Journal of Sustainability Science and Management
Volume 18 Number 4, April 2023: 120-138
eISSN: 2672-7226

IMPROVING LONG-TERM WAVE FORECASTING THROUGH SEASONAL ADJUSTMENT BASED ON STL AND CNN-GRU NETWORK

KHAN ABDUL REHMAN, AB RAZAK MOHD SHAHRIZAL AND MOHAMAD NOORASIAH*

Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia.

**Corresponding author: ar_shahrizal@upm.edu.my*
Submitted final draft: 12 January 2023 *Accepted: 7 March 2023* http://doi.org/10.46754/jssm.2023.04.009

Abstract: Most numerical models used to forecast wave parameters are time-consuming and computationally expensive. Currently, advanced machine learning techniques, such as artificial neural networks (ANN), provide a better alternative as they are substantially faster, more cost-efficient and more effective in handling non-linearity. In recent years, many ANN models have been developed to achieve satisfactory wave forecasting results. However, most of the research is limited to wave height forecasting and rarely any method that highlights the issue of seasonal fluctuation, which exists in time series data, is proposed. Keeping this in mind, this study proposes a hybrid convolutional neural network-gated recurrent network (CNN-GRU) model with a combination of seasonal adjustment based on seasonal-trend decomposition loess (STL) for wave parameters forecasting, including wave height and period. To evaluate model performance, error criteria methods, such as index of agreement (d), correlation coefficient (R) and root mean square error, were used. The results indicate that the proposed method outperformed every forecast horizon when compared with the model without seasonal adjustment with a degree of improvement ranging between 4% to 16% for wave height and 8% to 24% for wave period. Furthermore, the add-and-repeat prediction method is proposed in the study, where, after each prediction, the output of the model is added to the training set to produce a further prediction. The results from the proposed method indicate that predicted values follow the general trend to a great extent and there is a very small loss of accuracy between the first and final predictions with the R value reducing from 0.73 to 0.69 for wave height, and 0.63 to 0.61 for wave period.

Keywords: Artificial neural networks, machine learning, seasonal adjustment, wave parameters.

Introduction

A precise prediction of wave parameters is of great importance in the offshore and coastal environment due to the increase in coastal population and structures. Wave conditions not only influence marine inhabitants, but also play a vital role in implementing national security plans. Therefore, accurate forecasting of waves has always been a matter of major concern (Deo *et al*., 2001; Jain *et al*., 2006; 2011). However, it is challenging to accurately forecast wave characteristics due to the dynamic behaviour of coastal waves.

Wave parameters are forecasted generally by three mainstream methods involving numerical, empirical and machine learning techniques. Numerical models are the most widely used method to forecast waves. However, these models require substantial time and a high-performance processing system for computing large models due to the complexity of the calculation and the large size of data (Etemad-Shahidi & Mahjoobi, 2009; Wang *et al*., 2018). Therefore, these methods might not be suitable in cases where quick and reliable estimates are required. On the other hand, an empirical-based model, such as an autoregression moving average, cannot properly capture non-linearity and non-stationarities in data (Agrawal & Deo, 2002).

The rapid advancement in the field of machine learning has provided a great opportunity in the prediction of wave parameters within a shorter time and with high accuracy.

Machine learning methods, such as artificial neural networks (ANNs), mainly aim to identify the pattern in each data set and based on their understanding, produce a prediction of the desired target.

ANN models have been adopted to predict wave height. Deo and Sridhar Naidu (1998) conducted a study to predict wave height in real time based on a feedforward network (FFN) and autoregression model. The results showed that ANNs are more flexible, more general and have a better adaptive capability. Malekmohamadi *et al*. (2011) used several machine learning methods, including ANNs, support vector machine, adaptive neuro-fuzzy inference system and Bayesian network (BN) to check their wave height predictive capabilities. The results indicated that the prediction of these models is within the acceptable range, except that of the BN. Altunkaynak and Wang (2012) incorporated Kalman filtering with genetic programming to predict significant wave height. The results showed the superiority of this method over traditional ANNs. Mandal and Prabaharan (2006) predicted wave height using a recurrent neural network (RNN) and FFN. It was found that RNNs produce a higher correlation coefficient compared with a FFN. Nitsure *et al.* (2012) predicted wave height using genetic programming and wind data as input. The results were satisfactory and the correlation coefficient between the observed and predicted wave height was higher than 0.87. O'Donncha *et al*. (2019) provided an outline to integrate machine learning algorithms with physics-based models. The results of this study confirmed that machine learning can be a favourable tool for wave prediction on temporal and large spatial scales.

Kumar *et al.* (2017) performed a series of experiments by considering the differences in geographic locations using different ANNs. The results showed that minimal resource allocation network provides better wave height prediction in different geographic conditions. Wave data collected from adjacent buoys under two different situations was used in ANNs to predict wave height, where the results showed that the model produce better output for input consisting of extra data (Wei & Hsieh, 2018). A symbiotic organism search (SOS) was proposed by Akbarifard and Radmanesh (1997) to predict wave height from large data measured by buoys. The SOS algorithm outperformed other machine learning methods, including SVM, ANN and a dynamic model (SWAN). The research also introduced a SWAN-SOS hybrid model and achieved extraordinary results.

In time series data, the decomposition process is used to divide data into systematic components, including trend, seasonality, and non-systematic component residual. Burman (1980) recommended that time series data may be adjusted for seasonality by estimating and then removing seasonal components. The process of removing seasonal components is known as seasonal adjustment. The seasonal effect may hide the fundamental movement of the series and some interesting non-seasonal aspects that might be of interest. Seasonal adjustment may lead to a clear relationship between input and output variables.

Gardner and McKenzie (1989) suggested from the results of the renowned M-competition (Makridakis *et al*.,1982) that seasonal adjustment is an effective method for time series modelling. Nelson *et al.* (1999) discovered that ANNs trained with seasonally adjusted data produced considerably better predictions than those trained with non-seasonal adjusted data. Zhang and Qi (2005) also found similar results, where ANNs trained on seasonally adjusted data provided a better outcome. Mohanasundaram *et al.* (2019) used seasonally adjusted data with the autoregressive integrated moving average model to predict groundwater level and observed that the model with seasonally adjusted data performed better with R2 values of 0.82 and 0.93.

The STL decomposition procedure was chosen over the other decomposition techniques in the study because it can handle any type of seasonality, is robust to outliers and the smoothness of the trend is controlled by the user. The seasonal and trend decomposition using the loess (STL) procedure (Cleveland *et al.,* 1990) is used for the additive decomposition of the global time series. STL performs additive decomposition of the data through a sequence of applications of the loess smoother, which applies locally weighted polynomial regressions at each point in the data set, with the explanatory variables being the values closest to the point whose response is being estimated.

Past investigations were based on FFNs, feature selection, hybrid models and different learning algorithms to enhance model performance. But most of the suggested models were only tested for short-term or a small-range wave height prediction and models like feedforward neural networks or the ones consisting of classification algorithms were not well suited for time series problems (Burman, 1980). Furthermore, there is no research available that has attempted to improve wave forecasting through seasonal adjustment. Therefore, the aim of this research puts forward a seasonal adjustment method based on seasonal trend decomposition loess (STL), combined with a hybrid one-dimensional convolutional neural network and gated recurrent unit network (CNN-GRU) to improve long-term wave parameters forecasting. The research also proposed a new method of add-and-repeat prediction, where a small portion of data can be used to generate a considerably large span of prediction by combining model prediction results with actual data to generate an additional prediction.

