LONG-TERM TIDAL FORECASTING AND HINDCASTING USING QuickTIDE TIDAL SIMULATION PACKAGE

Lee Wei Koon
Water Resources and Environmental Systems Division, Faculty of Civil Engineering, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor.
Email: leewei994@salam.uitm.edu.my

ABSTRACT

A tidal simulation package built on MATLAB neural network toolbox with user-friendly graphical user interface (GUI) is presented. The application, named QuickTIDE, is modeled using a back-propagation neural network (BPN) with single hidden layer and five input constituents. Twenty-one built-in network models trained on one-year hourly tidal observations are able to generate reliable tidal simulation from year 1994 to 2004 corresponding to their respective locations in Malaysia. Further tests on the historical data of station Boston Harbour from year 1930 to 2000 show that good forecast and hindcast can be produced up to 70 years if the number of neuron in the hidden layer and the number of input argument are properly selected. Using 30-day hourly observations, the application is capable of producing satisfactory simulation with average correlation coefficient R in the order of 0.96 from year 1994 to 2004 for station Port Klang. For forecasting and hindcasting from year 1930 to 2000 for station Boston Harbour, R-values average at 0.92 and 0.90 respectively, with better results in the month the training data originates.

Keywords: *Artificial Neural Network, Back-Propagation, Graphical User Interface, Harmonic Analysis, Tidal Forecast, Tidal Hindcast, QuickTIDE*

1. INTRODUCTION

Tidal fluctuation is a natural phenomenon in coastal water. Driven primarily by the celestial gravitational forces, the daily high and low tides result in sea level difference in the order of a few meters typically, depending on the local shoreline features and bathymetry. In locations with gentle beach slope, these can translate into a few kilometers of alternating inundated and exposed shore. Furthermore, the tidal range varies considerably between the spring tide and the neap tide, which occur alternately every two weeks. Hence, accurate tidal forecasting is crucial in the planning and execution of any nearshore activities and development, such as determination of platform level, set back of buildings, crown level of sea wall etc.

The National Hydrographic Centre (NHC) under the Royal Malaysian Navy publishes the tide tables for over 71 standard ports in Malaysia, Singapore and Brunei, the information of which is generally intended for navigational use at the respective locations. Whereas, the Department of Survey and Mapping Malaysia (*Jabatan Ukur dan Pemetaan Malaysia*, JUPEM) also publishes the tide tables for the Tidal Observation Network (TON) which comprises 21 established tidal stations (Table 1) distributed along the coast, mainly located at the ports. While the locations of these gauging stations are selected such as to show typical characteristics of tides in the adjacent sea, they certainly do not represent the actual tidal variations in the specific location of interest which can be a great distance away with differing shoreline features and bathymetry. Direct interpolation of tidal range and highest water level between adjacent measurements is thus not generally without doubt.

Both NHS and JUPEM produce the tidal predictions using the conventional method of harmonic analysis which requires minimum one-year records. The long-term deployment and maintenance of the large number of tidal gauging stations for continuous data collection is an expensive endeavour and thus it may not be feasible to establish more tidal stations. Hence, ability to forecast tides for any location of interest using only short-term observations is very desirable.

Application of artificial neural network (ANN) in tidal forecasting using field data has been performed with satisfactory accuracy for hourly prediction up to one-month and one-year durations using only one-day records and fifteen-day records respectively for locations with diurnal and semi-diurnal tides [1] as well as locations with mixed tide condition [2]. The ANN can also be combined with the equation of harmonic analysis to produce similarly accurate hourly tidal prediction up to one-year duration using fifteen-day records [3,4]. More recently, water level prediction using functional and sequential learning neural networks [5] and radial basis function neural network [6] were also explored and reported.

Long-term simulation of coastal sea levels for a period of a decade [7] or more [8] has also been attempted to provide a background for related researches, which include but are not limited to, sediment transport and accretion, dispersion of pollutants plume, coastal morphology, bathymetry and shoreline changes etc. Along the same line, the availability of long-term past records will greatly facilitates the study of coastal changes which have taken place and hence led to the present day scenario. Therefore, while forecasting is essential to predict the future, hindcasting [8,10] plays an important role in supplementing past records which are not available.

In a recent paper by the author, the performance of a back-propagation neural network (BPN) in tidal forecasting was studied [9]. The neural network model developed was tested on three tidal stations with different types of tide. Reliable tidal forecasting was reported with up to ten years lead time using short-term observations in the order of 30 days or less [9]. The present paper further investigate the performance of the network model presented in [9] in long-term tidal forecasting and hindcasting up to a century. In addition, the model is developed into a desktop friendly application, named QuickTIDE. The paper is organized in the following manner: Section 2 describes the network model development and the computation algorithm; Section 3 the performance of QuickTIDE built-in network models; and Section 4 the custom network model training and simulation features.

2. MODEL DEVELOPMENT

A. Artificial Neural Network

Artificial neural network (ANN) is a parallel computing system emulating the ability of the biological nervous system by interconnecting many artificial neurons (Figure 1). It has been widely used to overcome problem of exclusive and non-linear relationships, and has found application in large number of disciplines. In particular, ANN can be used for tidal prediction using short-term observations as opposed to the conventional harmonic analysis, which requires long-term tidal records of over one-year duration. This enables predictions to be carried out for any specific location without having to install a permanent tidal station, thus reducing the prediction cost without compromise in accuracy.

A simple ANN comprises an input layer I and an output layer O. The number of inputs and outputs are usually governed by the nature of the problem. The network architecture describes how many hidden layers H a network has, the number of neuron and transfer function in each layer and how the layers are interconnected to each other. A feedforward (FF) network is one whereby the interconnections do not contain any feedback loop, i.e. *acyclic*. Typically the inputs are multiplied with the weight in the connections and summed up, together with the bias, at the receiving neuron, which produce an intermediate output by transformation using a differentiable transfer function. The computations proceed through the layers in a single direction and eventually generate the final outputs.

