WAVE HEIGHT PREDICTION WITH GRADIENT BOOSTING: CASE STUDY IN JAMAICA BAY, NEW YORK

Hoda El Safty (1) and Michael Ibrahim (2)

(1)Irrigation and Hydraulics Department, Faculty of Engineering, Cairo University, 12613, Giza, Egypt, hodael_safty@eng.cu.edu.eg

(2)Department of Computer Engineering, Faculty of Engineering, Cairo University, 12613, Giza, Egypt, michael.nawar@eng.cu.edu.eg

Keywords: Boosting trees, CatBoost model, Coupled ADCIRC + SWAN, Wind Wave Forecast, Back Barrier Bay, Jamaica Bay.

1. **ABSTRACT:** This study introduces a machine learning framework for forecasting wave heights in back- barrier bays using meteorological inputs, specifically wind and pressure data. A high-resolution coupled hydrodynamic-wave model, driven by local weather station data and ERA5 reanalysis, produced a 40-year (1980–2019) hindcast of wave heights in Jamaica Bay, New York. After validation and bias correction using field measurements, the model output was used to train a CatBoost model. Comparative analysis at two locations showed that machine learning predictions closely matched numerical model results while reducing computational cost. The machine learning model achieved slightly Root Mean Square Errors (RMSE) and Mean Absolute Errors (MAE) values, with RMSEs of 0.065 m and 0.11 m, and MAEs of 0.045 m and 0.08 m. To demonstrate practical application, a two-week hourly forecast was generated using GFS weather input. This study highlights the potential of machine learning as a fast, accurate, and efficient tool for wave height forecasting in data-scarce coastal environments.

2. INTRODUCTION

Wind waves play a crucial role in air-sea interactions, influencing momentum, energy, heat, and moisture exchanges between the ocean and atmosphere (Smith *et al*., [1]; Drennan *et al*., [2]). Additionally, wind-wave climate impacts coastal circulation, marine ecosystems, coastal morphology, and maritime safety (Reidenbach *et al*., [3] ; Heij and Knapp, [4]; Reguero *et al*. , [5]).

In limited-fetch environments, such as back-barrier bays, waves quickly reach equilibrium with the wind, often within minutes to an hour. As wind energy is transferred to the water surface, small, short-period waves develop but remain constrained by fetch and water depth. Consequently, wave height adjustments occur rapidly in response to wind variations, necessitating high-resolution wind data to capture these dynamics effectively.

Unlike open ocean conditions, where swells dominate, wind-driven waves in back-barrier bays are highly sensitive to local wind changes. In Jamaica Bay, for example, oceanic wave influence is minimal due to the shielding effect of Rockaway Inlet, and most waves originate from local wind forcing or vessel activity. The wind climate varies from calm conditions to severe storm events, influencing marsh stability, sediment transport, and coastal erosion. Wind, waves, and currents are crucial for the sustainability of marshes within Jamaica Bay, as the stability of vegetation and sediment retention depend significantly on these coastal processes.

Numerical models, such as ADCIRC+SWAN, are widely used for wave prediction, but their computational demands pose challenges, particularly for real-time applications. High- resolution wave modeling requires substantial computational resources due to the need for large domains to properly simulate wave propagation. Additionally, numerical models may not provide timely predictions in emergency scenarios.

With advancements in machine learning (ML), data-driven models have emerged as promising alternatives for wave forecasting. Several studies have demonstrated the effectiveness of ML-based approaches, particularly deep learning models such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs). For example, Fan *et al.* , [6] used LSTM networks with wind speed, wind direction, and wave height inputs to predict significant wave height (SWH) at multiple global locations, achieving accurate short-term forecasts. Kaloop *et al*., [7] combined Wavelet Transform, Particle Swarm Optimization, and Extreme Learning Machine techniques for enhanced SWH forecasting. Guan ,[8] developed a CNN-LSTM hybrid model, utilizing CNNs for feature extraction in short- term fore- casting. Pushpam and Enigo, [9] explored various architectures of the RNN-LSTM model and found that these configurations outperformed the persistence model, particularly at longer lead times. Zhou *et al.* ,[10] applied a ConvLSTM model with WaveWatch III (WW3) reanalysis data to forecast wave heights in the South and East China Seas under both normal and extreme conditions. Feng *et al* , [11] compared RNN, LSTM, and GRU models for site-specific wave prediction in offshore China.

These studies highlight the growing adoption of ML techniques for wave forecasting, offering improved accuracy and computational efficiency. However, back-barrier bays present unique challenges, as long-term wave records are often scarce compared to open-ocean regions where buoy data is available. This limitation makes traditional deep learning approaches, which often require large datasets, less practical in such environments.

Recently, boosting tree algorithms have gained popularity for wave height prediction due to their ability to produce accurate forecasts with lower computational costs. Unlike deep learning models, gradient boosting trees are easier to implement, require fewer computational resources, and offer higher interpretability. For instance, Hu and Chu, [12] used XGBoost to predict annual wave heights in Lake Erie within just 10 minutes on a single CPU, demonstrating the efficiency of boosting-based models. Boosting trees can effectively handle small datasets, unlike deep learning models, which often require large amounts of training data. They can incorporate lagged features and rolling statistics, improving predictive performance without the need for complex recurrent architectures like LSTMs. Given these advantages, gradient boosting trees provide a compelling alternative to traditional numerical and deep learning models for wave height forecasting in data-scarce, semi-enclosed regions.