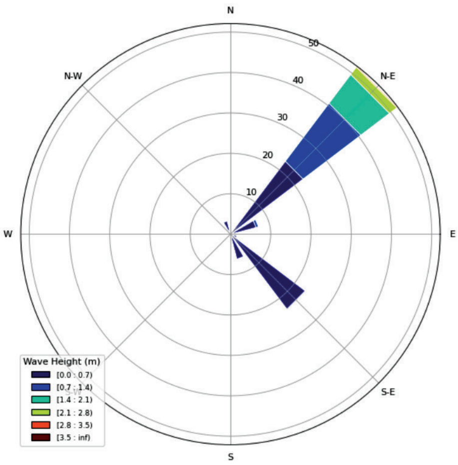
This research aimed to establish the accuracy of wave height and wave period prediction using seasonal adjustment methods like STL. The second objective was to develop a new method where long-term prediction could be possible using machine learning methods such as ANNs.

The research consists of three main stages. In stage one, raw data is cleaned and then STL is applied to the cleaned data for seasonal adjustment by decomposing it into its three components, i.e. trend, seasonality and residue. In stage two, the proposed CNN-GRU model is used to predict wave parameters, i.e. wave height and wave period based on seasonally adjusted data (trend component only) and non-seasonal adjusted data. In stage three, the proposed add-and-repeat prediction method is used to predict future wave height and period by combining actual data with predicted data. The advantage of the add-and-repeat prediction method proposed here is that it helps to achieve long-term prediction without significant loss of accuracy, which cannot be achieved by a single prediction using limited or small data. Finally, the same method is used to predict wave height and period for the year 2050. To our knowledge, no such research has been conducted, where the performance of wave predictions is done based on seasonal adjustment. Additionally, no prior studies predicted future wave height and wave period through a combination of repeated predicted data with actual data to produce long-term forecasted data. However, the study is limited to the use of only wave characteristics, such as wave height, wave period, and wave direction as input parameters to train the models and the use of a single seasonal adjustment method STL to determine the increase in model accuracy.

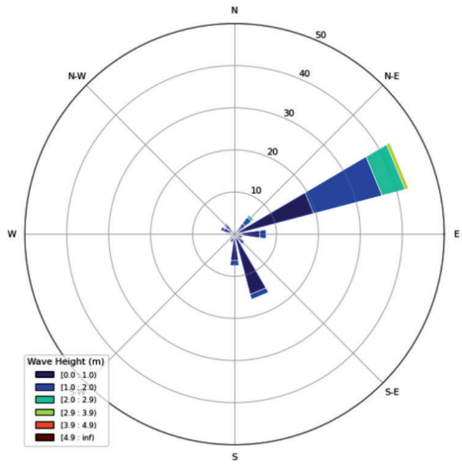
Data Collection

The research used hindcasted data collected from two sources for two different sites representing historical wave characteristics. Both data were obtained from global models. Figure 1 shows the wave rose diagrams, representing wave heights and associated directions for each site location.

Data Set 1**:** The wave data is obtained for the coast of Kuantan from the National Oceanic and Atmospheric Administration (NOAA). NOAA data are available from the year 1997 to 2010; a period of 14 years. The NOAA data was utilised to produce the offshore wave conditions. The wave data extraction point is located at 107E 5N. The data site is exposed to high waves from the South China Sea during the monsoon season. The associations of wave direction and wave height are shown in Figure 1 with the wave rose diagram, where the maximum wave



(a)

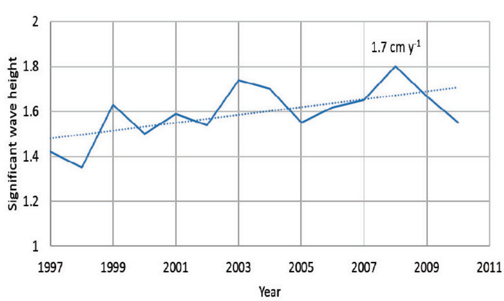


(b)

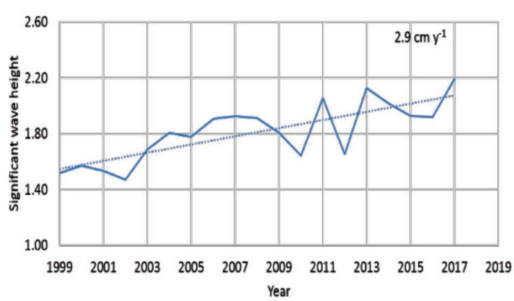
Figure 1: Wave rose for each data set used in the study. (a) Data set 1, (b) Data set 2

is generated from the northeast direction, with wave height reaching up to 3.51 m. Figure 2 reveals significant wave height annual variation for data set 1, where an increment of 1.7 cm in significant wave height can be observed. Many studies have shown a long-term increase in wave height due to climate change (Luijendijk *et al.,* 2018; Young & Ribal, 2019). Such variations in wave height can lead to dramatic consequences on off-shore works (Buizza *et al.,* 2018) and coastal populations (Ranasinghe, 2016).

***Data Set 2*: **The wave data used is collected from the Department of Irrigation and Drainage Malaysia. The location of data extraction is at 103.75E 6.39N, in the South China Sea. The duration of data collection is from 1999 to 2008 and the interval for data collection was 6 hours. The wave rose diagram for data set 2 is shown in Figure 5(b), with the maximum wave generated from the northeast direction and a wave height reaching 4.9 m. The annual variation of significant wave height reveals that there is a much greater increment, which is 2.9 cm per year when compared with data set 1 as shown in Figure 2, thus, proving that the site for data set 2 is more prone to be affected by future wave conditions.



(a)



(b)

Figure 2: Annual variations in significant wave height. (a) Data set 1, (b) Data set 2

Methodology

In this section, the details of the method adopted are explained, starting with data cleaning, STL, and data scaling followed by the proposed model. The subsection also includes detail about each layer of the proposed model. Finally, the model evaluation and experiment setup in an anaconda environment is provided. Figure 3 shows the flowchart of the proposed model.

Data Cleaning

Data cleaning involves the removal of null values, outliers, and errors from the data as shown in Figure 4. There are general data-cleaning procedures that can be performed, such as;

- Identifying outliers and defining normal data;
- Identifying and removing columns or rows with the same value;
- Marking null or empty cell as missing; and,
- Inputting missing values using statistics or a learned model.

Seasonal and Trend Decomposition using Loess (STL)

Cleveland *et al.* (1990) presented the STL method for time series data. In STL, time series data is decomposed using the loess method into three components, i.e. trend, seasonality, and residual. The method is based on additive decomposition, meaning that summing the components together will provide the original data. Therefore, for any time series data, Xt, decomposed using STL can be expressed as:

$$X_t = T_t + S_t + R_t \tag{1}$$

STL decomposition consists of two recursive procedures: an inner loop as shown in Figure 5 and an outer loop. The inner loop fits the trend and calculates the seasonal component. Every inner loop consists of six steps in total:

1. Detrending. An estimate of detrended sequence $X_t' = X_t - T_t^{(k)}$ at (k+1) iteration of the inner loop.
2. Seasonal smoothing. Using loess for detrended series, every cycle-subseries is smoothened and the result is given as $\tilde{S}_t^{(k+1)}$.

