Imperial College London
Department of Earth Science and Engineering
MSc in Environmental Data Science and Machine Learning

Independent Research Project
Project Plan

Estimating landfalling hurricane wave characteristics with parametric modelling

by
Sitong Mu

Email: sm1122@ic.ac.uk
GitHub username: edsml-sm1122
Repository: https://github.com/ese-msc-2022/irp-sm1122

Company: Moody's RMS

Company Address: The Minster Building, 21 Mincing Lane

London, EC3R 7AG, United Kingdom

Supervisors:
Dr. Christopher Thomas
Prof. Matthew Piggott

September 2023

Abstract

This study addresses the challenges in wave downscaling, a critical process in oceanography and coastal engineering, particularly for augmenting predicting nearshore wave fields during tropical cyclones and for long-term historical wave data. Employing deep learning techniques, this research focuses on hybrid downscaling that combines both dynamical and statistical methods. Two deep learning architectures are developed as surrogate models to accelerate the downscaling process: a multi-layer perception (MLP), and, uniquely, a Long Short-Term Memory (LSTM) architecture that incorporates temporal dependencies for enhanced prediction fidelity. In a pioneering approach, both models were rigorously assessed for their capacity to generalize across three critical dimensions: spatial locations, temporal intervals, and distinct storm events. Results unveil a novel 'context-sensitive' behaviour in both models, emphasizing their potential utility for extreme wave conditions. While LSTM shows better track generalization for new storm events, its performance is limited by short time sequences and high feature dimensions. Neither model could effectively generalize to neighbouring locations, presenting an area for future research. This work provides a foundation for efficient, high-resolution wave downscaling, with implications for coastal risk management and engineering design.

Table of Content

1 Introduction

1.1 Problem Description

Wave downscaling, the process of obtaining large scale, offshore wave data to a finer scale, localized wave data, is inevitable in oceanography and coastal engineering.

One primary usage of wave downscaling is to address tropical cyclones (TCs). TCs are significant perils to global coastal communities, inflicting extensive infrastructural damage amounting to billions of dollars every year. The increasing trend of coastal urbanization worldwide further amplifies the detrimental impact of TCs (Volton et al., 2019). Storm surge and damaging waves are amongst the most destructive hazards caused by TCs, which significantly influence the severity of coastal flooding (Hoeke et al., 2021). Consequentially, prediction of nearshore wave field from TCs becomes a paramount requirement, which lead to the use of wave downscaling.

Apart from addressing TCs, wave downscaling is also necessary in acquiring long-term historical wave data. The long-term historical wave data is useful under many circumstances. In marine structure designment, long-term database of spatial wave climate parameters is required for wave energy assessment (Camus et al., 2011). In risk management industry, insurance companies that cover nearshore property consistently demand flood risk probabilistic analyses which is based on long-term historical wave data (Volten et al., 2022). The analyses generally offer design wave height for more than 100 years return periods. For example, the United Kingdom, employs a 1000-year return value for coastal flood boundary conditions (Environment Agency, 2018).

However, long-term historical wave data suffers from a lack of observation. Buoy measurement of wave does not provide enough time span and can be far from place of interest, which make it difficult to represent local wave climate (Camus et al., 2011). Reanalysis technique can be used as augmentation to observation data, but their spatial and temporal resolution may not support local costal application, where wave downscaling is needed.

Given the paramount importance of wave downscaling, however, the current downscaling strategy presents substantial challenge. Numerical models are used to produce wave field near coastline from deep water wave field and large-scale ocean condition. Numerical models can generate wave field of good accuracy and fine spatial and temporal resolution at a cost of huge amount of computation. This computation intensity makes numerical models unsuitable when large amount model simulation is required (e.g., downscale long-term

historical wave data), or when model result is vital in limited time (e.g., real-time predictions of wave field caused by TCs).

Given the huge computational cost of traditional numerical downscaling method, there is a pressing need for new approaches that can generate high-resolution nearshore wave field at low computational cost.

1.2 Literature Review

In the coastal engineering domain, three general approaches have been developed to downscale large-scale wave information: dynamical, statistical, and hybrid approaches. Dynamical downscaling consists of a hierarchy of numerical models capable of simulating wave transformation processes (Rusu et al., 2008), while statistical approaches make use of empirical relationships between offshore ocean variables and nearshore local variables to obtain small-scale information (Browne et al., 2007; Kalra et al., 2005). Both dynamical and statistical approach has limitation. According to Camus et al., 2011, the dynamical downscaling is generally the most accurate, but require most computation; the statistical downscaling is faster than the dynamic but need observation data at target location to find the empirical relationship and the lack of observation data could be a problem.

Hybrid approach, as the name implies, combines numerical and statistical approaches. It integrates the strengths while mitigating the weakness. In a hybrid approach, part of the offshore ocean scenarios is numerically downscaled first. The downscaled offshore scenarios provide nearshore wave field, which functions as the local observation in statistical approach. A training dataset can then be built. The features are the offshore scenarios and the targets are the numerically downscaled local results (Groeneweg et al., 2007; Stansby et al., 2007). Statistical relationship between the feature and target is modelled. This model further takes the remaining offshore scenarios as input to estimate the full local wave field. The statistical relationship in this context, is referred as transfer function. The model that learning the transfer function, is the surrogate model.

In hybrid downscaling, after certain number of offshore scenarios are dynamically downscaled, several state-of-art methods are used to find a transfer function. The methods include typical machine learning (ML) based radial basis function (RBF) (Camus et al. 2011; Volten et al., 2022) and Gaussian Process Simulator (GPE) (Pullen et al., 2018, Parker et al., 2019), also deep learning (DL) based multi-layers perception (MLP) (James et al., 2018, Peach et al., 2023). A relatively traditional, non-ML based method is look-up table (LUT),

which was found to need more training cases to compete GPE (Malde et al., 2016). For RBF, Volten et al. 2022 makes improvement based on Camus et al. 2011. They compress training data using principal component analysis which significantly reduces the data dimension. For MLP, In Peach et al. 2023, a well-trained LUT method is compared with MLP and the MLP gives better result than LUT both in accuracy and efficiency. This finding also suggests the potential privilege of DL based MLP method, because the LUT in the study is trained on sufficient cases, competing RBF and GPE. Although MLP is a DL based method, the neural network in James et al. 2018 is a fully connected feed-forward network with 3 hidden layers, 20 nodes each layer, which is not very 'deep' while has good agreement with both observation and numerical output. This further motivates the study of DL based methods as surrogate models.