For a set of known inputs and outputs, the network can be trained to map the inputs to the targeted outputs for a selected network architecture. The back-propagation neural network (BPN) is the most representative learning model for ANN. The procedure of BPN repeatedly adjusts the weights and biases in the network so as to minimise the error function, typically the mean square error (MSE), defined as the averaged squared error between the network output and the target output. In each iteration (or epoch), the error function is determined at the output layer and back-propagated through the network to make necessary correction to the weights and biases until the error function reduces to a prescribed value or when the training process already reaches the maximum number of epoch specified. The calculated weights and biases can then be used to compute the output vector for the corresponding network architecture for any given input.

The training process is usually repeated using different network architecture and the best performing network architecture is ultimately determined using an appropriate performance function. For further reading on the fundamentals of ANN, interested readers are referred to established texts such as [11,12,13].

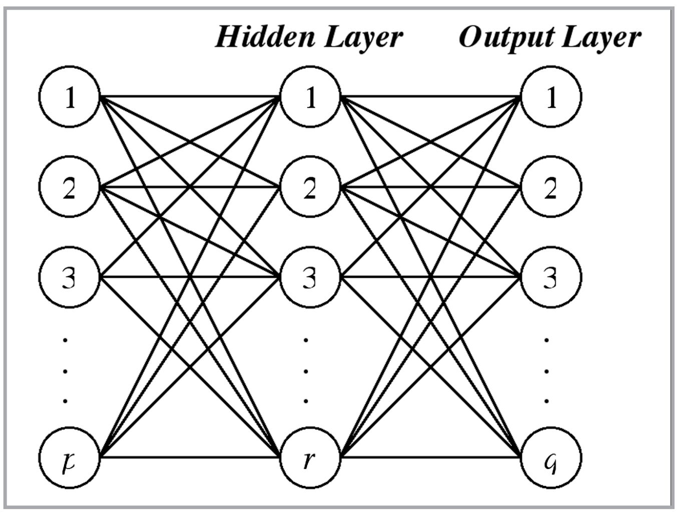

Figure 1: Example of an artificial neural network with one hidden layer, p number of input, q number of output and r number of neuron in the hidden layer

B. QuickTIDE Network Architecture

QuickTIDE is written in MATLAB programming language, using the MATLAB *Neural Network Toolbox* and MATLAB Graphical User Interface Development Environment (GUIDE). The feedforward back-propagation (FFBP) network model is used and the Levenberg-Marquardt (LM) algorithm based on numerical optimization techniques is selected for batch training of the network, i.e. the weights and biases are updated only after the entire training set has been applied to the network. The LM algorithm was designed to approach second-order training speed and can be ten to one hundred times faster than the conventional gradient descent algorithm, without the need to tune network learning rate and momentum coefficient as performed in [3.4]. As reported in [9], the LM algorithm is able to produce rapid convergence at well below 150 epochs for 30-day hourly input data, contrary to the gradient descent algorithm with momentum which requires over 5000 epochs for a learning rate η of 0.01 and momentum coefficient α of 0.8. Readers are referred to [9] for detailed description of the LM algorithm.

Sea level changes Y as a function of elapsed time t due to tidal effect at a particular location can be described by the sum of all harmonics in the form of

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

where A_0 is the mean sea level and N is the total number of constituents. For each constituent, the coefficients are given by A_i and B_i; the angular speed given by ω_i; the phase term given by $\varepsilon_i = \tan^{-1}(B_i/A_i)$; and the amplitude given by $h_i = \sqrt{A_i^2 = B_i^2}$. Equation (1) can be readily modelled using ANN by taking the $\cos\omega_i t$ and $\sin\omega_i t$ terms as the input variables and $Y(t)$ as the single output variable [3,4]. Hence, the total number of input variables is given by $2N$, depending on the total number of constituents considered. The coefficients correspond to the weights of the respective input variables whereas A_0 is the overall network bias.

Theoretically, network accuracy improves with the number of constituents. Nonetheless, as more and more components are included, the actual improvement eventually diminishes while computation time increases considerably. Based on the study in [9], it was concluded that the best results for tidal forecast up to ten-year could be obtained by using only the five primary constituents, namely O_1, K_1, N_2, M_2 and S_2, which have periods in the order of 24-hour or less and typically dominate over other less important components. These are the same constituents adopted in the study conducted by [3] and [4]. It was also reported in [9] that the performance of a network with single hidden layer is much better than that without hidden layer, and as good as one with dual hidden layer.

It is worth noting here that ANN is a nonlinear procedure whereby its performance is highly dependent on the network parameters which are themselves nonexclusive. Tuning these parameters to produce optimised network model is not an easy task due to the large number of possible combinations. Based on the previous trials performed as reported in [9], QuickTIDE is built on a FFBP network, performs batch training using LM algorithm, and contains a single-hidden layered with 5 predetermined input constituents aforementioned. The hyperbolic tangent sigmoid transfer function, which is typically used for multilayer network is adopted for both the hidden and output layer. The function is expressed as

$$f(x) = 2/[1+\exp(-2x)] - 1 \tag{4}$$

and is mathematically equivalent to hyperbolic tangent but can be executed more efficiently in MATLAB [14]. In addition, the maximum epoch is set to 500 and the target error function MSE is set to zero. This leaves the users with only one network parameter left to vary, i.e. the number of neuron in the hidden layer.

Hence, QuickTIDE essentially eliminates the tedious process of selecting the various network parameters which is time-consuming and can lead to no conclusive results if not well planned or directed. For occasional and general use, this feature greatly simplifies the application. With its user-friendly interface, users require only short learning curve, and are able to produce satisfactory results as shall be presented in the following sections. However, serious users and researchers will definitely find this a major constraint and may prefer to 'get their hands dirty' to explore the capabilities of ANN from scratch. By changing the network type, network architecture, transfer function, error function etc, there are infinite possible model which can be either better or poorer for the intended purpose and hence the search can by no means be exhaustive. Last but not least, computation speed is frequently also a major concern and hence a simple network could be much favourable compared to a complex one.

