H₂Open Journal

IWA PUBLISHING

Application of artificial neural network for predicting water levels in Hooghly estuary, India

Kalyan Kumar Bhar[a],* and Susmita Bakshi[b]

[a] Department of Civil Engineering, Indian Institute of Engineering Science and Technology, Shibpur, Howrah 711103, West Bengal, India

[b] Department of Civil Engineering, Meghnad Saha Institute of Technology, Kolkata 700150, West Bengal, India

*Corresponding author. E-mail: kalyan@civil.iiests.ac.in

Abstract

Hydrodynamic models for morphodynamic studies in estuaries require continuous tidal water level data as boundary conditions. However, for the Hooghly estuary in India, measurement of continuous tidal water elevation data at the most downstream point is a very difficult task because of the remote location and the confluence with the deep sea. The tidal water level data at this station are measured for a half tidal cycle which is not useful for hydrodynamic modeling. However, at other upstream stations, tide water level data are measured continuously. Accordingly, in this study, an attempt is made to generate continuous tidal water level data at the remote station, using the data of the neighboring stations as input to an artificial neural network (ANN) model. A three-layered feed-forward backpropagation (FFBP) network with two hidden layers is selected and five different combinations of input vectors are used. Simulated water level data obtained from each model are compared with the observed data graphically as well as by estimating the standard error parameters. The best model suitable for prediction of continuous tidal elevation during any time of the tidal cycle and applicable throughout the year is then identified. It is found that tidal data from the nearest neighboring station are more suitable for training.

Key words: artificial neural network, Hooghly estuary, tidal water level prediction

Highlights

- Application of ANN to predict tidal water levels at remote stations in an estuary where only 12-h levels can be recorded.
- Predicted water levels can be used in hydrodynamic modeling as a boundary condition.
- This is the first time this technique has been attempted for Hooghly estuary in India.
- The predicted water levels can also be used for forecasting tidal water levels.

INTRODUCTION

In an estuary, studies on characteristics of sediment transport processes and the resulting changes in the bed morphology are essential tasks for operation and management of ports, construction of jetties and other structures, maintenance dredging works to provide adequate navigational depths for ships, etc. Usually, for these purposes, an appropriate two-dimensional or three-dimensional hydrodynamic

model is used to assess the sediment transport phenomena and the resulting morphological changes. However, along with other data like bathymetry, sediment characteristics, etc., such a hydrodynamic model requires continuous tidal water level data over the period of simulation, to be provided at the key sections as boundary conditions. Unfortunately, in the large Hooghly estuary in India, it is extremely difficult to measure the tidal water level at the most downstream station near the sea continuously for 24 h because of the limited accessibility and proximity to the turbulent sea. Accordingly, tidal water level data at this station, which is an important boundary section for hydrodynamic modeling, are available only for daytime (12-h period). However, for neighboring upstream stations, tidal water level data are measured continuously. So, since a continuous tidal water level record at the remote station is essential for hydrodynamic modeling, an appropriate model needs to be developed which can simulate the continuous tidal water level at that station. This may preferably be done using the continuous water level data of the neighboring upstream stations, so that the characteristics of the tidal water wave during spring tides and neap tides, peaks, troughs, lags, etc., are adequately reflected in the simulated tidal water levels. Accordingly, in this study, an attempt is made to simulate the continuous tidal water level data at the remote station, by developing artificial neural network (ANN) models and training them with measured continuous water level data of the neighboring upstream stations as input.

Conventional tidal forecasting using least-squares method for the prediction of long-term tidal levels requires a large number of parameters for harmonic analysis. Least Square Estimation (LSE) plays a major role in harmonic analysis. Cai et al. (2018) and Li et al. (2019) used the Inaction Method (IM), to predict short-term and long-term tidal levels, respectively. A large number of tide gauge data comprising long and continuous records were required for the long-term prediction. ANN models can be used as an alternative to conventional harmonic analysis for forecasting tidal elevation where measured data are limited (Lee & Jeng 2002; Meena & Agrawal 2015; Salim et al. 2015).

ANN models have been successfully used in many studies related to hydraulics of flow. Examples include modeling stage discharge relationships (Jain & Chalisgaonkar 2000; Bhattacharya & Solomatine 2005; Habib & Meselhe 2006) and flood forecasting (Campolo et al. 1999; Shrestha et al. 2004; Peters et al. 2006). Important applications in sediment hydraulics include development of sediment rating curves (Jain 2008), sediment yield prediction (Sarangi et al. 2005), determination of scour (Akib et al. 2014), and sediment transport (Yitian & Gu 2003; Yang et al. 2009; Maanen 2010; Haddadchi et al. 2013); modeling of salinity in estuaries (Rath et al. 2017) and groundwater level predictions (Mohanasundaram et al. 2019).

ANNs have also been attempted for simulation of tidal water level, discharge, and wave height in coastal areas, as substitutes of numerical models. Tsai & Lee (1999) used backpropagation ANN to forecast tidal water level variations, using two advance consecutive tidal data and two corresponding shocks as inputs to obtain one prediction value. Lee & Jeng (2002) used ANN models to predict hourly tidal levels over a long duration using a short-term hourly tidal record. Ustoorikar & Deo (2008) used ANN and genetic programming for filling up gaps in wave height data. Panda et al. (2010) compared the river stage data obtained from an ANN model with a one-dimensional hydrodynamic model. Feed-forward neural network architecture with Levenberg–Marquardt (LM) backpropagation training algorithm was used to train the neural network model using hourly water level data. Chen et al. (2012) applied ANN as an alternative modeling approach to simulate the water stage time-series of the Danshui River estuary in northern Taiwan and compared the results with vertical (laterally averaged) 2D and 3D hydrodynamic models. Hidayat et al. (2014a) developed a tidal discharge forecast model for a site on the Mahakam River. Hidayat et al. (2014b) used an M5 model tree machine-learning technique to reconstruct a disrupted water level record of the Mahakam Delta in Indonesia. Meena & Agrawal (2015) used ANN models to predict the tidal level variations at stations situated on the southwest coast of India and fill the gap for the missing data. A sequential learning radial basis function (RBF) network was used by Yin et al. (2013) for accurate real-time

prediction of tidal level. Lee & Resdi (2014) presented a feed-forward backpropagation (FFBP) network and generalized regression neural network (GRNN) model for prediction of coastal high and low water, which provided effective prediction but was not suitable for one-year simulation. Salim et al. (2015) used FFBP and Nonlinear Auto Regressive with exogenous input (NARX) network to predict hourly tidal level variations at Mangalore, Karnataka, using a week's hourly tidal levels as input and found the NARX network performance better than the FFBP network.