The wave field data is time-series but none of the study above use a recurrent neural network (RNN) as surrogate model. Peach et al. 2023 has tried to stack a sequence of time series as feature to predict the target at the last timestep, but the model used is still MLP. LSTM is a type of RNN that shows great power in forecasting time-series data like solar irradiance (Srivastava and Lessmann, 2018), stock market (Shah et al., 2018) and meteorological elements (Miao et al., 2020). It is worth investing the ability of LSTM in wave downscaling. Furthermore, studies on wave downscaling tend to focus on the metrices of model result and the number of samples used to outperform other methods. Few of studies investigate the generalization of models and answer the question like how the model performs on neighbour location or wave field triggered by a different storm event. The generalization can be crucial when there is a lack of data at some time/place/event but have the data at time/place/event close to the target.

1.3 Objectives

To list the project objectives, it is necessary to state first that the project is external, which not only means there is a specific target to fulfil as part of the project, but also the research interest should be partly considered in light of the company.

For this project, the hybrid approach is the focus. Two reasons are listed. First, the research area does not have enough observation data to build a pure statistical model. Second, the company aims to build a surrogate model for their deep-water to shallow-water numerical spectral wave model and they offered numerical model data, which can be considered as the first step in hybrid downscaling. The methods used for surrogate modelling is decided to be DL based, considering the advantages of MLP method. The study will also explore the

possibility of downscaling wave field with LSTM and the generalization of both MLP and LSTM.

In conclusion, this study aims to build a surrogate model for a spectral wave numerical model that downscale offshore ocean variables to nearshore wave variables, particularly, significant wave height (SWH). The study will 1. Build a MLP based on former relative papers. 2.Invent a LSTM suitable for wave downscaling. 3. Test the model performance and model generalization.

2 Methodology

2.1 Data

2.1.1 Data description

The data is confidential, provided by RMS company. However, some illustration of data has been approved.

The data is extracted from simulation result of a hierarchy of numerical models. I will first briefly explain the hierarchy of numerical models. Initially, a storm model is used to generate synthesis storm tracks data (time series of storm eye location, max wind speed, max radius, lowest air pressure, etc.,) on North Atlantic Ocean. There are 188 storm events in total, randomly sampled during the period 1900-2020. For each storm event, the storm track data are fed into a 2D wind field model to reconstruct a 2D wind field. Then, there is a hydrodynamic model to simulate water currents and surface elevation. Finally, the wind field, water current and surface elevation are fed into a spectral wave model (Mike21) to run storm surge simulation. The result of Mike21 is extracted at three sets of locations along Florida coastline, as is shown in figure 1. For each point on Fig.1, 6 types of parameters are given, including 4 wave parameters, which are significant wave height (SWH), peak wave period (PWP), mean wave direction (MWD) and total water depth (TWD), from the spectral wave model Mike21, and 2 wind parameters, including wind speed and wind direction, from the 2D wind field model. Furthermore, for one storm track, data at one point of one parameter is a time series of 97 timestamps with a time interval of 30 minutes. In conclusion, there is three set of data, each of which can be described in 4 dimensions which are (storm track (188), time (97), space (155 for offshore, 145 for coastline), parameter (6)). It is worth mentioned

that there is a lack of observation data in the research for model verification. As the result, the surrogate model is focused on recreating the numerical model output only.

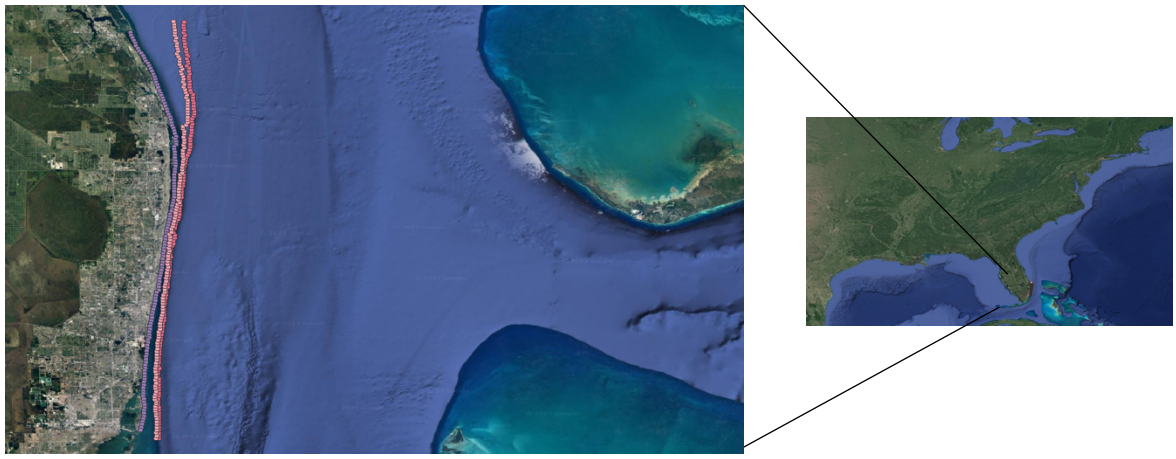

Fig 1, Research location, Florida coastline. Results of the spectral wave model (Mike21) is extracted at three sets of locations (purple, pink, red colour on the left image). The red set include 155 50m isobath points. The pink set include 155 25 isobath point. The purple set include 145 coastline points. For all the three sets, points are spaced at 0.01° (about 1km) resolution. Picture source from RMS.

2.1.2 Data Analysis

With the data given, to build a surrogate model for the wave spectral model to predict SWH at coastline from deeper water, several further considerations are necessary. First, as both the 50-meter contour data and 25-meter contour data are provided, I implement a correlation analysis to decide if both set of data should be used as feature. According to the analysis, data at 25m contour and 50m contour are highly correlated (Appendix, Fig. 1), the Pearson correlation range from 0.93 for wind speed to 0.99 for SWH. As a result, the data at 50m contour was excluded in further modelling. Second, to decide if all the 6 parameters should be used as features, I implement another correlation analysis (Appendix Fig. 2), according to which, the 6 parameters are not highly correlated and consequentially, all the parameters are included as features.

2.2.2 Data pre-processing

The data need to be pre-processed to improve model performance. The following steps has been taken: 1.nan value filling, 2. Feature engineering and 3. Normalization. Target and features are processed differently.

For the model target, SWH at coastline, about 26% of the data is missing. According to RMS, this could be the case if the numerical model recognizes a location to be dry (i.e.,

above the sea water level) and value 0 is filled to Nan. However, this is uncommon, as the coastline points are supposed to be on sea. The assumption that Nan value should be filled by 0 is further consolidated by investigating the TWD (total water depth) at Nan points. The 0 TWD points agree with 0 SWH points at coastline after filling Nan, indicating there is truly no water at some point. After filling Nan, there is no normalisation applied on target, as suggested in James et al., 2018.

