IMPERIAL

MOODY'S

Imperial College London

Department of Earth Science and Engineering

MSc in Environmental Data Science and Machine Learning

Independent Research Project

Final Report

Estimating landfalling Hurricane Wave Characteristics with Surrogate Modelling

by

Tianju (Tim) Du

Email: td323@ic.ac.uk

GitHub username: edsml-td323

Repository: https://github.com/ese-msc-2023/irp-td323.git

Supervisors:

Dr. Lluis Via Estrem

Prof. Matthew Piggott

30 Aug 2024

Contents

Table of Acronyms

Table 1: Table of Acronyms. This table provides a list of acronyms used in this report along with their full forms to aid in understanding the technical content.

Acronym	**Full Form**
TC	Tropical Cyclone
Hs	Significant Wave Height
Wspd	Wind Speed
Wdir	Wind Direction
SSE	Sea Surface Elevation
TWD	Total Water Depth
Tp	Peak Wave Period
FSE	Free Surface Elevation
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
PCA	Principal Component Analysis
TCN	Temporal Convolutional Network
MAE	Mean Absolute Error
MSE	Mean Squared Error
R^2	Coefficient of Determination
CoE	Coefficient of Efficiency

Abstract

Tropical cyclones (TCs) present significant risks to coastal regions, particularly in areas like Florida, where storm surges can lead to substantial economic losses. Traditional hydrodynamic models, such as MIKE 21, offer high accuracy in predicting significant wave heights (Hs) but at the cost of extensive computational resources. This report explores deep learning (DL) models as an alternative to these traditional methods, aiming to balance computational efficiency with predictive accuracy. A comparison between baseline LSTM, CNN-LSTM, and PCA-TCN-LSTM models revealed that while CNN-LSTM effectively captured spatial and temporal features, it struggled with peak value predictions. The PCA-TCN-LSTM model demonstrated superior performance in predicting extreme wave heights by leveraging dimensionality reduction through PCA and incorporating Temporal Convolutional Networks (TCNs) to capture spatial patterns. This research highlights the potential of DL models in disaster forecasting and risk management, offering a promising path toward faster and more accurate predictions of storm impacts.

1 Introduction

Recent tropical cyclones (TC), such as Irma in 2017 and Ian in 2022, have underscored the devastating financial impact of storm surges on Florida's coastline. These storms caused severe flooding in major urban areas such as Miami Beach, Key Biscayne, and South Dade, highlighting the region's vulnerability to rising waters [16]. The economic losses from these events are substantial, with Irma alone necessitating over $5.58 billion in recovery funds across Florida [6]. These events emphasized the need for accurate and efficient predictive models to aid in disaster management and mitigation.

To anticipate such disasters, hydrodynamic models are widely used in the insurance sector to predict the effects of tropical cyclones (TCs) on economic loss. Moody's RMS has chosen the MIKE 21 HD model to simulate hydrodynamics and the MIKE 21 SW model for spectral wave simulations. In this approach, the two models are run in a one-way coupled configuration: MIKE 21 HD first calculates sea surface elevation (SSE) and water currents using inputs such as tides and wind stress. These outputs are then used by MIKE 21 SW, along with additional wind stress data, to simulate wave characteristics such as significant wave height (Hs) and peak period (Tp). To enhance the prediction of storm impacts, synthetic track data are generated to address the lack of real storm data [5]. However, the computation of diffusion terms, which involve spatial derivatives, results in high computational costs and long simulation times (e.g., hydro simulations taking approximately 20 hours per storm and spectral wave simulations around 110 hours per storm). To mitigate this, research has focused on methods to reduce computational costs while maintaining accuracy, including hybrid efficient methods and the Probabilistic Coastal Hazard Analysis Framework [14, 15], which integrate machine learning into the modeling process. Consequently, This report is exploring an alternative machine learning model as a more efficient way to compute Hs caused by TCs near Florida.

2 Problem Description and Objectives

Numerical models deliver accurate predictions but require significant computational resources. In contrast, deep learning (DL) offers faster predictions, albeit with reduced accuracy. However, recent developments in DL have shown promising results. For instance, Google's MetNet demonstrated competitive performance compared to traditional numerical weather prediction models, excelling in both prediction length and accuracy [1, 12]. This advancement highlighted the growing potential of DL to bridge the gap between computational efficiency and predictive accuracy.

This research explored and developed DL approaches that could replicate and potentially enhance the functionality of the MIKE 21 software in predicting Hs from TC tracks. The project focused on leveraging the speed and efficiency of DL models while maintaining the accuracy of traditional methods. The goal was to create a DL model that competes with MIKE 21's Hs prediction, offering faster predictions with comparable accuracy, and to build a user-friendly surrogate model-based software.

The deep learning (DL) models not only replicate the functionality of MIKE 21 but also introduce new capabilities. When the DL model simply mimics MIKE 21, it functions as a regression model, learning the relationship between inputs and outputs directly. This approach is particularly useful for the re/insurance sector in conducting worst-case scenario analyses to predict peak Hs such as CNN + LSTM model and PCA model in the report. However, DL goes beyond this by reordering the inputs and outputs to enable forecasting capabilities, as discussed in the baseline LSTM model below. This forecasting application is crucial for disaster response, allowing for timely warnings and long-term predictions as soon as a storm forms, thereby protecting vulnerable populations.

The software developed through this research aimed to replicate the functionality of MIKE 21 for the Florida coastline region, with a focus on ease of use, flexibility, and robust performance. The software integrated MIKE 21's input parameters with the DL models, ensuring a seamless transition for users familiar with traditional methods. Additionally, the software included a testing unit ensuring reliable performance across a wide range of scenarios. This combination of features made the software a practical and powerful tool for coastal risk management and disaster response.

3 Literature Review

The application of DL models in predicting natural phenomena, particularly in the context of TCs, has been explored extensively. The LSTM model demonstrated strong predictive capabilities, particularly for short-term forecasts up to 12 hours. By using surface wind speed and Hs data as inputs, the LSTM model generates predicted Hs at various intervals, offering valuable real-time insights for disaster response and coastal risk management. However, it has limitations in predicting fine details and extreme wave height values [2].