This study develops a gradient boosting trees model to predict wave heights in Jamaica Bay, New York, where publicly available wave height data is limited. To create a robust dataset for ML training, we use a coupled ADCIRC+SWAN numerical model to generate a 40-year wave hindcast (1980–2019), driven by reanalysis data and local weather station observations. After validating and bias-correcting the numerical model results, we trained a CatBoost model for wave height forecasting. The remainder of this paper is structured as follows. Section 2 introduces the coupled ADCIRC+SWAN hydrodynamic-wave model. In section 3, details the CatBoost machine learning methodology. In section 4, compares the predicted wave heights from the ML model, the numerical model, and observed field data, discussing the results. Conclusions and future research directions are summarized in section 5.

3. METHODS

3.1 Study Area

Jamaica Bay, a back-barrier bay located in New York City at a latitude of 40.6 N, connects to the western North Atlantic Ocean via the Rockaway Inlet, which ranges from 1.2 to 2.5 km in width and is approximately 6 km long (Figure 1). The bay is shielded from the Atlantic Ocean by the Rockaway Peninsula and spans an area of about 54 km2, excluding its tributaries. Tides in the bay are semidiurnal, with a mean tidal range of 1.5–1.7 m, as reported by Marsooli *et al*, [13, 14] . The average water depth within the bay is approximately 4 m, with deeper shipping channels along the perimeter reaching depths of 10 m Marsooli *et al*, [13], and a large borrow pit in the eastern region, known as Grassy Bay, reaching 15 m in depth. The bay's center is shallow, hosting intertidal salt marsh islands.

Recent aerial photographs highlight a concerning trend: the bay's salt marshes have been disappearing at an annual rate of up to 1.4% over the past century. The loss of these marshes, which provide critical ecosystem services such as wave attenuation and sediment stabilization, may influence wave propagation and hydrodynamic processes, with potential impacts on erosion rates and coastal resilience. This environmental degradation underscores the importance of understanding the bay's wave dynamics and their potential impacts on shoreline stability.

3.2 Numerical Model and Model Evaluation

We employed the third-generation spectral wave model SWAN, coupled with the hydro-dynamic model ADCIRC (Dietrich *et al*. , [15]), to simulate wind-driven waves in Jamaica Bay. This coupled system is crucial for accurately capturing wave dynamics in shallow coastal environments, as it accounts for spatially and temporally varying currents and water depths. The models share an unstructured finite element mesh (Figure 1).

Tides are incorporated using the eight primary astronomical tidal constituents, with harmonic constants derived from the TPXO Global Tidal Dataset (Egbert and Erofeeva, [16]). The computational domain encompasses Jamaica Bay and extends into the deep ocean of the western North Atlantic (Figure 1). The unstructured mesh, adapted from Marsooli *et al*, [17], resulting in a mesh comprising 666,301 triangular elements and 338,414 nodes. Mesh resolution ranges from 5 km at the offshore boundary to 20 m within Jamaica Bay, providing sufficient detail to capture the bay's bathymetry. Manning's roughness coefficients were assigned based on land cover data from the frictional effects. The ADCIRC and SWAN models are configured with computational time steps of one second and 1800 seconds, respectively, with coupling occurring every 1800 seconds. In SWAN, the wave spectrum is resolved over a full circle with a directional resolution of 15 and frequencies ranging from 0.033 to1.00Hz, discretized into 23 logarithmic bins. The model's spectral domain resolution and time steps were optimized through sensitivity analysis to achieve a balance between computational efficiency and model accuracy.

The coupled model simulates wave characteristics from 1980 to 2019, providing hourly time series of significant wave heights at five hundred computational nodes evenly distributed across the bay. Forcing data, including hourly surface pressure and wind time series, were obtained from the ERA5 global reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasting (C3S 2017), with a horizontal resolution of 0.25 , which is adequate for capturing mesoscale processes. Within Jamaica Bay, local wind and pressure data from the JFK Airport weather station (Figure 1) were used to capture finer-scale dynamics that influence wave climate in this back-barrier setting.

Model accuracy was assessed by comparing modeled wave heights to measurements taken at two locations, Rockaway Point (RP) and Canarsie Pier (CP), from May to October 2015 (Marsooli *et al*, [13]; El Safty and Marsooli, [18]; Chant *et al* , [19]). Although no long-term wave measurement datasets are available for Jamaica Bay, these datasets provide useful validation. The comparison showed that the modeled and measured 95th percentile significant wave heights (Hs) at Canarsie Pier were 0.14 m and 0.2 m, respectively, and at Rockaway Point, they were 0.31 m and 0.36 m, respectively. The relative error for each Hs percentile was calculated, and an average relative error (averaged between RP and CP) was subtracted from the hindcasted wave data. This adjustment assumes consistent model performance during the hindcast period. The wave heights used for machine learning model training were based on these error-corrected wave heights, Hs.

