International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

AN EXPERT SYSTEM FOR PREDICTING AERATION PERFORMANCE OF WEIRS BY USING ANN AND RANDOM FOREST

[1]Mohd Wasi, [2]Dr.Subodh Ranjan, [3]Dr. Mahesh A , [4]Dr.N.K.Tiwari

[1]M .Tech scholar, Department of civil engineering, NIT Kurukshetra,Haryana,(India)

Professor, Department of civil engineering, NIT Kurukshetra, Haryana, (India)

[3]Assistant Professor, Department of Electrical engineering, NIT Kurukshetra,Haryana, (India)

[4]Associate Professor, Department of civil engineering, NIT Kurukshetra,Haryana, (India)

ABSTRACT

The paper investigates the modelling performance of experimentally observed aeration efficiency by multiple plunging jets having piano key weir varying no. of keys(2.5 and 3.5). The output values of aeration efficiency were calculated by using artificial neural network (ANN) and Random Forest techniques. The standard statistical performance evaluation measures, such as the coefficient of correlation (CC) and root mean square error (RMSE) have been utilized to compare the performance of modelling techniques. The coefficient of correlation and root mean square values of 0.9774 and 0.4105.respectively were achieved by ANN in comparison to values of 0.9890 and 0.2576 respectively achieved by Random forest. Furthermore, the performance of both these approaches in predicting the aeration efficiency was compared with the previously observed relationship for piano key weir available in literature. Random Forest gives the best results in comparison of ANN and two empirical equations.

Keywords: Artificial neural network, Piano key weir, Multi-linear Regression.

I INTRODUCTION

Water can hold a limited quantity of oxygen. That is determined by temperature, atmospheric pressure and salinity. In a natural setting, oxygen is added to water by atmospheric diffusion at the surface, circulation of wind and by photosynthesis. Photosynthesis accounts for most of the oxygen in water. The oxygen content of water increases with increasing atmospheric pressure and decreasing temperature and salinity. The amount of oxygen in water is measured as (mg/L) dissolved oxygen (DO). The minimum level of DO in river should not be below 4 mg/L for survival of aquatic life. Total DO concentrations in water should not exceed 110 percent. The concentrations above this level can be harmful to aquatic life. The physical process of transfer of oxygen from the atmosphere acts to replenish the used oxygen. This process is known as aeration. Hydraulic structures increase amount of DO in a river system, although the water is get in touch with the structure for only a short time. The amount of oxygen transfer

International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

that usually would take place over several kilometers in a river can occur at a single hydraulic structure. The most important reason for this accelerated transfer of oxygen is that entrainment of air into the flow in the form of a large number of bubbles. Wilhelms et al. (1992), Chanson (1995), Ervine (1998), and Gulliver et al. (1998) reviewed studies about aeration efficiency of hydraulic structures. Recently, Baylar and Bagatur (2000), Baylar and Bagatur (2001), Baylar et al. (2001), Baylar (2002), Baylar and Emiroglu (2002) and Baylar and Bagatur (2006) investigated sharp−crested weirs having different cross−sectional geometry and demonstrated that the air entrainment rate and the aeration efficiency of weirs changed depending on weir shapes. The two common methods used to develop models for aeration efficiency are ANN and Random forest. One of the advantages of Random forest and ANN develop input- output function relationship. The ANN system is a neural network is an attempt to build a mathematical model that supposedly works in an analogous way to human brain .A network consists of many elements or neurons that are connected by communication channels or connectors. These connectors hold numeric data arranged by a variety of means and organized into layers. Soft computing techniques such as ANN are recently developed methods which can be used for prediction of aeration efficiency. MLR techniques in developing prediction models for estimating the aeration efficiency. It learns features of the data set and adjust the system characteristics according to a given error criterion.