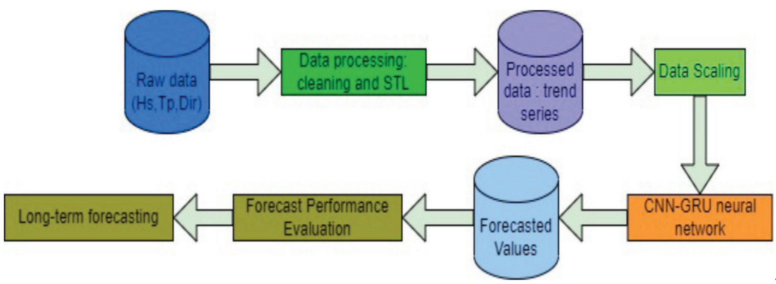


Figure 3: A flowchart of the proposed model

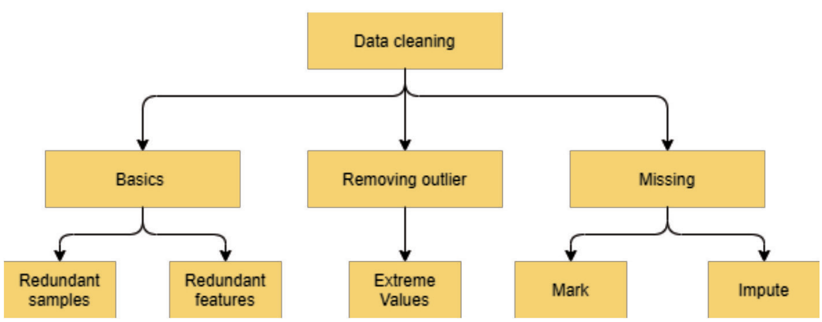


Figure 4: An overview of data cleaning

3. Low-pass filtering of smoothed cycle-subseries. Loess smoother and a low-pass filter are applied to $\tilde{S}_t^{(k+1)}$ to find any remaining trend. The result is marked as $\tilde{T}_t^{(k+1)}$.

4. Detrending of smoothed cycle-subseries. $S_t^{(k+1)} = \tilde{S}_t^{(k+1)}$ - $\tilde{T}_t^{(k+1)}$. $\tilde{T}_t^{(k+1)}$.

5. Deseasonalising. A deseasonalised series $X_t' = X_t - S_t^{(k+1)}$ is obtained by reducing the original series with seasonal components $S_t^{(k+1)}$.

The trend from deseasonalised series is smoothed using loess to get $T_t^{(k+1)}$ of the first loop (k+1).

Generally, to make sure the results converge, only a few iterations are needed. In case any anomaly is detected, robust loess will be applied through the outer loop by replacing loess in the second and sixth steps.

Data Scaling

Data scaling is the conversion of numerical input variables to a standard scaled range. Data scaling has been shown to increase the performance of many machine learning algorithms. Data scaling involves algorithms like linear regression that uses the input-weighted sum and k-nearest neighbours that utilise distance measures. Standardisation and normalisation are the two most used methods for scaling numerical data. In standardisation scaling, each input variable is deducted by the mean and divided by the standard deviation. This is done to achieve a mean of 0 for the attributes and a standard deviation of 1 for the resultant distribution. For normalisation scales, each numerical input is shifted and rescaled so that the resultant variable ranges between 0 and 1. Data standardisation techniques were adopted in this research, which can be expr

$$y = \frac{x - min}{standard\ deviation} \quad (2)$$

Model

When compared with traditional neural networks, CNN models provide high efficiency and better performance at automatically learning to detect and extract useful features from data. While CNN models are great at providing arbitrary mapping functions, GRU can offer effective and high performance for the learning of temporal dependencies, not only within the input sequence but also from the input sequence to the output. Combining both networks can help in harnessing the capabilities of both networks. A hybrid network combines the diverse capabilities of different architectures. A

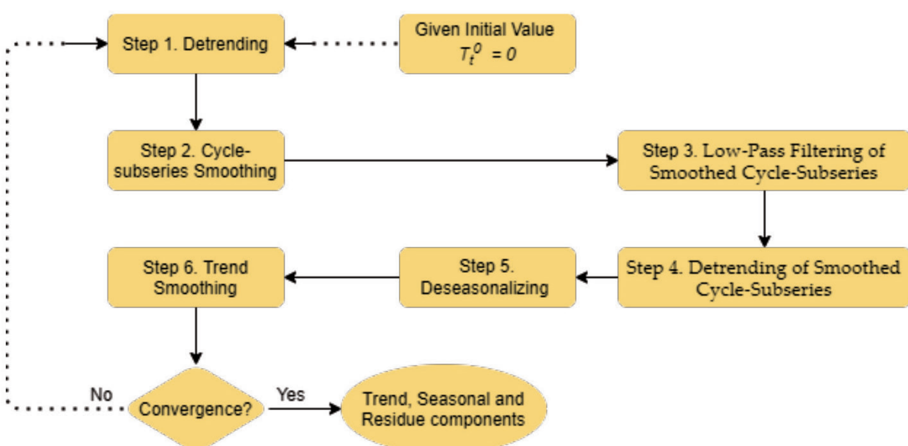


Figure 5: The STL inner loop procedure

hybrid model between CNN and GRU involves a CNN layer for feature extraction on input data, followed by a GRU layer to support the sequence.

The model's first layer consisted of a convolutional layer consisting of 64 filters followed by a max-pooling layer that distils the filter maps down to 1/4 of their size, which includes the most salient features. These structures are then flattened down to a single one-dimensional vector to be used as a single input time step to the GRU layer as shown in Figure 6. The entire CNN model is wrapped in a time-distribution wrapper, which allows the CNN model to read every sub-sequence of the data. The stacked GRU layers are followed by a dropout layer, which is also known as the regularisation layer that reduces the overfitting problem in the model. The convolutional layer reads across the subsequence with the help of several filters and kernel size defined. The number of filters is the number of reads or interpretations of the input sequence. The kernel size is the number of time steps included in each read operation of the input sequence.

Convolutional Layer

Recently, convolution neural networks have gained popularity due to their capabilities in image classification problems (Krizhevsky *et al*., 2012; Reichstein *et al*., 2019) and time series classification (Wang *et al.,* 2017). The extraordinary capability of CNN of parallel computation and feature extraction makes 1D (one-dimensional), 2D (two Dimensional and 3D (three-dimensional) CNN an extremely useful tool for image recognition, time series analysis and medical scanning image. Kim (2014) and Harbola and Coors (2019) both employed 1D CNN for sentence classification and wind prediction, respectively, where the results indicated that the 1D CNN models have higher efficiency and can provide similar results to the state-of-art model based on long-short term memory networks (LSTM). The basic concept behind the CNN architecture is that the model is trained to identify important features within each weight matrix and will then extract those features from the input.

In this research, 1D CNN is employed, rather than 2D or 3D CNN. The kernel in 1D CNN convolves along the time dimension, with the input sequence and the weights sheared, requiring fewer parameters to converge, and ensuring a faster and easier convergence. 1D CNN's one-dimensional features are extracted by kernels with specific characteristics from the input signal (Eren *et al*., 2019). Specific characteristics are located at any position by every kernel on the input features as shown in Figure 7. The number of parameters is reduced by weight-sharing on the same input feature map.

