https://doi.org/10.4038/sljastats.v24i2.8098

Sri Lankan Journal of Applied Statistics, Vol (24-2)

Prediction of Tidal Elevations at Eastern and Western Coastal Areas of Sri Lanka with Short-term Data

J.A.R.M. Perera*, P.A.D.A.N. Appuhamy, and E.M.P. Ekanayake

Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka.

*Corresponding author: malsharashini96@gmail.com

Received: 30th June 2023/ Revised: 5th September 2023/ Published: 30th September 2023

ABSTRACT

Prediction of tidal heights are increasingly beneficial for multitude of ocean functions such as port development, fishing industry, and safe movement of ships. As the general harmonic technique always needed great volumes of data for predicting tidal heights, Artificial Neural Networks (ANNs) emerged as a viable alternative for addressing diverse problems in the coastal engineering sector in recent decades. However, there has been no previous research to distinguish Harmonic Analysis from ANN models for predicting tidal heights around Sri Lanka by overcoming the rampant issue of data scarcity, which is the focus of the present study. Hourly tidal heights recorded in the Western (Colombo) and Eastern (Trincomalee) coastal areas of Sri Lanka were used in modelling. As tidal elevation is periodic in nature, it was expressed as Fourier Series with its coefficients (constituents) being determined by Harmonic Analysis, while the ANN technique employed the back-propagation procedure to forecast tidal heights. Harmonic Analysis displayed lesser prediction performance even with five months of data at Colombo (MSE=0.030 and MAPE=1.875) and Trincomalee (MSE=0.019 and MAPE=1.052), in contrast to the ANN models with only 7 days of data, which has much lower MSE and MAPE at Colombo (0.006 and 0.096) and (0.003 and 0.052) at Trincomalee respectively. Thus, the ANN model outperformed the Harmonic Analysis in terms of both accuracy and flexibility. Overall, this study demonstrated the potential of ANN modeling as a reliable, economical, and efficient alternative for predicting tidal heights to circumvent the dearth of tidal data on the

coastal Sri Lanka.

Keywords: Tide, Harmonic Analysis, Artificial Neural Network, Back Propagation.

1 Introduction

Tides can be classified into two major types viz. atmospheric tides and ocean tides of which atmospheric tides are global scale daily oscillations that are primarily driven by diurnal fluctuations in temperature caused by solar energy absorption in atmospheric ozone and water vapor. Movements of atmospheric tides have been studied extensively over the recent decades (e.g. Ekanayake et al., 1997) in terms of migrating and nonmigrating components with the sun. Ocean tides are influenced by the gravitational pull of the Moon and, to a lesser extent, by the gravity of the Sun. In some aspects, atmospheric tides are analogous to ocean tides. This study is focused on ocean tides because changing water levels can dramatically affect people, animals, plants, climate, coastal maritime operations, etc. Tides are responsible for the fluctuation of water levels and ocean tides can be described in general as stemming from gravitational interactions among the Sun, Moon, and Earth which cause a body of water to rise and fall on a regular basis. When tides change, vast volumes of water travel to or away from the shore creating tidal currents. There are two main ocean tides called high tide and low tide. As Sri Lanka is an equatorial country, most of its coastal areas have two high tides and two low tides during the day and night which can clearly be observed on every lunar day. This is because the Moon creates a much larger impact on the tides than the Sun does as it has a much lesser distance to Earth. The difference between high and low tides is called the tidal range.

As the relative locations of the Earth, Sun, and Moon change, so do the gravitational forces between them. The rising and setting of the Sun and Moon, the changing phases of the Moon, and the changing seasons of the year all reveal these positional changes. Each of these shifts is cyclical, recurs over time, and has a quantifiable impact on the tides seen along the ocean's edge. There are hundreds of periodic motions of the Earth, Sun, and the Moon that are identified by astronomy and each of these motions is called a constituent or a Harmonic constituent. A constituent has a measurable effect that describes how the Earth's, Sun's, and the Moon's cyclical motions affect the generation of ocean tides. According to Foreman's work in 1978, there are 69 standard tidal harmonic constituents that can be used to predict tidal generation at a particular site. These constituents include the major tidal constituents such as M2 (Principal lunar semi-diurnal), S2 (Principal solar semi-diurnal), N2 (Larger lunar elliptic semi-diurnal), K2 (Luni-solar semi-diurnal), K1 (Luni-solar declinational diurnal), O1 (Lunar declinational diurnal), P1 (Principal Solar diurnal), and Q1 (Larger lunar elliptic diurnal), as well as many minor constituents that contribute to the overall tidal pattern.

Tidal elevation can be predicted well in advance with a high degree of accuracy, which is different for each location. Lee & Jeng, (2002) proposed a new methodology for predicting tides using Artificial Neural Networks (ANNs) that were identified as promising tools for predicting tides, with a high degree of accuracy relative to traditional prediction methods. Further, it was suggested that ANNs could be used to predict tidal heights at diverse locations. However, they did not show any comparison between the results obtained from ANNs and those from the Harmonic Analysis. Traditionally, tidal heights used to be predicted using Harmonic Analysis, a technique that decomposes a tidal signal into its constituent harmonics forming Fourier series, which requires large volume of data to predict the tidal constituent accurately. However, with the advent of machine learning, Artificial Neural Network (ANN) modeling has emerged as a promising alternative for predicting tidal heights. Artificial Neural Networks (ANNs) are commonly used for modeling nonlinear dynamic systems including time series that depict physical characteristics of the relationship between input variables and the phenomenon to be predicted or estimated (output). Lee & Jeng (2002) further showed that ANN models could be used to predict hourly tidal levels over a long period of time based on a short-term hourly tidal record. More specifically, they predicted hourly tidal levels over a month using a day's data and over three months with half a month of observed data on the ANN model.