Combining convolutional neural networks (CNN) with dimensionality reduction techniques such as principal component analysis (PCA), the C1PK-Net model enhances the predictive accuracy of peak storm surge. It processes time-series data of TC parameters and outputs the predicted peak storm surge at multiple geographic locations. Unlike continuous forecasting models, C1PK-Net focuses on predicting peak values at a single moment, making it suitable for high-precision applications [10].

Another notable model is the WaveNet-based CNN (WCNN), which predicts river water levels using data from upstream gauging stations. This model, while not directly related to Hs prediction from TC tracks, illustrates the potential of combining different neural network architectures to improve accuracy, outperforming traditional methods such as LSTM and Gated recurrent units (GRU) [3]. These DL models, each with their strengths and limitations, provide a foundation for the development of more efficient and accurate predictive tools in disaster management.

4 Exploratory Data Analysis

4.1 MIKE21 Input and Output Data

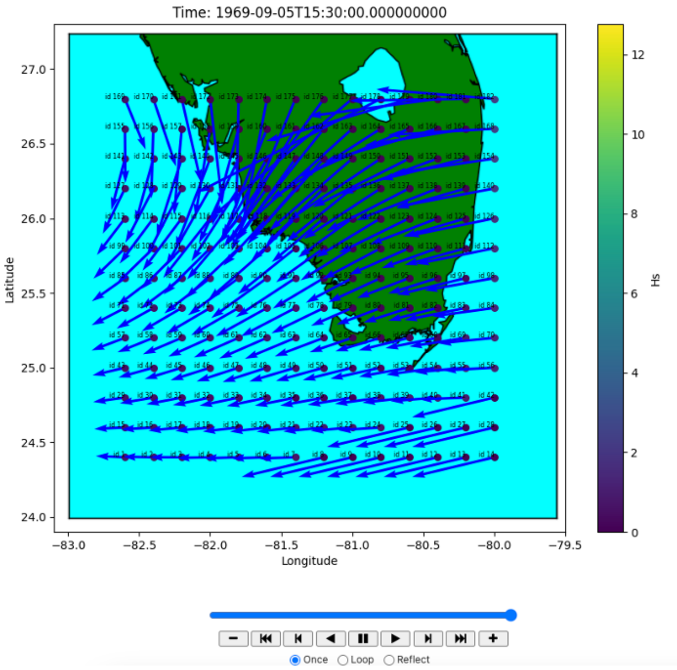
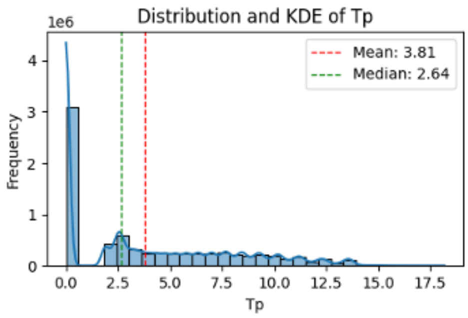


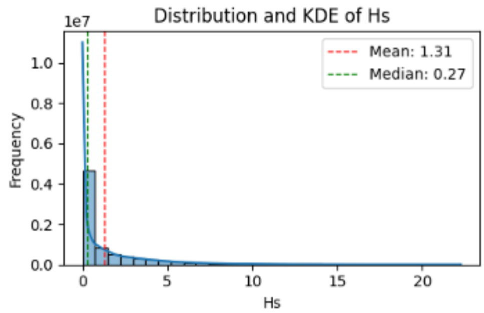
Figure 1: The grid visualization illustrates the arrangement of computational points in the MIKE 21 model's input and output data (refer to the GIF on GitHub). The green areas represent land, while the arrows indicate wind direction, with their length corresponding to wind speed. This GIFs is crucial for understanding the spatial and temporal resolution employed in wave predictions, providing insight into how the to build a model that captures and simulates the effects of wind and other factors on wave dynamics.

MIKE 21 Input and Output Data data consist of Hs, Maximum Wave Height (Hmax), Wspd, Wdir, TWD, Peak Wave Period (Tp), and Free Surface Elevation (FSE) in a 30-minute interval, 97-timestep length time series. Each location (grid point) has its set of the data. The data is right-skewed, particularly in wave heights (Figure 2) and Wspd because many of the given locations are inland (Figure 1), which are not vulnerable to flooding, and most of the track data record the storm from formation to dissipation, causing $0m$ to be the most common value for wave height.

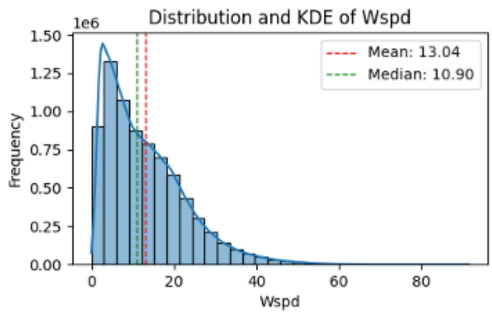
The correlation matrix (Figure 3) demonstrates the relationships between various parameters, with Hs and Hmax showing a strong positive correlation. This indicates that these wave metrics are closely related and can be predictive of each other. Tp also shows a strong correlation with both Hs and Hmax. Free Surface Elevation (FSE) is negatively correlated with wave characteristics including significant wave height (Hs) and peak period (Tp) because higher FSE levels often coincide with conditions that increase energy dissipation in the water [7]. As water levels rise, particularly during events like storm surges, the friction between the water and seabed increases, which can suppress wave growth by dispersing wave energy more rapidly.