Max Pooling and Flattened Layer

The pooling layer is an essential aspect of CNN (Huang *et al.,* 2020). Pooling layers are used to reduce feature map dimensions while preserving

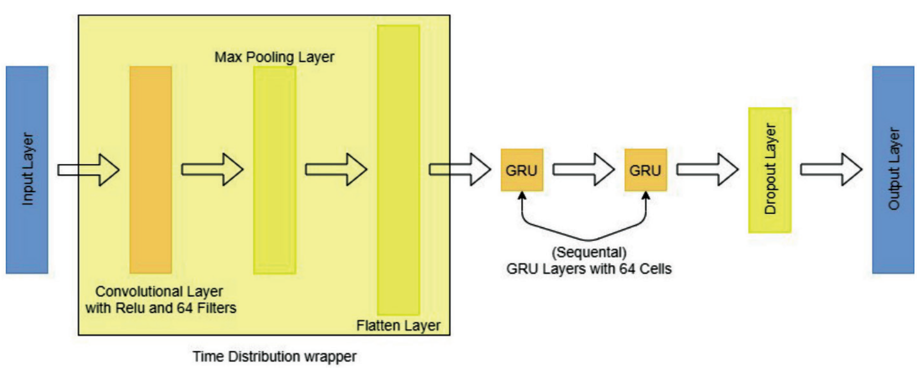


Figure 6: The proposed CNN-GRU model

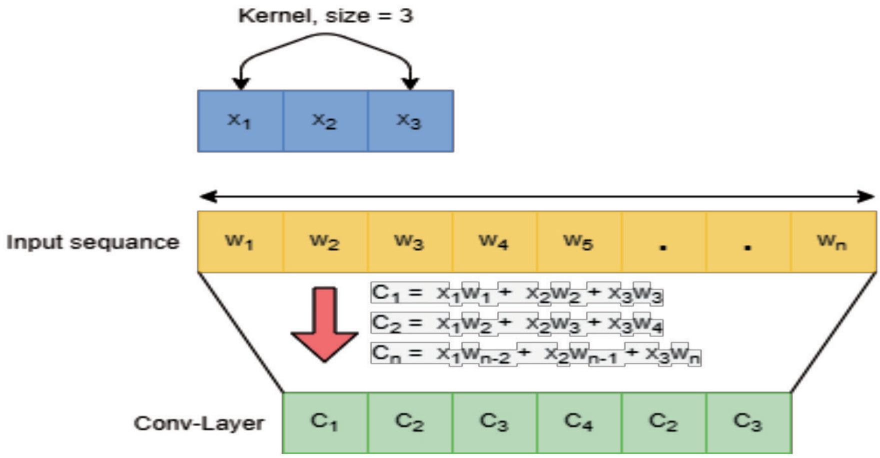


Figure 7: A simple representation of 1D convolutional operation

crucial features and helping in reducing network computation time. Pooling layers might also be used to solve the problem of overfitting. In this research, the max pooling layer was selected, which is one of the most used pooling methods (Zhao *et al.,* 2018). In max pooling, the highest parameters are selected by pooling operations from the feature map. Therefore, the output after the max pooling operation will contain the most prominent features as shown in Figure 8.

Flattened layers are used between CNN and GRU layers to reduce the feature map into a one-dimensional vector so that it can be used as a single input time step to the GRU layer.

Gated Recurrent Unit Network (GRU)

LSTM networks, which was developed in 1997, were initially intended for language processing because of their unique ability to remember long-term dependencies (Hochreiter & Schmidhuber, 1997) However, due to the complex structure of the model, it has a long processing time, making them unstable for very large data. To respond to this problem of decay in traditional RNNs and long computation complexity in LSTM, Chung *et al*., (2014) proposed the gated recurrent unit (GRU). Each GRU cell contains two gates: the reset gate rt and update gate zt, which track down the information towards the output gate as shown in Figure 9.

The primary function of the reset gate is to control how much information will flow in the current state from the previous state, whereas the updated gate acts as both an input and forget gate, and decides which information needs to be stored and which needs to be thrown away. GRU uses a hidden state to transfer information as it does not have a cell state like LSTM. The

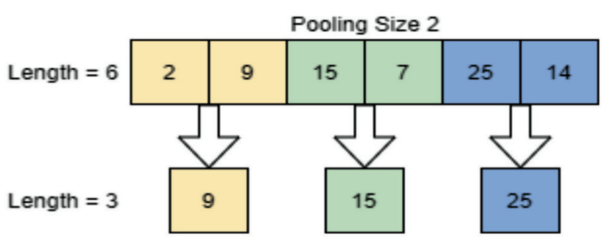


Figure 8: Simple representation of the 1D max pooling operation

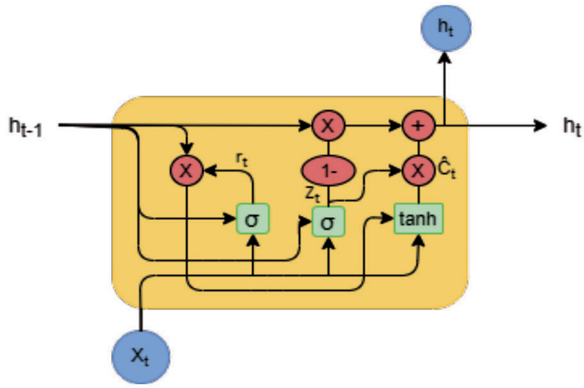


Figure 9: The structure of a GRU cell

following equations are used to determine the parameters of a GRU cell.

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (3)$$

$$Z_t = \sigma(W_Z \cdot [h_{t-1}, x_t] + b_Z) \quad (4)$$

$$\hat{h}_t = \tanh(W_c x_t + W_c(r_t * h_{t-1}) + b_c) \quad (5)$$

$$h_t = (1 - Z_t) * h_{t-1} + Z_t * \hat{h}_t \quad (6)$$

Wr, Wz, and Wc are weighted matrices of the network, br, bz, and bc are biased vectors and rt and Zt vectors are for the activation values of the update gate and reset gate.

Dropout Layer

In ANN, the term dropout means dropping out both hidden and visible units in the network. During the dropout process, the connection between neurons is temporarily disabled and there is no output given for those neurons. Thus, a dropout is a regularisation technique and is used to solve the issue of overfitting during the training stage of the model (Lv *et al*., 2019; Jeon *et al*.,2020) by slowing down the learning process.

Model Performance Evaluation

In this study, we have utilised the two most used methods to evaluate model performance, which is the cross-validation and out-of-sample (OOS) approaches. The primary reason for model performance evaluation in time series prediction is to handle dependence between observations. This can be achieved with the r by the OOS approach, in which data is split into training and testing sets and then a comparison is made between the model prediction and testing set, where cross-validation data is systematically split into a k-subset. Each subset is allowed to be used for testing, whereas k-1 subsets are used for training purposes.

Cross-validation Approach

The method is based on a single parameter called k, where the data is randomly split into k numbers of groups or folds (Bergmeir *et al*., 2018). Therefore, the method is mostly referred to as k-fold cross-validation. For a given dataset, it is important to carefully assign a *k* value as a poor selection might result in the misrepresented idea of the skill of the model.

Out-of-sample Approach

In the OOS approach, model performance is evaluated by splitting data into a training set and a testing set. The prediction of the proposed model is verified using three statistical error indexes, which are the index of agreement (d), root means square error (RMSE), and correlation coefficient (R). A standardised measure of the degree of model prediction error index of agreement was proposed by Pereira *et al*. (1981), in which the values range between 0 and 1, where the value of 0 indicates no agreement at all and 1 indicates perfect agreement. For the correlation coefficient (R), the value ranges between -1 and 1, where the value of -1 indicates a perfect correlation, but is negative and the correlation value of 1 shows perfect correlation positive. A correlation with a value of 0 indicates no correlation. For error criteria such as RMSE, there is no threshold of absolute good or bad value and the value depends on the range of the data set. The equation index of agreement and other error criteria are given as follows.

$$d = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(|y_i - \bar{x}| + |x_i - \bar{x}|)^2} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad (8)$$

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (9)$$

Where n is the total number of cases, x_i is the observed value, is the average of observed value, ȳ is the average of the predicted value and is the output/predicted value.