Predicted tidal heights are always important for people who look to the sea for their livelihood and for those involved in constructions in the ocean. Especially for fishermen who live in those areas, knowledge of predicted tidal heights can help manage their personal economy and use their time more productively. Moreover, for recreational beach-goers and surfers, knowledge of the predicted tides has also been helpful in looking after their safety. In any situation where there was a need to know the water level, the mariner typically used the tidal forecast, even in place of the actual observation of the water level. The most frequent use was for navigation, in particular, as having adequate water depth under its keel is important for a deep-draft vessel to ensure that it does not run aground. Tidal predictions are also used in emergency management and planning coastal engineering work, because projects must be scheduled well in advance if the area experiences large fluctuations in water levels during its tidal cycle. In the case of tsunami warning systems, observations at the water level station during a tsunami event are detected using predicted tides so that the signature of the tsunami wave can be more clearly seen in the data record. Therefore, it is very important to have a good knowledge of tidal predictions.

ANN is a deep learning method based on the concept of human brain's biological neural networks and its birth was the result of an attempt to imitate the functioning of the human brain. Forward propagation and back-propagation are the two phases in ANN, of which back-propagation is the most crucial as it entails determining optimal parameters for the model by propagating the neural network layers in the backward manner. Though tidal prediction models were mainly based on harmonic analysis in the past, the present researchers tend to use ANN models due to lack of data for analysis (Lee & Jeng, 2002; Meena & Agrawal, 2015) and the other concerns like simplicity of structure, acceptable performance (Bhara & Bakshi,2020), and effectiveness in modeling nonlinear systems, for instance time series.

This study shall compare the effectiveness of Harmonic Analysis and ANN modeling in predicting tidal heights when limited data are available for modeling. As Sri Lanka still face the challenge of collecting data for extended periods of time regularly, an economically feasible method for accurate prediction of tidal heights using small amounts of data would prove to be very significant. In the process, tidal data from two coastal stations were used in both methods and the results were evaluated using the Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) values. The study also investigated the flexibility of the methods in adapting to changes in the tidal signal. The comparison of these two methods will provide insight into their respective strengths and limitations, and help identify the most appropriate method for predicting tidal heights in different coastal areas.

Meena & Agrawal (2015) constructed an ANN model to predict tidal levels with different learning algorithms [e.g. Feed-Forward Back Propagation (FFBP) network with Levenberg-Marquardt (LM), Conjugate Gradient Fletcher Reeves update (CGF), and Broydan-Fletcher-Goldfarb Shanno (BFG)] by changing the number of neurons in its hidden layer and iterations, using limited measured data as an alternative to conventional harmonic analysis. They showed that ANN is a good technique for short-term prediction of tidal time series with site specific models. Lee & Jeng (2002) adopted the ANN model using back propagation with gradient descent algorithm for predicting the tidal level while applying the concept of the auto-regressive moving average (ARMA) into the ANN model. Lee (2004) presented an application of the back-propagation neural network using short-term measured data for long-term tidal predictions at Taichung Harbor in Taiwan and he concluded that Back Propagation Neural Network (BPN) for one-year tidal level prediction can be performed satisfactorily with fifteen days of observed data. The Feed Forward Back Propagation (FFBP) and Non-linear Auto Regressive with eXogenous input (NARX) network were used by Salim et al., (2015) to predict yearlong hourly tidal levels at Mangalore, Karnataka, using a week's hourly tidal levels as input and they showed that NARX network outperformed FFBP network in terms of data requirement, accuracy of predictions, and computational time. Overcoming the difficulty in finding a large volume of data for conventional harmonic analysis, Mandal (2001) used a BPN model with quickprop algorithm to predict time series data of hourly tides for a month at Gopalpur Port on the east coast of India with a greater accuracy indicated by a higher correlation coefficient. Lee et al., (2006) also applied BPN but with gradient descent algorithm to predict long term semi-diurnal tidal levels and concluded that the other distinct tidal types viz. diurnal and mixed, can be obtained for one year with fifteen days of observed data. In another application of BPN on long-term and short-term measured data, Jorge & Eduardo, (2009) assessed the model performance in comparison with harmonic method using onsite tidal level data at Ingeniero White harbor in Argentina and proved the dexterity of BPN in forecasting long term tidal levels with even half-a-month observed data.

Table 1 presents a summary of the most relevant work of some researchers who have used harmonic analysis and ANN architectures with different algorithms and performance indicators.According to Table 1, there is only one evidence-based study which compared the prediction accuracy of tidal elevations obtained from both ANN and harmonic methods of which the ANN technique was favored to produce better results. Moreover, from other studies too it can be understood that, for short temporal scales of data, ANN has performed with higher accuracy than the traditional Harmonic Analysis would provide based on long term records. However, evidence of comparison between the two methods is not presented in the above literature.

In Harmonic analysis, model accuracy entirely depends on the precision of observed data over a long-term tidal record, which is used to determine the coefficients of tidal constituents.This is the major disadvantage in using harmonic models, because gathering large volumes of reliable data, particularly in developing countries like Sri Lanka, necessitates the use of expensive devices over an extended period of time. In this study, we shall investigate spatial and temporal properties of tidal elevations and predicted the tidal heights using both traditional harmonic analysis and the ANN with BPN method.The accuracy of predictions made with limited tidal measurements at both eastern and western coastal areas of Sri Lanka shall be compared in terms of the MSE and MAPE values. Having observed that some researchers had resorted to BPN without a hidden layer to find major tidal constituents, R-Studio was utilized as a tool in this study to determine the significant constituents of hourly tidal heights. By applying both methods and evaluating their performance, the primary objective of the present work is to determine the most effective approach for predicting tidal heights at the two locations.