An adaptive variable-structure online sequential extreme learning machine (OS-ELM) was used by Yin et al. (2016) for online tidal level prediction purpose. Tidal prediction simulations were conducted based on the actual measured tidal data and meteorological data and OS-ELM was found effective for short-term tidal predictions in terms of accuracy and rapidness.

In this study, simulation of the continuous tidal water level data at the most remote station in the Hooghly estuary is done with an ANN (FFBP) model using observed continuous tidal water level data of the neighboring upstream stations. This simulated continuous tidal water level data at the remote station is proposed to be used with a hydrodynamic model as one of the important boundary conditions. The simulated water level data may also be used for navigation purposes. The ANN (FFBP) model is selected because of its simplicity of structure, minimum data requirement, good convergence, and acceptable performances.

The following aspects are investigated while developing the models: (i) which of the upstream neighboring stations are suitable for training and whether the inclusion of more stations improves the model's performance; (ii) whether the model's performance is the same for any part of the tidal cycle (i.e., during spring tide and neap tide); (iii) whether 1-day ahead prediction is possible; and (iv) whether increasing the number of lagged inputs actually improves the model's performance.

STUDY AREA

The Hooghly estuary (Figure 1) is the lowest portion of the Ganga–Bhagirathi–Hooghly river system which joins with the Bay of Bengal and lies approximately between 21°31′ N–23°20′ N and 87°45′ E–88°45′ E. Hooghly estuary is a funnel-shaped estuary. Width of this estuary varies from about 25 km at the mouth to about 6 km at the head. The estuary is very shallow with an average depth of 6 m and the maximum depth is about 20 m. It is a well-mixed estuary, and mixing zones of the estuary extend up to Diamond Harbour. The estuary is characterized by the presence of a large number of tidal bars and tidal islands. The large tidal variations, irregular coastlines, the presence of islands, and dredged navigational channels separated by shallow areas make the flow patterns quite complex.

The Hooghly estuary is the gateway to the two major riverine ports in India: the Haldia port and the Kolkata port. The Haldia port is shown in Figure 1, and the Kolkata port is situated further upstream of Diamond Harbour (at the top). The red lines shown in Figure 1 are the shipping channels. On the left side, the Eden Channel and the Jelligham channel take the ships to the Haldia port and the Rangafalla Channel on the right navigates the ships towards Kolkata port.

Operation and management of these two ports are totally dependent on the navigability of these shipping channels. Availability of adequate draft for movement of ships along these channels is largely affected by frequent sediment deposition. This poses a perennial problem to the Kolkata Port Trust (KoPT) Authority, and regular maintenance dredging requires to be employed, which is highly expensive. Disposing the dredged material at appropriate locations in the estuary is also difficult as tidal action brings the disposed dredged material back to the shipping channels. All these tasks require a proper understanding of the hydrodynamics of flow and characteristics of the morphological changes.
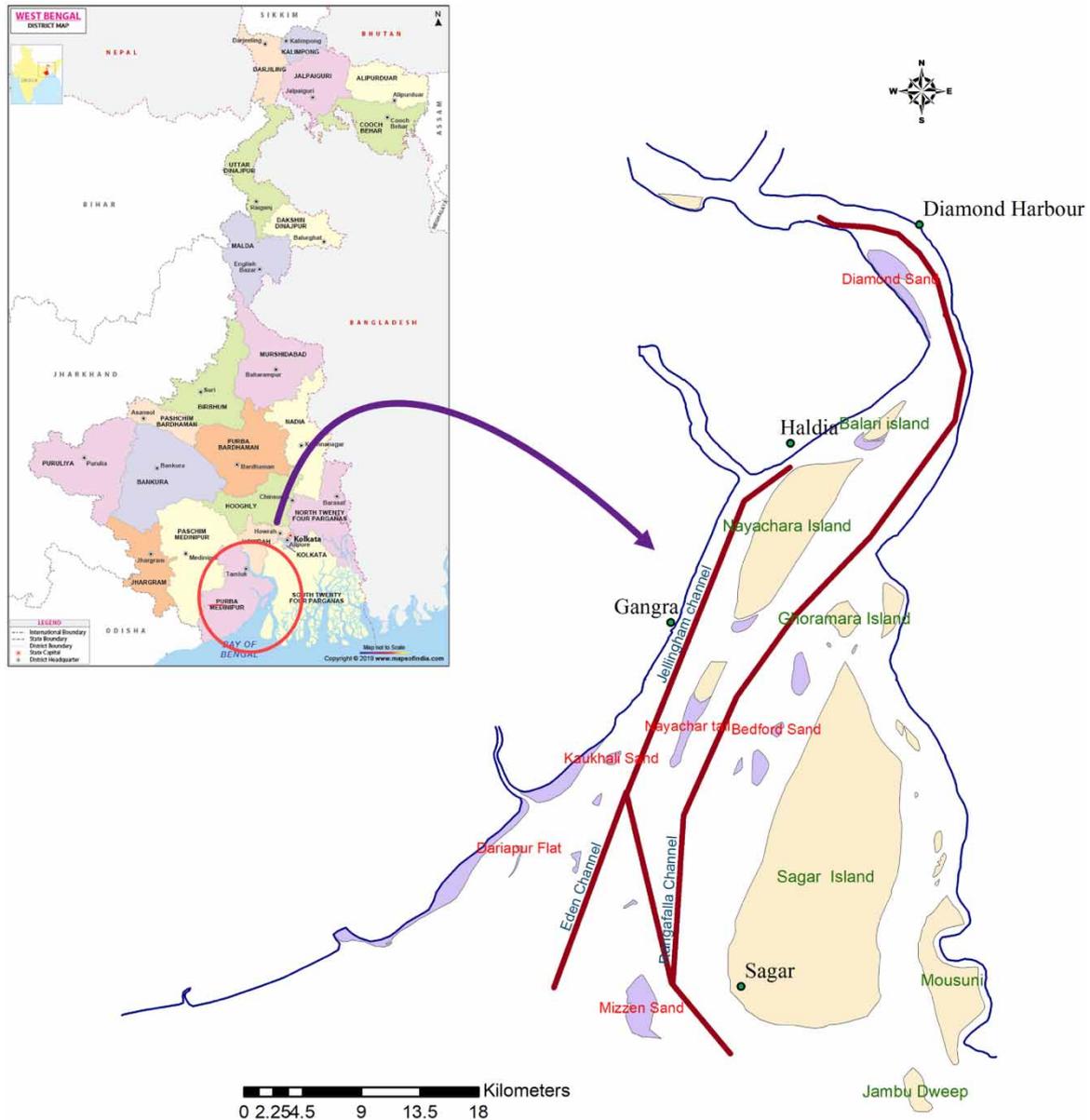
Figure 1 | Study area.