C. Network Performance Function

QuickTIDE network performance in tidal simulation is given by the normalised root mean square error RMS and correlation coefficient R which are defined as:

$$\text{RMS} = \sqrt{\sum_{j=1}^{M}(Y_j - O_j)^2 \Big/ \sum_{j=1}^{M} O_j^2} \tag{2}$$

$$R = \frac{\sum_{j=1}^{M}(Y_j - \bar{Y}_j)(O_j - \bar{O}_j)}{\sqrt{\sum_{j=1}^{M}(Y_j - \bar{Y}_j)^2 \sum_{j=1}^{M}(O_j - \bar{O}_j)^2}} \tag{3}$$

where Y_j and O_j are the predicted and observed values respectively, $\bar{Y}_j$ and $\bar{O}_j$ the arithmetic mean of the same. The two parameters can, of course, only be calculated if observations are known a priori during tidal simulation.

Network training, on the other hand, gives the values of MSE and R which form the basis for selection of the best performing network model. Note that users may also opt to use the RMS value in simulation to identify the best performing network model instead of the MSE which are generally well below 0.01 in all cases [9]. In this approach, two sets of observations are required: one for network training, and the other to validate the simulation. By varying the number of neuron in the hidden layer, the trained network is immediately used to perform simulation and the values of RMS and R of the simulated results compared to determine the optimum number of neuron required and hence the best performing network model.

Table 1: Details of tidal stations studied

Station Number	Station	Location	Latitude(N) Longitude(E)	Data Period	Type of Tide
48530	LANGKAWI ISLAND	*Jeti Telok Ewa*	06° 25' 51" N 99° 45' 51" E	1994-2004	Semi-diurnal
48509	LUMUT	*Pengkalan TLDM*	04° 14' 24" N 100° 36' 48" E	1994-2004	
48510	PENANG ISLAND	*Penang Yacht Club, George Town*	05° 25' 18" N 100° 20' 48" E	1994-2004	
48483	PORT KLANG	*Dermaga 25 Pelabuhan Utara*	03° 03' 00" N 101° 21' 30" E	1994-2004	
48390	TAWAU	*Pelabuhan*	04° 14' 00" N 117° 53' 00" E	1994-2004	
74005	BINTULU	*Pelabuhan*	03° 15' 44" N 133° 58' 20" E	1994-2004	Mixed, diurnal dominant
48507	CENDERING	*Kompleks LKIM*	05° 15' 54" N 103° 11' 12" E	1994-2004	
74007	MIRI	*Jeti Petronas*	04° 23' 28" N 113° 58' 20" E	1994-1998	
48558	KOTA KINABALU	*Pelabuhan*	05° 59' 00" N 116° 04' 00" E	1994-2004	
48549	GETING	*Kompleks LKIM*	06° 13' 35" N 102° 06' 24" E	1994-2004	Mixed, semi-diurnal dominant
48484	JOHOR BAHRU	*Jeti Kastam Johor Bahru*	01° 27' 42" N 103° 47' 30"E	1994-2004	
48529	KUKUP	*Jeti*	01° 19' 31" N 103° 26' 34" E	1994-2004	
48485	KUANTAN	*Pelabuhan Kuantan, Tanjung Gelang*	03° 58' 30" N 103° 25' 48" E	1994-2004	
74003	SANDAKAN	*Pelabuhan*	05° 48' 36" N 118° 04' 20" E	1994-2004	
48550	TANJUNG SEDILI	*Kompleks LKIM*	01° 55' 54" N 104° 06' 54" E	1994-2004	
48508	TANJUNG KELING	*Jeti Tanjung Bruas*	02° 12' 54" N 102° 09' 12" E	1994-2004	
48528	TIOMAN ISLAND	*Jeti Berjaya Resort*	02° 48' 26" N 104° 08' 24" E	1994-2004	
48004	KUCHING	*Pelabuhan, Sejingkat*	01° 34' 58" N 110° 25' 20" E	1996-2004	
74001	KUDAT	*Pelabuhan*	06° 52' 46" N 116° 50' 37" E	1996-2004	
74010	LABUAN	*Pelabuhan*	05° 16' 22" N 115° 15' 00" E	1996-2004	
74009	LAHAD DATU	*Pelabuhan*	05° 01' 08" N 118° 20' 46" E	1996-2004	

3. QUICKTIDE BUILT-IN NETWORK MODELS

The key function of QuickTIDE is to produce tidal forecast and hindcast readily using trained network models for the respective stations. QuickTIDE comes with 21 built-in network models corresponding to all the tidal stations under the Tidal Observation Network (TON) established by JUPEM. A summary of the stations is given in Table 1, which includes amongst others, the type of tide in each station and the data period used in the development of QuickTIDE.

The built-in network models are trained using the 11-years historical data obtained from JUPEM, employing hourly tidal observations up to one-year duration in each attempt and varying the number of neurons to identify the best performing network architecture based on the value of R primarily, supplemented by the value of RMS. The outcomes, as summarised in Table 2, show that most stations require 7 or 10 neurons for optimum performance. The year of which these results are obtained are as tabulated.

Note in particular that station Geting, which has a very strong tidal variation with a period in the order of one year, suffers from poor training results due to the fact that the five constituents selected do not account for such low frequency fluctuation. The records, subsequently trained with two additional constituents, namely Sa (Period = 8765.5h) and Ssa (Period = 4382.8h), produce results which are comparable with the other stations, as shown in Table 2. Nonetheless, as mentioned earlier, the current version of QuickTIDE does not yet incorporate the option for users to change the default 5 input constituents used in network training. Only the station Geting is designed to come with 7 input constituents among the built-in models.

While there is no restriction on the number of neuron assigned for the hidden layer, the value attempted is limited to 10 herein so as to avoid over-training and to reduce the network complexity which may lead to long execution time in tidal simulation. The R-values are generally higher for locations with semi-diurnal tide (average 0.9930) followed by locations with mixed, semi-diurnal dominant tide (average 0.9826) and locations with mixed, diurnal dominant tide (average 0.9776). The values of MSE are well below 0.0100 with the exception of station Chendering, and the number of epoch required for convergence are generally less than 300.

These trained built-in network models can be accessed using the graphical user interface (GUI) of QuickTIDE Tidal Simulation Module (Figure 2). The module allows user to select the network model corresponding to the location of interest (in abbreviation, e.g. BTU for Bintulu) from the popup menu, enter the starting date and the duration of simulation. By clicking the SIMULATE button, the tidal level is calculated and plotted on the figure pane to the right. If observations are available, user can choose to import the data (hourly) in text format before performing simulation. In this case, the observations will be plotted as well and the values of R (referred to as CC in the GUI) and RMS computed. Finally, user can select and save any of the simulation results in MATLAB MAT-file format for subsequent retrieval.