The results of the study could have significant implications for improving our understanding of tidal forecasting, which would be useful for various coastal and marine applications including navigation, resource management, and environmental monitoring. To the best of our knowledge, there are no previous studies on tidal prediction using both harmonic analysis and Artificial Neural

Table1: A summary of the most relevant studies for tidal predictions and their performances

Reference	ANN Architecture	Performance Measure	Results
Lee, T. L., & Jeng, D. S. (2002)	BPN network	Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), correlation coefficient (r)	The ANN was able to accurately predict tides up to 12 hours in advance, with RMSE values ranging from 0.13 to 0.22 meters, MAE values ranging from 0.09 to 0.16 meters, and correlation coefficients ranging from 0.97 to 0.99. Further they found that adding more hidden layers do not improve the accuracy of tidal prediction. The ANN outperformed traditional harmonic analysis methods for tide forecasting.
Meena & Agrawal (2015)	Feed-Forward Back Propagation (FFBP) network with Levenberg-Marquardt (LM), Conjugate Gradient Fletcher Reeves update (CGF), and Broydan-Fletcher-Goldfarb Shanno (BFG)	Root Mean Squared Error (RMSE), Mean Absolute Error (MAE)	The best performance was achieved by the FFBP network with LM, with an RMSE of 0.06 and an MAE of 0.04.
Lee, T.L., Tsai, C.P., and Shieh, R.J. (2006)	FFBP network	RMSE, MAE, R^2	Half month-The FFBP network achieved an RMSE of 0.07, an MAE of 0.03, and an R^2 of 0.975.
Lee, Tsong Lin (2004)	FFBP network	RMSE, MAE, Correlation Coefficient (CC)	15-days-The FFBP network achieved an RMSE of 0.05 m, an MAE of 0.03, and a CC of 0.97 for tidal level prediction, and an RMSE of 0.05, an MAE of 0.03, and a CC of 0.95 for tidal current velocity prediction.
Mandal, S. (2001)	BPN model with quickprop algorithm	RMSE, Correlation Coefficient (CC)	1 day-The BPN network achieved an RMSE of 0.0182, and CC of 0.9984.

Network (ANN) models covering the coastal areas of Sri Lanka. Therefore, this work could bridge that research gap by identifying the best statistical approach to predict tidal heights at two of the most dynamic Sri Lankan regions in terms of coastal activities, with limited data. Further, the findings shall contribute to enhance the understanding of tidal forecasting and provide valuable insights for future research in this field.

2 Methodology

2.1 Data Collection

Data sets of tidal heights at the two sites of Colombo and Trincomalee were collected from the National Aquatic Resources Research and Development Agency (NARA) in Sri Lanka for this study. Hourly tidal heights spanning from September 1, 2020 to January 31, 2021 were collected for Colombo and for Trincomalee. This specific period was chosen to mitigate the problem of intermittent missing data caused by instrumental errors and interruptions. As such, only a minimal number of data points were found missing during the selected period, which were imputed with the mean of the two neighboring values by considering the wave-like pattern characteristic of tidal data. Statistical software R-Studio (2021.08.0) and Python (3.9.4) were used for analyzing and predicting the tidal heights.

2.2 Distributional properties of tidal heights

Initially, the graphical and descriptive statistics of mean, median, variance, maximum, minimum, kurtosis, and skewness of tidal heights were examined to identify the distributional properties of hourly tidal heights at Colombo and Trincomalee. Moreover, tidal type at each location was identified using the value of the Form Number introduced by Dietrich (1963), which uses the amplitudes of major constituents (M2, S2, K1 and O1) at each location. The calculation and classification of tides into diurnal, semi-diurnal, and mixed constituents are based on the following Equation (1) of the Form Number.

$$F = \frac{K_1 + O_1}{M_2 + S_2} \tag{1}$$

where the tide is called,

- semidiurnal if $0 \leq F \leq 0.25$,
- mixed if $0.25 < F \leq 3.00$,
- diurnal if $F > 3.00$.

2.3 Significant Tidal Constituents

A thorough understanding of tidal constituents is essential not only for theoretical research and improving astronomical tidal predictions, but also for a more accurate assessment of variations in sea level caused by meteorological factors. A tidal signal is a combination of harmonic constituents. In this study, the 69 tidal constituents were obtained using R Studio open-source software with tidem building function in oce package with 1-month, 2-months, 3-months, 4-months and 5-months of hourly tidal measurements at both locations. Then, their p-values were considered to identify the significant constituents for each time period. The shortest of the periods which accommodated the constituents significant at 1% level was chosen so that all such constituents served as the main constituents at each location to proceed for tidal prediction under both techniques.

2.4 Harmonic Analysis

The harmonic analysis method is traditionally used for tidal prediction. It is a mathematical approach for analyzing a periodically recurrent nature of a phenomenon. In Mathematics, periodic function can be expressed as sum of the sine and cosine terms which forms a Fourier series introduced by Fourier in 1822. As tides may periodic in nature it could be considered as a periodic function which can therefore be expressed as a Fourier series as defined by Equation (2).

$$f(x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx) \tag{2}$$

where, $a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kxdx$ and $b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kxdx$, for $k \geq 0$.

Provided the function is single-valued, finite, and continuous except for a finite number of discontinuities and the terms $a_1 \cos\ x + b_1 \sin\ x, a_2 \cos\ 2x + b_2 \sin\ 2x, a_3 \cos\ 3x + b_3 \sin\ 3x$ and etc are called as the fundamental, second harmonic, third harmonic and etc respectively.