Setting up the Environment

A virtual environment in Anaconda was created to run the experiments for this research. Python 3.7.10 was used as the programming environment. In this virtual environment, the following packages were installed:

- Keras (2.4.3)
- Pandas (1.1.5)
- Numpy 1.19.5
- Matplotlib 3.3.2
- Statsmodels 0.12.2
- SKlearn 0.24.1

Results and Discussion

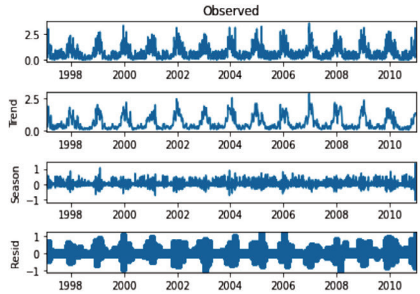
This section presents and discusses the results of the proposed research. Firstly, both data sets 1 and 2 were cleaned so all duplicates, outliers and missing data can be removed. Later, seasonal adjustments were made to both data with STL so a comparison could be made between model performance based on the data type. STL is based on locally weighted regression smoother (loess) and is well-suited for seasonal adjustment of data with high frequency (Ollech, 2018). Additionally, the fast computation of the STL algorithm makes it feasible to adjust different seasonal frequencies in an integrated iterative framework.

STL Decomposer

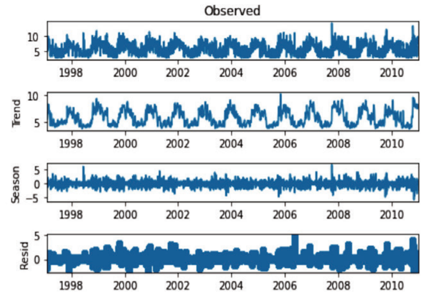
Figures 10 and 11 depict the result of STL decomposition for all three wave parameters (wave height, wave period and wave direction) of the entire data of data sets 1 and 2 used in the study. The reason for applying STL to wave direction is that in our first prediction, all three wave parameters are used as input to produce prediction outputs for wave height and wave period. Whereas in a subsequent stage, where the add-and-repeat prediction method is applied, the input consists of only wave height and wave period as no prediction will be produced for wave direction that can be added to the training data.

Model Evaluation through k-fold Cross-Validation

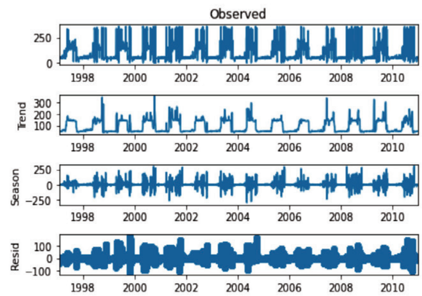
To establish an accurate estimate of the model, performance methods like cross-validation and OOS were adopted. For cross-validation, the proposed model was trained twice with two different inputs, i.e. (i) input consisting of actual data without any seasonal adjustment, and (ii) input consisting of data with seasonal adjustment. There is no formal rule for selecting the value of k, but is usually taken to be 5 or 10 (Kuhn & Johnson, 2013). By taking a large k value, the bias of the technique becomes smaller due to the difference between the subset and training set size becoming smaller. The value of k is selected to be 5 in this study so that the model can have half the run-time compared with when k=10, while still providing an accurate evaluation between models. For cross-validation, k= 5 means that the data set is split into 5-folds and in the first iteration, k=1 is used as the testing set whereas the remaining folds are used to train the model. In the second iteration, k = 2 is used as the testing set while the remaining fold is the training set and so on. The output from cross-validation for both data sets is shown in Table 1, where the mean absolute error (Eq. 10) is used to show model performance. The results suggest that for both data sets, the model trained with STL data provided a better significant error difference.



(a)

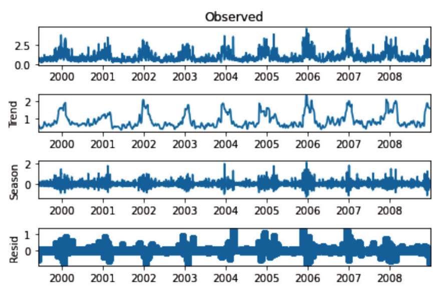


(b)

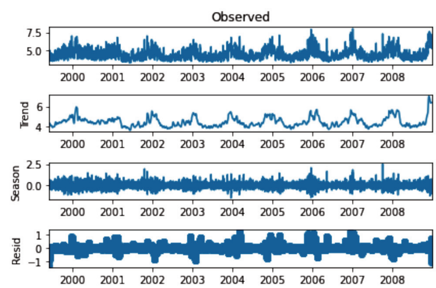


(c)

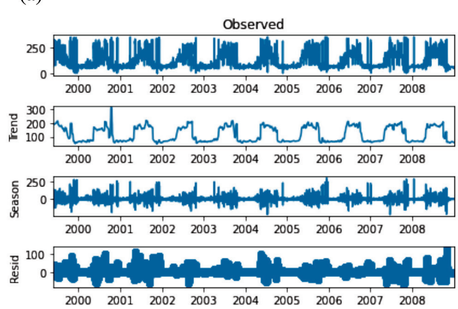
Figure 10: Seasonal decomposition results for the data set using wave height (a), wave period (b) and wave direction (c)



(a)



(b)



(c)

Figure 11: Seasonal decomposition results for data set 2 using STL. (a) wave height, (b) wave period and (c) wave direction

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (10)$$

Model Evaluation through OOS

For the OOS approach, before training the model on seasonal adjusted and non-seasonal adjusted data, both were split into a training set and a testing set. A total of 70% of the wave data from data set 1 (14 years) and data set 2 (7 years) were used to train the model and the prediction was made for the remaining 4 years and 3 years of data, respectively. Figure 12 shows a time series plot between observed and predicted wave height and wave period for both data sets. It can be seen in the figure that the model trained with seasonally adjusted data, i.e. SLT-CNN-GRU, follows the general trend of the data more closely and have fewer non-linear spikes compared with the model trained with non-seasonal adjusted data, i.e. CNN-GRU.

Figure 13 shows the error distribution between seasonally adjusted data and the non-seasonal-adjusted model for wave height and wave period. The plot signifies that model trained with seasonally adjusted data provided much better output, with error density showing more concentration close to 0, meaning the difference between the actual values and predicted values is much less compared with the model trained with non-seasonal adjusted data.

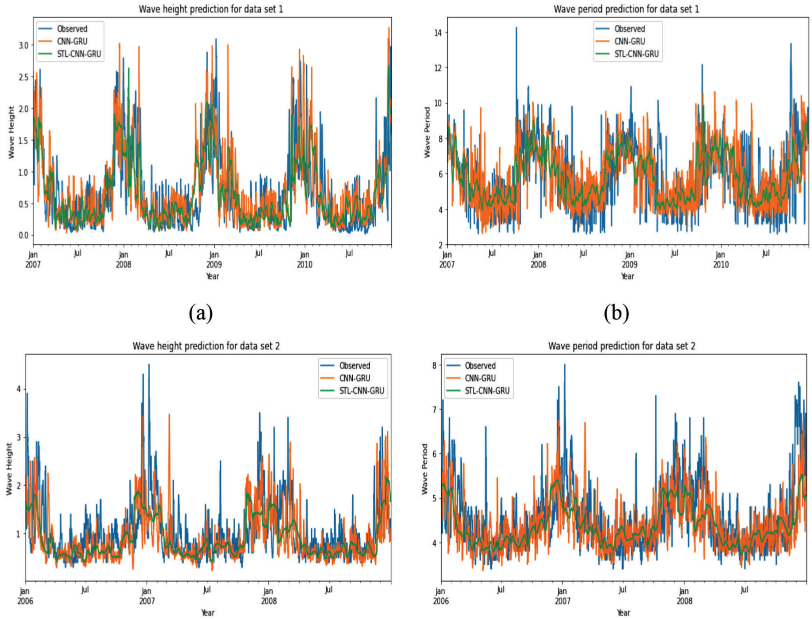
This can be further noticed more clearly in Table 2, where the accuracy of the model based on data type is evaluated using different error criteria. Table 2 also shows the degree of improvement in model prediction using STL data, where the degree of improvement determines the increase from one value to another in terms of percentages. For all cases, the seasonally adjusted model provided better prediction with a degree of improvement for