The morphology of the estuary is extremely dynamic with high rates of erosion and accretion, which makes it essential to develop appropriate hydrodynamic models for assessing the morphological behavior of the estuary. These hydrodynamic models require continuous tidal water level data at the key stations as boundary conditions. However, continuous record of tidal water level data is not available at some of the remote stations in the estuary, like Sagar (Figure 1), Dadanpatra and Fraserganj (not shown in Figure 1). The tidal elevation data measured at these stations are usually for a 12-h stretch (daytime). However, at other upstream stations like Gangra, Haldia, Diamond Harbour (Figure 1), tidal water level data are measured continuously for 24 h. Accordingly, in this study, it is attempted to simulate the continuous tidal water level data of the remote gauging station Sagar, situated at the mouth of the estuary, using continuous tidal water level data of the neighboring upstream stations, Gangra, Haldia, and Diamond Harbour, by developing a number of ANN models.

METHODS

ANN models consist of interconnected nodes which can extract complex nonlinear relationships from a set of input and output data. A typical network consists of one input layer, one or more hidden layers, and one output layer. Each layer is made up of interconnected nodes with a set of associated weights. Multilayer feed-forward network (MLFF) is the commonly used ANN model (Maier *et al.* 2010) in water resources applications. The MLFF network has a unidirectional flow of information. The output from nodes in one layer is used as inputs to nodes in the next layer. Different learning rules are available to train the network, e.g., Delta Learning Rule, Memory-based learning rule, Hebbian Learning Rule, and Boltzman Learning Rule. The learning algorithm is the specific mathematical method, used to update the synaptic weights during each training iteration. A set of optimal values of connection weights and thresholds are thus found that minimize a predetermined error function between the ANN model simulated output and the target values. In the present study, the nonlinear least squares LM algorithm is adopted for this training, which has been found to have a faster convergence rate, compared to the standard gradient descent BP algorithm (Habib & Meselhe 2006; Chen *et al.* 2012). As mentioned by Wu *et al.* (2014), the steps followed in ANN model development are a selection of input variables, division of data, determination of best network architecture, model calibration, and finally model validation.

Selection of input variables

Tidal water levels of the three upstream stations at Gangra, Haldia, and Diamond Harbour are selected as the input variables, and the tidal water at Sagar is taken as the output variable. Five different models are developed and tested, which have the following input variables as shown in Table 1. The measured water level data at all four stations are at half-hourly intervals. Hence, in the second column of Table 1, level at $(t-1)$ denotes level at half an hour before t; level at $(t-2)$ denotes level at 1 h before t, and so on. Similarly, level at $(t-48)$ indicates level at 24 h before t, i.e., 1-day before t. It may be mentioned here that Gangra is the nearest station and Diamond Harbour is the farthest station from Sagar.

Table 1 | Input variables used in different models

Model	Input variables	Model architecture
I	Water level at Gangra at current time, t	1-3-3-1
II	Water levels at Gangra, Haldia, and Diamond Harbour at current time, t	3-5-5-1
III	Water levels at Gangra at time t, $(t-1)$, $(t-2)$, and $(t-3)$	4-6-6-1
IV	Water levels at Gangra at time $(t-48)$, $(t-49)$, $(t-50)$, and $(t-51)$	4-7-7-1
V	Water levels at Gangra at time, t, $(t-1)$, $(t-2)$, ..., $(t-12)$	13-9-9-1

Performance evaluation of the models

Performances of the developed ANN models are measured with respect to the following statistical parameters, which are calculated from the predicted values of ANN and the target values.

Root mean square error (RMSE)

It is one of the most widely used functions for assessing the performance of a predictive model and is defined as the square root of the average sum of squares of the difference between the target and

predicted values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_i - P_i)^2} \tag{1}$$

where T_i is the target data, P_i is the corresponding value predicted by the ANN model, and n is the number of data.

Nash–Sutcliffe model efficiency coefficient (E)

This coefficient, E, measures the differences between the observations and predictions relative to the variability in the observed data itself.

$$E = 1 - \frac{\sum_{i=1}^{n}(T_i - P_i)^2}{\sum_{i=1}^{n}(T_i - \bar{T})^2} \tag{2}$$

where $\bar{T}$ denotes the mean of T_i. As per Equation (2), E may range from $-\infty$ to 1.0. A value of $E = 1.0$ indicates a perfect model, whereas $E = 0.0$ indicates that the observed mean is as good as the model predicted, and $E < 0$ indicates that the model is worse than using mean as a predictor.

Correlation coefficient (R)

This is a measure of the linear correlation between the target values and predicted values:

$$R = \frac{n\left(\sum_{i=1}^{n}T_iP\right)_i - \left(\sum_{i=1}^{n}T_i\right)\left(\sum_{i=1}^{n}P_i\right)}{\sqrt{\left[n\sum_{i=1}^{n}T_i^2 - \left(\sum_{i=1}^{n}T_i\right)^2\right]\left[n\sum_{i=1}^{n}P_i^2 - \left(\sum_{i=1}^{n}P_i\right)^2\right]}} \tag{3}$$

Value of R ranges between 0 and 1, with a value of 1 indicating a perfect linear relation.

Mean absolute percentage error (MAPE)

This is an unbiased statistical measure for the predictive capability of a model and is a measure of the relative error in prediction with respect to the actual value of the variable. MAPE calculates the error as a percentage of the actual value.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|T_i - P_i|}{T_i} \times 100 \tag{4}$$

The combined use of R, RMSE, MAPE, and E helps to assess the performance of a model and compare the accuracy of any two modeling approaches. Statistically, ANN performance can be considered good when the values of R, MAPE, MSE, and RMSE are close to 1, 0, 0, and 0, respectively (Habib & Meselhe 2006).