The process of determination of the coefficients of these sine and cosine terms are said to be harmonic analysis. Then these estimated coefficients can be used to predict tidal elevation at a particular location. The accurate estimation of harmonic constituents for tidal prediction at a particular location through conventional methods such as harmonic analysis requires more than one year of hourly data (Foreman and Neufeld, 1991; Reid, 1990; Lee, 2004). The accuracy of tidal prediction through traditional methods can be increased with a higher number of harmonic constituents, however, this increases computation time and requires more data. Some researchers including Foreman (1998) used only the standard constituents (69) which are sufficient for marine tidal prediction.

In relation to the tide, the vertical tidal level at any time t, $Y(t)$, at any position is expressed in terms of a sum of harmonic terms as follows:

$$Y(t) = A_0 + \sum_{i=1}^{N} h_i \cos(\omega_i t + \epsilon_i) \tag{3}$$

where, A_0 is the mean height of water level used for prediction, N is the total number of constituents, and $h_i, \omega_i, \epsilon_i$ are the amplitude, speed, and phase of the ith constituent respectively. In general, determination of the total number of constituents using power spectral analysis depends on long-term tidal level records (more than one year), as reported by Reid (1990).

For simplification ϵ_i is often assumed as zero and Equation (3) can also be expressed using a standard trigonometric identity:

$$Y(t) = A_0 + \sum_{i=1}^{N}(A_i \cos\omega_i t + B_i \sin\omega_i t) \tag{4}$$

in which $A_i = \frac{\sum_{i=1}^{m} \eta \sin \omega_i t}{m}$ and $B_i = \frac{\sum_{i=1}^{m} \eta \cos \omega_i t}{m}$ are the coefficients of constituents satisfying the relation $h_i = \sqrt{A_i^2 + B_i^2}$ and $\epsilon_i = \tan^{-1} \frac{B_i}{A_i}$. Further, η– tidal water level at a particular time t, ω_i- angular frequency at a particular time t, m- number of observations in the training set, h_i- amplitude, and ϵ_i- phase angle. In addition, the parameters A_i and B_i could be conventionally derived from a long duration (more than one month) of tidal measurements, before predicting tides with the least squares method. Accurate predictions of tidal levels could be obtained with an adequate number of constituents. Selecting more constituents would result in much accurate predictions of tidal levels but it will complicate operation problems such as growing memory, calculation time and requires long-term data. In order to avoid the necessity of long-term data to find the adequate number of constituents for accurate tidal prediction under harmonic analysis, only the major constituents, which were determined using the oce package mentioned above, were considered in this study. Once the significant tidal constituents are determined, the A_i and B_i values shall be computed first using the training data (defined under ANN technique below) separately for each different time period (last 7days;10 and 15 days; 1 month; 2, 3, 4, and 5 months). Thereafter, the mean of tidal heights A_0 for each training set shall be found separately. Eventually, those values would be substituted into Equation (4) to find tidal predictions, and the MSE and MAPE values for each time period were calculated for the testing data using Equations (5) and (6).

$$\text{MSE} = \frac{\sum_{k=1}^{n} (y_k - \hat{y}_k)^2}{n} \tag{5}$$

$$\text{MAPE} = \frac{\left| \sum_{k=1}^{n} \frac{y_k - \hat{y}_k}{y_k} \right|}{n} 100\% \tag{6}$$

where, y_k is the actual tidal height, $\hat{y}_k$ is the predicted height and n is the number of observations.

2.5 ANN to establish a tidal prediction model

Upon constructing the tidal prediction model by Harmonic analysis, a distinct tidal prediction model was developed utilizing an Artificial Neural Network, details of which are explained in the following sub-section.

2.5.1 Constructing the Artificial Neural Network (ANN) for tidal elevation

In this study, an ANN was used next to model tidal elevation pattern at each location and thereby to predict tidal heights at the corresponding venues. In order to find the optimal artificial neural networks for tidal heights at Colombo and Trincomalee, the Stochastic Gradient Descent with Back-propagation method developed by Rumelhart et al. (1986) was used. In BPN, the error at the output layer propagates backwards to the input layer via the network's hidden layer to produce the final output. The gradient descent approach calculates the weight of the network and fine-tunes the weight of interconnected neurons to minimize the output error, which is illustrated in Figure 1.

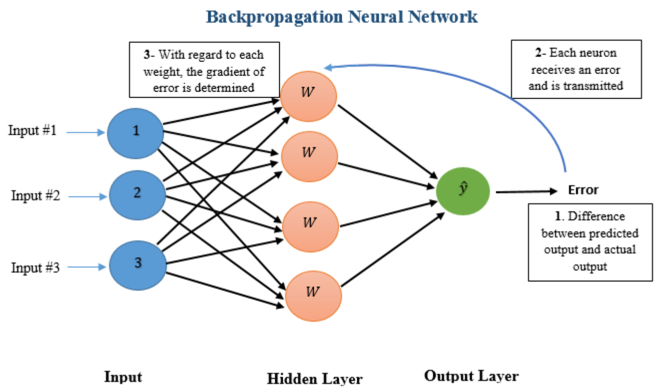

Figure 1: An illustration of the Backpropagation process

Thus, back propagation neural networks were constructed for hourly tidal heights for different time periods (last 7 days; 10 and 15 days; 1 month; 2, 3, 4, and 5 months) considered under harmonic analysis method, for identifying the optimum network with minimum data requirement by avoiding the data collection issue encountered in considering long periods. Moreover, the hourly data for each time period at each location was apportioned into 70%, 15% and 15% for training, validating and testing respectively. The Non-linear Autoregressive neural network (NAR) can be used as a prediction method in time series when the series is nonlinear and non-stationary. However, in this study, NAR was not employed to model tidal elevation, as similar studies could be found in literature, for instance, the work by Lee & Jeng, (2002). In order to enhance the novelty of the present study, sine and cosine terms of the significant constituents were considered as inputs to the neural network.

