Desalination and Water Treatment
www.deswater.com

doi:10.5004/dwt.2018.23054

132 (2018) 75–78
November

Empirical modelling of total suspended solids from turbidity by polynomial neural network in North Eastern India

Ritabrata Roy[a,*], Mrinmoy Majumder[b]

Hydro-Informatics Engineering (under Civil Engineering Dept), National Institute of Technology Agartala, Barjala, Jirania, Tripura (W) – 799046, India, Tel. +918414934874, email: ritroy@gmail.com (R. Roy), Tel. +919612946791, email: mmajumder15@gmail.com (M. Majumder)

Received 30 December 2017; Accepted 29 August 2018

ABSTRACT

Suspended solids in water cause turbidity. Suspended solids reduce dissolved oxygen,hinder the respiration of aquatic organisms and induce bacterial growth. Due to such significant impact on water quality and aquatic ecosystem, turbidity and total suspended solids are important water quality parameters. Now, turbidity of water can be estimated easily and instantly by modern devices, but total suspended solids can be measured only by conventional analytical procedure. Such analytical procedure is troublesome, time consuming and requires laboratory setup. This study aims the modelling of total suspended solids from turbidity through polynomial neural network to save the time, effort and cost of analytical method to estimate total suspended solids. Different polynomial neural network methods were used for the modelling. Field data from 176 water bodies of North Eastern India were used for the training and testing of those models. An empirical equation was developed from those models which can calculate total suspended solids from turbidity. The equation was tested to be valid for its high correlation (correlation coefficient and coefficient of determination = 0.999) and very low deviation (mean absolute percentage error = 3.477 or less and root mean square error = 1.904 or less) from the actual values.

Keywords: Total suspended solids; Turbidity; Modelling; Polynomial neural network; North Eastern India; India

1. Introduction

Suspended solids in water has a considerable impact on the quality of water and aquatic ecosystem. Suspended particles reduce the dissolved gases in water to some extent. As such particles block the sunlight from penetration into water; photosynthesis by the aquatic plants also gets reduced [1]. Suspended particles also absorb more heat to increase the temperature of water, which, in turn, reduces the capacity of water to dissolve gases. Thus, presence of suspended particles in water causes decrease in dissolved oxygen. Low level of dissolved oxygen makes the respiration of aquatic organisms difficult and may also cause their death. So, suspended solids are one of the most harmful factors for aquatic ecosystem [2] and may cause significant depletion of aquatic biota [3]. Suspended particles also hinder respiration of aquatic organisms and induce bacterial growth and sustenance [4].

Turbidity of water is for the particles suspended in water. Turbidity is estimated by the scattering of light in a water sample, which can be accomplished easily and instantly by modern sensor based devices. Total suspended solids (TSS), on the other hand, cannot be measured so easily as it requires filtration, drying and weighing of water samples. Alternatively, total dissolved solids (TDS) can be estimated by sensor based devices and Total Solids (TS) can be measured by evaporation of water. Then deduction of TDS from TS will give the amount of TSS. Both the procedures are not

*Corresponding author.

only troublesome and time consuming, but also require a standard water quality laboratory setup. As turbidity and TSS are, in fact, different aspects of the same characteristic, they are highly correlated. So, TSS can be effectively modelled by turbidity to predict it easily and instantly [5].

Polynomial neural network (PNN) is a computational model inspired by the schematic functioning of the biological neural network. It is an unsupervised machine learning model, which can be trained with a set of inputs and their corresponding outputs [6]. A relation between input(s) and output(s) is developed from iteration modeling by putting adaptive weights in the multiple neural layers [7]. PNN can be used to predict the output from input(s) when the relation between input and output is not known as it can develop a mathematical relation (a model) between the given set of input(s) and output(s) through regression analysis [8]. PNN has been used for forecasting of water quality [9] and prediction of a WQP from other WQP [10–12] including TSS [13].

176 water bodies of the North Eastern India were sampled to estimate turbidity and TSS. The water bodies are situated in the Indian states of Tripura, Manipur, Meghalaya and Assam. The data were then fed into PNN for the empirical modelling of TSS.

2. Methodology

Turbidity and TSS of 176 water bodies in North Eastern India were estimated for the modelling purpose. These field data were used for training and validation of the PNN models.