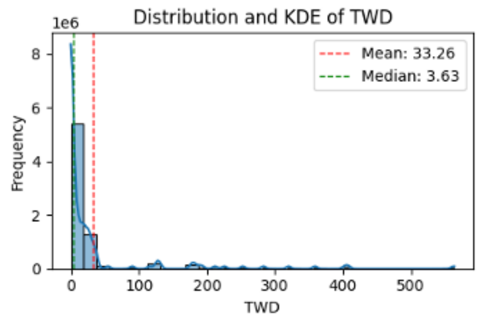
(a) Tp



(b) Hs



(c) Wspd



(d) TWD

Figure 2: Distribution analysis of Wind Field data. These histograms illustrate the distribution of peak wave period (Tp), significant wave height (Hs), wind speed (Wspd), and total water depth (TWD) across the grid points. Notably, the data is right-skewed, with many points having low wave heights, reflecting the inland nature of several grid points.

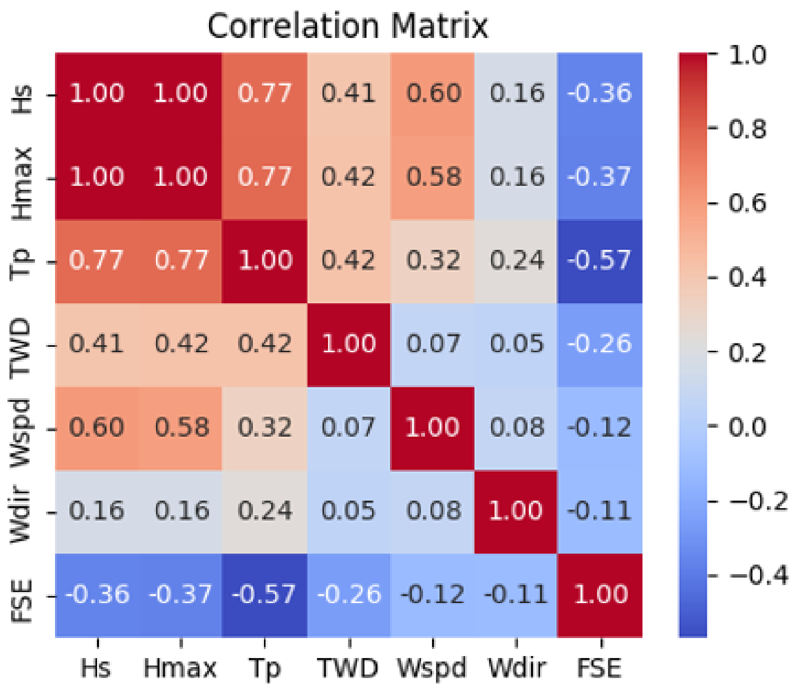


Figure 3: Correlation Matrix of Wind Field data. The matrix demonstrates the relationships between various parameters, with Hs (significant wave height) and Hmax (maximum wave height) showing the highest correlation, indicating that these wave metrics are closely related.

4.2 Bathymetry and Grid Points of Florida South M08FLS

Figure 4 illustrates the 3D bathymetry surface and the grid points used in the MIKE 21. The areas with higher bathymetry values (indicating greater depth) are represented by the peaks and are colored in green to yellow. The higher bathymetry values are concentrated around the region near longitude $-80°$, particularly between latitudes $24.5°$ and $25.5°$. High bathymetry values, indicating deeper waters, are often associated with larger Hs because waves in deeper water experience less frictional energy loss with the seabed, allowing them to maintain or grow in height.

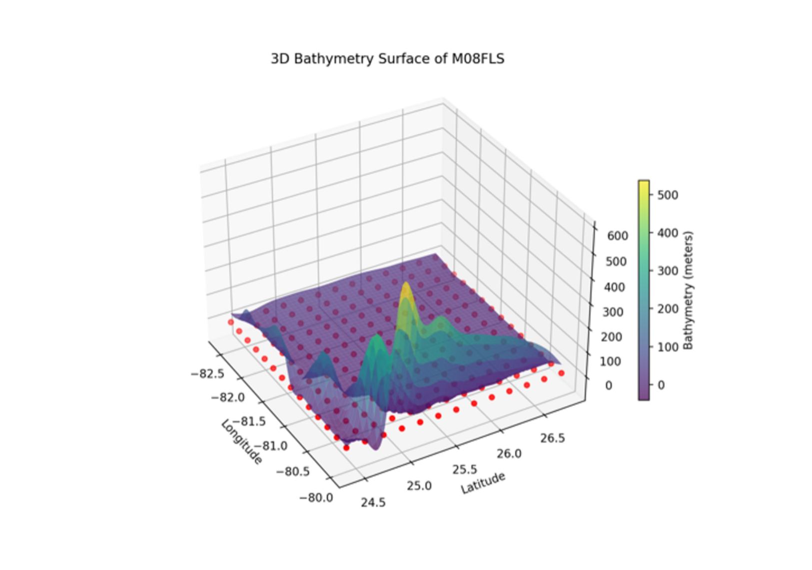


Figure 4: Bathymetry and grid points. The 3D bathymetry map illustrates the variation in sea floor depth across the study area, with deeper regions near Florida's coast highlighted. The grid visualization on the right shows the arrangement of computational points used in the MIKE 21 simulations. The distribution of grid points helps in understanding the spatial resolution used for wave predictions.

5 Baseline Model Implementation and Limitations

5.1 "1D input to Hs" LSTM model

Objective of the models

The primary objective of the simple model was to forecast Hs for one specific grid point near Florida's coastline using Wspd and Wdir as inputs. A single grid point (pt_id = 53), located near the shore at the southern tip of Florida is selected, because this point encompasses coastal areas, including residential regions such as Key West, which are particularly susceptible to the impacts of storm surges. This area frequently experiences flooding events due to its geographical location and proximity to the ocean, making it a critical point for analysis.

Data Preparation

The raw data were provided in CSV format, sourced from grids labeled as "Florida South M08FLS." The dataset included columns for Hs, Hmax, Tp, Wspd, Wdir, FSE, TWD, time, pt_id, and track_id. Each CSV file corresponded to a unique storm (identified by track_id) and a specific grid point (identified by pt_id). After initial filtering, only the Hs, Wspd, and Wdir columns were retained before the data was processed by the dataloader, as FSE - TWD was almost constant for a fixed grid, and Tp was not the focus of this study.