The learning rate η and momentum factor α are significant parameters in minimizing the network error. The present study started with moderate learning rates and momentum values as too high values can lead to overshooting and instability, while too low values can slow down convergence. During training, the random search technique was employed to expedite the process of fine-tuning these parameters and consequently determining the optimal values for the hyperparameters. Further, the common activation function called the Sigmoid function, $f(x) = (1 + e^{-x})^{-1}$, which has the characteristic of $\frac{df}{dx} = f(x)[1 - f(x)]$, was considered as the activation function for hidden neurons and the linear function, $f(x) = x$, was applied as the activation function for the output neurons which produced the output from inputs through an appropriate transformation. Once the training and validation are over, optimal network structure for each data set was tested for prediction accuracy with the allocated test data (testing data) in terms of the same performance indicators used in harmonic analysis viz. MSE and the MAPE.

3 Results and Discussion

3.1 Distributional properties of tidal heights

Figure 2 illustrates the behavior of hourly tidal elevation for the one-month period of January 2021 at Colombo (Fig. 2a) and Trincomalee (Fig. 2b) coastal areas.

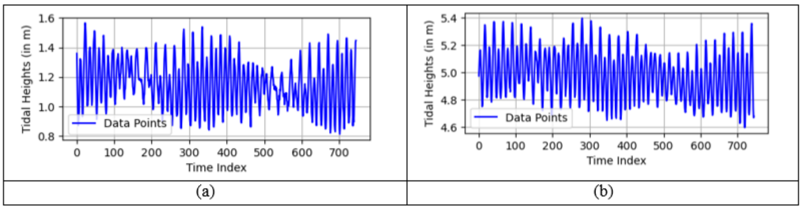

Figure 2: Behavior of the hourly tidal elevations in January 2021 for (a) Colombo and (b) Trincomalee

It can be seen that periodic rise and fall of sea level occur at both locations. This phenomenon involves two high-tides and two low-tides at approximately every 24 hours and 50 minutes. Furthermore, the tidal heights were higher at Trincomalee compared to that at Colombo coastal area.

Table 2: Descriptive statistics of daily tide elevations at Colombo and Trincomalee

Descriptive statistics	Colombo	Trincomalee
Mean	1.0726	4.9376
Variance	0.0352	0.0365
Minimum	0.5450	4.4740
Median	1.0740	4.9320
Maximum	1.6480	5.5340
Skewness	0.02	0.11
Kurtosis	-0.29	-0.61

According to Table 2, which presents descriptive statistics of daily tidal elevations at the two locations of Colombo and Trincomalee, the lowest and the highest daily tidal heights turn out to be 0.545m and 1.648m in Colombo and 4.474m and 5.534m in Trincomalee, respectively. However, the variance indicate that there is only a little difference in the observed values at the two locations. The tidal height at Trincomalee is positively skewed than that at Colombo. Further, there is a substantial difference between the mean tidal elevations at eastern and western coastal areas during the period of study. In general, equatorial countries like Sri Lanka experience semi-diurnal tidal patterns. However, the present study used main tidal constituents to identify the tidal type at each location based on the Form Number. As the Form Numbers at Colombo and Trincomalee were 0.265767 and 0.2472725 respectively, the tidal type at Colombo could be identified as mixed semi-diurnal and that at Trincomalee as semi-diurnal. This implies that the tidal system at Colombo is somewhat different from the standard tidal type in equatorial countries.

3.2 Harmonic analysis and ANN to establish a tidal prediction model

This section presents the results obtained from each method: Harmonic analysis and ANN, in the prediction of tidal elevation at the western and eastern coastal areas of Sri Lanka.

3.2.1 Determination of main tidal constituents

The significant tidal constituents obtained for the 2-month period were considered as minimum length of data required to identify the main tidal constituents at Colombo and Trincomalee as they were common for two other periods longer than this(3-months, 5-months) . The tidal constituents identified with1-month period of hourly data at both locations did not contain a one major tidal constituent observed in the 2-month period. This indicated the necessity of data for at least a 2-month period to determine the main significant constituents. Table 3 summarizes the frequencies and p-values of significant constituents obtained from two months of hourly data at Colombo and Trincomalee respectively.

Table 3: Significant tidal constituents at Colombo and Trincomalee

Constituents	Frequency	P-value (Colombo)	P-value (Trincomalee)
Z0	0	< 2e-16	< 2e-16
MM	0.001512	0.0094	-
MSF	0.002822	0.0003	-
O1	0.038731	1.90E-08	0.00201
K1	0.041781	< 2e-16	0.00014
N2	0.078999	< 2e-16	0.00171
M2	0.080511	< 2e-16	7.70E-08
S2	0.083333	< 2e-16	< 2e-16

It can be seen that there are 8 significant tidal constituents for Colombo at 1% level of significance. However, due to low frequencies of the constituents: Z0, MM, and MSF, their contribution to tidal generation becomes minimal and so the sine and cosine values of them were not considered as inputs to the neural network and for the harmonic analysis. Therefore, only 5 significant tidal constituents viz. O1, K1, N2, M2, and S2 were considered as the main tidal constituents for Colombo. Similarly, there are 6 significant tidal constituents for Trincomalee at 1% level of significance but due to negligible frequency of Z0, its contribution to the tidal generation is marginal and therefore the sine and cosine values of this constituent was also not considered for both techniques for tidal prediction. Thus, only 5 significant tidal constituents viz. O1, K1, N2, M2, and S2, similar to the constituents identified for Colombo, were retained as the main tidal constituents at Trincomalee. This identification reveals that the main constituents for both locations are similar. The main tidal constituents emerged from this study agree well with those identified in past studies too (De Vos el al.,2014).These five constituents were then used to forecast tidal heights by subjecting them to Fourier formula derived in Harmonic analysis for the eight time periods, wherein the sin $(\omega_i t_i)$ and cos $(\omega_i t_i)$ values of each main constituent were calculated separately for each location using relevant frequencies and time, thus making those values as input neurons to the neural network.