For model features, there is no Nan value and significant outliers. Further feature engineering was applied, following Peach et al. 2023. The two direction related parameters, MWD and wind direction are converted to sin(direction) and cos(direction) respectively, leading to the number of feature types increasing from 6 to 8. After the feature engineering, normalization was applied to features such that each feature has 0 mean and 1 standard deviation.

2.2 Surrogate Models

Two supervised deep learning models are used to perform the same task: to reconstruct the Mike21 model output SWH at coastline from 25m contour SWH, PWP, MWD, TWD, wind direction and wind speed. Both models are developed as standalone codes and are not extension of pre-existing code. The first model is a MLP following Peach et al. 2023. The second model is a Long Short-Term Memory (LSTM) model. The LSTM is chosen as the memory capability of LSTM should be able to learn time dependency, making LSTM more suitable to model time-series data.

2.2.1 Background

Deep learning, a specialized subset of machine learning, has shown significant potential for pattern recognition in extensive data sets. Utilizing artificial neural networks with multiple layers, it mimics the human brain's structure, enabling complex data processing.

A physics-oriented model like Mike21can be considered a nonlinear function that transmutes input variables such as tidal boundary conditions, ocean currents, and wind field into outputs like SWH and PWP. These input and output datasets can be structured into respective vectors, $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$.

The aim of the surrogate modelling is to construct a deep learning framework to serve as a stand-in for the numerical model (Mike21), seeking a nonlinear function that aligns the inputs $\underline{\boldsymbol{x}}$ to the optimal depiction of the outputs $\underline{\boldsymbol{y}}$., represented by:

$$g\left(\underline{\mathbf{x}};\ \underline{\underline{\mathbf{\Theta}}}\right) = \underline{\hat{\mathbf{y}}}\ . \tag{1}$$

Where the $\underline{\hat{\mathbf{y}}}$ represent the optimal depiction and $\underline{\underline{\mathbf{\Theta}}}$, the matrix of model parameters.

A deep learning model that has undergone adequate training will generate a mapping that can function as a proxy for the numerical model. This enables the bypassing of the numerical model, replacing the expensive solution of the partial differential equation with a data-driven machine learning model defined by vector-matrix operations as outlined in equation (1).

In the development of the deep learning models, Google Colab was utilized as the computational platform. Google Colab provides access to high performance GPUs for efficient computation and facilitate modelling through a browser-based environment with pre-installed libraries. It offers flexibility and scalability in model development, with compatibility to Jupyter Notebooks and optional access to enhanced resources.

2.2.2 The Multi-Layer Perceptron (MLP) network

A typical MLP has one input layer, several hidden layers, and one output layer, as shown in Fig. 2. The layers are made up with neurons and are fully connected. For each neuron in hidden layers and output layer, the neuron output y is calculated as:

$$y = \sigma(\textstyle\sum_i w_i x_i + b) \tag{2}$$

which is the value of an activation function σ applied on weighted (w_i) sum of outputs from previous layer neurons (x_i) plus a bias term (b). This is shown in the left of Fig. 2.

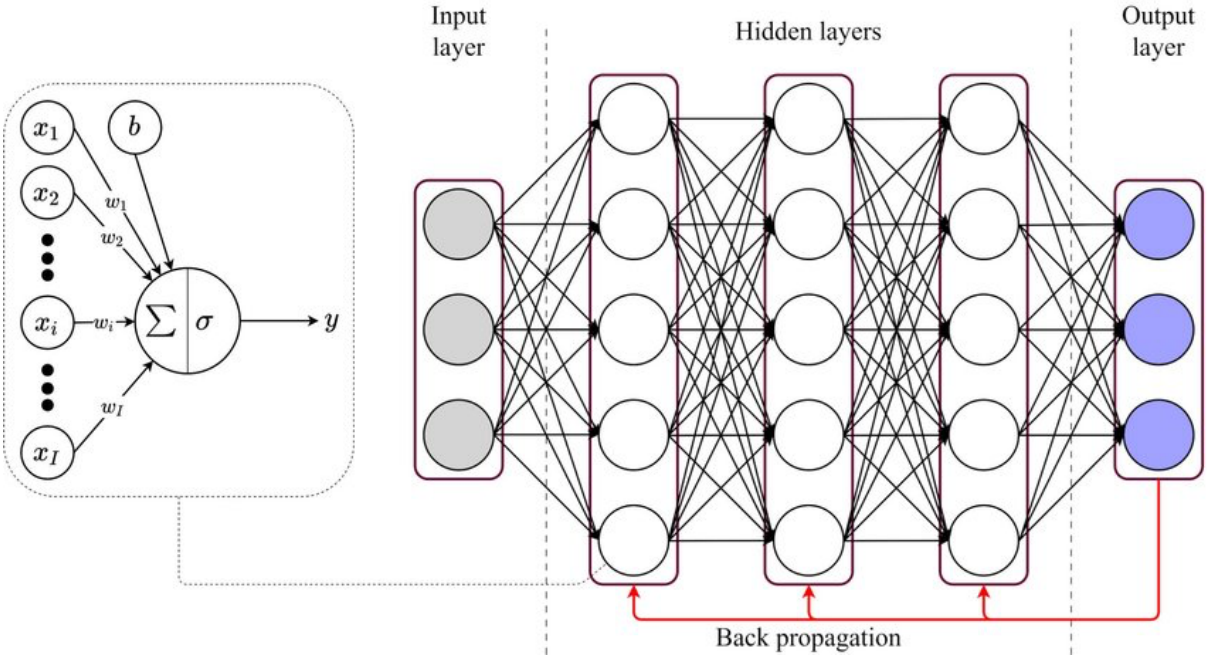

Fig. 2 basic structure of MLP. (Wang & Wu, 2020)

The MLP in this study consists of three hidden layers, each with 50 neurons, built in Pytorch. The activation function is set to LeakyRelu, which is:

$$\text{LeakyReLU(x)} = max(0, \text{x}) + \text{negativeslope} * min(0, \text{x}) \quad (3)$$

The negative slope is chosen as default, which is 0.01. In both study (James et al., 2018, Peach et al. 2023), the activation function was set to Relu. However, the Relu activation is know for 'dying neuron' issue, If a neuron's output is negative during the forward pass, the gradient becomes zero during backpropagation, effectively killing the neuron, and the neuron may never activate again. This was observed when I first using Relu training the model and has caused higher train and validation loss compared to using LeakyRelu, leading to a surge of error at some fragment of model prediction (Appendix, Fig. 3).

The loss function (L) is set to be mean squared error between the model prediction ($\underline{\boldsymbol{y}}$) and the numerical model output ($\underline{\boldsymbol{y}}$), plus a penalty term for L2 regularization, which gives:

$$L = |\underline{\boldsymbol{y}} - \underline{\hat{\mathbf{y}}}|_2^2 + \alpha|\underline{\Theta}|_2^2 \quad (4)$$

where α is a coefficient indicating the strength of penalty. The application of a penalty term can prevent the model parameter $\underline{\boldsymbol{\Theta}}$ become too large and avoid overfitting (Goodfellow et al., 2016).