Table 1: The MAE value of cross-validation for k = 5

5-fold CV	**Data set 1**				**Data set 2**			
	CNN-GRU		**STL-CNN-GRU**		**CNN-GRU**		**STL-CNN-GRU**	
	MAE (%)		**MAE (%)**		**MAE (%)**		**MAE (%)**	
	Wave height	Wave Period	Wave height	Wave Period	Wave height	Wave Period	Wave height	Wave Period
Lag 1	5.25	16.27	5.61	3.67	15.92	19.73	8.30	4.95
Lag 2	8.35	16.28	5.41	2.33	15.65	18.82	8.07	6.13
Lag 3	7.10	15.29	4.03	5.29	16.95	19.57	8.07	4.7
Lag 4	5.85	16.51	4.81	4.36	16.70	19.44	5.07	7.25
Lag 5	5.85	16.72	8.11	3.41	20.00	19.27	5.73	5.31
Mean	6.48	16.21	5.60	3.81	17.04	19.37	6.82	5.67

Table 2: CNN-GRU and STL-CNN-GRU results for wave height and period

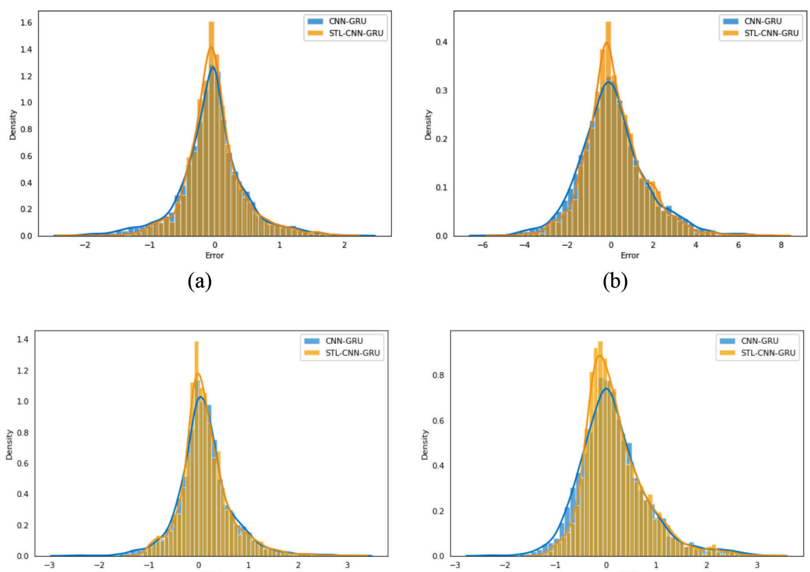
Data set	**Parameters**	**CNN-GRU**			**STL-CNN-GRU**			**Degree of Improvement (%)**		
		d	**R**	**RMSE**	**d**	**R**	**RMSE**	**d**	**R**	**RMSE**
1	Wave height	0.78	0.64	0.51	0.82	0.71	0.43	5.13	11	15.7
	Wave period	0.67	0.51	1.59	0.73	0.62	1.40	8.95	21.6	11.9
2	Wave height	0.63	0.48	0.57	0.66	0.55	0.51	4.8	14.6	10.5
	Wave period	0.64	0.51	0.7	0.68	0.63	0.63	6.25	23.5	10



(a) (b)

(c) (d)

Figure 12: The results for data sets 1 and 2. (a) and (b) are time series plots for observed and predicted wave height and wave period for data set 1, while (c) and (d) are time series plots for observed and predicted wave height and wave period for data set 2



(a) (b)

(c) (d)

Figure 13: The results for data sets 1 and 2 through the error distribution plot. (a) and (b) show the rrror distribution of predicted wave height and wave period for data set 1 by seasonal-adjusted and non-seasonal models, while (c) and (d) show the rrror distribution of predicted wave height and wave period for data set 2 by seasonal-adjusted and non-seasonal model

data set 1, ranging between 5% and 16% for wave height, and 8 %to 22% for wave period, respectively. For data set 2, similar results are observed with a degree of improvement ranging between 4% and 15% for wave height, and 6% to 24% for wave period, respectively. These results are in good agreement with the results observed by Nelson *et al*. (1999) and Mohanasundaram *et al*. (2019), where the model with seasonally adjusted data showed better prediction than the model without seasonal adjustment.

Long-term Wave Forecasting

To produce long-term wave forecasting, it is important first to understand the limitation of the data set. There is a certain extent to which any given data set can produce prediction without much loss of accuracy. Furthermore, the ANN model used in this study can only produce a prediction size equal to that of the training set minus the number of time steps. Since training data size plays a vital role in not only prediction quality, but quantity as well (Markham & Rakes, 1998; Fan *et al*., 2020), it is important to establish how many predictions can be generated with minimum loss of accuracy. To address this issue, here we suggest using the add-and-repeat prediction method. The method is based on a simple idea which is adding the prediction produced by the model to the training set to generate an additional prediction. To check the accuracy of the method, data set 1 was selected due to the shorter interval between observed readings, thus providing more data to train.

The first step to adopting the add-and-repeat method is to establish model accuracy based on the time interval, for which it can provide the most accurate results. Therefore, first, the model is trained with two different training set sizes, i.e. 4 years and 6 years. In both cases, the data used is seasonally adjusted with STL. The model is then used to make future predictions with time intervals ranging from 1 to 3 years to establish the time length for which the model can provide the best output. Since the prediction is only made for wave height and wave period, therefore, the input data also consists of wave height and period. Table 3 shows the model performance based on different periods of prediction.

Based on the result observed in Table 3, it can be assumed that the model produces the best output when the prediction size is almost half the training data size. Thus, for the add-and-repeat method, the prediction size each time will be half of that of the training data set.

The second step of the add-and-repeat method is to determine the overall loss in the accuracy of prediction. This can be done with the following stages:

- In the first stage, the model is trained with 4 years (1997-2000) of seasonally adjusted data from data set 1 to generate 2 years (2001-2002) of prediction.
- Then, the 2-year data is added to the training set, making a total of 6-year (1997-2002) training data to generate the next 3 years' data (2003-2005).
- Finally, the result from stage 2 is then added to the training data set to produce 5 years (2005-2010) of prediction. In stage 3, five years predictions are made instead of four-and-a-half years because generating a five-

Table 3: Seasonally adjusted model results for wave height and period

Data set size	Parameters	1-year Prediction			2-year Prediction			3-year Prediction		
		d	R	RMSE	d	R	RMSE	d	R	RMSE
4 years	Wave height	0.75	0.63	0.47	0.84	0.73	0.39	0.75	0.63	0.47
	Wave period	0.75	0.61	1.30	0.77	0.63	1.2	0.72	0.57	1.32
6-years	Wave height	0.78	0.66	0.47	0.79	0.68	0.46	0.81	0.71	0.45
	Wave period	0.73	0.60	1.32	0.74	0.61	1.24	0.77	0.65	1.19

year prediction will cover the entire data set.

Table 4 shows the prediction accuracy for each stage based on the error criteria used in this study. It can be observed that during the first stage where the model used four years of seasonally adjusted data without any addition of predicted data, the model produces a slightly higher quality prediction for both wave height and wave period. But the overall loss in accuracy is very minimum, which can be seen in stage 3, where the training data set consists of 4 years of actual data with seasonal adjustment and 5 years of prediction produced in stages 1 and 2. This indicates that the add-and-repeat method is capable of producing significantly accurate results without much loss in accuracy.