2.1. Collection of data

Turbidity and TDS were measured with sensor based device (Horiba U50 Multiparameter Water Quality Analyzer). TS were measured by conventional evaporation method. TSS were then calculated by deducting TDS from TS [14].

2.2. Correlation analysis

Though it is known that turbidity and TSS are closely related as TSS is the main cause of turbidity, still correlation between the two was analyzed for the sake of accuracy of the methodology. The correlation between turbidity and TSS was measured by Pearson product-moment correlation coefficient method [15].

2.3. Development of prediction model

Different PNN algorithms like combinatorial (COM), stepwise forward selection (SFS), stepwise mixed selection (SMS) and GMDH neural network (GNN) were used for development of different models for the prediction of TSS from turbidity.

In COM method, the model is developed by forming polynomial functions of linear parameters [16]. SFS is a regression analysis where a model is being optimized by inclusion of suitable variables until it becomes statistically significant [17]. SMS, on the other hand, optimizes a model by both inclusion and exclusion of variables [16]. GNN develops a model by iteratively creating layers of neurons where the neurone connections are being optimized by COM algorithm [16]. All these algorithms being automated regression analysis, they can be used to rapid development of optimized PNN models inattentively.

The accuracy of the PNN Models were determined by different statistical methods like mean absolute percentage error (MAPE) [Eq. (1)], root mean square error (RMSE) [Eq. (2)], coefficient of determination (R^2) [Eq. (3)] and correlation coefficient (r) [Eq. (4)] of the predictions with actual data.

$$\mathit{Mean\,Absolute\,Percentage\,Error}\left(MAPE\right)=\frac{100}{n}\sum_{i=1}^{n}\left|\frac{\mathrm{A_i}-\mathrm{P_i}}{\mathrm{A_i}}\right| \quad (1)$$

$$\mathit{Root\,Mean\,Square\,Error}\left(RMSE\right)=\sqrt{\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{P_i}-\mathrm{A_i}\right)^2}{\mathrm{n}}} \quad (2)$$

$$\mathit{Coefficient\,of\,Determination}\left(R^2\right)=1-\frac{\sum_{i=1}^{n}\left(A_i-P_i\right)^2}{\sum_{i=1}^{n}\left(A_i-\overline{A}\right)^2} \quad (3)$$

$$\mathit{Correlation\,Coefficient}\left(r\right)=\frac{\sum_{i=1}^{n}\left(A_i-\overline{A}\right)\left(P_i-\overline{P}\right)}{\sqrt{\sum_{i=1}^{n}\left(A_i-\overline{A}\right)^2\sum_{i=1}^{n}\left(P_i-\overline{P}\right)^2}} \quad (4)$$

where $\overline{A}=\frac{1}{n}\sum_{i=1}^{n}A_i$, $\overline{\mathrm{P}}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{P_i}$, n = number of samples, A = actual values, P = predicted values.

2.4. Validation of models

80% of the field data were used for the training of the PNN Models while 20% were used for testing purpose. Significant correlation between actual and predicted values of TSS ensured the validation of the models.

3. Results and discussion

3.1. Correlation analysis

The correlation coefficient between turbidity and TSS values of the sampled water bodies was found 0.97. This is a very high value to establish the correlation between those parameters statistically.

3.2. Development of prediction model

The predicted TSS values were found to be very close to their actual values during both training (Fig. 1) and testing (Fig. 2) phases of the development of PNN Models. All the model predictions were very accurate with high correlations and negligible deviations from the actual values (Table 1 and Table 2).

During training of the PNN, predictions from all the models were found be highly correlated with the actual (R^2 = 0.96, r = 0.98) and having little deviation (MAPE = 4.04–4.08, RMSE = 2.32–2.34) (Table 1).

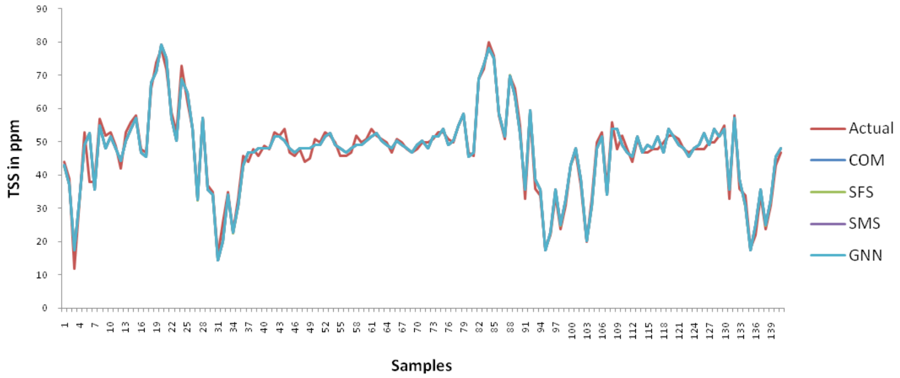

Fig. 1. Actual and predicted values of TSS were very close during training of ANN models.