The MLP network in this study is decided to have three layers with 50 neurons each layer, after parameter tuning. The input layer has a dimension of 1240 (8 features * 155 points at 25m contour) and the output layer has a dimension of 145 (145 points at coastline). The optimizer for loss function is Adam (Kingma and Ba, 2014), with the learning rate = 10^{-4} and default α =10^{-4}. It's worth mentioned that the implementation of the penalty term in loss function (Eq. 4) using Pytorch is by adding a weight decay factor in the Adam optimizer and leave the loss function without the penalty term. The weight decay factor is equivalent to the α of L2 penalty (see this documentation).

2.2.3 The Long Short-Term Memory (LSTM) network

Recurrent neural networks (RNNs) are well-suited for time-series prediction tasks due to their ability to maintain historical context through loops and memory elements. However, RNN suffers from the vanishing gradient problem, affecting their learning and memory retention. To address this, Long Short-Term Memory (LSTM) structure can be used

In a LSTM cell, the inputs are cell state and hidden state, and the outputs of one cell are fed back as the inputs to the next cell recurrently. A LSTM cell has input, output, and forget gate that selectively allow information flow, making them effective for handling vanishing gradients. Fig. 3 shows the structure of the LSTM cell used in this study. First, a forget gate f_t decides what to forget from the last cell state based on the input x_t and last hidden state h_{t-1} and give the ratio (0-1) of h_{t-1} to remain:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{5}$$

where σ is a sigmoid activation function, W_f and b_f are weights and bias of forget gate. Second, an input gate decides the information to be added to new cell state C_t based on two parts:

$$\widetilde{C_t} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

where $\widetilde{C_t}$ is the new information and i_t is the ratio of $\widetilde{C_t}$ to update C_t. $\tanh$ is the activation function used to decide $\widetilde{C_t}$. W_c and W_i are weights. b_c and b_i are bias.

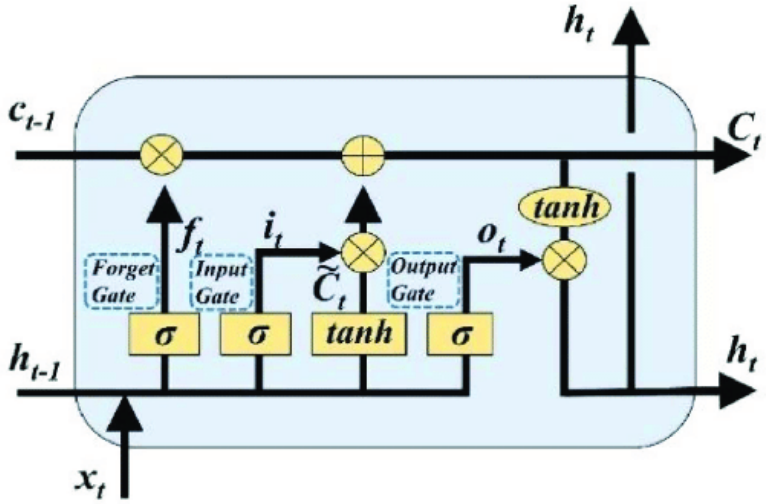

Fig. 3 Basic structure of a LSTM cell. (Jiang et al., 2020)

The updated cell state C_t can then be expressed as weighted sum of C_{t-1} and $\widetilde{C_t}$:

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \quad (8)$$

where $\odot$ is Hadamard product, which is an element-wise multiplication. Third, an output gate updates hidden state h_t, which essentially decides which parts of the cell state are going to be outputted:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$

$$h_t = o_t \odot \tanh(C_t) \quad (10)$$

Where W_o and b_o are weights and bias of the output gate.

The implemented LSTM use a Loss function same to Eq. 4, with an Adam optimizer. The model has an input dimension of 1450(8 features * 155 25m contour points) and output dimension of 145 (SWH of 145 points at coastline). After parameter tuning, the model is configured to have 2 layers of LSTM cells with a hidden state dimension of 64 and sequence length of input of 9. The sequence length is decided by analysing the autocorrelation of time-series data (Appendix, Fig 4).

2.3 Model generalization

The study has tested the models' configuration on unseen data and attempted to improve the model generalization. As part of the model testing is not straightforward, the strategy used is explained in this part.

The data has 4 dimensions which are track, space, time, and wave related parameters. The generalization of the two types of models is tested on unseen track, unseen time, and unseen space respectively. For most of the testing cases, new models are trained on designed training set and testing set.

To test the models on unseen tracks, the models are trained on 80% (148) of all the tracks (185) and tested on the 20% unseen tracks. It is worth mentioned that, a new LSTM model is not trained in the track generalization test. This is because when the LSTM model was built, the training data need to be consecutive and the 80% training set is the same as using the first 148 tracks to train the model.

To test the model on unseen times, the training data are selected to be the first 79 time steps of the 97 time steps in total. The remaining data at the18 timesteps consists test set.

To test the model on unseen space. There are 155 points at 25m contour and 145 points at coastline, distributed in a north-south band alone Florida coastline, as shown in Fig.1. The data is sliced into 5 consecutive chunks by point location. In each chunk, there is 31 points at 25m contour and 29 points at coastline. Models are trained on the southernmost, northernmost, and middle chunks respectively, and are tested on the other 4 chunks.

3 Comparisons and Results

To measure the model performance, several metrics are selected based on the literatures that had wave model performance analysis (Zieger and Greenslade, 2021; Morim et al., 2016). The metrics are explained in Table 1.

Table 1. The 7 metrics used. M_i stands for model result and O_i stands for observation, which is numerical model output in this study. Bias, RMSE, MAE are expressed in the units of the parameters (m), whereas the Scatter Index is a unitless error metric, often represented as a percentage. The Correlation Coefficient ranges from 0 to 1, with 1 indicating flawless linear correlation. Similarly, both the Coefficient of Efficiency and the Index of Agreement are dimensionless metrics that span from 0 to 1, where a score of 1 signifies optimal performance.