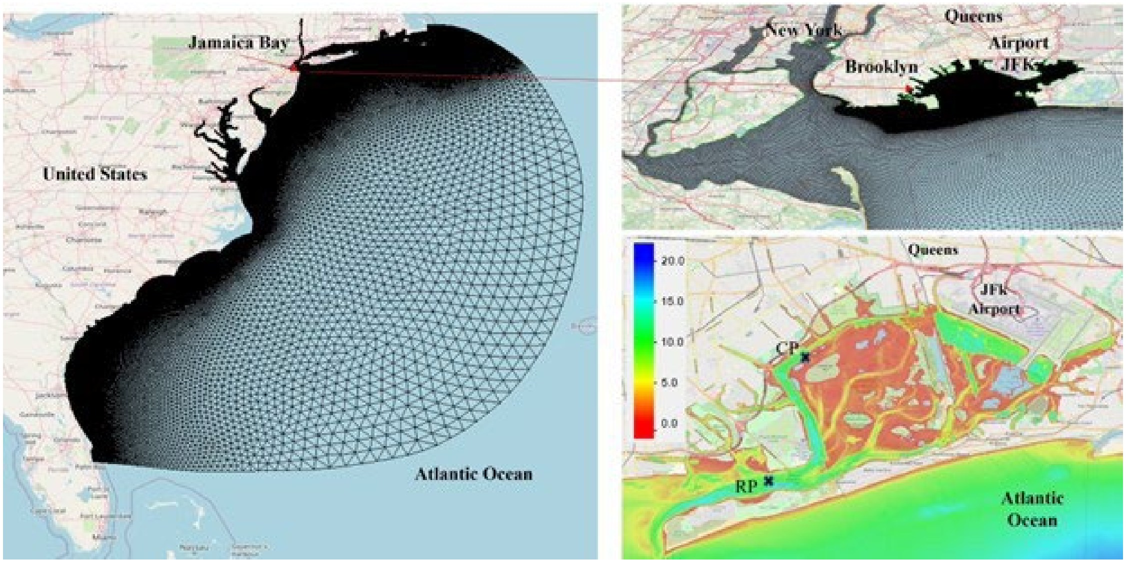

Figure 1: Computational domain of coupled ADCIRC+SWAN model. Wave measurement locations at two stations named Rockaway Point (RP) and Canarsie Pier (CP).

3.3 Analyzing Numerical Model Results

The numerical model demonstrated a high correlation between current and previous time observations, with correlation coefficients of 0.972 and 0.978 for Canarsie Pier and Rockaway Point, respectively. This strong correlation indicates that recent wave heights are good predictors for short-term forecasts. However, the autocorrelation analysis revealed that the data exhibited high autocorrelation at small lags, meaning that immediate past wave heights have a strong influence on the current value, even after accounting for the effects of intervening observations. This presents a challenge for long-term forecasting, as models relying heavily on recent trends tend to neglect broader temporal dynamics.

High autocorrelation at small lags can lead to overly focused predictions on the near past, resulting in poor long-term forecasts that fail to capture structural changes in the data over time Petropoulos *et al*., [20]. Research has shown that models using a limited number of lags may underperform compared to naive benchmarks due to their inability to fully capture the complexities of time series data (Makridakis and Hibon , [21]). Additionally, high autocorrelation can lead to overfitting, where the model becomes too closely tailored to the training data and captures noise rather than underlying trends. This is particularly problematic in long-term forecasting, where the dynamics of the data can evolve. To address these challenges, we decided to employ a machine learning approach for fore- casting wave heights that does not rely on previous wave height observations. This method is designed to capture the underlying patterns of wave behavior without being constrained by the immediate past, allowing for more accurate long-term predictions. By avoiding the pitfalls of overfitting and short-term bias, the machine learning model is expected to pro- vide a more reliable tool for forecasting wave heights in Jamaica Bay, with applications for coastal management and erosion control.

In this study, we have presented a comprehensive numerical modeling approach to simulate wave dynamics in Jamaica Bay, a critical back-barrier coastal system. By combining SWAN and ADCIRC models with a machine learning technique to address autocorrelation issues, we aim to improve the accuracy of long-term wave forecasts. This work provides valuable insights into the bay's hydrodynamics, which will be crucial for effective coastal management and conservation efforts, particularly in the context of ongoing environmental changes such as salt marsh loss.

4. MACHINE LEARNING MODEL

Machine learning has gained widespread use for forecasting complex datasets, particularly when external variables are involved. Unlike traditional statistical methods, which rely on predefined relationships and assumptions, machine learning techniques learn patterns directly from data. This ability to model non-linear relationships and handle high-dimensional data makes machine learning particularly valuable for complex forecasting tasks.

A powerful machine learning technique is boosting trees, an ensemble learning method that improves predictive accuracy by combining multiple weak learners into a stronger model. Boosting trees are particularly effective for both classification and regression tasks, using a sequential learning process. In each iteration, a new weak learner is trained to correct the mistakes of the previous model, ultimately forming a robust ensemble model. These weak learners are typically shallow decision trees, known as decision stumps, when they have one split, or shallow trees when they have a limited number of levels. Their simplicity allows them to extract only the most fundamental data patterns, providing a stable foundation for further refinement.