3.2.2 Results of the harmonic analysis prediction

The MSEs of tidal predictions obtained from the harmonic analysis is summarized in Table 4.

Table 4: MSEs of tidal predictions from harmonic analysis for Colombo and Trincomalee

	MSE	
Time Periods	Colombo	Trincomalee
7 days	0.072456	0.056345
10 days	0.061356	0.042049
15 days	0.048959	0.041961
1 month	0.044037	0.040775
2 months	0.038030	0.026190
3 months	0.032172	0.022565
4 months	0.031793	0.020345
5 months	0.030195	0.019403

Table 4 shows the performance of tidal prediction of Harmonic analysis for Colombo and Trincomalee in terms of the mean squared error. It can be seen that as the time period gets larger, the MSE values become smaller at both locations. This confirms the fact that in harmonic analysis predictions become more accurate when long term measurable data are available for the calculation. Moreover, the mean squared error obtained for Trincomalee is somewhat smaller compared to that for Colombo for all time periods, due to mixed semi-diurnal pattern that prevails in Colombo. Therefore more data would be required to identify such mixed patterns accurately compared to Trincomalee where the semi-diurnal pattern exhibits.

3.2.3 Effects of neural network structure

This section presents the results derived from varying ANN structures across different time periods for both locations.

3.2.3.1. Effect of the number of hidden layers in training the network The number of neurons in the hidden layer also affects the performance of the ANN. There are usually one or more hidden layers between the input and output layers, where the main data processing activity is carried out. The forecasting performance of BPN is improved for all training sets when the number of hidden neurons increases. In the present study, networks were trained with both a single and two hidden layers to identify the optimal network structure.

Table 5: The effect of the number of hidden layers in training the network for Colombo and Trincomalee

Training set	Number of hidden layers	MSE (Colombo)	MSE (Trincomalee)
7 days	1	0.005952	0.003231
	2	0.004278	0.003984
10 days	1	0.006430	0.002856
	2	0.007236	0.003387
15 days	1	0.015272	0.001182
	2	0.012966	0.001211
1 month	1	0.004089	0.005906
	2	0.004123	0.004678
2 months	1	0.002573	0.009538
	2	0.002765	0.009867
3 months	1	0.003429	0.007613
	2	0.003498	0.007568
4 months	1	0.007170	0.004729
	2	0.007654	0.005678
5 months	1	0.012793	0.008145
	2	0.011689	0.008439

Table 5 shows the performance of the neural network structures in terms of MSE when one and two hidden layers were introduced for each time period (Training set) at Colombo and Trincomalee. It can be seen that adding two layers does not enhance the accuracy of the network except for three time periods at Colombo and two time periods at Trincomalee. Overall, MSE values indicate that one hidden layer yields a better outcome for all time periods at both locations. The input neurons that do not have hidden layers are not mapped with non-linear transformation, resulting in a higher inaccuracy than the input neurons with one or two single hidden layers. Therefore, the performance of networks without hidden layer was not considered for this comparison. A learning rate (η) of 0.01 and a momentum factor (α) of 0.8 were selected to achieve higher performance and convergence during training. In all cases, the number of training iterations was set to 1000.

3.2.3.2. Effect of the number of neurons in the hidden layer and tidal prediction performance of optimum network structure

The number of neurons in the hidden layer also affects the performance of the ANN. During training, this study tested different network structures with 8 different time periods which included one or two hidden layers with the number of neurons varying from 1 to 10. The number of neurons to be included in the hidden layer for each network was determined based on the MSE. The testing data from each time period (actual data) and the predicted values from each network structure for each period were considered to calculate the MSE. The network with the minimum MSE was used to determine the number of hidden neurons and the prediction performance of each optimal network for Colombo and Trincomalee (Table 6). It showed that, in both locations, the forecasting performance of BPN improved for all training sets with more neurons being included in the hidden layer. However, if overlearned, a large number of neurons in the hidden layer could also reduce the ability of generalization. Because of the good prediction performance from training sets with 2-months and 15-days, which yielded the smallest MSE values and higher R^2 values than the other training sets, the number of neurons in the hidden layer could be recommended to be 5 and 8 in Colombo and Trincomalee respectively.

Table 6: Effect of # neurons in the hidden layer and MSE for different testing sets at Colombo and Trincomalee

	Colombo		Trincomalee	
Time Periods	# neurons in the hidden layer	MSE	# neurons in the hidden layer	MSE
7 days	6	0.005952	3	0.003231
10 days	9	0.006430	8	0.002856
15 days	2	0.015272	8	0.001182
1 month	2	0.004089	10	0.005374
2 months	5	0.002573	2	0.009538
3 months	8	0.003429	2	0.007613
4 months	8	0.007170	4	0.004729
5 months	8	0.012794	5	0.008145

Table 7: The optimal structures of the ANN when $\eta = 0.01, \alpha = 0.8$ and number of iterations=1000

Coastal Area	Required Time period	Number of inputs	Number of hidden layers	Number of hidden neurons	MSE	R^2
Colombo	2-months	10	1	5	0.002573	0.90165
Trincomalee	15-days	10	1	8	0.001182	0.96398

Table 7 shows the optimal network structures of the ANN for Colombo and Trincomalee for the specified η, α, and 1000 iterations. It is evident that the ANN constructed from 2-months of hourly data with one hidden layer and 5 neurons has emerged as the optimal structure to predict tidal elevations at Colombo while that constructed from 15-days of hourly data with one hidden layer and 8 neurons being the best BPN network for the prediction of tidal heights at Trincomalee.