The 21 trained network models are used to simulate hourly tidal level for the year 1994 and 2004 specifically, some exception due to availability of observed data are as detailed in Table 3. The R-values are generally above 0.9500, with an average of 0.9606 except for station Geting and Chendering and few isolated cases. The average value for locations with semi-diurnal tide is the highest at 0.9892, followed by locations with mixed, semi-diurnal dominant tide (average 0.9606) and locations with mixed, diurnal dominant tide (average 0.9524). The root mean square error RMS is well below 0.1000 except a case for station Miri. Figure 3(a) shows the plot of the forecasted and observed water level for

Table 2: Training of built-in network models

Station	Year of Training Data	Number of Neuron in hidden layer	R	MSE
Langkawi	1994	10	0.9934	0.0022
Lumut	1999	10	0.9915	0.0024
Penang	2000	10	0.9879	0.0036
Pelabuhan Klang	1994	10	0.9964	0.0013
Tawau	2004	10	0.9959	0.0009
Bintulu	2004	10	0.9874	0.0034
Chendering	2002	7	0.9549	0.0119
Miri1997	10	0.9857	0.0038	
Kota Kinabalu	1997	9	0.9823	0.0015
Geting*	2003	7	0.9646	0.0081
Johor Bahru	1998	10	0.9865	0.0043
Kukup	1994	10	0.9823	0.0015
Kuantan	1998	10	0.9713	0.0070
Sandakan	2004	10	0.9947	0.0008
Tanjung Sedili	1997	7	0.9684	0.0081
Tanjung Keling	1994	10	0.9925	0.0027
Tioman	2002	10	0.9774	0.0063
Kuching	1997	10	0.9954	0.0017
Kudat	2004	7	0.9885	0.0013
Labuan	2001	7	0.9863	0.0035
Lahad Datu	2001	10	0.9836	0.0020
		Average	**0.9627**	**0.0045**

* Trained using 7 constituents as opposed to 5 (refer Section 3)

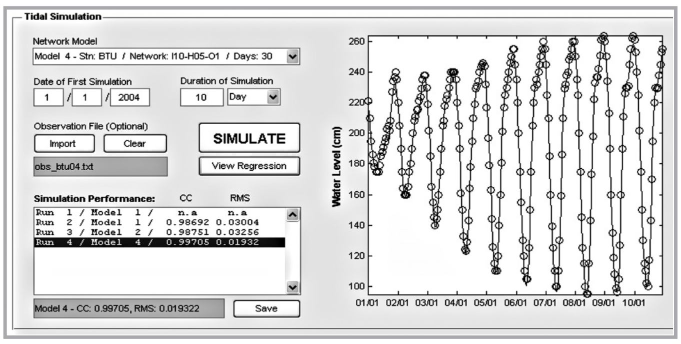

Figure 2: QuickTIDE tidal simulation module

station Port Klang from 1-15 January 2004 using the built-in model trained on 1994 data, whereas Figure 3(b) shows the plot of the hindcasted and observed water level for station Tawau from 1 - 15 January 1994 using the built-in model trained on 2004 data. Excellent agreement are obtained with R = 0.98684, RMS = 0.05269 in the former, and R = 0.98876, RMS = 0.04775 in the latter. The plots are saved from the QuickTIDE GUI by right-click on the figure pane.

The simulation results in Table 3 show that these models are able to produce good estimates of the tidal levels regardless of whether the simulated time period is before or after the year of tidal data used in network training (Table 2). This is attributed to the elapsed time which is calculated with reference to 1 January 1900, 00:00am in both the training and simulation module in order to compensate for the astronomical arguments and Greenwich phase lag which are omitted from the harmonic equation for convenience of network implementation. Hence, QuickTIDE can be readily used for both forecasting and hindcasting.

To further verify the applicability of the network model over a longer time span, additional tests are carried out using tidal observations extracted from National Oceanic and Atmospheric Administration (NOAA) Tides and Currents. The portal (http://tidesandcurrents.noaa.gov), managed by the Center for Operational Oceanographic Products and Services (CO-OPS), houses vast collection of historical and real-time oceanographic and meteorological data, predictions, nowcasts and forecasts. Hourly tidal records in local time with reference to the station datum for Boston Harbour (Station Number: 8443970; 42º21.3'N,

Table 3: Tidal simulation using built-in network models

Station	Simulation for year 1994*		Simulation for year 2004**	
	R	**RMS**	**R**	**RMS**
Langkawi	0.9932	0.0358	0.9746	0.0703
Lumut	0.9867	0.0661	0.9841	0.0540
Penang	0.9813	0.0582	0.9221	0.0834
Pelabuhan Klang	0.9963	0.0264	0.9774	0.0664
Tawau	0.9884	0.0438	0.9959	0.0248
Bintulu	0.9616	0.0615	0.9870	0.0366
Chendering	0.9038	0.0983	0.9456	0.0791
Miri	0.9816	0.0445	0.9220	0.1305
Kota Kinabalu	0.9626	0.0437	0.9934	0.0505
Geting	0.8595	0.0756	0.9284	0.0534
Johor Bahru	0.9678	0.0651	0.9644	0.0701
Kukup	0.9947	0.0204	0.9845	0.0356
Kuantan	0.9509	0.0685	0.9510	0.0735
Sandakan	0.9706	0.0470	0.9945	0.0198
Tanjung Sedili	0.9515	0.0695	0.9368	0.0838
Tanjung Keling	0.9924	0.0229	0.9722	0.0457
Tioman	0.9468	0.0710	0.9689	0.0562
Kuching	0.9958	0.0366	0.9938	0.0325
Kudat	0.9577	0.0521	0.9883	0.0266
Labuan	0.9552	0.0455	0.9829	0.0408
Lahad Datu	0.9839	0.0315	0.9869	0.0418
Average	**0.9601**	**0.0505**	**0.9710**	**0.0486**

* Except Kuching, Kudat Labuan, dan Lahad Datu for year 1996.