The boosting process begins by training an initial weak model using equal weights for all data points. In each subsequent iteration, the algorithm adjusts the weights of misclassified instances, training new models that focus on these errors. Finally, the predictions from all weak learners are combined—often using weighted voting or averaging—to form a final prediction. This iterative process allows boosting to gradually refine the model's errors, improving its accuracy with each step.

There are two key variants of boosting trees: AdaBoost (Freund and Schapire ,[22]) and Gradient Boosting, (Friedman [23]). AdaBoost adjusts the weights of misclassified instances to focus learning on the harder examples, while Gradient Boosting optimizes a loss function through gradient descent. Both approaches aim to build an ensemble model that integrates the strengths of individual learners, ensuring high accuracy and generalization.

Recent advancements in boosting algorithms have led to the development of XGBoost (Chen and Guestrin , [24])and CatBoost (Dorogush *et al*., [25]), which enhance the performance and efficiency of boosting models in real-world applications. XGBoost builds upon Gradient Boosting by introducing features such as parallel processing for faster training and handling missing values during training. This capability reduces the need for external data preprocessing and improves robustness. Additionally, XGBoost employs a distributed computing framework to manage computational resources efficiently, making it particularly suitable for large datasets.

CatBoost, on the other hand, is designed to handle categorical features more effectively. Using balanced symmetric trees, it improves both training efficiency and model generalization, preventing overfitting. CatBoost also employs advanced optimization strategies, such as more efficient gradient descent techniques, which contribute to its high performance and ease of use. It excels in datasets with many categorical features, offering strong predictive power and requiring less extensive preprocessing compared to other boosting methods. Both XGBoost and CatBoost have become popular for their ability to handle complex datasets and deliver state-of-the-art performance across a range of tasks, from forecasting to classification.

5. RESULTS

We applied the CatBoost machine learning model to predict wave heights within Jamaica Bay, New York, based on a 40-year dataset derived from numerical simulations using the coupled ADCIRC and SWAN models, as described earlier. The machine learning model was trained using input features such as wind speed, wind direction, and mean sea level pressure from the JFK weather station, with the output being the predicted wave height at various locations within the bay. We validated the model using observational data from three field deployments in 2015: May 21st–July 9th, July 10th–August 29th, and September 1st–October 29th. As illustrated in the figures, wave heights are higher in winter compared to summer.

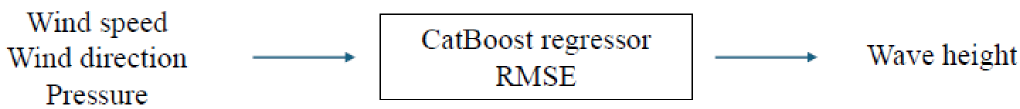

Figure 2: A diagram explaining the methodology used in this study.

Figures 3 and 4 compare the time series of significant wave heights (Hs) at two locations within the bay, Canarsie Pier (CP) and Rockaway Point (RP), using data from the coupled numerical model, the machine learning model, and actual field measurements. The machine learning model closely replicates the wave heights predicted by the numerical model while offering a significant reduction in computational time and cost. Specifically, the machine learning model takes less than 5 minutes to train, whereas the numerical simulation requires approximately 24 hours using 40 cores.

This drastic reduction in computational resources aligns with findings from other studies. For example, Hu and Chu, [12] demonstrated that their machine learning model could complete annual wave height predictions in just 10 minutes on a single CPU, compared to 12 hours using 60 CPUs for traditional numerical simulations. Similarly, Mohaghegh *et al*., [26] reported that their machine learning approach achieved phase-resolved wave predictions over two orders of magnitude faster than conventional numerical methods. These examples emphasize the efficiency and accuracy of machine learning models for wave height forecasting. An analysis of the feature importance in the CatBoost model reveals that wind speed and wind direction are the most influential factors in predicting wave heights. The model assigns high importance to these meteorological variables, reflecting their significant role in wave dynamics.

Mean sea level pressure, while important, has a lower contribution, which suggests that the model relies more on dynamic atmospheric conditions than on static pressure patterns. This interpretability allows for insights into the driving forces behind wave height predictions and confirms the model's ability to capture the complex interactions in the bay's environment.

While the CatBoost model generally performs well, certain conditions lead to higher prediction errors. Notably, extreme wave events, such as those caused by winter storms or high winds, exhibit larger discrepancies between predicted and measured wave heights. At Rockaway Point (RP), the model tends to underestimate wave heights during such events,

which is likely due to the high-frequency variability in wave dynamics that the model struggles to capture. Future improvements could focus on increasing the model's sensitivity to these extreme conditions by incorporating additional environmental variables or adjusting the training data to better represent such events.