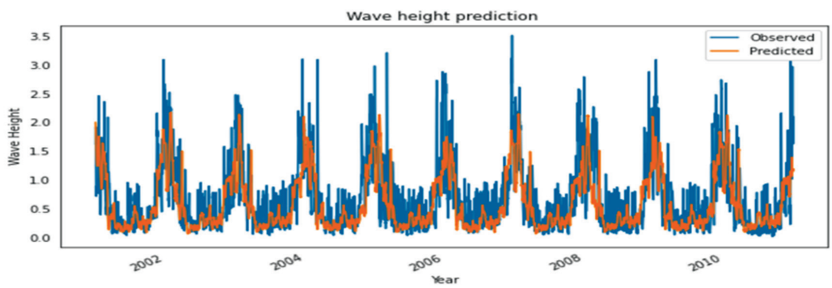
The results of the add-and-repeat prediction method are also shown in Figure 14 with the help of time series plots. The time series plots indicate that the model can accurately predict the trend in data and is also able to follow non-linear spikes in data. Since the model prediction was based on the trend of data without any seasonality and noise, the model prediction is also much smoother, representing the general trend in data to a very accurate extent.

Based on the accuracy of the results (Figure 14 and Table 4), the analysis was extended to generate long-term wave height and wave period prediction for the year 2050 using the same method. During the first run, the entire 14 years of seasonally adjusted data is used to train the model and produce a prediction for the next seven years (2011-2017), followed by the combination of predicted results with actual data and producing further prediction that is equal to half of the training data set size. This process is repeated until the prediction for 2050

Table 4: The add-and-repeat prediction result of wave height and period using the seasonally adjusted model

Parameters	**Stage 1**			**Stage 2**			**Stage 3**		
	d	**R**	**RMSE**	**d**	**R**	**RMSE**	**d**	**R**	**RMSE**
Wave height	0.84	0.73	0.39	0.75	0.63	0.47	0.79	0.69	0.44
Wave period	0.77	0.63	1.20	0.74	0.6	1.31	0.74	0.61	1.32

(a)



(b)

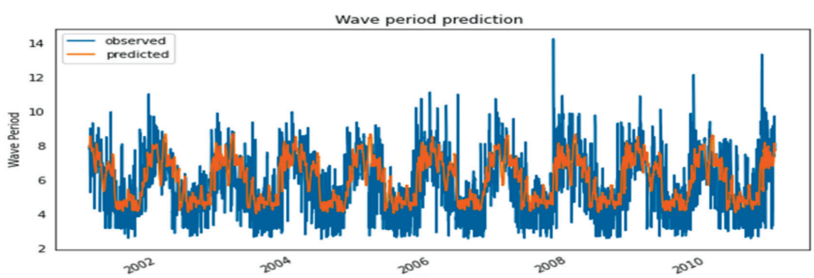
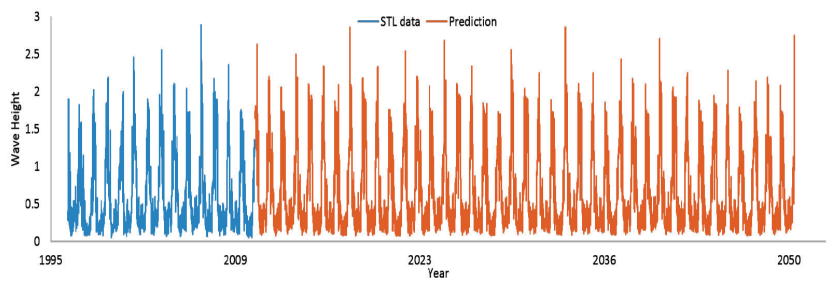


Figure 14: The add-and-repeat prediction results of data set 1. (a) The time series plots for observed and predicted wave height, and (b) the time series plots for observed and predicted wave period

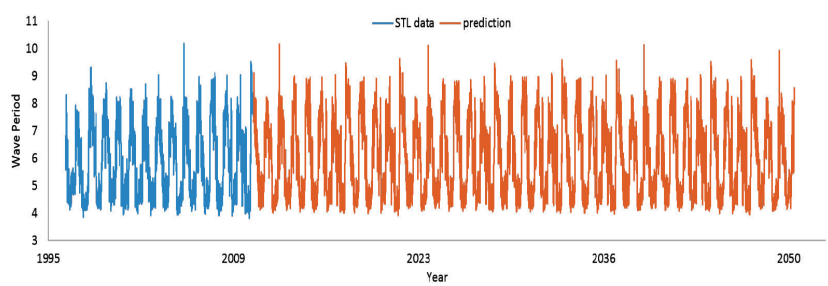
is achieved. The results of wave height and wave period prediction for the year 2050 are shown in Figure 15, along with actual seasonal adjusted data represented by the blue line. It can be observed that the general trend in actual data is similar to that produced by the model. The interval of non-linear spikes is also almost at the same interval, suggesting that the model has high accuracy.

Conclusion

The current research is designed to establish the effect of seasonal adjustment through STL decomposition as a means of improving ANN wave-forecasting capabilities. To the bset of the authors' knowledge, this study is the first attempt to improve wave parameter predictions through seasonal adjustment of data. The results show a good increment in prediction quality when seasonal effects and noise are removed from wave data. The research further provides a new method to generate long-term wave height and wave period data through the proposed add-and-repeat prediction method, where prediction from the previous step is added to the original data to provide a successive prediction. Both methods tested in this study provided satisfactory results when evaluated through the error difference criteria. However, there is room for improvements and future studies can focus on predicting wave height and wave period with a combination of different input parameters, such as wave direction, atmospheric pressure, atmospheric temperature, wind speed, and wind duration, to establish an ideal combination of parameters. Furthermore, research can be done using different types of machine learning methods with different architectures and combinations of hyperparameters. Additionally, a different technique, like X-11, or software packages, like X-13ARIMEA-SEATS(X-13), for seasonal adjustment could be adopted to determine the effects on prediction quality.



(a)



(b)

Figure 15: Prediction for the year 2050 using the add-and-repeat prediction method and seasonally adjusted data. (a) Time series plots for observed (STL data) and predicted wave height, and (a) time series plots for observed (STL data) and predicted wave period

Acknowledgements

The author would like to acknowledge the Department of Irrigation and Drainage Malaysia for providing the data sets.

References

Agrawal, J. D., & Deo, M. C. (2002). Online wave prediction. *Marine Structures*, *15*(1), 57-74. https://doi.org/10.1016/S0951-8339(01)00014-4

Akbarifard, S., & Radmanesh, F. (2018). Predicting sea wave height using Symbiotic Organisms Search (SOS) algorithm. *Ocean Engineering*, *167*, 348-356. https://doi.org/10.1016/j.oceaneng.2018.04.092

Altunkaynak, A., & Wang, K.-H. (2012). Estimation of significant wave height in shallow lakes using expert system techniques. *Expert Systems with Applications*, *39*(3), 2549-2559. https://doi.org/10.1016/j.eswa.2011.08.106

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70-83. https://doi.org/10.1016/j.csda.2017.11.003

Buizza, R., Brönnimann, S., Haimberger, L., Laoyaux, P., Martin, M. J., Fuentes, M., Alonso-balmaseda, M., Becker, A., Blaschek, M., Dahlgren, P., Boisseson, E. De, Dee, D., Doutriaux-boucher, M., Feng, X., John, V. O., Haines, K., Jourdain, S., Kosaka, Y., & Lea, D. (2018). Contribution to advancing science and production of earth system climate. *Bulletion of the American Meteorological Society*, *99*(5), 1003-1014. https://doi.org/10.1175/BAMS-D-17-0199.1

Burman, J. P. (1980). Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society. Series A (General)*, *143*(3), 321. https://doi.org/10.2307/2982132

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of Gated Recurrent Neural Networks on Sequence Modeling. http://arxiv.org/abs/1412.3555

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on Leoss. *Journal of Official Statistics*, *6*(1), 3-73. https://doi.org/10.1007/978-1-4613-4499-5_24