Loading Data

A CustomDataset class in PyTorch was utilized to load and preprocess the time-series data for the LSTM model. This class normalized Wspd and Hs, converted Wdir into its sine and cosine components as is common in research [11], and prepared sequences of inputs with corresponding targets (as illustrated in Figure 5). The CustomDataset considered time gaps and sequential inputs to structure the data for LSTM training. Specifically, the model was fed sequences of inputs (e.g., Wspd and Wdir at time steps T=0, T=2, and T=4), capturing temporal dependencies. The target output for the model was the value at the next time step (e.g., Hs at T=6), enabling the LSTM to predict future outcomes based on historical data. This method allowed the model to learn patterns over time, using spaced time steps to predict subsequent events in the sequence. It is important to note that true values were always used for forecasting.

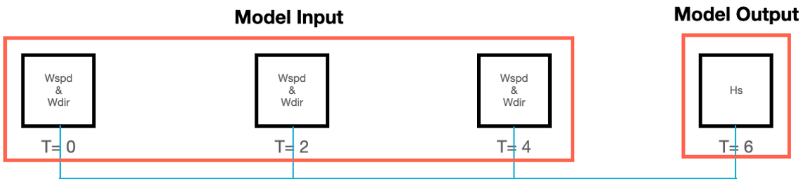


Figure 5: LSTM Data Inputs and Target. The figure shows how input sequences (Wspd, Wdir) are structured for the LSTM model training, with the goal of predicting the next time step's Hs. The structure captures the temporal dependencies necessary for accurate wave height forecasting.

DL Model Design

The baseline LSTM (Long Short-Term Memory) model consisted of 3 layers and a hidden size of 256, followed by a fully connected layer and a ReLU activation function because the Hs is always a non-negative number.

Results and Limitations

Forecasting Time (Hidden Size of LSTM)	MSE (m^2)	CoE	R^2
2 hours (256)	0.0147	0.971	0.89
4 hours (256)	0.0233	0.947	0.81
6 hours (256)	0.0372	0.936	0.77
8 hours (512)	0.0451	0.899	0.69

Table 2: Performance Metrics of LSTM Model on Test Dataset. The table summarizes the Mean Squared Error (MSE), Coefficient of Efficiency (CoE), and Coefficient of Determination (R^2) values for different forecasting times.

While the model performed reasonably well at predicting for short-term forecasts, such as 2 hours ahead, with a Mean Squared Error (MSE) of 0.0147 m^2 and a Coefficient of Determination (R^2) of 0.89, its performance declined for longer forecasting intervals. The Coefficient of Efficiency (CoE) and R^2 are both indicators of the model's predictive power. CoE measures how well the model's predictions capture the variance in the observed data, while R^2 indicates the proportion of variance in the dependent variable that is predictable from the independent variables.

The observed inconsistency in the hidden size (512) of the LSTM during the 8-hour forecast highlights a limitation of LSTM models when applied to long-duration predictions. LSTM networks are designed to learn long-range dependencies in data, which are crucial for accurate long-term forecasting. However, when the hidden size is insufficient, the model struggles to capture these dependencies, resulting in increased error and lower efficiency metrics for instance CoE and R^2. This is a well-documented issue in the literature [4], where larger hidden sizes are often recommended for handling longer prediction horizons, as they provide the model with the capacity to better capture complex temporal patterns [13].

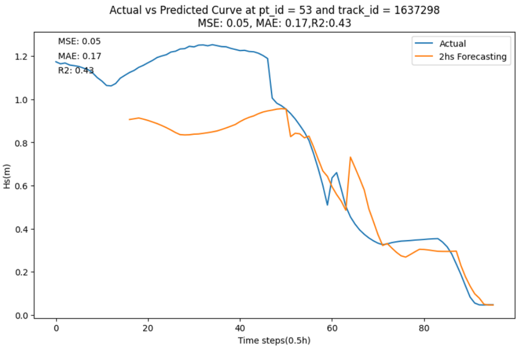
The results clearly indicate the model's declining accuracy with longer forecasting intervals, underscoring the importance of appropriately sizing the hidden layers in LSTM models for long-term predictions. Despite the inherent strengths of LSTM networks in forecasting, there are significant limitations of the baseline model:

- **Inability to Generalize Across Different Bathymetry and Storm Characteristics:** The baseline model, trained on data from a single grid point (pt_id = 53), exhibited significant limitations in its ability to predict Hs at other grid points. Bathymetry, which critically influences wave heights, presented a challenge for the simple LSTM model, which struggled to learn the complex interactions between bathymetry and storm characteristics. For example, track id 1637298 was selected because it triggered the highest Hs scenario at grid point 53. However, while the model performed satisfactorily at this grid point and at nearby grid points with similar bathymetry (Figure 8c pt_id = 54), it produced larger errors at more distant points (e.g., pt_id = 113) that have different underwater topographies.

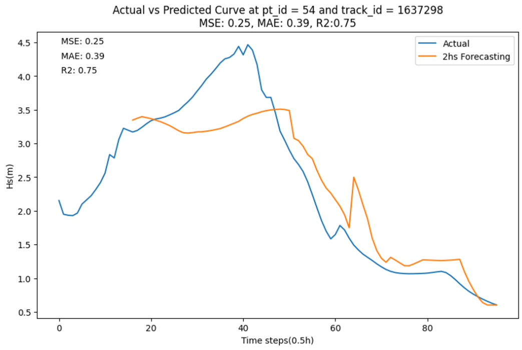
- **Lack of Spatial Feature Learning:** The LSTM model was designed to learn temporal dependencies but lacked the capability to capture spatial features that are critical for accurate Hs prediction. Hs at a given point is influenced by conditions at neighboring grid points. However, the simple LSTM model did not incorporate information from these neighboring points, leading to inaccuracies in predictions, particularly for regions where spatial variation is pronounced. This limitation is evident (Figure 8c) when the model was applied to grid points other than the one it was trained on, where performance declined sharply.