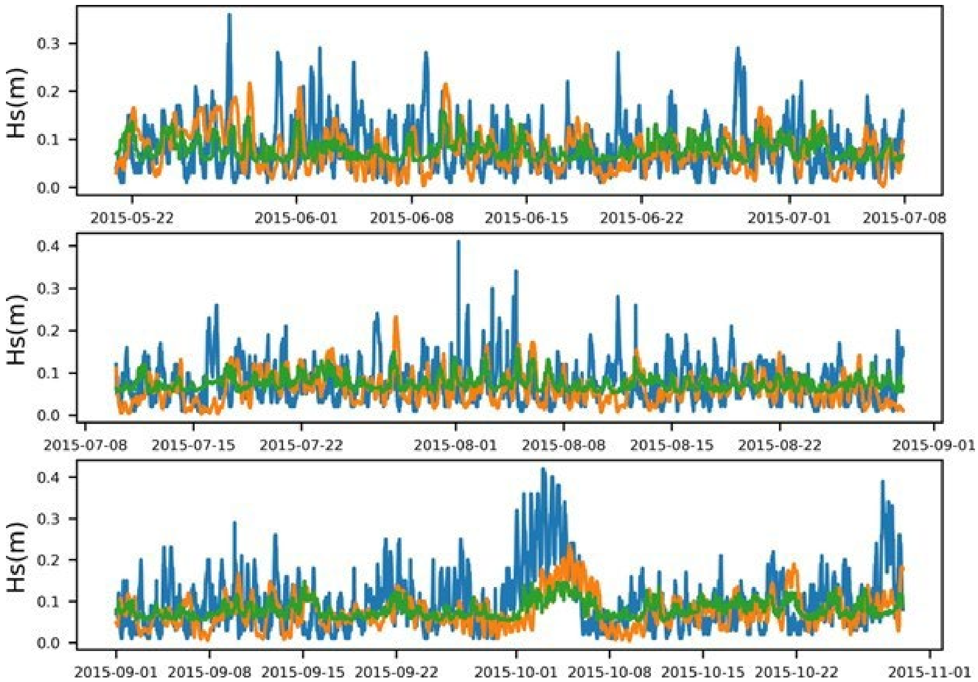

Figure 3: Comparison of wave measurements at Canarsie Pier (CP) point. Orange: numerical model; Green: machine learning CatBoost; Blue: Observations. Top: first deployment; Middel: second deployment; Bottom: third deployment.

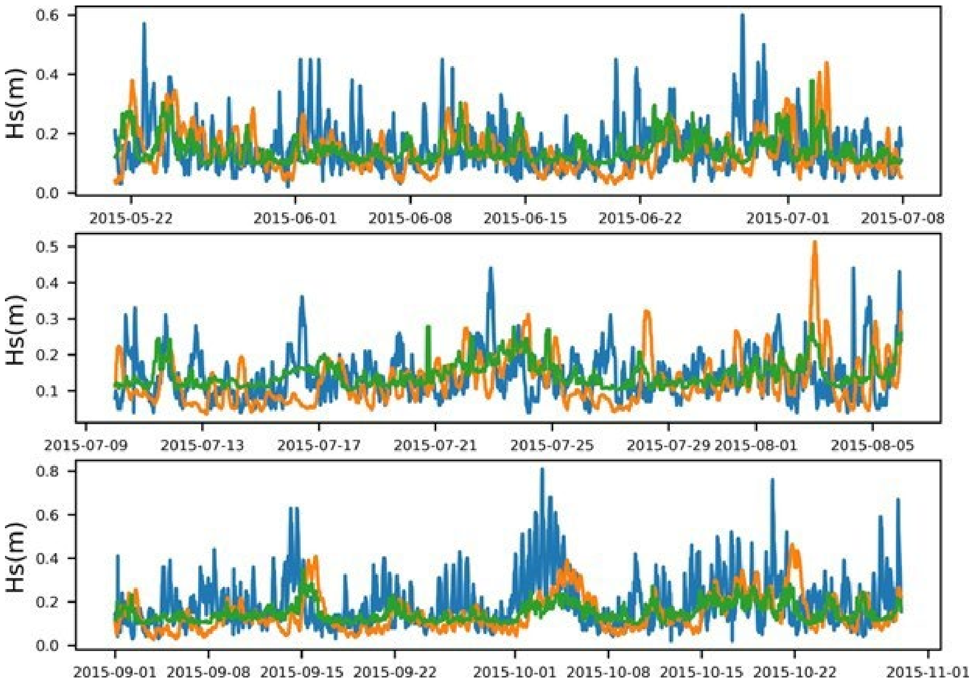

Figure 4: Comparison of wave measurements at Canarsie Pier (RP). Orange: numerical model; Green: machine learning CatBoost; Blue: Observations. Top: first deployment; Middel: second deployment; Bottom: third deployment

Field measurement data exhibit strong peaks and spikes, which make comparison with both numerical and machine learning models challenging. To assess model performance, we use two statistical metrics: mean absolute error (MAE) and root mean square error (RMSE). As shown in Table 1, the CatBoost model results in smaller error metrics compared to the numerical model across the three different deployment periods.

Table 1. Comparison of statistics at points CP and RP

Deployment Date in 2015	*CatBoost (CP)*		*Numerical (CP)*		*CatBoost (RP)*		*Numerical (RP)*	
	MAE	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*
May 21st–July 9th	0.078	0.056	0.076	0.067	0.143	0.091	0.137	0.106
July 10th–August 29th	0.076	0.049	0.063	0.062	0.146	0.066	0.129	0.097
September 1st–October 29th	0.08	0.071	0.076	0.077	0.149	0.12	0.146	0.14
All	0.044	0.06	0.052	0.07	0.069	0.10	0.088	0.12

Improving the alignment between the numerical model's training data and field observations would enhance the machine learning model's ability to replicate real-world data. High-quality training data are crucial for developing reliable machine learning models, as inaccuracies or inconsistencies can impair performance.