Bias	$\left(\frac{1}{N}\sum_{i=1}^{N} M_i\right) - \left(\frac{1}{N}\sum_{i=1}^{N} O_i\right)$
Root Mean Squared Error (RMSE)	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}$
Mean Absolute Error (MAE)	$\frac{1}{N}\sum_{i=1}^{N}\lvert M_i - O_i\rvert$
Scatter Index (SI)	$\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}}{\overline{O}}$
Correlation Coefficient (CC)	$\frac{\sum(M - \overline{M})(O - \overline{O})}{\sqrt{\sum(M - \overline{M})^2}\sqrt{\sum(O - \overline{O})^2}}$
Coefficient of Efficiency (CoE)	$1 - \frac{\sum_{t=1}^{T}(M_t - O_t)^2}{\sum_{t=1}^{T}(O_t - \overline{O})^2}$
Index of Agreement (IoA)	$1 - \frac{\sum_{i=1}^{N}(O_i - M_i)^2}{\sum_{i=1}^{N}(M_i - \overline{O}) + (O_i - \overline{O})^2}$

3.1 Model Accuracy

Figure 4 shows LSTM model prediction (left) and MLP model prediction (right) of SWH at coastline. The figure shows 9 randomly selected cases from validation dataset that is unseen when the model was trained. The figure aims to provide a relatively straightforward feeling about the model performance, with the concern that metrics alone could be abstract. The figure should be combined with Table 2 to better understand model performance. The Metrices in table 2 conclude model performance regarding all the data, which has 17945 cases in total.

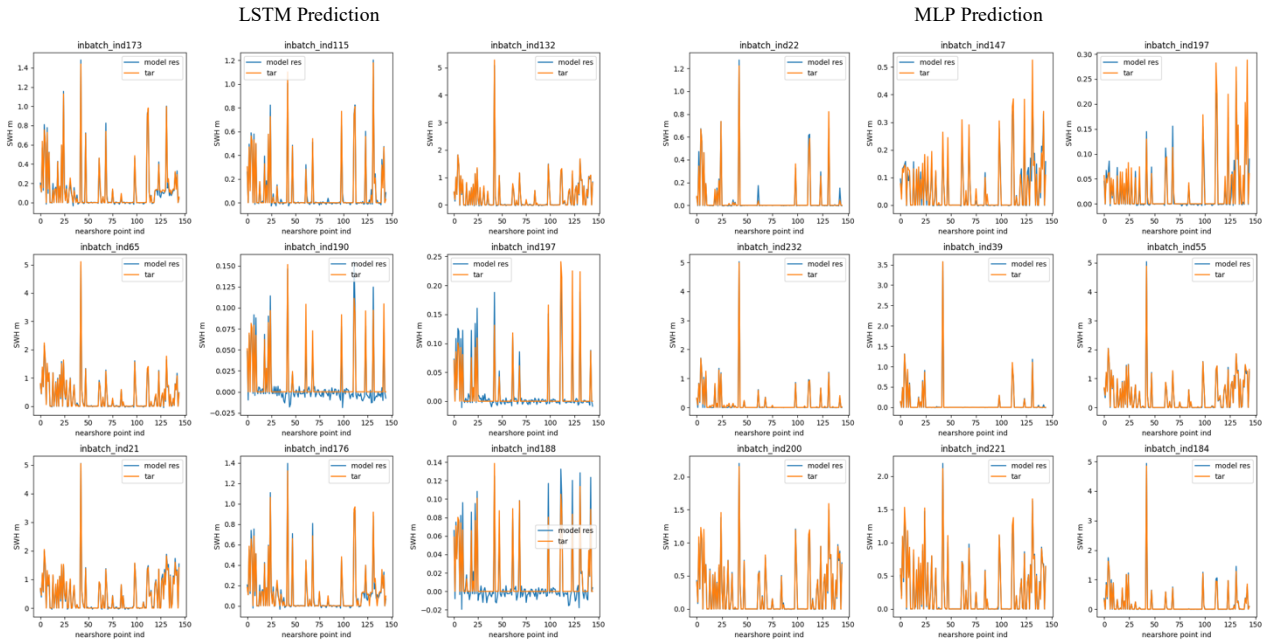

Fig. 4 Model estimation (blue line) and true value (orange line) of SWH at coastline. The Left image shows LSTM results. The right image shows MLP results. For each model, 9 cases are randomly selected from validation dataset. The x axis stands for coastline points index, the y axis is SWH in meter.

Both models show good agreement with the target numerical model output. Prediction of both models is systematically lower than the actual values, suggested by the negative bias as in table 2. This general underestimation of true value is not significant considering the 10^{-3} magnitude of bias and the 10^{-1} magnitude of true value mean (0.156 m). Both models tend to overestimate true value when SWH is large and underestimate when SWH is close to zero, which can be directly observed from figure 4 and can also be reflected by the about 10 times larger RMSE compared with bias.

Compare the LSTM result with the MLP result, the MLP outperform the LSTM, indicated by the higher CC, CoE, IoA, the lower SI and the smaller RMSE and MAE. The LSTM model has negative SWH prediction which is physically impossible.

Table 2. Table of performance metrices for estimation of SWH at coastline against numerical model output. The metrices are calculated based on all the numerical model cases.

	Bias (10^{-3} m)	RMSE (10^{-2} m)	MAE (10^{-2} m)	SI	CC	CoE	IoA
LSTM	-2.25	3.68	1.98	0.22	0.996	0.992	0.998
MLP	-2.84	2.67	1.15	0.17	0.998	0.995	0.999

3.2 Model Generalization

Table 3. Table of performance metrices for estimation of SWH at coastline against numerical model output. Metrices are calculated based on test data.

Dimension to generalize: Track

	Bias (10^{-3} m)	RMSE (10^{-2} m)	MAE (10^{-2} m)	SI	CC	CoE	IoA
LSTM	-3.33	3.18	1.78	0.18	0.997	0.994	0.999
MLP	-1.37	4.05	1.70	0.25	0.995	0.990	0.997

Dimension to generalize: Time

	Bias (10^{-3} m)	RMSE (10^{-2} m)	MAE (10^{-2} m)	SI	CC	CoE	IoA
LSTM	390	59	39	5.27	0.648	-3.30	0.589
MLP	1.02	4.48	2.03	0.37	0.990	0.979	0.995

3.2.1 Track Generalization

As shown in upper part of table 3, both models agree with true value when tested on unseen tracks. The LSTM model generally outperforms the MLP, with a leading advantage in the RMSE and SI.

3.2.2 Time Generalization

As shown in lower part of table 3, The MLP model prediction generally match the true value. The LSTM model fails to model the true value on unseen times. The CoE of LSTM is smaller than 0, which suggests that the mean of the true value is a better predictor than the model itself.