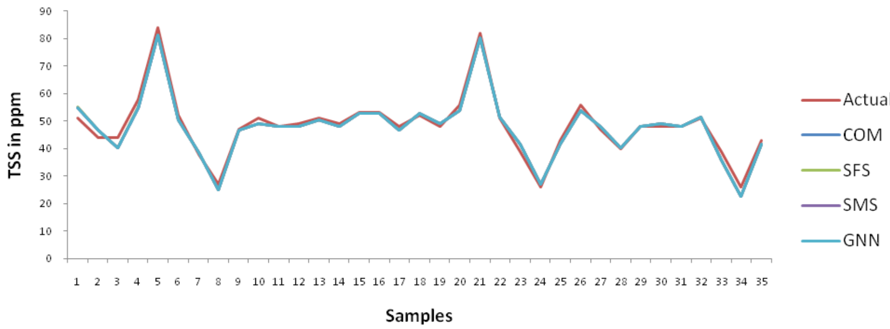

Fig. 2. Actual and predicted values of TSS were very close during testing of ANN models.

Table 1
Accuracy of the model predictions during training

Model	Mean absolute percentage error (MAPE)	Root mean square error (RMSE)	Coefficient of determination (R^2)	Correlation coefficient (r)
COM	4.07835	2.33589	0.96377	0.98172
SFS	4.07148	2.33580	0.96378	0.98172
SMS	4.03686	2.32420	0.96412	0.98190
GNN	4.07835	2.33589	0.96377	0.98172

Table 2
Closeness between actual and predicted BOD during testing

Model	Mean absolute percentage error (MAPE)	Root mean square error (RMSE)	Coefficient of determination (R^2)	Correlation coefficient (r)
COM	3.31012	1.84581	0.99862	0.99936
SFS	3.31363	1.85307	0.99861	0.99935
SMS	3.47726	1.90433	0.99852	0.99930
GNN	3.31012	1.84581	0.99862	0.99936

3.3. Validation of models

20% of the field data were used for the testing of the PNN Models. Actual data set were compared with the predictions from the PNN Models to find the accuracy of those models.

During testing, the predicted values were found to be even closer to the actual values than those during training. Very high correlation (R^2 = 0.999, r = 0.999) and low deviation (MAPE = 3.310–3.477, RMSE = 1.846–1.904) (Table 2) indicates that predictions are almost same as the actual values (Fig. 2).

3.4. Selection of best model

During both training and testing phases, all the PNN Models were found to be fairly accurate in predicting the TSS from turbidity. The four procedures, viz. COM, SFS, SMS and GNN, were found to predict TSS with almost equal accuracy (Tables 1 and 2). The predictions from all the four procedures were found almost equally correlated and deviating equally from the actual values.

During training, SMS values were found to be least deviated and highest correlated with the actual values (Table1). During testing, however, SMS values were found highest deviated and least correlated and COM and GNN values were least deviated and highest correlated (Table 2). Thus, no single method can be selected as the best method among the four PNN methods used. Any of them can be used for the prediction of TSS from turbidity.

The empirical equations for calculation of TSS from turbidity by different PNN methods are summarized in Eqs. (5)–(8). Eqs. (5)–(8) were generated by PNN methods COM, SFS, SMS and GNN respectively. As Eqs.(5), (7) and (8) are the same; this equation can be used to predict TSS from turbidity.

$$TSS = 0.00411561T^2 + 25.0847T^{1/3} - 25.5003. \quad (5)$$

$$TSS = 0.00376798T^2 + 0.335244T^{2/3} + 23.6626T^{1/3} - 23.9845 \quad 6)$$

$$TSS = 0.00411561T^2 + 25.0847T^{1/3} - 25.5003 \quad (7)$$

$$TSS = 0.00411561T^2 + 25.0847T^{1/3} - 25.5003 \quad 8)$$

where T = turbidity.