** Except Lumut for year 2003, and Miri for year 1998

71º3.1'W) for selected number of years are downloaded for the above purpose. The 1-year data in year 1930 and 2000 are used to train the network model using 10 neurons and 7 input constituents (Sa, Ssa, M_2 , K_1, S_2, O_1, and N_2), giving R values of 0.9877 and 0.9917 respectively for the station which has semi-diurnal tide. If the Sa and Ssa components are excluded, the R-values reduce to the order of 0.9500.

The two trained network models are then used to generate 1-year tidal forecast from year 1930 to 2000 for every 10-year. Figure 4 shows the 15-day plot from 1 -15 January of the simulated and observed values for year 1930 hindcast using the year 2000 model, and the year 2000 forecast using a year 1930 model respectively. The values of R and RMS, as tabulated in Table 4, show an average R-value of 0.9795 and 0.9739 for tidal forecasting and hindcasting respectively. The NOAA predictions using the conventional harmonic analysis, on the other hand, have an average R-value of 0.9904, which is 2% higher. This shows that the built-in network models in QuickTIDE, which are trained using 1-year data, are able to produce tidal simulation with relatively good accuracy in the order of those generated from analysis of long-term observations.

4. CUSTOM NETWORK MODEL TRAINING

While the built-in network models in QuickTIDE are trained using data as long as 1-year, users can train hourly observations of any length to produce custom network model for specific location of interest.

Figure 5 shows the GUI of QuickTIDE Network Training Module. User can create new station and import the hourly observations in text format for network training. Date and time of the first observation must be specified. Entering appropriate number of neuron in the hidden layer, just click the TRAIN button to execute network training of which the results will be displayed in the list box on the right. Notice that each output line comprises the station abbreviation, the network architecture in standard notation (e.g. *I*10-*H*03-*O*1 means 10 inputs, 3 hidden neurons and 1 output), length of the training data (in days), number of epochs executed, the values of MSE and R (referred to as CC in the GUI). After a number of trial, which basically involves varying only the number of neuron in the hidden layer, user can select the best result from the list to save the corresponding network model. The saved model is then immediately accessible from the QuickTIDE Tidal Simulation module.

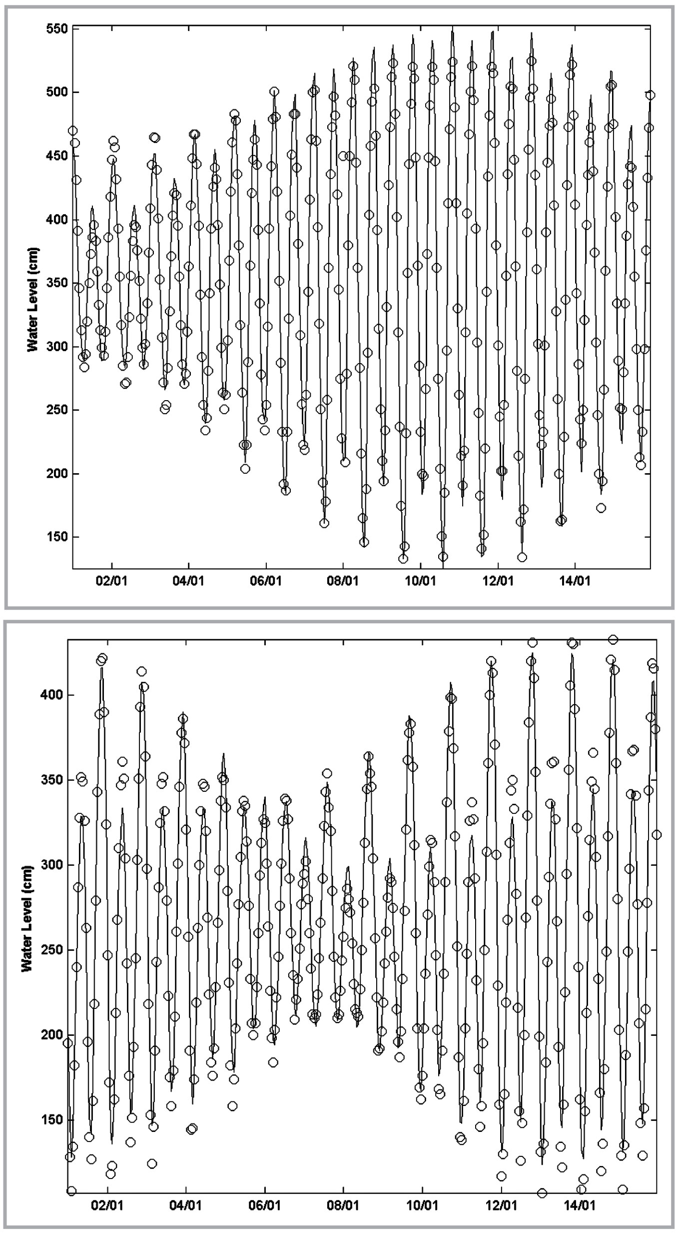

Figure 3: Plot of the simulated and observed water level for
a) station Port Klang from 1 - 15 January 2004 using the built-in model trained on 1994 data, and
(b) station Tawau from 1 - 15 January 1994 using the built-in model trained on 2004 data