3.3 Comparison of harmonic analysis and the neural network structure

Table 8 compares the prediction performance of both tidal prediction models in terms of MSE and MAPE for all time periods at both locations.It shows that both MSE and MAPE values of harmonic analysis are greater than those of the corresponding optimal ANN models for both locations. It can also be seen that as the time period gets larger, both MSE and MAPE values become smaller at both locations in Harmonic Analysis, but not in the ANN models. This confirms the fact that in harmonic analysis predictions become more accurate when long term measurable data are available for the calculation. Therefore, it is evident numerically that the standard BPN model with short periods of data provides more accurate forecasts than those from conventional models like harmonic analysis for the corresponding time period.

Table 8: MSE and MAPE values of harmonic analysis and optimal ANNs at both locations

Required Time period	Harmonic analysis				Optimal ANN models			
	Colombo		Trincomalee		Colombo		Trincomalee	
	MSE	MAPE(%)	MSE	MAPE(%)	MSE	MAPE(%)	MSE	MAPE(%)
7 days	0.0725	5.048	0.0563	3.926	0.006	0.096	0.0032	0.052
10 days	0.0614	4.275	0.0420	2.930	0.0064	0.104	0.0029	0.046
15 days	0.049	3.411	0.0419	2.924	0.0153	0.246	0.0012	0.026
1 month	0.044	3.069	0.0408	2.842	0.0041	0.066	0.0054	0.087
2 months	0.038	2.649	0.0262	1.825	0.0026	0.042	0.0095	0.154
3 months	0.0322	2.242	0.0226	1.573	0.0034	0.055	0.0076	0.123
4 months	0.0318	2.215	0.0203	1.418	0.0072	0.115	0.0047	0.076
5 months	0.0302	1.875	0.0194	1.052	0.0128	0.206	0.0081	0.132

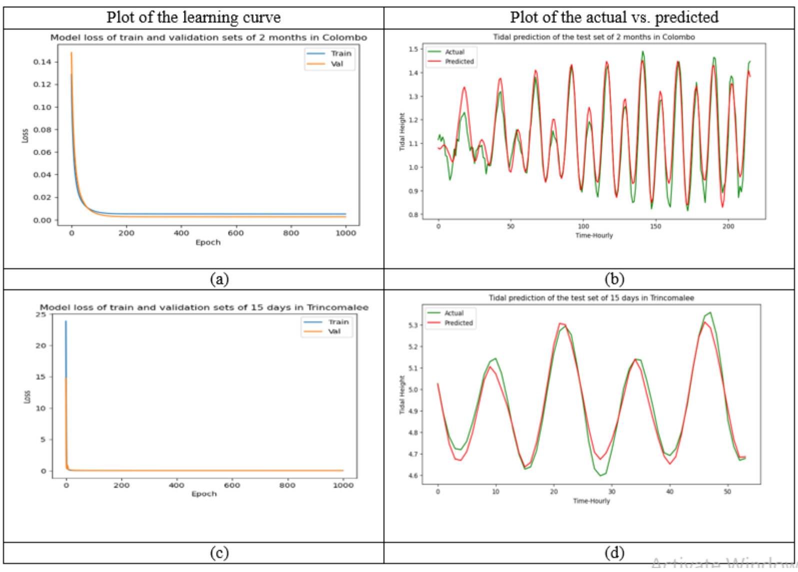

Figure 3: Prediction performance of optimal ANN structures for Colombo (a) Plot of the learning curve (b) Actual vs. predicted tidal heights and for Trincomalee (c) Plot of the learning curve (d) Actual vs. predicted tidal height

Supplementing this information, Figure 3 illustrates the prediction accuracy of the optimal neural networks for both sites with plots of the learning curve and actual vs. predicted values drawn using the test data. It can be seen that both actual and predicted tidal heights (Fig. 3b and 3d) at both locations are very close to each other except at a few time points, which means that the predictions from the proposed BPN models constructed for short time periods agree well with the actual observations at both coastal areas. Further, plots of the learning curves (Fig. 3a and 3c) confirm that both models have comparable performance in both training and validation datasets.

3.4 Discussion

One objective behind this study was to identify an accurate methodology to predict tidal elevations at eastern and western coastal areas of Sri Lanka with minimum usage of data to avoid issues of missing data due to instrumental errors and interruptions. Further, the study addressed the need of a reliable method to forecast tidal heights in the absence of large volumes of data, as an alternative to the conventional approach of Harmonic analysis. Initially, for this study, 5 complete months of data spanning from September, 2020 to January, 2021 were available. The traditional Harmonic analysis usually requires a large amount of data, more than one-year hourly data, to achieve precise harmonic constituents and thereby carry out predictions. This study used both Harmonic analysis and neural network structures on short term data with a comparison of prediction accuracy of tidal heights in terms of MSE and MAPE. The prediction accuracy of tidal heights could be enhanced when large numbers of tidal constituents were considered in both methods. However, this increases network complexity and takes considerable time for network convergence. Therefore, in the present study the tidal generation potential was described by using the most significant tidal constituents with the largest amplitudes, out of the 69 originally identified by Foreman (1978). The impact of the remaining constituents on tidal elevation has little significance and therefore could be ignored. Moreover, in order to select the minimum data requirement to obtain accurate predictions for tidal heights at each location, the entire period of available data was divided into eight intervals of different temporal lengths and both procedures were carried out for each interval separately. The values of the learning rate and momentum factor considered here are the same for many studies including that adopted by Lee, 2004.