After validating the machine learning model, we used weather forecasts from Global Forecast System (GFS) at coordinates 40.5◦N latitude, 73.75◦W longitude to predict wave heights for the period starting January 1, 2025, over the course of two weeks. Figure 5 shows the forecasted wave heights across the bay, displayed at 6-hour intervals over a 5-day period. The machine learning model effectively captures the variations in wave height, demonstrating its capacity for accurate forecasting—a crucial aspect of wave prediction. Similar success was observed in Scott *et al*, [27], who developed a machine learning framework for Monterey Bay, achieving a root-mean-squared error of 9 cm in wave height predictions. Similarly, a study by Wang *et al*., [28] utilized a stacking ensemble learning method to improve wave height prediction accuracy by analyzing the correlation between wave heights and other oceanic features.

Figure 6 presents the hourly wave height time series for CP and RP. However, in-situ wave height data within Jamaica Bay are unavailable for direct comparison. The closest National Data Buoy Center (NDBC) station, buoy 44065, is located approximately 15 nautical miles southeast of Breezy Point, NY at coordinates 40.368 N, 73.701 W—outside the bay. Likewise, the Wave Information Studies (WIS) model data for station ST63126 do not include Jamaica Bay. During the two-week forecast period starting January 1, 2025, predicted wave heights at CP ranged from approximately 0.05 to 0.2 meters, while at RP, near the Atlantic Ocean, they ranged from 0.15 to 0.4 meters. Correspondingly, wind speeds fluctuated between 5 and 20 meters per second, predominantly coming from the north and northwest directions.

It is important to note that the accuracy of the GFS forecasts, used as input for the machine learning model, can influence the model's predictions. While the GFS data generally provides reliable forecasts for the region, there is inherent uncertainty in long-range weather predictions, especially in terms of wind speed and direction. A sensitivity analysis shows that small deviations in the forecasted wind speed can lead to noticeable changes in predicted wave heights. Future work could focus on improving the handling of forecast uncertainty by incorporating ensemble weather forecasts. The significant reduction in computational resources offered by the CatBoost mode, requiring less than 5 minutes for training versus 24 hours for numerical simulation, has important real-world implications. In coastal management scenarios, this speed allows for near real-time predictions, making the model an invaluable tool for decision-makers who need to rapidly assess wave conditions, especially during extreme weather events. Moreover, the model's efficiency allows for frequent updates of forecasts, which is crucial for informing disaster response strategies and protecting vulnerable coastal infrastructure.

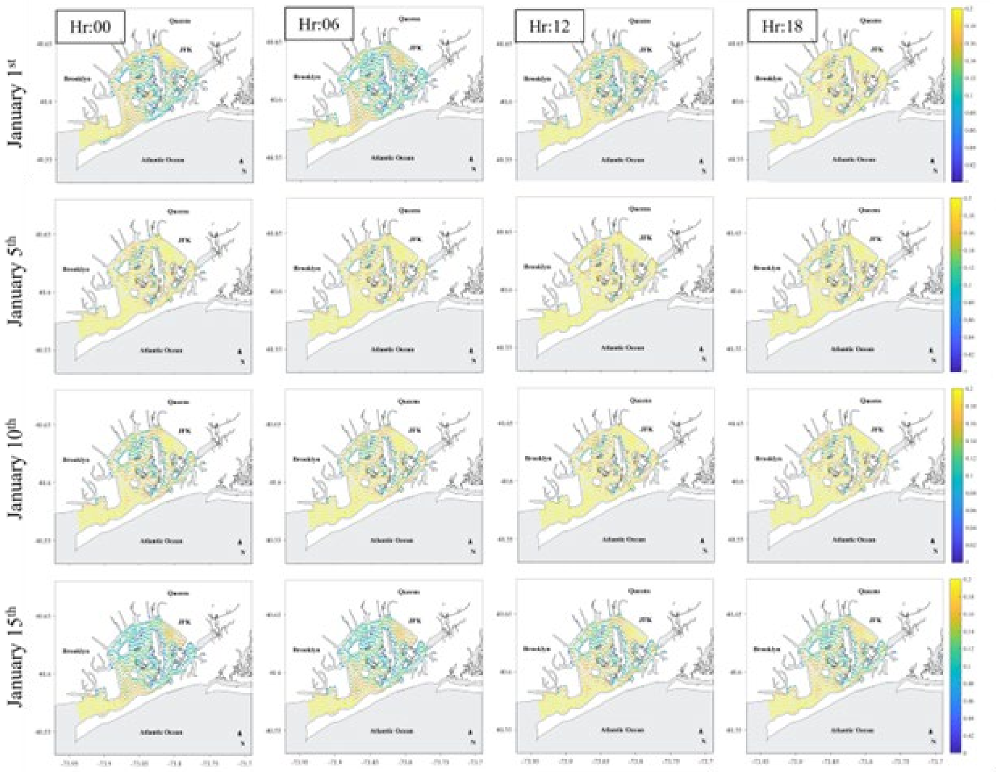

Figure 5: Map of wave height hindcasted using GFS weather for different periods during January 2025 every 6 hours

