o RMSE é superior a 0,3 m e o *Skill* inferior a 0,85. Após a aplicação das técnicas de aprendizado de máquina, o RMSE foi inferior a 0,15 m e o *Skill* superior a 0,97. A Figura 2 apresenta diagramas de dispersão do modelo CMS-Wave antes e após a aplicação do método de aprendizado de máquina *Gradient Boosting Regressor*. Ficando evidente a relativa melhoria das previsões ajustadas após a aplicação da técnica de aprendizado de máquina durante o período analisado. No entanto, vale destacar que com a aplicação das técnicas de aprendizado de máquina o resultado apresentou tendência de redução da altura significativa da onda nos eventos com ondas superiores a 2 metros.

Tabela 2. Avaliação dos resultados obtidos através dos índices estatísticos, para a altura significativa das ondas na estação Ilha das Palmas.

Modelo de Previsão		**Viés [m]**	**RMSE [m]**	**R [0-1]**	***Skill* [0-1]**
Modelo Global	CMS-Wave	0.548	0.606	0.854	0.720
	Wave Watch	0.169	0.328	0.827	0.834
Previsão Ajustada	Random Forest	0.003	0.147	0.953	0.974
	Extra Trees	0.012	0.151	0.954	0.971
	Gradient Boosting	0.005	0.146	0.953	0.974

Figura 2. Diagramas de dispersão da altura significativa da onda (em metros) no ponto Ilha das Palmas, para o modelo global de ondas *CMS-Wave* antes (à esquerda) e após a aplicação de técnica de aprendizado de máquina *Gradient Boosting Regressor*.



4 CONCLUSÃO

A partir dos resultados obtidos, é possível concluir que obter previsões com maior confiabilidade a partir de resultados de modelos globais, utilizando técnicas de aprendizado de máquina é uma realidade. Foram obtidos resultados com elevados índices de correlação e Skill e com baixos viés e RMSE. Dentre os modelos aplicados, *Gradient Boosting Regressor* e *Random Forest Regressor* se apresentaram como os mais eficazes. Entretanto, é importante destacar que ainda há limitações nos resultados obtidos, principalmente na previsão de eventos com altura de ondas superiores a 2m na região. Pretende-se como continuidade em trabalhos futuros avaliar outras técnicas de aprendizado de máquina para a melhoria da previsão deste e de outros parâmetros de onda, e também operacionalizar esta previsão, como forma de fornecer resultados cada vez melhores para a região.

Este estudo faz parte do Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação (PIBITI-CNPq) e contou com apoio financeiro deste programa. Agradecimentos especiais para o Núcleo de Pesquisas Hidrodinâmicas (NPH) e a Universidade Santa Cecilia (UNISANTA).

5 REFERÊNCIAS

Ardhuin, F., Rogers, E., Babanin, A. V., Filipot, J., Magne, R., Roland, A., van der Westhuysen, A., Queffeulou, P., Lefevre, J., Aouf, L., & Collard, F. (2010). Semiempirical Dissipation Source Functions for Ocean Waves. Part I: Definition, Calibration, and Validation. Journal of Physical Oceanography, 40(9), 1917-1941.

Camargo, R. D., & Harari, J. (1994). Modelagem numérica de ressacas na plataforma sudeste do Brasil a partir de cartas sinóticas de pressão atmosférica na superfície. *Boletim do Instituto Oceanográfico*, *42*, 19-34.

Campos, R. M., Camargo, R. D., & Harari, J. (2010). Caracterização de eventos extremos do nível do mar em Santos e sua correspondência com as reanálises do modelo do NCEP no sudoeste do Atlântico Sul. Revista Brasileira de Meteorologia, 25, 175-184.

Callens, A., Morichon, D., Abadie, S., Delpey, M., & Liquet, B. (2020). Using Random forest and Gradient boosting trees to improve wave forecast at a specific location. Applied Ocean Research, 104, 102339.

Cheung, K. F. (2010). WaveWatch III (WW3) Global Wave Model. The Pacific Islands Ocean Observing System (PacIOOS).

Londhe, S. N., Shah, S., Dixit, P. R., Nair, T. B., Sirisha, P., & Jain, R. (2016). A coupled numerical and artificial neural network model for improving location specific wave forecast. Applied Ocean Research, 59, 483-491.

Ribeiro, R. B., Sampaio, A. F. P., Ruiz, M. S., Leitão, J. C., & Leitão, P. C. (2019). First approach of a storm surge early warning system for Santos region. In Climate change in Santos Brazil: projections, impacts and adaptation options (pp. 135-157).

Willmott, C. J. (1982). Some comments on the evaluation of model performance. Bulletin of the American Meteorological Society, 63(11), 1309-1313.

Wojtas, M., & Chen, K. (2020). Feature importance ranking for deep learning. Advances in Neural Information Processing Systems, 33, 5105-5114.