II OXYGEN TRANSFER EFFICIENCY

Oxygen is a very highly volatile compound with rate of air water exchange that is controlled totally by the liquid phase. The change in concentration of oxygen over time in a parcel of water as the parcel travels through a hydraulic structure can be expressed as:

$$\frac{dm}{dt} = V\frac{dC}{dt} = K_L A(C_S - C) \qquad (2.1)$$

where K_L = bulk liquid film coefficient. C_S and C are the saturation concentration of oxygen in water at prevailing ambient conditions and the actual concentration of oxygen respectively in the water at time t−difference being proportional to the concentration gradient. A is the air−water contact area and V is the volume of water related with this. Equation (2.1) does not consider sources and sinks of oxygen in the water body because their rates are comparatively slow compared to the oxygen transfer that occurs at mainly hydraulic structures due to the increase in free−surface turbulence and the large quantity of air that is normally entrained into the flow. The predictive relations assume that C_S is constant and determined by the water−atmosphere partitioning. If that assumption is made, C_S is constant with respect to time, and the aeration efficiency, E may be defined as :

$$E = \frac{C_d - C_u}{C_s - C_u} \qquad (2.2)$$

International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

where u and d = subscripts which indicate upstream and downstream locations, respectively. A value of E > 1 means the downstream water has become supersaturated (i.e.,C_d>C_s). A transfer Efficiency value of 1.0 means that the full transfer up to the saturation value has occurred at the structure. No transfer would correspond to E = 0. For hydraulic structures.Gulliver et al. [1990] applied the theories of Levich [1962],Hinze [1955], and Azbel [1981] to mass transfer similitude and developed the relationship:

$1 - E_{20} = (1 - E)^{\frac{1}{f}}$ (2.3)

where E = efficiency of aeration at the water temperature;

E20 = aeration efficiency at the 20 °C; and f = the exponent described by

$f = 1.0 + 0.02103(T-20) + 8.261 \times 10^{-5}(T-20)^2$ (2.4)

where T = Temperature of water.

III METHODOLOGY AND DATA SET

Experiments were conducted in a hydraulic laboratory located at National Institute of Technology, Kurukshetra, India. The study was conducted in a main channel having a length of 4 m and cross section 25cm wide, 15cm deep. The maximum supplying discharge capacity of the pump is 5.84 l/sec. The discharge in the main channel is regulated by valve which is fitted in the water supply pipe an overhead water tank and discharge as well as the velocity is controlled in the main channel by regulating gate.

Data Set

Data set consisting of 30 observations were used and obtained from the laboratory experiments and used for testing the models. Input data set consists of velocity, No of keys, Head; whereas Aeration efficiency was considered as output. The characteristics of experimental data are specified in Table No. 2.

Table 1: The characteristics of testing data set.

Input	Units	Testing Data			
Parameters		Min	Max	Mean	St.Dev.
Velocity	m/sec	0.3125	1.1320	0.7981	0.28621
No of keys		2.5	3.5	3	0.516
Head	cm	0.96	2.38	1.7969	0.452
E_{20}		1.61	6.51	4.1138	1.6924

IV ARTIFICIAL NEURAL NETWORK

An artificial neuron network (ANN) is a computational model based on the structure and elements of biological neural systems. Data that moves through the system influences the structure of the ANN in light of the fact that a neural system changes - or learns, as it were - based on that input and output. Artificial neural Network (ANN) depend on the present understanding of biological nervous system, however a great part of the biological detail is ignored. ANN are hugely parallel frameworks made out of many processing elements associated by link of variable weights. Of the numerous ANN models, the multi-layer back propagation network(MLP) is the most well known . The system comprises of layers of parallel processing elements, known as neurons, with each layer being completely associated with the proceeding layer by interconnection completely associated with the procedure layer by interconnection strengths. The versatile learning rates were utilized with the for the purpose of faster training speed and solving local minima problem. For every epoch, if performance decreases toward the goal, then the learning rate is increased by the factor learning increment. If performance increases, the learning rate is adjusted by the factor learning decrement. The numbers of hidden layer neurons were found using simple trial error method.The determination of optimal number of hidden layers and hidden neurons is usually cumbersome, as no general methodology is available for their determination. These networks learn from the training data by adjusting the connection weights. A three layer feed forward ANN consists of three layers known as input, hidden and output layers (Goel, 2014). The sum of inputs and their weights lead to a summation function. The output of a neuron is decided by an activation function, which can be step, sigmoid, threshold and linear etc. A large number of trials have been carried out to find the optimal values of parameter of ANN. In the last decade, research into ANNs has shown explosive growth. They are often applied in physics research like speech recognition (Chu, 1998) and image recognition (Dekruger and Hunt, 1994; Kung and Taur, 1995).A large number of trials have been carried out to find the optimal values of parameter of ANN. The optimal parameters were selected on the basis of statistical parameters (CC,R^2, and RMSE).