3.2.3 space Generalization

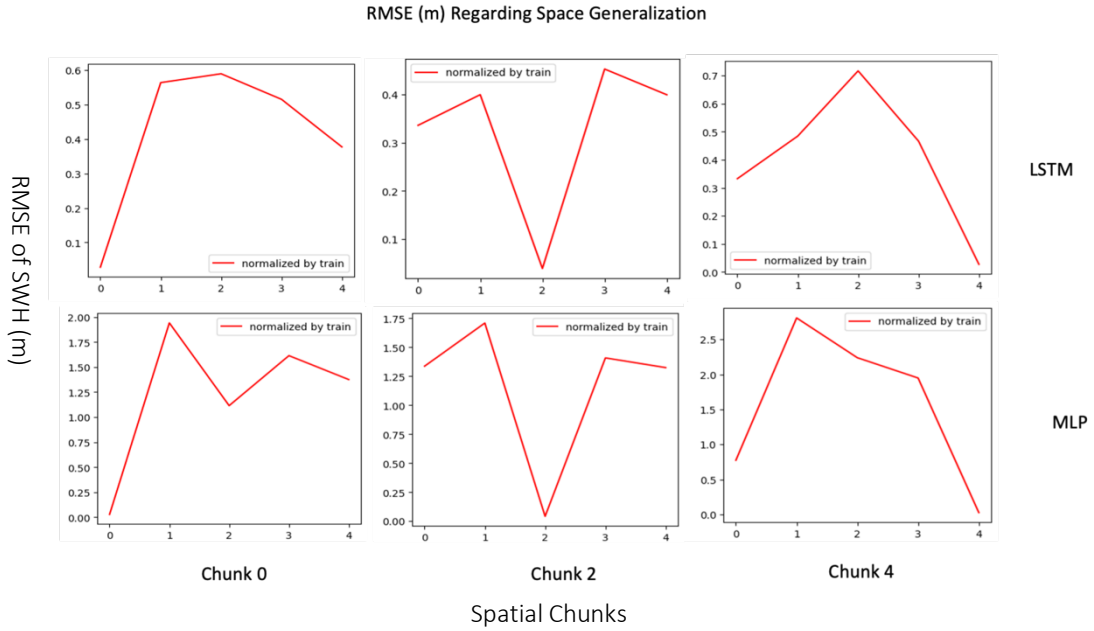

Figure 5. The RMSE of LSTM and MLP trained on one chunk of data and tested on all the 5 chunks of data. The bottom x label shows the chunk of data that the model is trained on. Testing chunks are normalized by the mean and standard deviation of the training chunk.

The RMSE of both models explode on unseen locations. Considering the true value has a mean about 0.15, the model cannot be used on neighbouring location.

4 Discussion & Conclusions

4.1 Result Discussion

The study has improved model accuracy compared with former research. The RMSE of SWH at coastline (2.7 cm for MLP, 3.7 cm for LSTM) is better than the one in James et al., 2018, using a MLP (6 cm), and comparable to the one in Malde et al., 2016 (about 3 cm), using a GPE with 80% of training events. However, detailed comparison is difficult to make as 1. Most papers only compare models' prediction with observation data, which is missing for this study. 2. The magnitude of SWH data and the performance metrics are not consistent for different studies.

The model accuracy result shown in Fig. 4 and Table 2 suggest a 'context-sensitive' model behaviour, i.e., the model tends to underestimate SWH when SWH is close to 0 and overestimate SWH when SWH is large. This is important for the company like RMS. In flood disasters, it is the extreme high value of SWH that has most impact and the low SWH value are usually not of serious concern. Thus, it is crucial to understand the model performance

when SWH is high and have the model data more representative to extreme conditions. The 'context-sensitive' behaviour of deep leaning surrogate modelling in this study is aligned with the most popular sample selection algorithm in coastline downscaling, the MDA, proposed in Camus et al., 2014. The selected samples of MDA, like the models in this study, stress on the extreme boundary situations that tends to be disastrous (Volten et al., 2022).

The comparative performance of MLP and LSTM reconstructing numerical model SWH result and the better track generalization result of LSTM, compared with MLP, suggests the potential of using LSTM for surrogate modelling. However, the advantage of LSTM in time series modelling has not been observed. Sutskever et al., 2014 demonstrates the efficiency of LSTMs for long sequence of data, while also suggesting that LSTM might be less efficient for shorter sequences. Pascanu et al., 2013 argues that when the feature dimension is large, it could be more challengeable to train a recurrent neural network. The data used in this study has a feature dimension of 1240, with only 97 timesteps and the LSTM model does no better than MLP. When the timesteps further reduce to 79 when tested on unseen timesteps, the LSTM dramatically underperform MLP. Apart from the relatively large feature dimension and small sequence length, another property of the data shall also limit the performance of LSTM. There is a lack of time dependency in model target. Although the features of model present considerable time dependency, the model target SWH exhibits no signs of autocorrelation as confirmed by the ACF plot (Appendix, Fig. 4).

The most challenging part is to generalize the model on unseen space, which is also the most serious limitation to the models. Current structures of both the models fail to predict SWH on neighbouring location. This does partly make sense as the bathymetry of coastline must have some difference. The bathymetry is part of the input to the hierarchy of numerical models that produce the target SWH but is not fed to the surrogate model because the bathymetry data is not provided. I have attempted to incorporate spatial characteristics to the models, specifically, elevation levels at coastlines sourced from Google Maps, in addition to the distance between contour points at 25m contour and points along the coastline. However, the model generalization does not improve. Relative work of this part is not presented here and can be found in code.

4.2 Future work

Suggested future work include 1. Find the bathymetry data to add to the model. 2. Adjust model structure to be more flexible in predicting different space. The implemented models predict 145 coastline point from 155 offline points. New model can try to predict one

coastline point from a few close offline points. 3. Compare the traditional ML surrogate model with DL based surrogate models regarding the ability to predict extreme wave conditions. 4. Explore the influence of feature dimension and time series length to LSTM model performance.

4.3 Conclusion

In light of the computational limitations of traditional numerical methods for wave downscaling, this study employed MLP and LSTM as surrogate models to enhance computational efficiency and to leverage their strengths in time-series prediction. The study builds a MLP model and a LSTM model function as surrogate model in hybrid wave downscaling. The models predict SWH at coastline from 6 wave related parameters at offshore 25m contour points. Both models' results show strong alignment with numerical model output. The models are tested ability to generalize on unseen tracks, times, and locations. It is discovered that 1. MLP and LSTM both tend to lightly overestimate extreme SWH and underestimate SWH. 2. LSTM is more flexible than MLP in downscaling wave field for new storm events. 3. It is not appropriate to use LSTM when the feature dimension is high, and length of time series is short. 4. Both DL based models are difficult to generalize to nearby location.

Word count (excluding captions and tables): 4808

Acknowledgements

I extend my deepest gratitude to Professor Matthew Piggott and Dr. Christopher Thomas for invaluable guidance and expert insights. Special thanks to Dr. Parastoo Salahfor her technical support in modelling methods. I also acknowledge Moodys' RMS for providing the great opportunity and the data for this study.