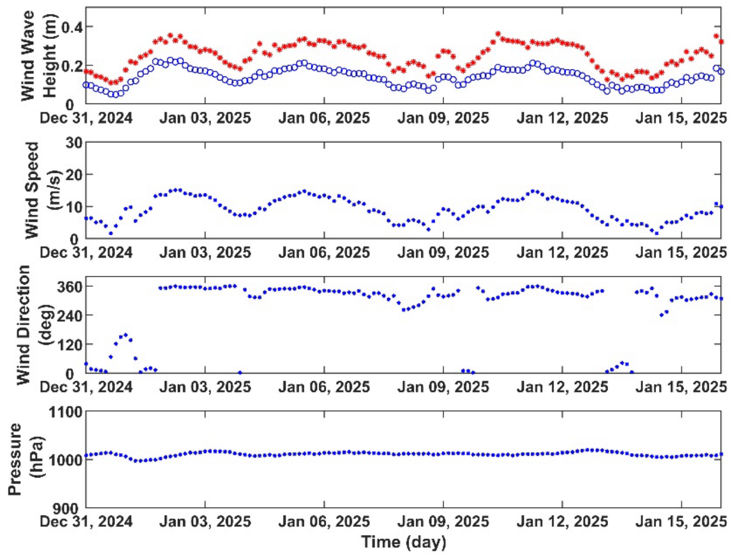

Figure 6: Wind wave height forecasted using ML model at the two points RP and CP and the GFS weather forcings. Top: Blue wave height at CP; Red wave height at RP

6. CONCLUSIONS

(1) This study presents an innovative machine learning framework for forecasting wave heights in Jamaica Bay, New York
(2) The approach leverages meteorological data (wind speed, wind direction, and atmospheric pressure) combined with a high-resolution coupled hydrodynamic-wave model and the CatBoost algorithm
(3) The model accurately predicted significant wave heights at two key locations within the bay.
(4) Achieved RMSE values of 0.07 m and 0.12 m, and MAE values of 0.05 m and 0.08 m, demonstrating strong predictive performance.
(5) The machine learning–based method proves effective for wave height forecasting in data-scarce regions.
(6) Reliable forecasts enhance coastal management, infrastructure planning, and environmental protection by improving understanding of local wave dynamics.
(7) Use enhanced predictive models to support resilient coastal infrastructure and long-term adaptation strategies

7. REFERENCES

[1] Smith, Stuart D. (1988). "Coefficients for Sea Surface Wind Stress, Heat Flux, and Wind Profiles as a Function of Wind Speed and Temperature." Journal of Geophysical Research: Oceans 93, no. C12 (1988): 15,467-15,482. https://doi.org/10.1029/JC093iC12p15467.

[2] Drennan, W. M., K. K. Kahma, and M. A. Donelan. 1999. "On Momentum Flux and Velocity Spectra over Waves." Boundary-Layer Meteorology 92 (1999): 489-515. https://doi.org/10.1023/A:1002054820455.

[3] Reidenbach, M. A., J. R. Koseff, and M. A. Koehl. 2009. "Hydrodynamic Forces on Larvae Affect Their Settlement on Coral Reefs in Turbulent, Wave-driven Flow." Limnology and Oceanography 54, no. 1 (2009): 318-330. https://doi.org/10.4319/lo.2009.54.1.0318

[4] Heij, C., and S. Knapp. 2014. "Effects of Wind Strength and Wave Height on Ship Incident Risk: Regional Trends and Seasonality." Econometric Institute Research Papers (2014). 10.13140/RG.2.1.1946.6407

[5] Reguero, B. G., Iñigo J. Losada, and F. J. Méndez. 2019. "A Recent Increase in Global Wave Power as a Consequence of Oceanic Warming." Nature Communications 10, no. 1 (2019): 1-14. https://doi.org/10.1038/s41467-018-08066-0

[6] Fan, S., N. Xiao, and S. Dong. 2020. "A Novel Model to Predict Significant Wave Height Based on Long Short-Term Memory Network." Ocean Engineering 205 (2020): 107298. https://doi.org/10.1016/j.oceaneng.2020.107298

[7] Drennan, Kaloop, M. R., D. Kumar, F. Zarzoura, B. Roy, and J. W. Hu. 2020. "A Wavelet-Particle Swarm Optimization Extreme Learning Machine Hybrid Modeling for Significant Wave Height Prediction." Ocean Engineering 213 (2020). https://doi.org/10.1016/j.oceaneng.2020.107777

[8] Guan, X. 2020. "Wave Height Prediction Based on CNN-LSTM." In Proceedings of the 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), 10-17, 2020. 10.1109/MLBDBI51377.2020.00009

[9] Pushpam, P. M., and V. S. Enigo. 2020. "Forecasting Significant Wave Height Using RNN-LSTM Models." In Proceedings of the 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Vol. 2020, 1141-1146. 10.1109/ICICCS48265.2020.9121040