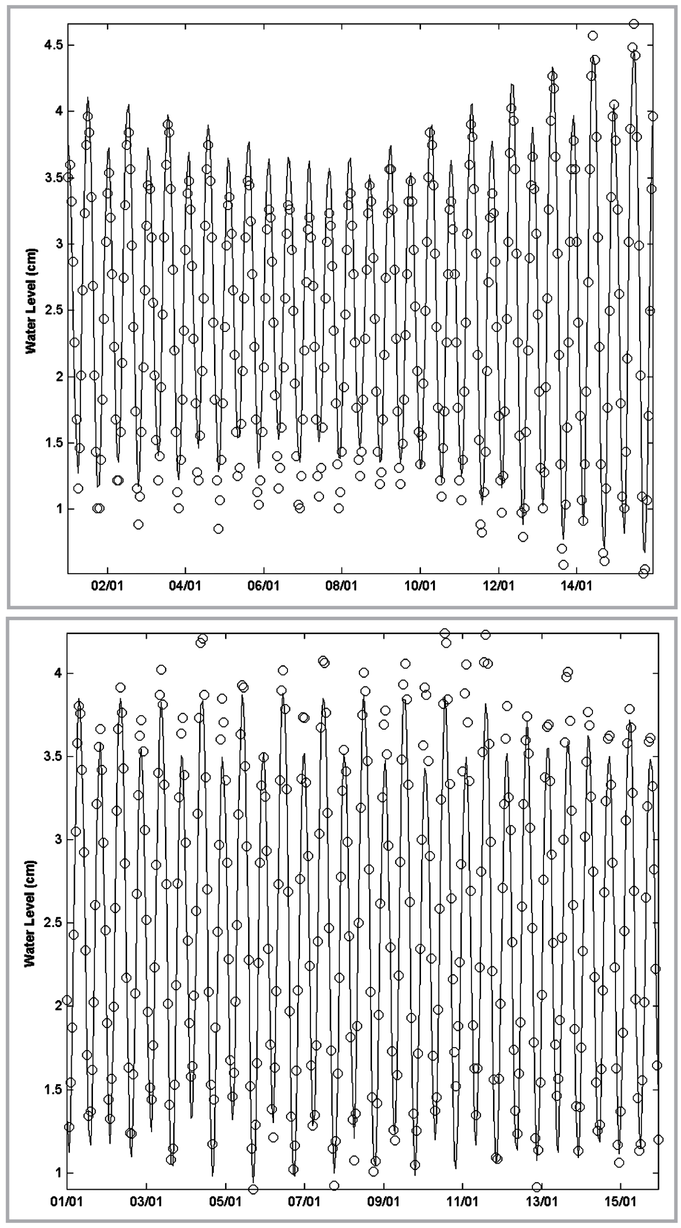

Figure 4: Plot of the simulated and observed water level for station Boston
a) from 1 - 15 January 1930 using the built-in model trained on 2000 data, and
(b) from 1 - 15 January 2000 using the built-in model trained on 1930 data

Table 4: Tidal forecast and hindcast for station Boston Harbour (Station Number: 8443970)

Year	Tidal Forecast using year 1930 model		Tidal Hindcast using year 2000 model		NOAA Forecast usingHarmonic Analysis	
	R	**RMS**	**R**	**RMS**	**R**	**RMS**
1930	0.9868	0.0603	0.9730	0.1188	0.9893	0.0927
1940	0.9810	0.0811	0.9713	0.1021	0.9903	0.0693
1950	0.9776	0.0834	0.9689	0.1039	0.9883	0.0702
1960	0.9793	0.0965	0.9752	0.0830	0.9913	0.0513
1970	0.9780	0.0949	0.9677	0.0936	0.9910	0.0494
1980	0.9803	0.0920	0.9746	0.0874	0.9915	0.0532
1990	0.9769	0.0954	0.9691	0.0923	0.9909	0.0509
2000	0.9757	0.1092	0.9915	0.0474	0.9910	0.0492
Average	**0.9795**	**0.0891**	**0.9739**	**0.091**	**0.9904**	**0.0608**

It has been shown that satisfactory tidal forecast can be produced using short-term observations of 30 days or less [1,2,3,4,9]. Here, 30-day hourly tidal observations, dated 1 -30 January for station Port Klang for the year 1994 and 2004 respectively are used to train two network models which are subsequently used to simulate the tides on a yearly basis from year 1994 to 2004 (Figure 6). The results are compared with simulation generated by QuickTIDE built-in model which was trained using 1-year observations. It can be seen that the network models trained using 30-day data is able to produce better forecast (average R = 0.9756) than hindcast (average R = 0.9573), the later of which is 3.2% inferior to the simulation generated by the built-in model (average R = 0.9891). The average RMS-values are 0.07 and 0.10 for forecasting and hindcasting respectively using the 30-day models, and 0.05 for the built-in model.

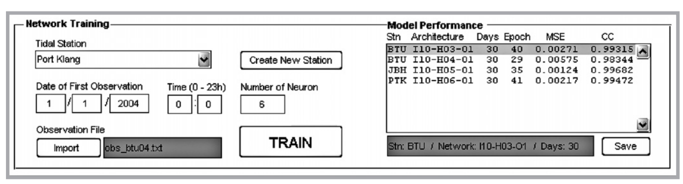

Figure 5: QuickTIDE network training module

Figure 7 shows the variation of R-values on a monthly basis from January to December for the tidal level in year 1994 and year 2004 respectively in Port Klang, simulated using the two models above. The plot shows that the higher and lower R-values alternate

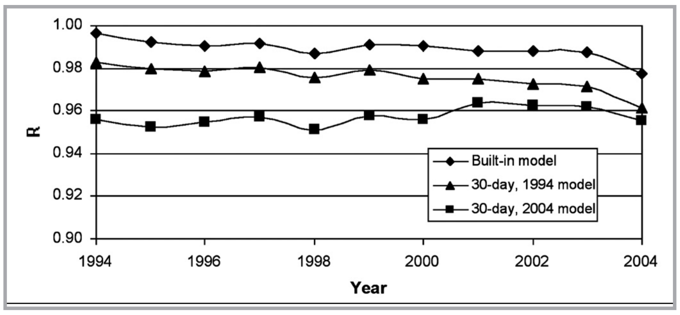

Figure 6: Tidal simulation for station Port Klang using 30-days (1 - 30 January) training data

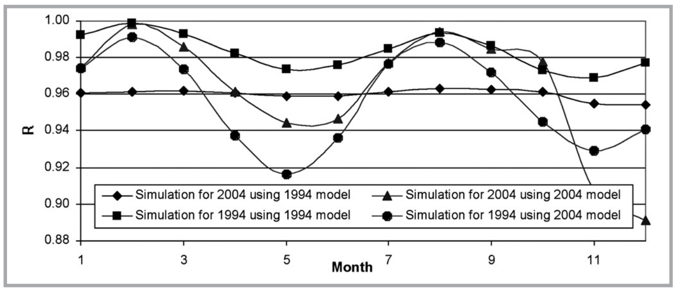

Figure 7: Variation of R-valueby month using 30-days training data dated 1 - 30 January (Port Klang)