4 Conclusion

Changing water levels can dramatically affect people, animals, plants, climate, and coastal maritime operations. These changing water levels in sea are caused by gravitational impact of the sun and moon, which form tides. Accurate information on tidal heights is always vital as it is beneficial to ensure the livelihood of fishermen, offshore constructions, and to mitigate flooding of low lands etc. It was found that the tidal height at Trincomalee is significantly higher than that at Colombo. Tidal type at Trincomalee was found to be semi-diurnal which is in agreement with the type for an equatorial country, whereas it was mixed-semi diurnal in Colombo. The main tidal constituents are similar at both locations which are O1, K1, N2, M2, and S2. Both harmonic analysis and ANN modeling can be effectively used for tidal prediction but their performance depends on the number of data points available for modelling. It is concluded that the standard BPN model with short period of data provides precise forecasts than those resulting from conventional models like harmonic analysis for the same period with same constituents. That is, the proposed standard BPN models are more appropriate to predict tidal elevations at Western and Eastern coastal areas of Sri Lanka with short-term data. More specifically, it was identified that 15 days and 2 months of hourly data are sufficient to produce accurate tidal predictions at Trincomalee and Colombo respectively. Colombo on the western coast required little more data than Trincomalee on the Eastern coast due to the presence of mixed semi-diurnal tide type which is not the standard type for equatorial countries. Thus, this study has shown that 2-months of data are adequate for the identification of tidal constituents and to ensure reliable predictions, which enables to implement activities such as road & rail constructions, repair drainage systems, generate tidal power as an alternative source of energy, and mitigate the risk of flooding low lands etc.

Acknowledgement
The authors would like to acknowledge the National Aquatic Resources Research and Development Agency (NARA) in Sri Lanka for providing the hourly tidal heights data for this study.

References

Boon J. D. and Kiley K. P. (1978) 'Harmonic Analysis and Tidal Prediction by the Method of Least Squares',*A User's Manual. W&M ScholarWorks* , 186. DOI:https://doi.org/10.21220/V5PF2T

Bhara K.K. and Bakshi S. (2020) 'Application of artificial neural network for predicting water levels in Hooghly estuary, India', *H2Open Journal*,3(1),401-415. DOI: https://doi.org/10.2166/h2oj.2020.041

De Vos, A., Pattiaratchi, C. B., and Wijeratne, E.M.S.(2014) 'Surface circulation and upwelling patterns around Sri Lanka',*Biogeosciences*, 11, 5909–5930. DOI: https://doi.org/10.5194/bg-11-5909-2014

Dietrich G. (1963) *General Oceanography*, Inter-science Publishers: New York.

Doodson A. T. (1921) The Harmonic Development of the Tide-Generating Potential, *Proceeding of the Royal Society of London.Series A*, 100(107), 305-329.

Ekanayaka E. M. P., Aso T. and Miyahara S. (1997) 'Background Wind Effect on Propagation of Nonmigrating Diurnal Tides in the Middle Atmosphere', *Journal of Atmospheric and Solar-Terrestrial Physics*, Vol. 59, No 4,401-429.

Foreman M. G. G. and Canada O. (2017) 'The harmonic analysis of tidal model time series', 1708(October). DOI:https://doi.org/10.1016/0309-1708(89)90017-1

Foreman M. G. G. and Neufeld E. T. (1991) 'Harmonic Tidal Analysis of Long Time Series'. *International Hydrographic Review*, 86-108.

Foreman M. G. G. (1978) 'Manual for Tidal Currents Analysis and Prediction', *Pacific Marine Science Report*, Institute of Ocean Sciences: Patricia Bay, British Columbia, Canada.

Foreman M. G. G. (1998) 'Manual for Tidal Heights Analysis and Prediction', *Pacific Marine Science Report*, No.98-1, Institute of Ocean Sciences: Sidney, BC, Canada.

Jorge O. and Eduardo A. (2009) 'Tidal forecasting in the Bahia Blanca Estuary', Argentina.

Lee T. L. and Jeng D. S. (2002) 'Application of artificial neural networks in tide forecasting', *Ocean Engineering*, 29(9), 1003–1022. DOI:https://doi.org/10.1016/S0029-8018(01)00068-3

Lee T. L., Tsai C. P., and Shieh R. J. (2006) 'Applied the Back-Propagation Neural Network to Predict Long-term Tidal Level', *Asian Journal of Information Technology*, 5(4), 396–401.

Lee T. L. (2004) 'Back-propagation neural network for long-term tidal predictions', *Ocean Engineering*, 31(2), 225–238. DOI:https://doi.org/10.1016/S0029-8018(03)00115-X

Mandal S. (2001) 'Prediction of tides using back-propagation neural networks', Vol 1–6, Issue Penerapan Backpropagation, p. 6.

Meena B. L. and Agrawal J. D. (2015) 'Tidal level forecasting using ANN', *Procedia Engineering*, 116(1), 607–614. DOI:https://doi.org/10.1016/j.proeng.2015.08.332

Reid R.O. (1990) 'Tide and storm surges', In: Herbich, J.B. (Ed.), Handbook of Coastal and OceanEngineering, 11, 534–590.

Rumelhart D. E., Hinton G. E. and Williams R. J. (1986) 'Learning representations by back-propagating errors', Nature 323, 533–536.

Salim, A. M., Dwarakish, G. S., Liju, K. V., Thomas, J., Devi, G. and Rajeesh, R. (2015) 'Weekly prediction of tides using neural networks', *Procedia Engineering*,116(1),678–682. DOI:https://doi.org/10.1016/j.proeng.2015.08.351