- **Failure to Address the Skewed Distribution of Hs:** Another major limitation of the baseline model was its inability to handle the skewed distribution of Hs values effectively. The distribution of Hs is right-skewed, with many 0 values and fewer, but more significant, higher values. The simple LSTM model tended to underpredict the higher values of Hs (Figure 8), as it did not explicitly account for this skewness in the data. This underestimation was problematic in forecasting extreme wave heights, which are critical for insurance applications.

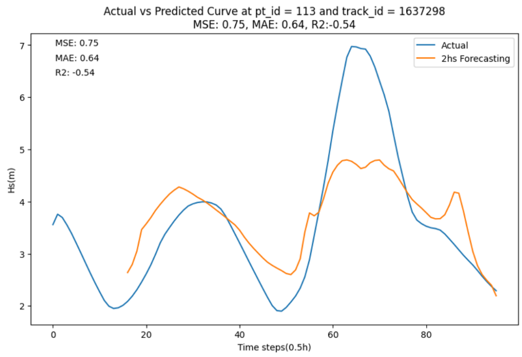
- **Decreased Prediction Performance with Longer Forecasting Time:** The performance of the baseline LSTM model deteriorated as the forecasting time increased (Table 2). Longer forecasting intervals resulted in poorer predictions, as the model was forced to generalize over larger time gaps. This issue arose because the model, when predicting over longer time intervals (e.g., across 4 * 16 time-steps for 8 hours prediction), tended to learn only the generalized features of storm development rather than the detailed characteristics of the storm at each time step. This loss of detail during training is consistent with findings from other studies [13], which also noted that large prediction gaps lead to decreased model performance.



(a) pt_id = 53



(b) pt_id = 54



(c) pt_id = 113

Figure 6: True vs Predicted Hs values at different grid points. The comparison between grid points shows that while the LSTM model can capture the general trend of Hs, it struggles with accuracy across different bathymetric conditions, highlighting its limitations in generalization.

6 Advanced Models

6.1 "2D input to Hs" model: CNN+LSTM

Model Objectives

The "2D input to Hs" model was designed to address the spatial deficiencies observed in the baseline "1D input to Hs" LSTM model. Specifically, it aimed to improve the model's ability to predict Hs by incorporating spatial features that the previous model could not capture. This approach involved converting each grid point into a pixel, and encoding the variables Hs, Wspd, and Wdir as separate channels within the image.

Data Preparation

The 2D input to Hs model involved converting the numerical data into an image format. Each storm within the M08FLS grid was represented as a sequence of 97 images, with each image corresponding to one of the time-steps in the storm's progression. The transformation resulted in a total of 430 sets of storms and 41710 images. After creating the images, a custom dataset was used to apply data normalization techniques similar to those used in the baseline model.

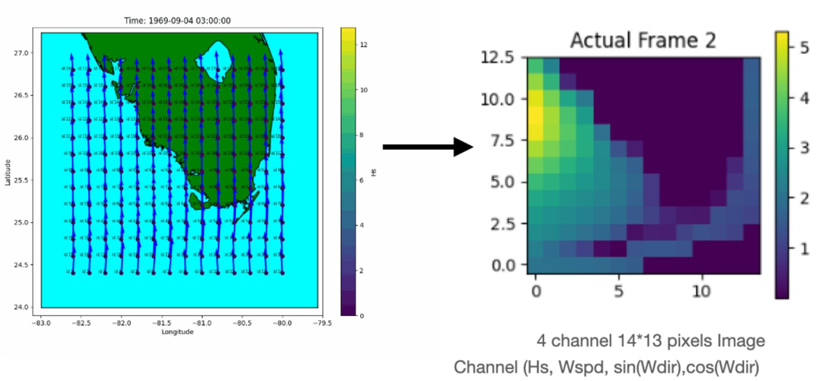


Figure 7: Convert 1D input to 2D image input. The figure illustrates the process of transforming numerical data into images, where each pixel represents a grid point, and each channel corresponds to a different variable, namely, Hs, Wspd, sin(Wdir), cos(Wdir). This approach allows the CNN to capture spatial relationships in the data.

DL Model Design

The 2D input to Hs model employed a hybrid architecture combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to enhance the accuracy of Hs predictions by capturing both spatial and temporal features.

- **Convolutional Layers:** These layers extracted spatial features from the input images, using reflection padding, batch normalization, and ReLU activation to ensure stable learning. The model progressively increased the number of filters (32, 64, 128) across three convolutional layers, each followed by max-pooling, to produce a compact representation of spatial dependencies.
- **LSTM Layers:** The 256-dimensional feature vector was fed into LSTM layers with a hidden size of 512, capturing temporal dependencies across the image sequences.

6.2 "Latent space" model: PCA Temporal Convolutional Networks LSTM (PCA-TCN-LSTM) Model

Model Objective

The PCA-TCN-LSTM model was designed to improve the accuracy of peak Hs predictions, addressing key limitations observed in previous models. Inspired by the C1PK-Net approach [10], which demonstrated that PCA can enhance peak wave predictions, this model applied PCA to compress image sequence data into a lower-dimensional latent space. By learning features within this latent space, the model focused on the most significant patterns, thereby improving its ability to predict extreme values more effectively.

Data Preparation

PCA reduced both the input sequence (97, 13*14*3) and the target sequence (97, 13*14*1) to a uniform shape of (97, 24), where 24 components captured 95% of the variance of the input. This dimensionality reduction ensured that the input and target sequences were aligned in the same latent space, enabling efficient backpropagation of the loss function during training. This compression retained the most critical information while reducing redundancy, especially the replicated features where inland pixel values are frequently equal to 0. This process helped mitigate the right-skewed distribution of Hs, allowing the model to focus on the most relevant features during training.

In this model, Wdir was kept in its original form instead of being split into sine and cosine components before applying PCA. By preserving the original form, the model leveraged PCA's capability to extract the most relevant features from the raw data, leading to improved overall performance. This choice simplified the input data structure and allowed PCA to more effectively capture the inherent relationships among the variables.