Thus, the empirical equations were found valid for prediction of TSS from turbidity. As the field data were sampled over a large spatial and temporal space, it should be accurate over time and in different conditions. However, the accuracy may be decreased if any remarkable change occurs in the present equilibrium of surface water condition in the study area.

4. Conclusion

Empirical equations were developed to predict TSS from turbidity with PNN. As turbidity can be estimated instantly with the aid of sensor based electronic devices, TSS can also be predicted quickly with this empirical equation. Thus, a considerable amount of time, effort and infrastructure (i.e. cost) can be saved if this model is used to predict TSS instead of conventional methods of estimating TSS. However, the equation being empirical, it may not be accurate for a different region. As TSS and turbidity are very closely related, the equation should be accurate in temporal variations and for other sets of data. If this model is used for routine monitoring, the best practice will be to calibrate it in regular intervals with the results from conventional methods.

References

[1] G. Hoetzel, R. Croome, Long-term phytoplankton monitoring of the Darling River at Burtundy, New South Wales: Incidence and significance of cyanobacterial blooms, Marine Freshwater Res., 45 (1994) 747–759.

[2] R.D. JudyJr, P.N. Seeley, T.M. Murray, S.C. Svirsky, M.R. Whitworth, 1982 National Fisheries Survey. Volume I. Technical Report: Initial Findings (No. FWS/OBS-84/06). Engineering-Science Inc Denver Co., 1984.

[3] W.F. Henley, M.A. Patterson, R.J. Neves, A.D. Lemly, Effects of sedimentation and turbidity on lotic food webs: a concise review for natural resource managers, Rev. Fish. Sci., 8 (2000) 125–139.

[4] H. Perlman 2014 Turbidity, In: The USGS Water Science School. Retrieved from http://water.usgs.gov/edu/turbidity.html on 22 December 2017.

[5] J. Packman, K. Comings, D. Booth, Using turbidity to determine total suspended solids in urbanizing streams in the Puget Lowlands, 1999.

[6] J.L. McClelland, D.E. Rumelhart, G.E. Hinton, The Appeal of Parallel Distributed Processing. MIT Press, Cambridge MA, 1986, pp. 3–44.

[7] S. Agatonovic-Kustrin, R. Beresford, Basic concepts of polynomial neural network (PNN) modeling and its application in pharmaceutical research, J. Pharm. Biomed. Anal., 22 (2000) 717–727.

[8] D.F. Specht, A general regression neural network, IEEE Trans. Neural Networks, 2 (1991) 568–576.

[9] R.R. Goyal, H. Patel, S.J. Mane, Polynomial neural network: an effective tool for predicting water quality for Kalyan-Dombivali Municipal Corporation, Int. J. Sci. Res., 4 (2015) 2863–2866.

[10] M. Khandelwal, T.N. Singh, Prediction of mine water quality by physical parameters, J. Scient. Ind. Res., 64 (2005) 564–570.

[11] L. Swathi, B. Lokeshappa, Polynomial neural networks application in prediction of water quality, Int. J.Innov. Res. Sci. Eng. Technol., 4 (2015) 6911–6916.

[12] E. Dogan, B. Sengorur, R. Koklu, Modeling biological oxygen demand of the Melen River in Turkey using a polynomial neural network technique, J. Environ. Manage., 90 (2009) 1229–1235.

[13] A. Verma, X. Wei, A. Kusiak, Predicting the total suspended solids in wastewater: a data-mining approach, Eng. Applic. Artif. Intell., 26 (2013) 1366–1372.

[14] American Public Health Association and American Water Works Association, Standard Methods for the Examination of Water and Wastewater. American Public Health Association (APHA), 20th Ed, Washington, D.C., USA, 1999.

[15] L. Rodgers, W.A. Joseph, Nicewander, Thirteen ways to look at the correlation coefficient, Amer. Stat., 42 (1988) 59–66.

[16] L. Anastasakis, N. Mort, The development of self-organization techniques in modelling: a review of the group method of data handling (GMDH). Research Report 813, Department of Automatic Control & Systems Engineering, The University of Sheffield, UK, 2001.

[17] L.J. Lancashirea, R.C. Reesb, G.R. Ball, Identification of gene transcript signatures predictive for estrogen receptor and lymph node status using a stepwise forward selection Polynomial Neural Network modelling approach, Artif. Intell. Medicine, 43(2008) 99–111.