Network architecture

A typical MLP has an input layer, an output layer, and one or two 'hidden' layers of neurons to connect the input and output layers. The neurons are connected by excitation functions whose weights and biases are estimated during training. The best network architecture for an ANN model is

determined considering the number of input variables used. Appropriate training parameters are selected and then the network is trained using pre-processed training data. Identification of the best network architecture involves determining the number of hidden nodes, the number of iterations, the learning rate, and the momentum coefficient. This is done by trial and error, based on the mean square error (MSE) value. Parameters like learning rate, epoch, show, goal, and momentum coefficient are used to control the rate of convergence and are obtained by trial and error, based on the MSE values. In this study, a learning rate of 0.1 is used in all models. A maximum epoch of 1,500 is assigned during training and testing of the ANN models. The value of show is taken as 500, that is, the training result is shown after every 500th iteration. In all models, a momentum coefficient of 0.3 is used. For convergence, the acceptable value of MSE is taken as 10^{-6} which is termed as the goal.

Most of the reported works have used one or two hidden layers. The number of neurons in the hidden layer of an MLP determines the potential complexity of the model estuaries (Hidayat *et al.* 2014a; Rath *et al.* 2017). In this study, it is found that using two hidden layers yielded much better results than using a single hidden layer, keeping the total number of neurons in the hidden layers the same. So, two hidden layers with an equal number of nodes in each hidden layer are adopted for further use. A three-layer fully connected feed-forward network (Figure 2) is thus used in all the models with two hidden layers and a single output layer, and different activation function for different layers. Tan-sigmoid function is used in the first hidden layer, log-sigmoid function is used in the second hidden layer, and pure-line function is used in the output layer.
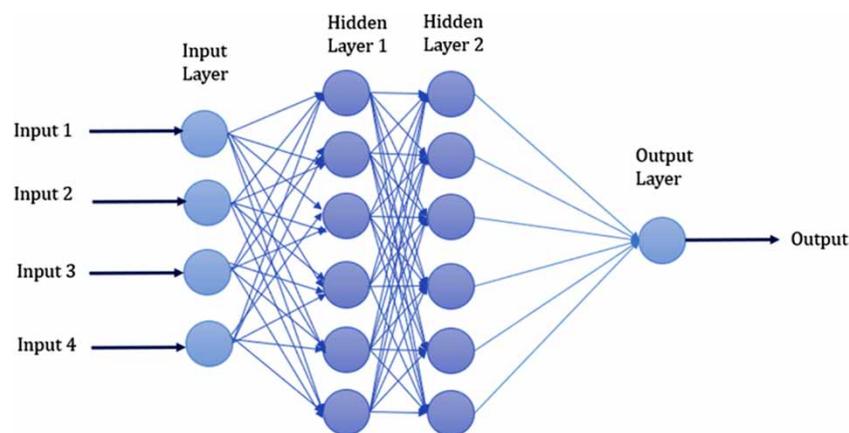

Figure 2 | Network architecture (Model III).

The optimum number of nodes in the hidden layer is selected by varying them and checking the improvement in the *R* value (Figure 3).

In Table 1, the column 'Model architecture' indicates the type of layers and nodes used in the particular model. For example, in Model III, the term '4-6-6-1' indicates four inputs in the input layer, two hidden layers with six nodes, and one output layer. In the present study, design of the network, its training and simulation are performed using the Neural Network Toolbox MATLAB 13.0 developed by MathWorks®, Inc.

The backpropagation algorithm requires initial values of layer weights and biases to start the training process. In the Neural Network Toolbox, the layer weights and biases are initialized using the Nguyen-Widrow method. Once a network is properly trained with acceptable levels of accuracy using a known dataset of input and target, it can be used for simulation. This implies that the trained network can be used to predict possible outputs for any other dataset of input, not used in the training
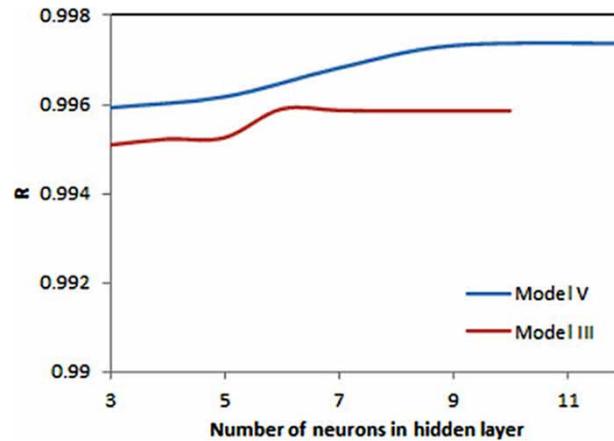
Figure 3 | Effects of variation in the number of neurons in the hidden layer.

process. Levels of acceptance of the predicted output can be verified from the parameters R, RMSE, MAPE, and E, as defined in Equations (1)–(4).

Dataset

Three different datasets are used in the study. Dataset 1 and Dataset 2 contain 14 days' tidal water level data measured continuously at an interval of 0.5 h at all the four stations: Gangra, Haldia, Diamond Harbour, and Sagar. Dataset 1 was for spring tide measured during April 2014 and Dataset 2 was for neap tide measure during January 2015. On the other hand, Dataset 3 (spring tide measured during November 2016) contains 14 days' tidal water level data measured continuously at an interval of 0.5 h at stations Gangra, Haldia, and Diamond Harbour, but only 12-h data (during daytime) at an interval of 0.5 h over a period of 14 days measured at station Sagar. So, Dataset 3 contains half-cycle data at Sagar, whereas Dataset 1 and Dataset 2 contain full-cycle data at Sagar. In fact, Dataset 1 and Dataset 2 are the only two available datasets with full-cycle data at Sagar.

In this study, Dataset 1 is used for training the networks, Dataset 2 is used for validation of the network outputs, and Dataset 3 is used for simulation (prediction). To improve the forecasting capability of ANN outside the range of training datasets, the input data are generally normalized in the range of 0 to +1 or −1 to +1 (Kisi 2004; Panda *et al.* 2010). Standardization of the input and output vector is, therefore, done to keep the data within the interval of [−1, 1].

RESULTS AND DISCUSSION

During training with Dataset 1, following R values are obtained from the five models: (i) $R = 0.8732$ for Model I; (ii) $R = 0.8667$ for Model II; (iii) $R = 0.9993$ for Model III; (iv) $R = 0.9907$ for Model IV; and (v) $R = 0.9994$ for Model V. Since Model I and Model II yielded low values of R, these two models were not considered for further analysis. Since Model I contains only one data as input, it might not be able to represent the variabilities in the output water level. Model II has three input variables which include data of stations which are located further upstream and away from the station Sagar and in these stations, there are appreciable lags in the tidal data and the tidal ranges are also larger compared to that of Sagar. These might have resulted in a lower R value.