Deo, M. C., Jha, A., Chaphekar, A. S., & Ravikant, K. (2001). Neural networks for wave forecasting. *Ocean Engineering*, *28*(7), 889-898. https://doi.org/10.1016/S0029-8018(00)00027-5

Deo, M. C., & Sridhar Naidu, C. (1998). Real-time wave forecasting using neural networks. *Ocean Engineering*, *26*(3), 191-203. https://doi.org/10.1016/S0029-8018(97)10025-7

Eren, L., Ince, T., & Kiranyaz, S. (2019). A Generic Intelligent Bearing Fault Diagnosis System using Compact Adaptive 1D CNN Classifier. *Journal of Signal Processing Systems*, *91*(2), 179-189. https://doi.org/10.1007/s11265-018-1378-3

Etemad-Shahidi, A., & Mahjoobi, J. (2009). Comparison between M5′ model tree and neural networks for prediction of significant wave height in Lake Superior. *Ocean Engineering*, *36*(15-16), 1175-1181. https://doi.org/10.1016/j.oceaneng.2009.08.008

Fan, S., Xiao, N., & Dong, S. (2020). A novel model to predict significant wave height based on a long short-term memory network. *Ocean Engineering*, *205*(December 2019), 107298. https://doi.org/10.1016/j.oceaneng.2020.107298

Gardner, E. S., & McKenzie, E. (1989). Note—Seasonal exponential smoothing with damped trends. *Management Science*, *35*(3), 372-376. https://doi.org/10.1287/mnsc.35.3.372

Harbola, S., & Coors, V. (2019). One-dimensional convolutional neural network architectures for wind prediction. *Energy Conversion and Management*, *195*, 70-75. https://doi.org/10.1016/j.enconman.2019.05.007

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, S., Tang, J., Dai, J., Wang, Y., & Dong, J. (2020). 1DCNN Fault Diagnosis based on cubic spline interpolation pooling. *Shock and Vibration*, *2020*, 1-13. https://doi.org/10.1155/2020/1949863

Jain, P., & Deo, M. C. (2006). Neural networks in ocean engineering. *Ships and Offshore Structures*, *1*(1), 25-35. https://doi.org/10.1533/saos.2004.0005

Jain, Pooja, Deo, M. C., Latha, G., & Rajendran, V. (2011). Real-time wave forecasting using wind time history and numerical model. *Ocean Modelling*, *36*(1-2), 26-39. https://doi.org/10.1016/j.ocemod.2010.07.006

Jeon, B., Park, N., & Bang, S. (2020). *Dropout Prediction over Weeks in MOOCs via Interpretable Multi-Layer Representation Learning*. http://arxiv.org/abs/2002.01598

Kim, Y. (2014). Convolutional neural networks for sentence classification. EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 1746-1751. https://doi.org/10.3115/v1/d14-1181

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90. https://doi.org/10.1145/3065386

Kuhn, M., & Johnson, K. (2013). Applied predictive modelling. In *Applied Predictive Modeling*. https://doi.org/10.1007/978-1-4614-6849-3

Kumar, N. K., Savitha, R., & Mamun, A. Al. (2017). Regional ocean wave height prediction using sequential learning neural networks. *Ocean Engineering*, 129(April 2016), 605-612. https://doi.org/10.1016/j.oceaneng.2016.10.033

Luijendijk, A., Hagenaars, G., Ranasinghe, R., Baart, F., Donchyts, G., & Aarninkhof, S. (2018). The state of the world's beaches. *Scientific Reports*, *8*(6641), 1-11. https://doi.org/10.1038/s41598-018-24630-6

Lv, M., Xu, W., & Chen, T. (2019). A hybrid deep convolutional and recurrent neural network for complex activity recognition using multimodal sensors. *Neurocomputing*, 362, 33-40. https://doi.org/10.1016/j.neucom.2019.06.051

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*(2), 111-153. https://doi.org/10.1002/for.3980010202

Malekmohamadi, I., Bazargan-Lari, M. R., Kerachian, R., Nikoo, M. R., & Fallahnia, M. (2011). Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction. *Ocean Engineering*, *38*(2-3), 487-497. https://doi.org/10.1016/j.oceaneng.2010.11.020

Mandal, S., & Prabaharan, N. (2006). Ocean wave forecasting using recurrent neural networks. *Ocean Engineering*, *33*(10), 1401-1410. https://doi.org/10.1016/j.oceaneng.2005.08.007

Markham, I. S., & Rakes, T. R. (1998). The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression. *Computers and Operations Research*, *25*(4), 251-263. https://doi.org/10.1016/S0305-0548(97)00074-9

Mohanasundaram, S., Kumar, G. S., & Narasimhan, B. (2019). A novel deseasonalized time series model with an improved seasonal estimate for groundwater level predictions. *H2Open Journal*, *2*(1), 25-44. https://doi.org/10.2166/H2OJ.2019.022

Nelson, M., Hill, T., Remus, W., & O'Connor, M. (1999). Time series forecasting using neural networks: should the data be deseasonalized first? *Journal of Forecasting*, *18*(5), 359-367. https://doi.org/10.1002/(SICI)1099-131X(199909)18:5<359::AID-FOR746>3.0.CO;2-P

Nitsure, S. P., Londhe, S. N., & Khare, K. C. (2012). Wave forecasts using wind information and genetic programming. *Ocean Engineering*, *54*, 61-69. https://doi.org/10.1016/j.oceaneng.2012.07.017

O'Donncha, F., Zhang, Y., Chen, B., & James, S. C. (2019). Ensemble model aggregation using a computationally lightweight machine-learning model to forecast ocean waves. *Journal of Marine Systems*, *199*(June), 103206. https://doi.org/10.1016/j.jmarsys.2019.103206

Ollech, D. (2018). Seasonal adjustment of short time series. *Deutsche Bundesbank*, *41*.

Pereira, H. R., Meschiatti, M. C., Pires, R. C. de M., & Blain, G. C. (2018). On the performance of three indices of agreement: An easy-to-use r-code for calculating the Willmott indices. *Bragantia*, *77*(2), 394-403. https://doi.org/10.1590/1678-4499.2017054

Ranasinghe, R. (2016). Assessing climate change impacts on open sandy coasts: A review. *Earth-Science Reviews*, *160*, 320-332. https://doi.org/10.1016/j.earscirev.2016.07.011

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195-204. https://doi.org/10.1038/s41586-019-0912-1

R ibal, A., & Young, I. R. (2019). 33 years of globally calibrated wave height and wind speed data based on altimeter observations. *Scientific Data*, *6*(1), 1-15. https://doi.org/10.1038/s41597-019-0083-9

Wang, X., Zhou, J.-F., Wang, Z., & You, Y.-X. (2018). A numerical and experimental study of internal solitary wave loads on semi-submersible platforms. *Ocean Engineering*, *150*, 298-308. https://doi.org/10.1016/j.oceaneng.2017.12.042

Wang, Z., Yan, W., & Oates, T. (2017). Time series classification from scratch with deep neural networks: A strong baseline. *2017 International Joint Conference on Neural Networks (IJCNN)*, 1578-1585. https://doi.org/10.1109/IJCNN.2017.7966039

Wei, C.-C., & Hsieh, C.-J. (2018). Using Adjacent Buoy Information to Predict Wave Heights of Typhoons Offshore of North-eastern Taiwan. *Water*, *10*(12), 1800. https://doi.org/10.3390/w10121800

Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, *160*(2), 501-514. https://doi.org/10.1016/j.ejor.2003.08.037

Zhao, J., Mao, X., & Chen, L. (2018). Learning deep features to recognise speech emotion using merged deep CNN. *IET Signal Processing*, *12*(6), 713-721. https://doi.org/10.1049/iet-spr.2017.0320