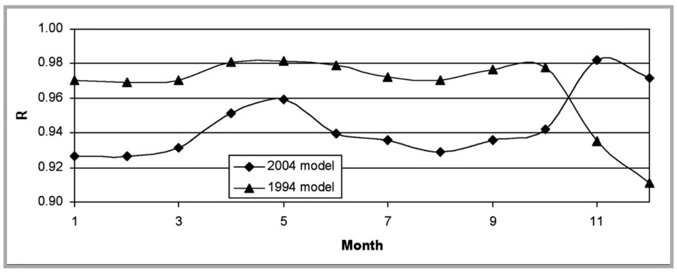

Figure 8: Variation of R-value by month using 30-days training data dated 1 - 30 May (Port Klang)

at a period of six months. To verify the reason for such fluctuation, the tidal level in year 2004 is again simulated using 30-day observations dated 1-30 May for the year 1994 and 2004 (Figure 8). It is then clear that the higher R-value corresponds to the date of the data used for network training, and recurs at an interval of approximately six months. The trend prevails, as can be seen in Figure 7, whereby the year 1994 model and year 2004 model shows similar pattern over a time span of 10 years. The RMS-values show similar trend with better results in the time vicinity of the training data, averaging 0.1.

Next, the ability of the 30-day trained network model to produce tidal simulation over a longer time span is considered. Using the station Boston Harbour as before, two network models are trained using 30-day observations taken from 1 - 30 January for the year 1930 and 2000 respectively. 1-year simulation is generated for every ten years from 1930 to 2000 as shown in Figure 9. Generally, forecasting using the year 1930 model yields R-values well above 0.90 with an average of 0.9216 but hindcasting using the year 2000 model is about 2% poorer with an average R-value of 0.9045. The RMS-values average at 0.16.

Further check on the variation of R-values by month as shown in Figure 10 for the simulation of year 2000 tides again shows that better results are obtained in the time vicinity of the training data. It is however interesting to note that if 30-day observations in January are used, good simulations are obtained for five months from November through March with January in the middle and being the best; whereas if the 30-day observations are taken from May, the better simulations fall in the month April through October, with the month of May being the best but not in the middle. Last but not least, if the training dataset is less than 30 days, forecasting and hindcasting results generally deteriorates.

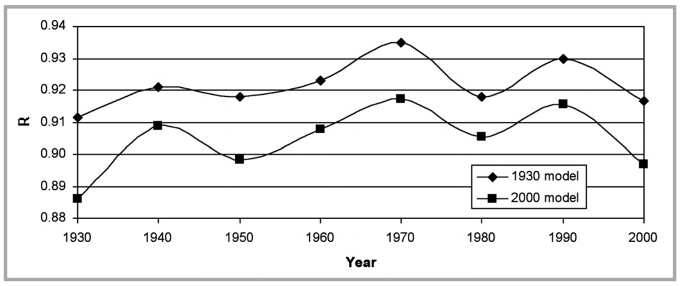

Figure 9: Variation of R-value in long-term simulation using 30-days training data (Boston Harbour)

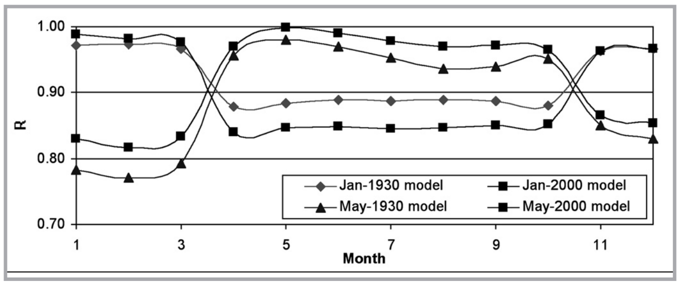

Figure 10: Variation of R-value by month in year 2000 using 30-days training data (Boston Harbour)

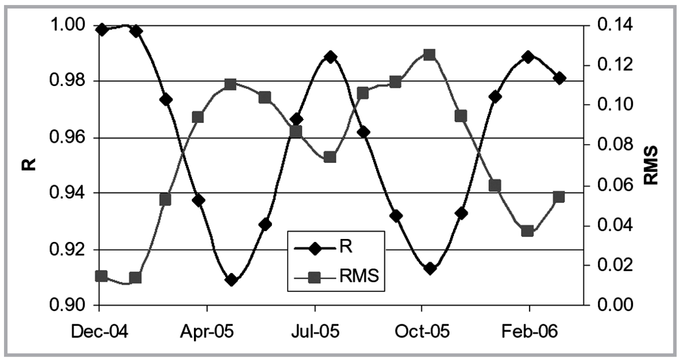

Figure 11: Variation of R and RMS values by month for station Kuah using 30-days trained network model

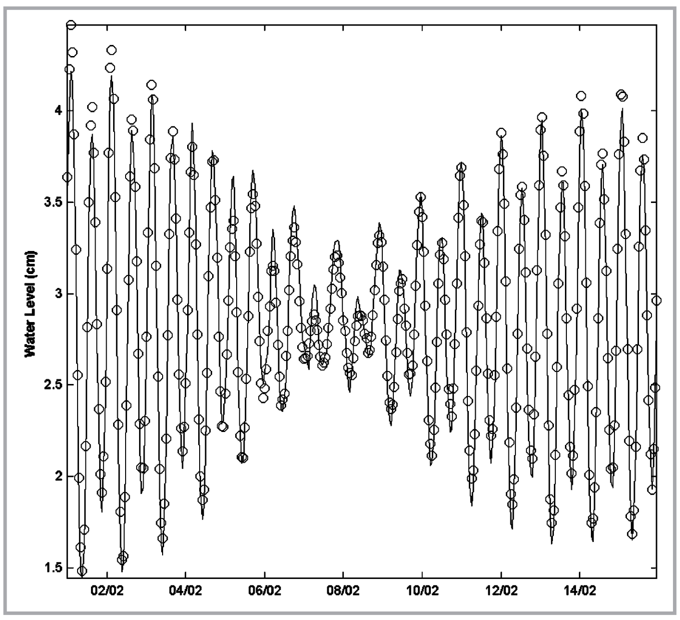

Figure 12: Plot of the forecast and observed water level for station Kuah 1 - 15 February 2006 using 30-days trained network model (R = 0.9849, RMS = 0.0431)

Last but not least, QuickTIDE is further tested on tidal data provided by National Hydrographic Centre of the Royal Malaysian Navy. Using 30-day (27 January - 25 February 2005) hourly observations from station Kuah, Langkawi (Station Number: 4654, 6∞18'3.7"N, 99∞57'0.45"E), the network model is trained using 10 neuron, giving R = 0.9989 and MSE = 0.0004. The values of R and RMS for monthly forecast from January 2005 to March 2006 are as shown in Figure 11, with an average of 0.9590 and 0.0758 respectively. Again, it is evident that better results are obtained near the month the training data is extracted from and and recurs at an interval of approximately six months. The plot of the simulated and observed water level in Figure 12 shows that excellent agreement is obtained.