DL Model Design

In the PCA-TCN-LSTM model, Temporal Convolutional Networks (TCNs) play a crucial role in processing the latent space features derived from the 2D input data. This component is responsible for capturing the spatial nature of storm-related features within the compressed data. TCNs are particularly well-suited for this task due to their use of dilated convolutions, a technique that allows the model to capture information over a broader context without increasing the computational load.

Dilated convolutions operate by applying a standard convolutional filter with gaps, or dilations, between the filter elements. For instance, with a dilation rate of 2, the filter interacts with every second data point, effectively "skipping" one in between. This design enables the filter to cover a larger area of the input data, allowing it to capture more global patterns while maintaining efficiency. This capability is essential for managing the spatial complexity of storm-related features in the compressed data.

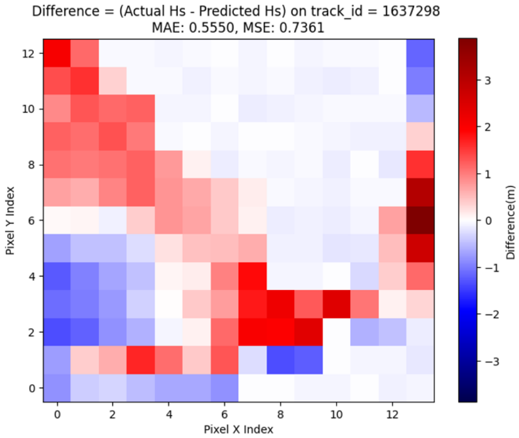
In this model, TCNs are applied after PCA compresses the input data, focusing on identifying significant spatial patterns within the latent space. These extracted features are then passed to an LSTM, which is responsible for capturing temporal dependencies within the time series data. The combination of TCNs for spatial pattern recognition and LSTMs for temporal dependency capture ensures that the model can effectively handle both spatial and temporal complexities, ultimately enhancing the accuracy and robustness of the predictions. This integrated approach allows the model to perform well in forecasting tasks that involve both complex spatial structures and long-term temporal dependencies [9].

7 Results and Discussion

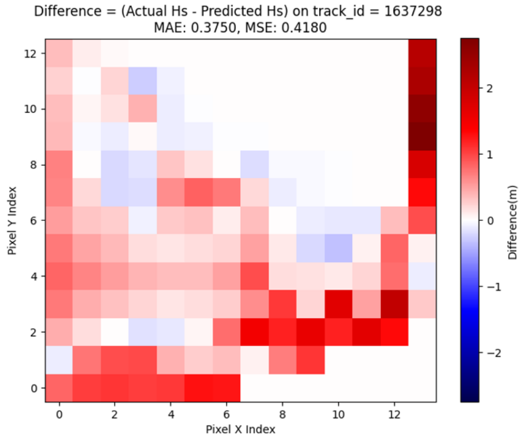
Model Results

The CNN-LSTM model demonstrated superiority over the baseline LSTM by effectively capturing both spatial and temporal features, which are crucial for accurate prediction of Hs. The convolutional layers of the CNN-LSTM model successfully extracted vital spatial relationships between grid points, particularly in regions with complex bathymetry, such as Florida's coastline. This spatial modeling, combined with the LSTM's ability to capture temporal dynamics, enabled the model to achieve robust performance metrics, including an R^2 of 0.937, an MSE of 0.226 m^2, and a percentage of large errors (actual $\geq$ 2 m and error $>$ 1 m) of 2.92%. While the MSE was higher than that of the baseline LSTM, it was computed on a pixel-by-pixel basis, offering a comprehensive assessment across all grid points and storm tracks. Cropping the outer pixels, which are prone to edge effects in CNNs, further reduced the MSE to 0.162, enhancing the overall accuracy.

Despite its strong performance, the CNN-LSTM model exhibited notable limitations in predicting peak Hs values, a critical factor for reinsurance applications, such as those by Moody's RMS. Although



(a) Comparison of Actual and Predicted Max Hs Using the CNN-LSTM Model



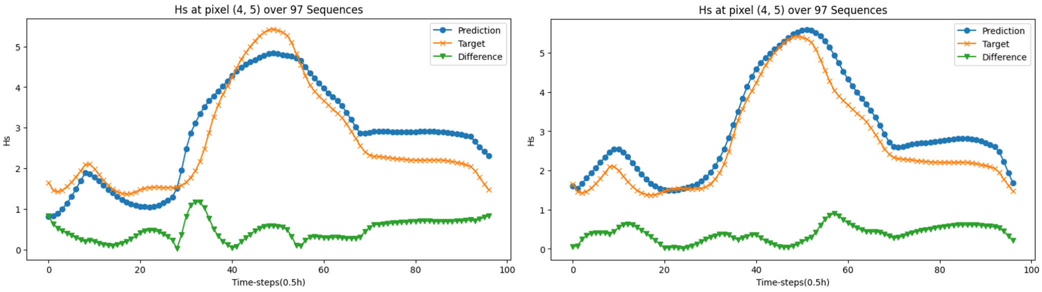
(b) Comparison of Actual and Predicted Max Hs Using the PCA Model

Figure 8: The figures compare the predictions from the physical model and the latent space model for peak Hs values. The latent space model demonstrated better alignment with observed data, particularly in capturing extreme values, which are critical for accurate coastal risk management.

the model captured general trends effectively, it underperformed in predicting peak Hs values, as demonstrated by a mean absolute error (MAE) of 0.555 and an MSE of 0.736 for track id 1637298 which is a strong storm that triggers high Hs values of grids (Figure 8a). The underprediction of extreme values, highlighted in red in the corresponding figure, points to the limitations of CNNs when applied to 2D image data, where convolutional layers may inadvertently smooth out important input features, such as wind speed and direction. This smoothing effect compromises the model's ability to accurately predict sharp peaks in wave height, underlining the need for more specialized approaches for extreme value prediction.