Table 2: Optimal values of user-defined parameters of ANN

Learning rate	**Momentum**	**Hidden layer**	**No. of iterations**
0.2	0.1	4	100

V RANDOM FOREST REGRESSION

Breiman (2001) proposed random forests, which add an additional layer of randomness to bagging. In addition to constructing each tree using a different bootstrap sample of the data, random forests change

International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

how the classification or regression trees are constructed. In standard trees, each node is split using the best split among all variables. In a random forest, each node is split using the best among a subset of predictors randomly chosen at that node. This somewhat counterintuitive strategy turns out to perform very well compared to many other classifiers, including discriminant analysis, support vector machines and neural networks, and is robust against over fitting (Breiman, 2001). In addition, it is very user-friendly in the sense that it has only two parameters (the number of variables in the random subset at each node and the number of trees in the forest), and is usually not very sensitive to their values.The design of random forest regression permits a tree to develop to the maximum depth of new training data using the mixing of variables. These full-developed trees are not pruned back. This is one of the main advantages of random forest regression over other tree techniques like M5P tree model (Quinlan 1992). Studies propose that the decision of the pruning process and not the variable determination, influence the performance of tree based algorithms (Pal and Mather, 2003). Breiman (1999) suggests that as the quantity of trees increases, the speculation error always converges even without pruning the tree and over fitting is not an issue in view of the Strong Law of large numbers (Feller, 1968). Number of variable used (m) at every node to grow a tree and the quantity of trees to be developed (k) are two user-defined parameters required for random forest regression (Breiman, 1999). At every node, just selected variables are found through for the best split. Thus, the random forest regression consists of k trees, where k is the quantity of trees to be developed which can be any value defined by the user. In the random forest based regression, the yield values are numeric, thus the mean-squared generalization error for any numerical predictor is acquired. The random forest predictor is developed by taking the average of generalization error over k trees. In present study random forest and bagged random forest are used.

VI ANALYSIS OF RESULTS

To evaluate the value of Random forest model and ANN model in predicting the aeration efficiency. The coefficient correlation (CC) and root mean square error (RMSE) values obtained with test data set was used to compare the performance of ANN with Random forest. Table 3 provides the values of the correlation coefficient, coefficient correlation and RMSE values provided by different analytical approaches. Comparison of results suggests improved performance by Random forest in terms of CC, and RMSE values used with this data set. A substantial improvement in predictive accuracy of Random forest approach indicates that it can be effectively used in predicting of aeration efficiency. Fig. 1and 2 shows the agreement of actual and predicted

Modelling Approach	CC	RMSE

International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

efficiency of aerationof for data set using Random forest and ANN respectively. Predicted values of oxygen transfer coefficient using Random forest within an error range of 15% suggest better accuracy of Random forest in prediction, whereas the values predicted values by ANN within error range of 40% depict inferior performance.

ANN	0.9774	0.4105
Random Forest	0.9890	0.2576
Eq. of Tojo & Mayami(1982)	0.9178	0.0055

Table 3: Detail of performance evaluation parameters using ANN, Random forest, Equation of Tojo & Miyami (1982) testing data set.