Reference

Browne, M., Castelle, B., Strauss, D., Tomlinson, R., Blumenstein, M., Lane, C., 2007. Near-shore swell estimation from a global wind-wave model: Spectral process, linear, and artificial neural network models. Coast. Eng. 54, 445–460. http://dx.doi.org/10.1016/j.coastaleng.2006.11.007.

Camus, P., Menéndez, M., Méndez, F. J., Izaguirre, C., Espejo, A., Cánovas, V., Pérez, J., Rueda, A., Losada, I. J., and Medina, R. (2014), A weather-type statistical downscaling framework for ocean wave climate, *J. Geophys. Res. Oceans*, 119, 7389–7405. https://doi.org/10.1002/2014JC010141

Camus, P., Mendez, F. J., & Medina, R. (2011). A hybrid efficient method to downscale wave climate to coastal areas. Coastal Engineering, 58(9), 851–862. https://doi.org/10.1016/j.coastaleng.2011.05.007.

Environment Agency, 2018. Coastal flood boundary conditions for the UK: update 2018. Retrieved from https://assets.publishing.service.gov.uk/media/60365303e90e0740ac3ea1b5/Coastal_flood_boundary_conditions_for_the_UK_2018_update_-_user_guide.pdf

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Groeneweg, J., van Ledden, M., & Zijlema, M. (2007). Wave transformation in front of the Dutch Coast. In J. M. Smith (Ed.), Proceedings of the 30th International Conference on Coastal Engineering (pp. 552–564). ASCE. https://doi.org/10.1142/9789812709554_0048

Hoeke, R.K., Damlamian, H., Aucan, J. and Wandres, M. (2021) Severe flooding in the atoll nations of Tuvalu and Kiribati triggered by a distant tropical cyclone pam. Frontiers in Marine Science, 7, 1–12. https://doi.org/10.3389/fmars.2020.539646

James, S. C., Zhang, Y., & O'Donncha, F. (2018). A machine learning framework to forecast wave conditions. Coastal Engineering, 137, 1–10. https://doi.org/10.1016/j.coastaleng.2018.03.004.

Kalra, R., Deo, M.C., Kumar, R., Agarwal, V.K., 2005. Artificial neural network to translate offshore satellite wave to data to coastal locations. Ocean Engineering 32, 1917–1932. http://dx.doi.org/10.1016/j.oceaneng.2005.01.007

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

Malde, S., Wyncoll, D., Oakley, J., Tozer, N., & Gouldby, B. (2016). Applying emulators for improved flood risk analysis. E3S Web of Conferences, 7. https://doi.org/10.1051/e3sconf/20160704002

Miao, K., Han, T., Yao, Y., Lu, H., Chen, P., Wang, B., & Zhang, J. (2020). Application of LSTM for short term fog forecasting based on meteorological elements. Neurocomputing, 408, 285–291. https://doi.org/10.1016/j.neucom.2019.12.129.

Morim, J., Cartwright, N., Etemad-Shahidi, A., Strauss, D., Hemer, M., 2016. Wave energy resource assessment along the southeast coast of Australia on the basis of a 31-year hindcast. Appl. Energy 184, 276–297. http://dx.doi.org/10.1016/j. apenergy.2016.09.064.

Parker, K., Ruggiero, P., Serafin, K. A., & Hill, D. F. (2019). Emulation as an approach for rapid estuarine modelling. Coastal Engineering, 150, 79–93. https://doi.org/10.1016/j.coastaleng.2019.03.004

Pascanu, R., Mikolov, T. & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. Proceedings of the 30th International Conference on Machine Learning, 28(3), 1310-1318. https://proceedings.mlr.press/v28/pascanu13.html

Peach, L., Vieira da Silva, G., Cartwright, N., Strauss, D. (2023). A comparison of process-based and data-driven techniques for downscaling offshore wave forecasts to the nearshore. Ocean Modelling, 182, 102168. https://doi.org/10.1016/j.ocemod.2023.102168.

Pullen, T., Liu, Y., Otinar Morillas, P., Wyncoll, D., Malde, S., & Gouldby, B. (2018). Bayonet GPE: A generic and practical overtopping model that includes uncertainty. Maritime Engineering, 171(3). https://doi.org/10.1680/jmaen.2017.31

Rogers, R. F., Velden, C. S., Zawislak, J., & Zhang, J. A. (2019). Tropical Cyclones and Hurricanes: Observations. Reference Module in Earth Systems and Environmental Sciences, Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.12065-2

Rusu, L., Pilar, P., & Guedes Soares, C. (2008). Hindcast of the wave conditions along the west Iberian coast. Coastal Engineering, 55(11), 906–919. https://doi.org/10.1016/j.coastaleng.2008.02.029.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. Proceedings of the Twenty-Eighth Conference on Neural Information Processing Systems, Montréal, QC, Canada, p. 27. https://arxiv.org/abs/1409.3215

Shah, D., Campbell, W., & Zulkernine, F. H. (2018). A Comparative Study of LSTM and DNN for Stock Market Forecasting. Proceedings of the 2018 IEEE International Conference on Big Data, Seattle, WA, USA, pp. 4148–4155. https://doi.org/10.1109/BigData.2018.8622462

Srivastava, S., & Lessmann, S. (2018). A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. Solar Energy, 162, 232–247. https://doi.org/10.1016/j.solener.2018.01.005.

Stansby, P., Zhou, J., Kuang, C., Walkden, M., & Hall, J. (2007). Long-term prediction of nearshore wave climate with an application to cliff erosion. In J. M. Smith (Ed.), Proceedings of the 30th International Conference on Coastal Engineering (pp. 616–627). https://doi.org/10.1142/9789812709554_0053

Van Vloten, S. O., Cagigal, L., Rueda, A., Ripoll, N., & Méndez, F. J. (2022). HyTCWaves: A Hybrid model for downscaling Tropical Cyclone induced extreme Waves climate. Ocean Modelling, 178, 102100. https://doi.org/10.1016/j.ocemod.2022.102100

Wang, H., & Wu, T. (2020). Knowledge-Enhanced Deep Learning for Wind-Induced Nonlinear Structural Dynamic Analysis. Journal of Structural Engineering, 146. 146.10.1061/(ASCE)ST.1943-541X.0002802.

Zieger, S., Greenslade, D.J.M., 2021. A Multiple-Resolution Global Wave Model – AUSWAVE-G3. Technical Report, Bureau of Meteorology, Melbourne, http://www.bom.gov.au/research/publications/researchreports/BRR- 051.pdf.