[10] Zhou, S., W. Xie, Y. Lu, Y. Wang, Y. Zhou, N. Hui, and C. Dong. 2021. "ConvLSTM-Based Wave Forecasts in the South and East China Seas." Frontiers in Marine Science 8 (2021): 680079. https://doi.org/10.3389/fmars.2021.680079

[11] Feng, Z., P. Hu, S. Li, and D. Mo. 2022. "Prediction of Significant Wave Height in Offshore China Based on the Machine Learning Method." Journal of Marine Science and Engineering 6 (2022): 10. https://doi.org/10.3390/jmse10060836

[12] Hu, H., and P. Chu. 2021. "Predicting Lake Erie Wave Heights and Periods Using XGBoost and LSTM." Ocean Modelling 164 (2021). https://doi.org/10.1016/j.ocemod.2021.101832

[13] Marsooli, R., P. M. Orton, J. Fitzpatrick, and H. Smith. 2018. "Residence Time of a Highly Urbanized Estuary." Journal of Marine Science and Engineering 6 (2018). https://doi.org/10.3390/jmse6020044

[14] Marsooli, R., P. M. Orton, and G. Mellor. 2017. "Modeling Wave Attenuation by Salt Marshes in Jamaica Bay Using a New Rapid Wave Model." Journal of Geophysical Research: Oceans 122 (2017): 5,689-5,707. https://doi.org/10.1002/2016JC012546.

[15] Dietrich, J. C., M. Zijlema, J. J. Westerink, L. H. Holthuijsen, C. Dawson, R. A. Luettich, W.-G. Stone, and others. 2011. "Modeling Hurricane Waves and Storm Surge Using Integrally-Coupled, Scalable Computations." Coastal Engineering 58, no. 1 (2011): 45-65. https://doi.org/10.1016/j.coastaleng.2010.08.001

[16] Etemad-Shahidi, A., and J. Mahjoobi. 2009. "Comparison between M5j Model Tree and Neural Networks for Prediction of Significant Wave Height in Lake Superior." Ocean Engineering 36 (2009): 1175-1181. https://doi.org/10.1016/j.oceaneng.2009.08.008

[17] Marsooli, R., and N. Lin. 2020. "Impacts of Climate Change on Hurricane Flood Hazards in Jamaica Bay." Climatic Change 163, no. 4 (2020): 2153-2171. 10.1007/s10584-020-02932-x

[18] Safty, H. E., & Marsooli, R. (2020). "Ship Wakes and Their Potential Impacts on Salt Marshes in Jamaica Bay, New York. Journal of Marine Science and Engineering 2020. Vol. 8, Page 325, 8(5), 325 (2020). doi:10.3390/JMSE8050325

[19] Chant, R. J., D. K. Ralston, N. K. Ganju, C. Pianca, A. E. Simonson, and R. A. Cartwright. 2021. "Sediment Budget Estimates for a Highly Impacted Embayment with Extensive Wetland Loss." Estuaries and Coasts 44, no. 3 (2021): 608-626. 10.1007/s12237-020-00784-3

[20] Petropoulos, F., D. Apiletti, V. Assimakopoulos, M. Z. Babai, D. K. Barrow, S. B. Taieb, and Ziel. [no initials given]. "Forecasting: Theory and Practice." International Journal of Forecasting 38, no. 3 (Year?). [705-871]. https://doi.org/10.1016/j.ijforecast.2021.11.001

[21] Makridakis, S., and M. Hibon. 2000. "The M3-Competition: Results, Conclusions and Implications." International Journal of Forecasting 16, no. 4 (2000): 451-476. https://doi.org/10.1016/S0169-2070(00)00057-1

[22] Freund, Y., and R. E. Schapire. 1997. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting." Journal of Computer and System Sciences 55, no. 1 (1997): 119-139. https://doi.org/10.1006/jcss.1997.1504

[23] Friedman, J. H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." Annals of Statistics 29 (2001): 1189-1232. 10.1214/aos/1013203451

[24] Chen, T., and C. Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2016): 785-794. https://doi.org/10.1145/2939672.2939785.

[25] Dorogush, A. V., V. Ershov, and A. G. Yandex. 2018. "CatBoost: Gradient Boosting with Categorical Features Support." [Preprint?] (2018). https://doi.org/10.48550/arXiv.1810.11363.

[26] Mohaghegh, F., M. R. Alam, and J. Murthy. 2021. "Rapid Phase-Resolved Prediction of Nonlinear Dispersive Waves Using Machine Learning." Applied Ocean Research 117 (2021). https://doi.org/10.1016/j.apor.2021.102634

[27] Scott, C., Y. James, Zhang, O. Fearghal, "A Machine Learning Framework to Forecast Wave Conditions." Journal of Coastal Engineering (2017). https://doi.org/10.1016/j.coastaleng.2018.03.004

[28] Wang, W., R. Tang, C. Li, P. Liu, and L. Luo. 2018. "A BP Neural Network Model Optimized by Mind Evolutionary Algorithm for Predicting the Ocean Wave Heights." Ocean Engineering 162 (2018): 98-107. https://doi.org/10.1016/j.oceaneng.2018.02.025