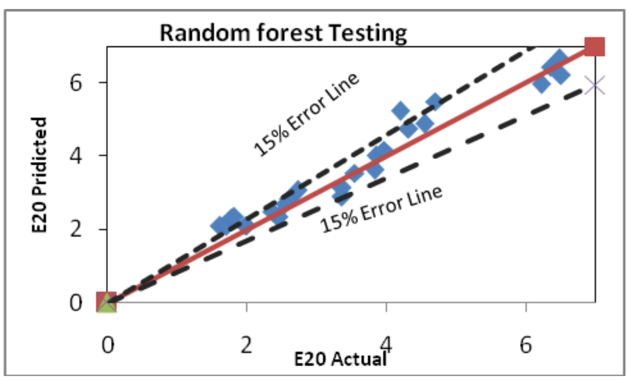

Figure 1: Actual vs. Predicted values of oxygen transfer coefficient using MLR Model

Fig. 3 show that the variation of experimental and predicted aeration efficiency by surface jet aerators with the no. of data set. It is apparent from Fig. that oxygen transfer coefficient predicted by Random forest regression technique is in very good contract with actual experiment values; whereas, that is not the case with ANN regression approach as the predicted values by this technique are deviating at few of the test data. Thus, suggesting a better performance of Random forest in comparison to ANN regression.

International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

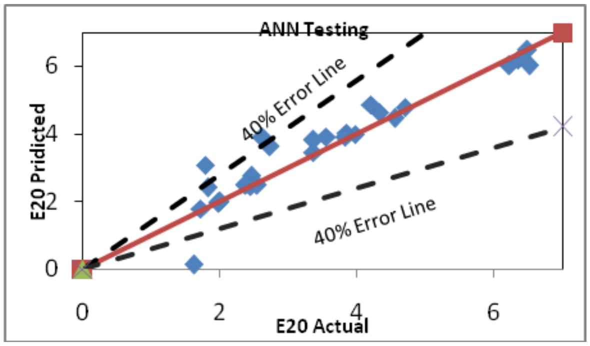

Figure 2: Actual vs. Predicted values of efficiency of aeration using ANN Model.

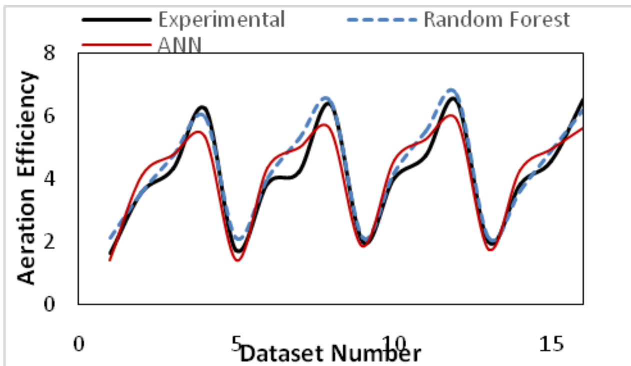

Fig. 3: Variation in predicted values by Observed values, ANN and Random forest in comparison to the actual values of efficiency of aeration by piano key weir.

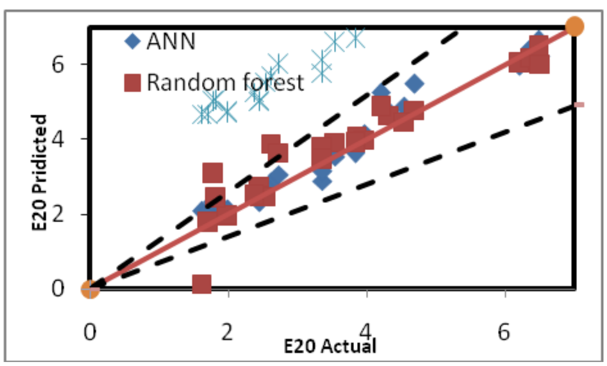

Fig. 4 Comparison between actual and predicted values of efficiency of aeration.

As depicted from Figure 4 ANN and Random forest values are well with the error band of 30%. The correlation proposed by Tojo and Miyami (1982) approximates the oxygen transfer coefficient of multiple jets to some extent and underperformed for the lower values of oxygen transfer coefficient. However the predicted points by this eq. lie outside the 30 % error band mostly at the lower and higher values of efficiency of aeration E_{20}. The overall performance of Random forest is best in prediction of efficiency of aeration of Artificial Neural Network.