The latent space model (PCA-TCN-LSTM) demonstrated clear advantages over the CNN-LSTM model, particularly in its ability to predict peak values. This model achieved a higher R^2 score of 0.966 and a lower MSE of 0.120 m^2, indicating a more accurate fit across the dataset. Moreover, the percentage of large errors (actual ≥ 2 m and error > 1 m) was reduced to 1.51 %, underscoring the model's improved capability in predicting extreme values, which is essential for accurate disaster prediction and insurance assessments. However, both models struggled at the edge pixels in the right column, correlating with regions of high bathymetry near longitude $-80°$, as shown in Figure 4. This suggests that the models' performance is challenged by outlier locations, which present extreme conditions.

The latent space model's superior performance is attributed to its effective use of PCA, which compressed both the input and target sequences, thereby focusing the model on the most critical features for accurate Hs prediction. This compression was particularly advantageous for the target sequence, where many middle pixels contained low and repetitive Hs values that were less informative. By concentrating on components that captured significant variance, PCA reduced redundancy and noise, enabling the model to generalize better across diverse storm scenarios and improve its prediction of extreme wave heights. As Jolliffe and Cadima [8] noted, PCA simplifies complex datasets by retaining the most crucial information, enhancing the model's precision and robustness in handling high-dimensional data. This approach resulted in the latent space model achieving lower MSE, higher R^2 scores, and a reduced percentage of large errors, making it a superior tool for precise wave height forecasting.



(a) Physical model whole track prediction (b) Latent space model whole track prediction

Figure 9: Whole track predictions comparison. The latent space model outperformed the CNN-LSTM model in predicting the entire storm track, as seen in the closer alignment of predicted (blue) and actual (orange) values. The flatter difference line in the latent space model highlights its superior accuracy in capturing the storm's progression.

The latent space model's superiority was further substantiated through a comprehensive analysis of maximum peak value predictions and the entire storm track predictions (Figures 9a and 9b). In these evaluations, the latent space model consistently outperformed the CNN-LSTM model, particularly in capturing the nuances of storm progression. The model demonstrated a notably flatter difference line between observed and predicted values, which is indicative of its higher accuracy and precision in tracking the storm's development. This level of performance suggests that the latent space model is more adept at maintaining consistent accuracy across varying storm conditions, especially in predicting peak Hs values that are critical for disaster forecasting and risk management.

However, despite these significant improvements, the latent space model still exhibited limitations, particularly in scenarios involving data imbalance. The model's predictive capability was less reliable in extreme conditions, where the distribution of Hs values is highly skewed, leading to potential underestimation of the most severe wave heights. This ongoing challenge highlights the need for further refinement in the model's handling of outlier data to ensure robust performance across all possible storm scenarios.

Forecasting vs Regression

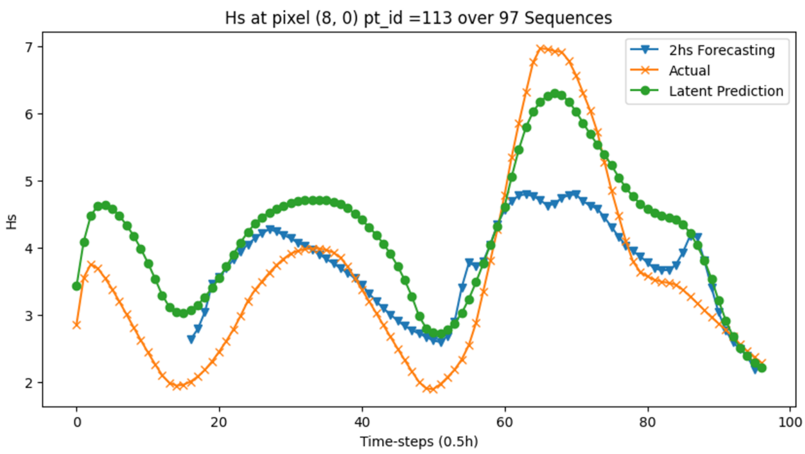


Figure 10: Regression vs Forecasting model performance. This figure compares the accuracy of the regression model and the forecasting model in predicting extreme wave heights. The regression model showed superior performance in capturing peak values, making it more suitable for applications requiring high precision in extreme scenarios.

The comparison between forecasting and regression models reveals significant differences in their performance and suitability for different tasks. Forecasting models (baseline model) are typically designed to predict future values based on past data, handling sequential data where each prediction depends on the previous ones. However, in the testing approach applied here, real data is used at each step, preventing the accumulation of errors typically associated with forecasting models. Although both models show similar performance in terms of overall error—with the regression model having an MSE of 0.75, MAE of 0.64, and R^2 of -0.54, compared to the forecasting model's MSE of 0.75, MAE of 0.7935, and R^2 of 0.2204—the regression model performs better in predicting peak values (Figure 10). This advantage arises because regression models predict each point independently, without relying on sequential dependencies, which allows for more accurate peak value predictions, avoiding potential error propagation common in forecasting models. For applications focused on achieving accurate peak value predictions, particularly in scenarios like predicting extreme wave heights, the PCA regression model is the preferred choice due to its ability to handle critical outliers effectively. Conversely, forecasting models are better suited for capturing the evolution of events over time, making them ideal for disaster response scenarios where understanding the temporal progression of a sequence is crucial.

8 Conclusion

This research has demonstrated the potential of deep learning models in predicting significant wave heights (Hs) induced by tropical cyclones, providing a viable alternative to traditional hydrodynamic models. The baseline LSTM model, while effective for short-term forecasts, was limited by its inability to generalize across different bathymetric conditions and its struggle with skewed data distributions. The CNN-LSTM model improved upon this by incorporating spatial features, significantly enhancing predictive accuracy. However, it still faced challenges in predicting peak Hs values, a critical requirement for applications such as reinsurance.