Accordingly, it may be concluded that not all of the upstream stations are suitable for training and inclusion of more stations does not necessarily improve the performance of the model.

Next, validation is done for Models III, IV, and V, with Dataset 2. The statistical parameters R, MAPE, RMSE, and E estimated from the results of this validation are shown in Table 2.

Table 2 | Statistical parameters estimated from validation and prediction results

Model	Validation with Dataset 2				Prediction with Dataset 3			
	R	MAPE	RMSE	*E*	*R*	MAPE	RMSE	*E*
III	0.9336	15.91	0.5184	0.8212	0.9495	20.29	0.5994	0.8373
IV	0.9256	15.57	0.5053	0.8246	0.9297	14.94	0.7914	0.7527
V	0.9349	15.96	0.6096	0.7793	0.9551	11.75	0.3654	0.9103

It may be seen from Table 2 that all three models yielded very close values of parameters *R*, MAPE, RMSE, and *E*, with Model IV resulting in marginally better values.

The water levels obtained from these three models with Dataset 2 are plotted with the corresponding observed levels at Sagar, and these are shown in Figures 4(a), 5(a), and 6(a), for Models III, IV, and V, respectively. It may be observed from these figures that all the three models resulted in water levels very close to the observed levels. However, the peak values of the simulated levels are a little lower than the observed levels. One probable reason for this could be that the Dataset 1 which is used in training, contains the water level data of spring tides, whereas Dataset 2 used in validation has the water level data of neap tides (lower peaks). The purpose behind the use of these two datasets of different tides (spring and neap) was to examine whether the trained model could be used to predict any part of the tidal cycle (spring or neap), which was one of the objectives of this study. Based on the results, it may be stated that the trained network can be used to predict any part of the tidal cycle satisfactorily.
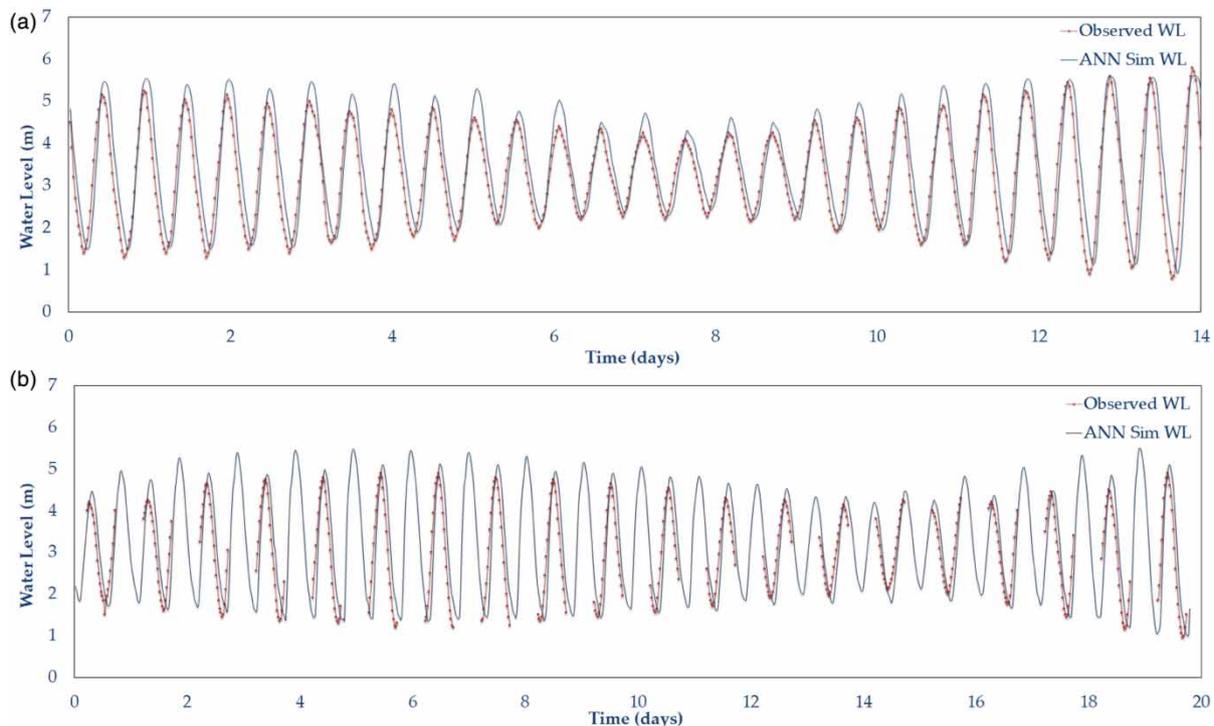
Figure 4 | Comparison of observed and ANN predicted level by Model III: (a) validation with Dataset 2 and (b) simulation with Dataset 3.