Appendix

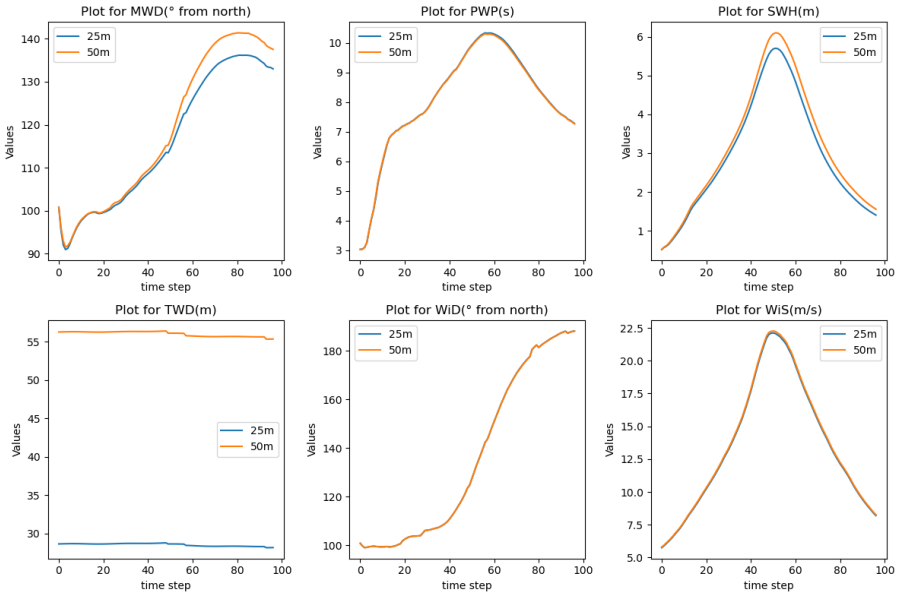

Fig. 1. Mean value plots for 6 wave related parameters at 25m contour (blue line) and 50m contour (orange line). Visual inspection of the plots suggests a strong correlation between each pair of parameters. The TWD shows a significance difference in magnitude. The remaining 5 parameters indicate high similarity. The numerical model resolution might not be high enough to clearly differentiate 25m contour and 50m contour.

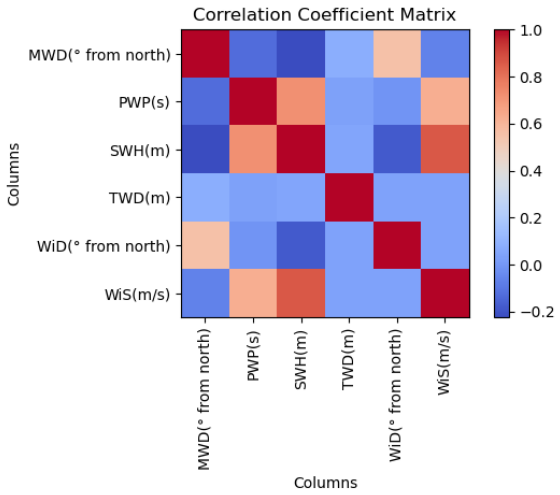

Fig. 2. Cross plot of Pearson correlation coefficient for the 6 parameters at 25m contour. One pair of parameters show high correlation, which is WiS vs. SWH.

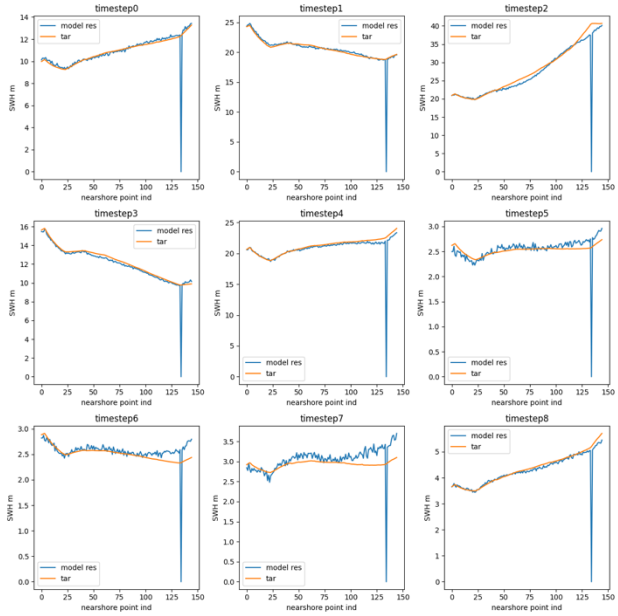

Fig. 3. MLP model prediction of wind speed (note, it is now SWH, the y labels in the plots are wrong) using an activation function of Relu. The model was built as an initial step to test the idea of predicting a set of coastline points from a set of offshore points. Wind speed was chosen as the target considering wind speed's much more consistent distribution for offshore and coastline, compared with SWH. The Relu activation function has caused difficulty in optimizing loss function and the error is very high at some samples. The LeakyRelu is more appropriate.

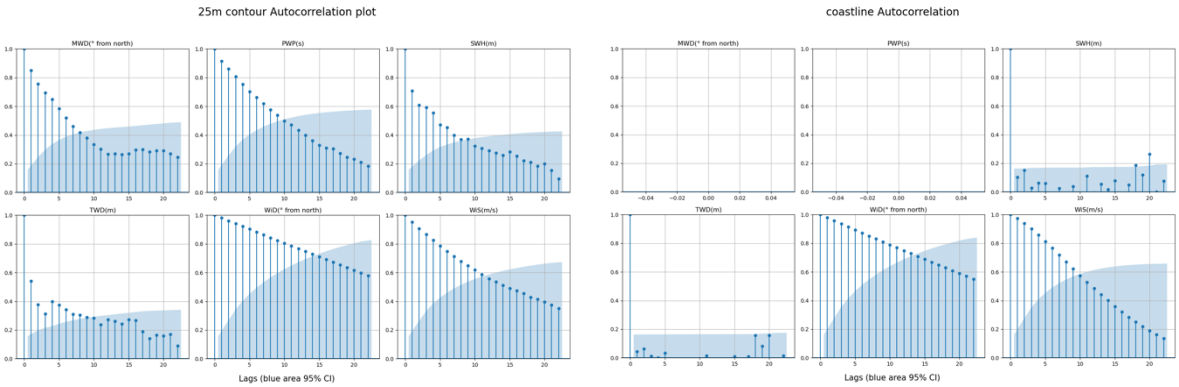

Fig. 4. Autocorrelation plot for the 6 parameters at coastline (right) and at 25m contour (left). The blue area indicates 95% confidence interval. The plot is used to find appropriate sequence length for LSTM model. After a lag of 9, the time series data is not significantly autocorrelated. For the coastline plot, the first two figures have no data due to the missing value and. The third and fourth images of right plot are of interest, suggesting time irrelevance of SWH and TED at coastline, which is odd. Both SWH and TWD have shown time dependency at 25m contour, as in left plot, but their auto correlation at a lag >1 is insignificant when at coastline, as in right plot.