The introduction of the PCA-TCN-LSTM model marked a significant advancement, leveraging Principal Component Analysis (PCA) to compress data and focus on the most critical features for accurate prediction. This model outperformed the CNN-LSTM model, particularly in its ability to predict extreme wave heights, which are vital for effective disaster management and risk assessment. Despite these advances, challenges remain, particularly in handling data imbalances that affect model performance in extreme conditions.

Overall, this study highlights the importance of combining spatial and temporal features in predictive models and the value of dimensionality reduction techniques like PCA. Future work will focus on further refining these models, particularly in addressing data imbalances, to enhance their robustness and reliability across all storm scenarios. These improvements will be crucial in advancing the application of deep learning in coastal risk management and disaster response, offering faster and more accurate predictions that can better protect vulnerable coastal communities.

9 Future Work

Future work will explore the feasibility of transitioning from a regression-based model to a forecasting-based model for predicting significant wave heights (Hs). Given that both the current PCA-TCN-LSTM model and the baseline forecasting model utilize LSTM in their final layers, this transition is not only possible but also promising. By adopting a forecasting approach, the model could provide real-time predictive capabilities, continuously updating Hs predictions as new data becomes available. This shift could enhance the model's effectiveness in dynamic storm scenarios, where conditions change rapidly, and timely, accurate predictions are crucial for disaster response and risk management.

Another important area for future exploration is the application of transformer-based models in the latent space. Although the current report did not include results from previous attempts with Vision Transformers (ViT) in the physical 2D image space, these experiments showed that ViT did not significantly outperform the CNN-LSTM model. However, applying the transformer's attention mechanism within the latent space, rather than relying solely on TCN's dilated convolutions, could potentially improve the model's ability to capture critical features, leading to better performance in predicting peak Hs values.

Bibliography

[1] Andrychowicz, M., Espeholt, L., Li, D., Merchant, S., Merose, A., Zyda, F., Agrawal, S., and Kalchbrenner, N. (2023). Deep Learning for Day Forecasts from Sparse Observations. *arXiv.org*. doi:10.48550/arXiv.2306.06079.

[2] Bethel, B.J., Sun, W., Dong, C., and Wang, D. (2022). Forecasting hurricane-forced significant wave heights using a long short-term memory network in the Caribbean Sea. *Ocean Science*, 18(2), pp.419–436. doi:10.5194/os-18-419-2022.

[3] Chen, J., Huang, Y., Wu, T., and Yan, J. (2023). A WaveNet-based convolutional neural network for river water level prediction. *Journal of Hydroinformatics*, 25(6), pp.2606–2624. doi:10.2166/hydro.2023.174.

[4] Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv.org*. Available at: https://arxiv.org/abs/1412.3555.

[5] Emanuel, K., Ravela, S., Vivant, E., and Risi, C. (2006). A Statistical Deterministic Approach to Hurricane Risk Assessment. *Bulletin of the American Meteorological Society*, 87(3), pp.299–314. doi:10.1175/bams-87-3-299.

[6] FEMA Gov. (2022). 5 Years Later: Hurricane Irma in Florida. Available at: https://www.fema.gov/fact-sheet/5-years-later-hurricane-irma-florida.

[7] Idier, D., Bertin, X., Thompson, P., and Pickering, M.D. (2019). Interactions Between Mean Sea Level, Tide, Surge, Waves and Flooding: Mechanisms and Contributions to Sea Level Variations at the Coast. *Surveys in Geophysics*, 40(6), pp.1603–1630. doi:10.1007/s10712-019-09549-5.

[8] Jolliffe, I.T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), p.20150202. doi:10.1098/rsta.2015.0202.

[9] Lea, C., Vidal, R., Reiter, A., and Hager, G.D. (2016). Temporal Convolutional Networks: A Unified Approach to Action Segmentation. *Computer Vision and Pattern Recognition*. doi:10.48550/arxiv.1608.08242.

[10] Lee, J.-W., Irish, J., Bensi, M., and Marcy, D. (2021). Rapid prediction of peak storm surge from tropical cyclone track time series using machine learning. *ScienceDirect*. Available at: https://www.sciencedirect.com/science/article/pii/S0378383921001691.

[11] Meng, A., and Chen, S. (2022). A hybrid deep learning architecture for wind power prediction based on bi-attention mechanism and crisscross optimization. *Energy*, 238(PB). Available at: https://ideas.repec.org/a/eee/energy/v238y2022ipbs0360544221020430.html [Accessed 29 Aug. 2024].

[12] Merchant, S. (2023). MetNet-3: A state-of-the-art neural weather model available in Google products. *Research.google*. Available at: http://research.google/blog/metnet-3-a-state-of-the-art-neural-weather-model-available-in-google-products/ [Accessed 13 Jun. 2024].

[13] Minuzzi, F.C., and Farina, L. (2023). A deep learning approach to predict significant wave height using long short-term memory. *Ocean Modelling*, 181, p.102151. doi:10.1016/j.ocemod.2022.102151.

[14] Nadal-Caraballo, N., Yawn, M., Taflanidis, A., Kyprioti, A., Carr, M., Bensi, M., Liu, Z., and Gonzalez, V. (2021). Probabilistic Coastal Hazard Analysis: A Comprehensive Framework for Quantification of Coastal Storm Hazards Integrating Synthetic Hurricanes, Joint Distributions, Numerical Modeling, and Machine Learning. *NASA ADS*, 2021(0749-0208), pp.NH32A01.

[15] van Vloten, S.O., Cagigal, L., Rueda, A., Ripoll, N., and Méndez, F.J. (2022). HyTCWaves: A Hybrid model for downscaling Tropical Cyclone induced extreme Waves climate. *Ocean Modelling*, 178, p.102100. doi:10.1016/j.ocemod.2022.102100.

[16] Viglucci, A., Rivero, N., and Herald, M. (2022). If an Ian slammed South Florida: Nine-foot coastal surge, flooding inland to Hialeah. *WUSF*. Available at: https://www.wusf.org/environment/2022-11-05/if-ian-slammed-south-florida-nine-foot-coastal-surge-flooding-inland-to-hialeah.