5. CONCLUSIONS

A user-friendly tidal simulation package named QuickTIDE is presented. The application is built on MATLAB neural network toolbox with an easy-to-use graphical user interface. It comes with 21 built-in network models corresponding to the locations of the tidal stations under the TON managed by JUPEM. These models readily generate tidal forecast and hindcast within the period of 1994 to 2004 with relatively good accuracy, giving an average correlation coefficient R of 0.96 or better. Using historical data for station Boston Harbour from year 1930 to 2000, it is shown that the built-in models trained on 1-year data can produce reliable estimates up to 70 years if the number of neuron in the hidden layer and the number of input constituents are properly selected. The results are also found to be best for locations with semi-diurnal tide, followed by locations with mixed, semi-diurnal dominant tide and locations with mixed, diurnal dominant tide.

QuickTIDE also allows users to train hourly tidal data of any length. This feature enables users to simulate tides for any location of interest using short-term observations. It is shown that for a data length of 30-day, QuickTIDE is able to produce satisfactory forecast and hindcast with an average R-value of 0.96 within the period of 1994 to 2004 for station Port Klang. For simulation up to 70 years at Boston Harbour, R-values average at 0.92 and 0.90 for forecasting and hindcasting respectively. Better results are obtained in and around the month the training data originates.

It is thus concluded that QuickTIDE tidal simulation package can be potentially used for the following applications:

- To perform instance tidal forecast for up to 70 years using the 21 built-in network models.
- To perform tidal hindcast for up to 70 years using the 21 built-in network models as a way to supplement unavailable pass records or missing data.
- To perform tidal simulation for any location of interest using tidal observations in the order of 30 days.

ACKNOWLEDGEMENT

This work was funded by Institute of Research, Development and Commercialisation (IRDC), UiTM (Project No: 600-IRDC/ST 5/3/1094). The author acknowledges the support extended by Jabatan Ukur dan Pemetaan Malaysia (JUPEM) on the provision of extensive historical tidal data. Appreciation also goes to Lt. Cdr. Zaharuddin Hj Maideen of National Hydrographic Centre for vetting the application. ■

REFERENCES

[1] C.P. Tsai and T.L. Lee, "Back-propagation neural network in tidal level forecasting". Journal of Waterway, Port and Coastal Engineering. Vol. 12, No. 4, pp. 195-202, ASCE, 1999.

[2] T.L. Lee, and D.S. Jeng, "Application of artificial neural networks in tide-forecasting". Ocean Engineering. Vol. 29, pp. 1003-1022, Elsevier, 2002.

[3] T.L. Lee, C.P. Tsai, D.S. Jeng, and R.J. Shieh, "Neural network for the prediction and supplement of tidal record in Taichung, Taiwan". Advanced Engineering Software. Vol. 33, pp. 329-338, Elsevier, 2002.

[4] T.S. Lee, "Back-propagation neural network for long-term tidal predictions". Ocean Engineering. Vol. 31, pp. 225-238, Elsevier, 2004.

[5] S. Rajasekaran, K. Thiruvenkatasamy, and T.S. Lee, "Tidal level forecasting using functional and sequential learning neural networks". Applied mathematical Modelling. Vol. 30, pp. 85-103, Elsevier, 2006.

[6] F.J. Chang and Y.C. Chen, "Estuary water-stage forecasting by using radial basis function neural network". Journal of Hydrology. Vol. 270, pp. 158-166, Elsevier, 2003.

[7] D. Hainbucher, W. Hao, T. Pohlmann, J. Sundermann and S. Feng, "Variability of the Bohai Sea circulation based on model calculations". J. Marine Systems, Vol. 44, pp. 153-174, 2004.

[8] C.G. Soares, R. Weisse, J.C. Carretero and E. Alvarez, "A 40 years hindcast of wind, sea level and waves in European waters". Proc. Offshore Mechanics and Arctic Engineering, ASME, OMAE2002-28604, June 2002.

[9] W.K. Lee, "Long-term tidal forecasting using back-propagation neural network". Journal – The Institution of Engineers, Malaysia (accepted for publication).

[10] S. Rao and S. Mandal, "Hindcasting of storm waves using neural networks". Ocean Engineering, Vol. 32, pp. 667-684, 2005.

[11] S. Haykin, Neural networks: A comprehensive foundation. 2nd Ed. Prentice Hall, 1999.

[12] K. Swingler, Applying neural networks: A practical guide. Academic Press, 1996.

[13] P.J. Werbos, The roots of backpropagation: From ordered derivatives to neural networks and political forecasting. Wiley Series on Adaptive & Learning Systems for Signal Processing, Communications and Control. Wiley-Interscience, 1994.

[14] H. Demuth, M. Beale, & M. Hagan, Neural network toolbox for use with MATLAB: User's Guide Version 4. The Mathworks, 1992.

PROFILE

Lee Wei Koon

The author is a lecturer in hydraulics and hydrology in Faculty of Civil Engineering, Universiti Teknologi MARA. Graduated from University of Technology Malaysia in year 1999, he pursued his master of engineering in Nanyang Technological University, Singapore as a full time research scholar in coastal engineering and was conferred in year 2002. In the same year, he was awarded the Ir Thean Lip Thong Prize by IEM on outstanding technical paper on engineering subject. His research interest is in coastal dynamics, computational hydraulics and water resources engineering.