VII CONCLUSIONS

This paper examines the potential of artificial neural network (ANN) and Random forest approaches in predicting the aeration efficiency. From the comparison of performance evaluation parameters, it has been observed that Random forest approach works well than ANN. It can be successfully used in estimation of aeration efficiency of piano key weir with rectangular and trapezoidal geometry plunging into the water pool. The equations proposed by different researchers are compared with the current applied modelling approaches and Random forest is concluded to be the most effective modelling technique in approximation of the volumetric oxygen transfer coefficient.

REFERENCES

[1]Azbel,D. *Two phase flows in chemical engineering*. Cambridge University Press Chicago (1981).

[2] A. Goel, 2014,neural network technique for prediction of discharge coefficient and discharge for a weir,IWRS, Vol.34,NO.2, April2014,pp 25-31.

[3] Baylar, A., Emiroglu, M.E., "The Effect of Sharp-Crested Weir Shape on Air Entrainment", Canadian Journal of Civil Engineering, 29 (3):375-383 (2002).

[4] Baylar, A., Bagatur, T., "Experimental Studies on Air Entrainment and Oxygen Content Downstream of Sharp-Crested Weirs", Water & Environment Journal, 20 (4): 210-216 (2006).

[5] Baylar, Ahmet, and Tamer Bagatur. "Aeration performance of weirs." *WATER SA-PRETORIA-* 26.4 (2000): 521-526.

[6] Baylar, Ahmet, DavutHanbay, and EmrahOzpolat. "An expert system for predicting aeration performance of weirs by using ANFIS." *Expert Systems with Applications* 35.3 (2008): 1214-1222.

[7]Breiman, Leo. "1 RANDOM FORESTS--RANDOM FEATURES." (1999).

[8]Breiman, Leo, Jerome Friedman, Charles J. Stone, and Richard A. Olshen. *Classification and regression trees*. CRC press, 1984.

[9]Chu, S. (1998). Nobel Lecture: The manipulation of neutral particles. *Reviews of Modern Physics*, *70*(3), 685.

[10]DE-Kruger, D., & Hunt, B. R. (1994).Image processing and neural networks for recognition of cartographic area features. *Pattern Recognition*, *27*(4), 461-483.

[11] Chanson, H., "Predicting Oxygen Content Downstream of Weirs", Spillways and Waterways. Proc. Instn Civ. EngrsWat., Marit. & Energy,112, March, 20-30 (1995).

[12] Ervine, D.A., "Air Entrainment in Hydraulic Structures: A Review", Proc. Instn Civ. EngrsnWat., Marit. & Energy, 130, September, 142-153(1998).

International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

[13]Feller, William. *An introduction to probability theory and its applications: volume I*. Vol. 3. New York: John Wiley & Sons, 1968.

[14] GULLIVER JS and RINDELS AJ Measurement of air-water oxygen transfer at *hydraulics* structures. *J. Hydraul. Eng., ASCE* 119(3) 328-349.GULLIVER JS, THENE JR(1993).

[15] Hinze, J.O., “Fundamentals of the Hydrodynamic Mechanism of Splitting in Dispersion Processes”, Am. Inst. of Chem. Eng., 1 (3): 289-295 (1955).

[16]Kung, S. Y., &Taur, J. S. (1995).Decision-based neural networks with signal/image classification applications. *IEEE Transactions on Neural Networks*, *6*(1), 170-181.

[17] Levich, V.G., “Physicochemical Hydrodynamics”, Prentice-Hall, Englewood Cliffs, N.J, (1962).

[18]Pal, Mahesh, and Paul M. Mather. "Assessments of the effectiveness of decision tree methods for land cover classification." *Remote sensing of environment* 86, no. 4 (2003): 554-565.

[19]Quinlan, John R. "Learning with continuous classes." In *5th Australian joint conference on artificial intelligence*, vol. 92, pp. 343-348. 1992

[20] Tojo K., and Miyanami K., 1982, Oxygen transfer in jet mixers, Chem. Eng. J. Netherlands, 24, pp 89-97.