All the three models are considered for simulation of the discontinuous Dataset 3. The simulated water levels are plotted with the observed water levels at Sagar for assessing the performances of the three models in terms of predicting full-cycle data. These plots are shown in Figures 3(b), 4(b), and 5(b) for Models III, IV, and V, respectively.
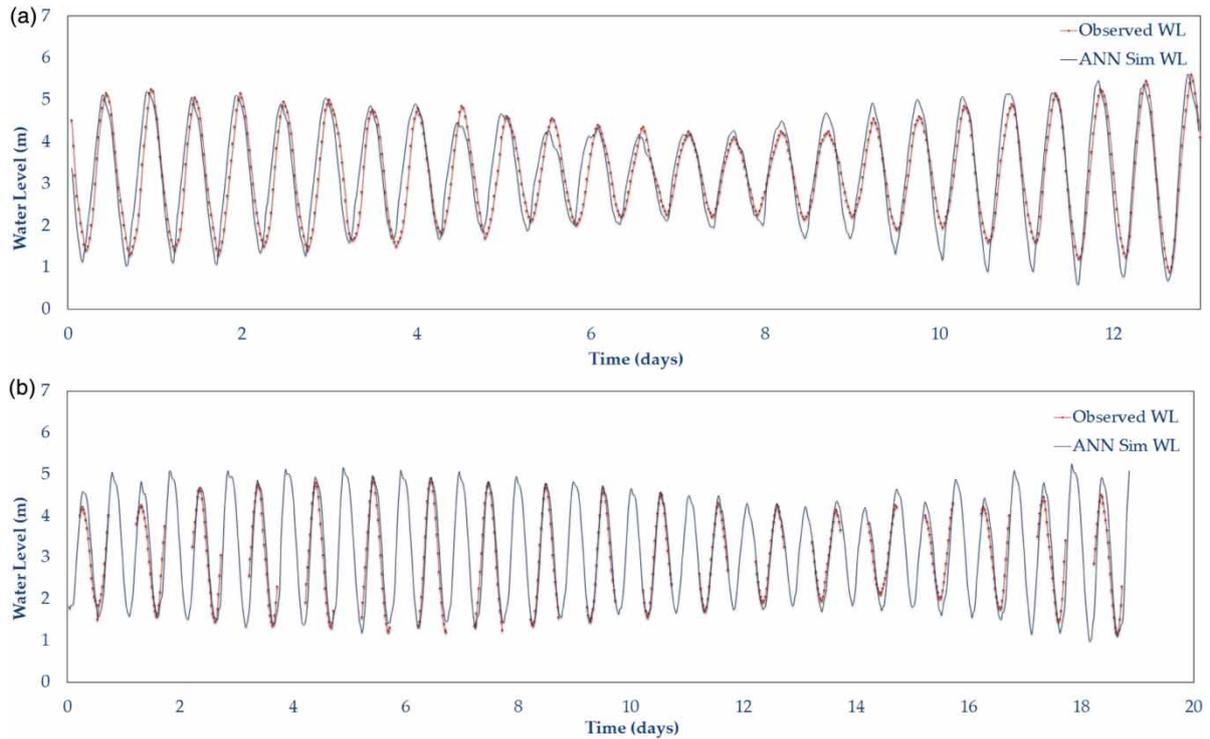
Figure 5 | Comparison of observed and ANN predicted Level by Model IV: (a) validation with Dataset 2 and (b) simulation with Dataset 3.
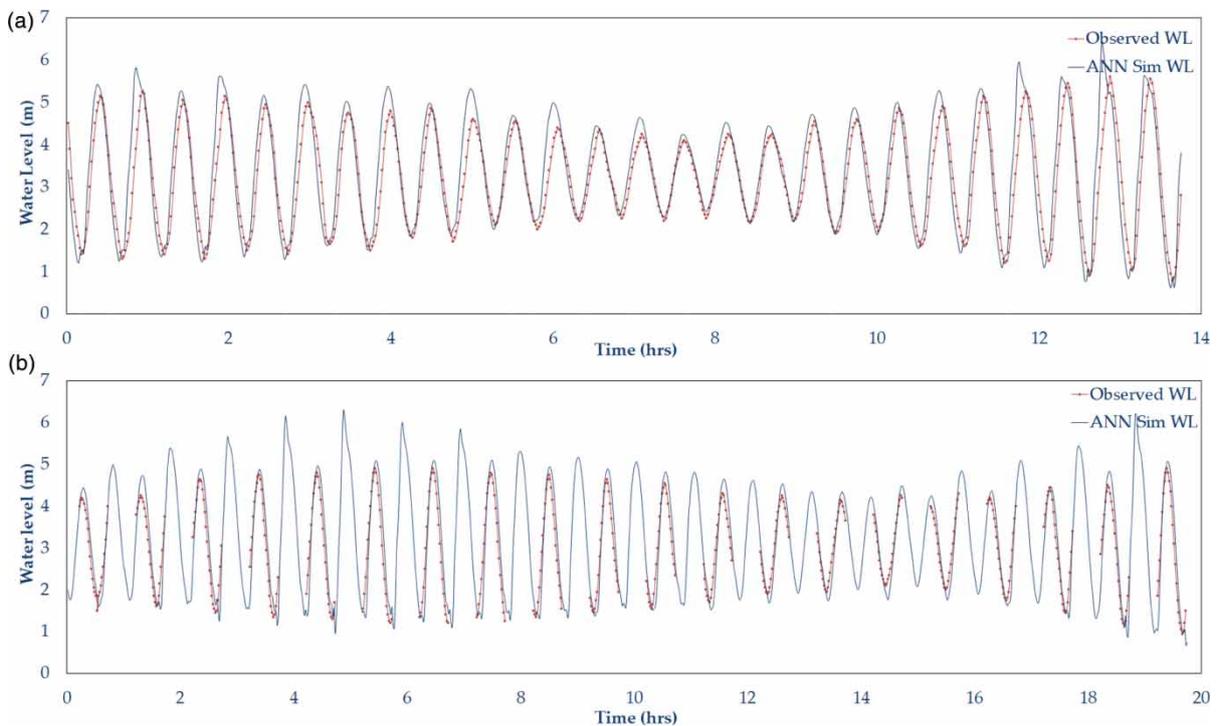

Figure 6 | Comparison of observed and ANN predicted level by Model V: (a) validation with Dataset 2 and (b) simulation with Dataset 3.

Scatterplots of observed vs. ANN predicted levels are developed for all the three models, and these are shown in Figure 7 for Dataset 2 and in Figure 8 for Dataset 3. It may be observed that in case of Model III, the scatter and dispersion around the regression line is less, and the points are more
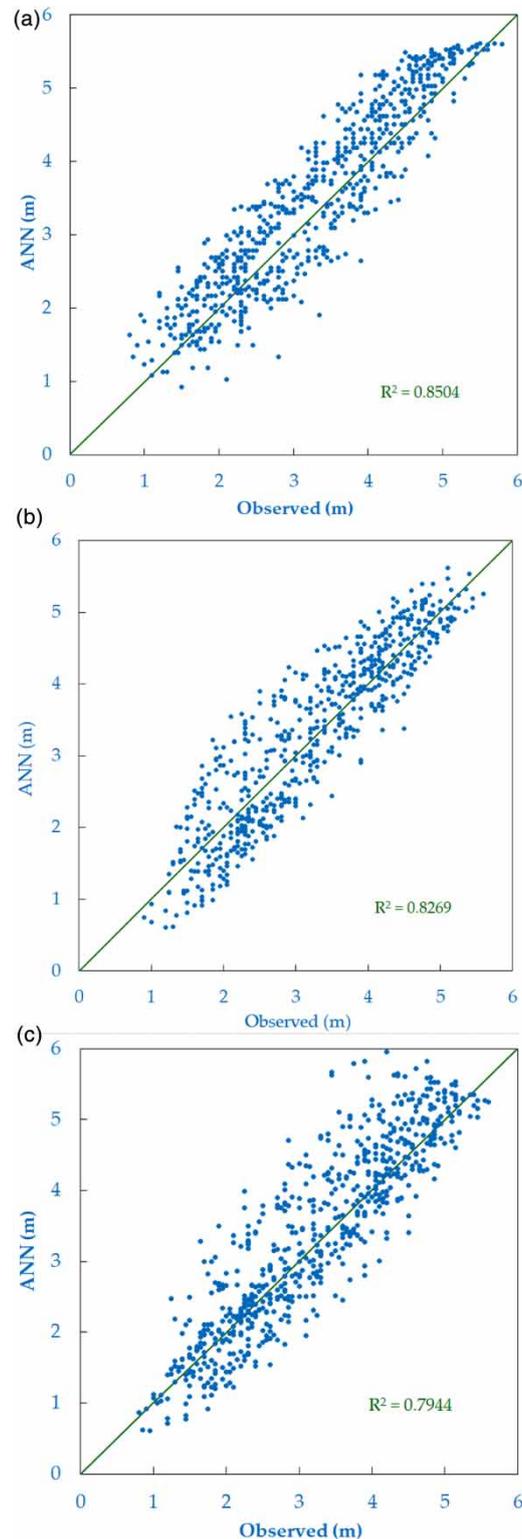
Figure 7 | Scattered plots with Dataset 2 and ANN results: (a) Model III, (b) Model IV, and (c) Model V.

cohesive compared to those of Model IV and Model V. Scatter is more prominent in the case of Model V. The situation is reversed in case of simulation with Dataset 3 (Figure 8), with Model V having more cohesive scatter, closer to the regression line. Consistent and low scatter is observed for Model IV, in both Figures 7 and 8.
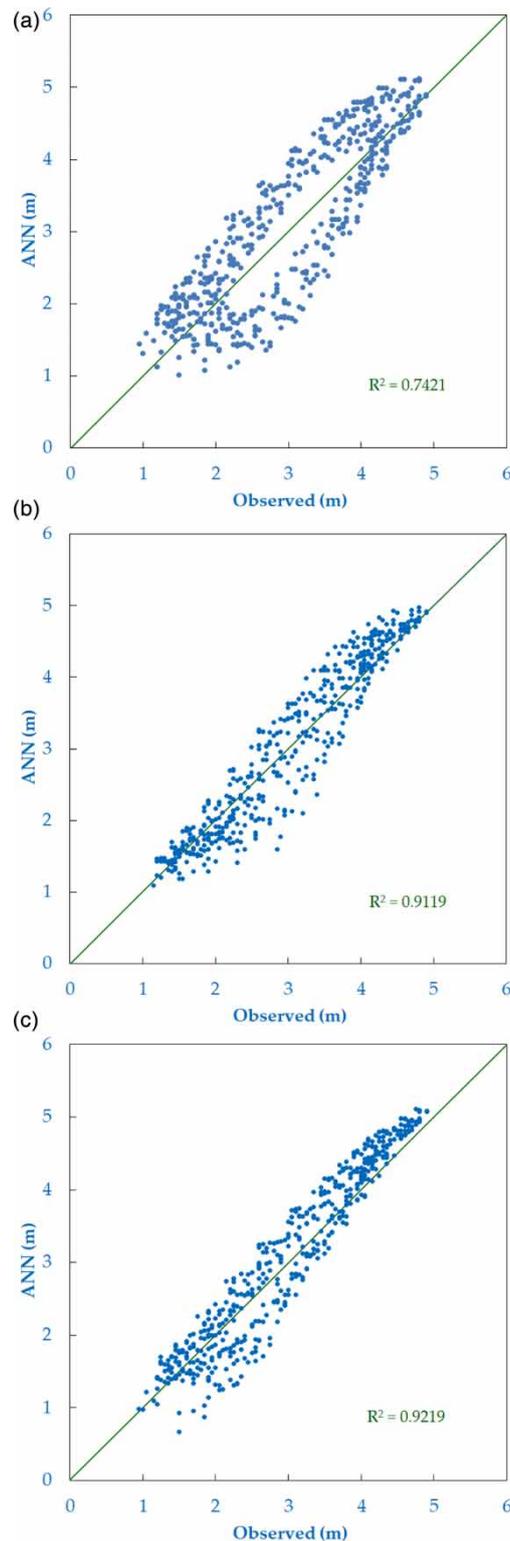
Figure 8 | Scattered plots with Dataset 3 and ANN results: (a) Model III, (b) Model IV, and (c) Model V.

It may be noted that, both Model III and Model IV have four input variables, and in terms of statistical parameters mentioned in Table 2, Model IV apparently performed better. Considering the variations in the tidal range, as well as high water levels and low water levels (peaks and troughs), it can be seen from Figure 5 that Model IV performed better. Model III marginally overpredicted the high water levels compared to Model IV.

Model V overpredicted the high water levels (Figure 6(b)), more than that of Model III (Figure 4(b)). It may be noted that Model V has input variables similar to those of Model III but has nine additional variables. Inclusion of these additional variables was expected to improve the performance, but it did not happen. This indicates that the increasing number of variables does not necessarily improve the performance of an ANN model.

Based on the validation and simulation results, Model IV may be considered as the preferred one, with the advantage of having 1 day before data (i.e., already observed data) of the neighboring station as input. Hence, it may also be used for 1-day ahead forecast.

CONCLUSION

In this paper, an ANN-based approach is presented for prediction of complete cycles of tidal water levels at a remote station in an estuary where only half-cycle data can be physically measured. Based on the objectives of this study, the following conclusions made from the results obtained.

- Comparisons of the graphical plots of observed and predicted tidal water levels, and the comparisons of statistical errors parameters (R, MAPE, RMSE, and E) estimated from the predicted and observed tidal water levels, indicated that the developed ANN models may be accepted for tidal water level prediction.
- Water level data of the nearest upstream station are more suitable for training. Inclusion of data of additional upstream stations further away from the station concerned not necessarily improve the performance of the model.
- The developed ANN models can be used to generate complete cycles of water level data at the remote station, for both spring and neap cycles.
- Performance of Model IV indicated that 1-day ahead prediction is possible.
- Performance of Model V indicated that increasing the number of lagged input variables does not improve the performance of the model.
- The generated tidal data can be used as a boundary condition in the hydrodynamic models.

However, it should be noted that, because ANN is a data-driven model, the above conclusions are primarily related to the system under study. Also, it should be mentioned here that availability of adequate and synchronous data at different stations is a concern. In this study, only two sets of continuous tidal water level data were available at Sagar, which were used in training and validation. With more continuous dataset available, additional training or validation could have been done. Data during the monsoon period were not available because the sea becomes very rough and there are administrative restrictions in movements at that time. Nevertheless, this study successfully identified the applicability of ANN in such field conditions.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Akib, S., Mohammadhassani, M. & Jahangirzadeh, A. 2014 Application of ANFIS and LR in prediction of scour depth in bridges. *Computers and Fluids* **91**, 77–86.

Bhattacharya, B. & Solomatine, D. P. 2005 Neural networks and M5 model trees in modelling water level-discharge relationship. *Neurocomputing* **63**, 381–396.

Cai, S., Liu, L. & Wang, G. 2018 Short-term tidal level prediction using normal time-frequency transform. *Ocean Engineering* **156**, 489–499.

Campolo, M., Andreussi, P. & Soldati, A. 1999 River flood forecasting with a neural network model. *Water Resources Research* **35** (4), 1191–1197.

Chen, W. B., Liu, W. C. & Hsu, M. H. 2012 Comparison of ANN approach with 2D and 3D hydrodynamic models for simulating estuary water stage. *Advances in Engineering Software* **45**, 69–79.

Habib, E. H. & Meselhe, E. A. 2006 Stage-discharge relations for low-gradient tidal streams using data-driven models. *Journal of Hydraulic Engineering* **132** (5), 482–492.

Haddadchi, A., Mohavedi, N., Vahidi, E., Omid, M. H. & Dehghani, A. A. 2013 Evaluation of suspended load transport rate using transport formulas and artificial neural network models (Case study: Chelchay catchment). *Journal of Hydrodynamics, Series B* **25** (3), 459–470.

Hidayat, H., Hoitink, A. J. F., Sassi, M. G. & Torfs, P. J. J. F. 2014a Prediction of discharge in a tidal river using artificial neural networks. *Journal of Hydrologic Engineering* **19** (8), 04014006-1-8.

Hidayat, H., Setiawan, F. & Handoko, U. 2014b Reconstructing disrupted water level records in a tide dominated region using data mining technique. *Indonesian Journal of Geospatial* **3** (2), 9–16.

Jain, S. K. 2008 Development of integrated discharge and sediment rating relation using a compound neural network. *Journal of Hydrologic Engineering* **13** (3), 124–131.

Jain, S. K. & Chalisgaonkar, D. 2000 Setting up stage-discharge relations using ANN. *Journal of Hydrologic Engineering* **5** (4), 428–433.

Kisi, O. 2004 River flow modeling using artificial neural networks. *Journal of Hydrologic Engineering* **9** (1), 60–63.

Lee, T. L. & Jeng, D. S. 2002 Application of artificial neural networks in tide forecasting. *Ocean Engineering* **29**, 1003–1022.

Lee, W.-K. & Resdi, T. A. B. T. 2014 Neural network approach to coastal high and low water level prediction. R. Hassan, M. Yusoff, Z. Ismail, N. Amin & M. Fadzil (eds). Proceeding of the International Civil and Infrastructure Engineering Conference (InCIEC 2013), Kuching, Malaysia, pp. 121–125.

Li, S., Liu, L., Cai, S. & Wang, G. 2019 Tidal harmonic analysis and prediction with least-squares estimation and inaction method. *Estuarine, Coastal and Shelf Science* **220**, 196–208.

Maanen, B. 2010 The use of artificial neural networks to analyze and predict alongshore sediment transport. *Nonlinear Processes in Geophysics* **17**, 395–404.

Maier, H. R., Jain, A., Dandy, G. C. & Sudheer, K. P. 2010 Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environmental Modelling Software* **25**, 891–909.

Meena, B. L. & Agrawal, J. D. 2015 Tidal level forecasting using ANN. *Procedia Engineering* **116**, 607–614.

Mohanasundaram, S., Suresh Kumar, G. & Narasimhan, B. 2019 A novel deseasonalized time series model with an improved seasonal estimate for groundwater level predictions. *H2Open Journal* **2** (1), 25–44.

Panda, R. K., Pramanik, N. & Bala, B. 2010 Simulation of river stage using artificial neural network and MIKE11 hydrodynamic model. *Computers and Geosciences* **36**, 735–745.

Peters, R., Schmitz, G. H. & Cullmann, J. 2006 Flood routing modelling with artificial neural networks. *Advances in Geosciences* **9**, 131–136.

Rath, J. S., Hutton, P. H., Chen, L. & Roy, S. B. 2017 A hybrid empirical-Bayesian artificial neural network model of salinity in the San Francisco Bay-Delta estuary. *Environmental Modelling & Software* **93**, 193–208.

Salim, A. M., Dwarakisha, G. S., Liju, K. V., Thomas, J., Devi, G. & Rajeesh, R. 2015 Weekly prediction of tides using neural networks. *Procedia Engineering* **116**, 678–682.

Sarangi, A., Madramootoo, C. A., Enright, P. & Prasher, S. 2005 Performance evaluation of ANN and geomorphology-based models for runoff and sediment yield prediction for a Canadian watershed. *Current Science* **89** (12), 2022–2033.

Shrestha, R. R., Theobald, S. & Nestmann, F. 2004 Combined hydrodynamic-neural network model for flood flow simulation. In *Proceedings of the 6th International Conference on Hydroinformatics*, Vol. 2, 21–24 June 2004, Singapore, pp. 1489–1496.

Tsai, C. P. & Lee, T. L. 1999 Back-propagation neural network in tidal level forecasting. *Journal of Waterway, Port, Coastal and Ocean Engineering* **125** (4), 195–202.

Ustoorikar, K. & Deo, M. C. 2008 Filling up gaps in wave data with genetic programming. *Marine Structures* **21**, 177–195.

Wu, W., Dandy, G. C. & Maier, H. R. 2014 Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environmental Modelling Software* **54**, 108–127.

Yang, C. T., Marsooli, R. & Aalmi, M. T. 2009 Evaluation of total load sediment transport formulas using ANN. *International Journal of Sediment Research* **24** (3), 274–286.

Yin, J., Zou, Z. & Xu, F. 2013 Sequential learning radial basis function network for real-time tidal level predictions. *Ocean Engineering* **57**, 49–55.

Yin, J., Li, L., Cao, Y. & Zhao, J. 2016 An adaptive online sequential extreme learning machine for real-time tidal level prediction. In: *Proceedings of ELM-2015 Volume 2. Proceedings in Adaptation, Learning and Optimization* (J. Cao, K. Mao, J. Wu & A. Lendasse, eds), Hangzhou, China, pp. 55–66.

Yitian, L. & Gu, R. R. 2003 Modeling flow and sediment transport in a river system using an artificial neural network. *Environmental Management* **31** (1), 122–134.