

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

Computation of Water Quality Parameters and Prediction Tool ANN for Modeling of Water Quality of Reservoir

Purushottam Sarda¹, Dr. Parag Sadgir²

Research Scholar, Dept. of Civil Engineering, Government College of Engineering, Aurangabad, M.S., India¹

Associate Professor, Dept. of Civil Engineering, Government College of Engineering, Aurangabad, M.S., India²

ABSTRACT: Major Sources of water are surface and subsurface. Surface water includes River, Reservoir, Creek, Streams, etc. The main objective of paper was to study the important, broad water quality parameters such as COD, EC, TDS, BOD, Temperature, DO and pH those covered maximum water quality for various purposes i.e. multipurpose and to investigate the potential of the ANNs for modeling. Paper also includes the one of details of prediction model i.e. Artificial Neural Network (ANN) used for water quality parameters. These include; the determination of appropriate network and training functions, different transfer function, division of data, pre-processing of the available data, model inputs, optimization of the connection weights, network geometry, epoch sizes and performance indicators has been used in last two decades.

KEYWORDS: Algorithms, Architectures, Model development, Networks, Prediction, Training Function

I. INTRODUCTION

The water is a primary natural resource for people for different consumptions such as drinking, irrigation, hydro-electricity, fish fostering and recreation; therefore, it requires at least an acceptable level of water quality [1]. Surface water quality is one of the major issues today due to increasing the load of pollution from industrial, commercial and residential with its effects on human health and aquatic ecosystems [2]; [3]. Water quality measurements include a variety of physical, chemical and biological parameters. Basic problem in the case of water quality monitoring is the complexity associated with analyzing the large number of variables [4]; [5]. Various physico-chemical parameters such as Chemical Oxygen Demand (COD), Electrical Conductivity, Total Dissolved Solids, Biological Oxygen Demand (BOD), Temperature, Dissolved Oxygen and pH are studied to know its importance for water quality parameters. Predicting the water quality is also the key factor in the water quality management of surface water, and it enables a manager to choose an option that satisfies a large number of identified conditions [6].

II. LITERATURE SURVEY

Various researchers explained their views about water quality parameters and ANN Architectures that are explained below.

Physico- Chemical Parameters of Surface Water

Following are the broad parameters for surface water quality. These are studied and discussed here.

2.1 Chemical Oxygen Demand is an index of organic content of water because the most common substance oxidized by dissolve oxygen in water is organic matter having biological origin i.e. dead plant and animal wastes [7]. It is a valuable parameter of water quality assessment which measures oxygen demand of biodegradable pollutants as well as non-biodegradable pollutants [3]. Alum showed that mortality due to liver cancer for men and women was positively correlated with the COD in drinking water.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

2.2 Electrical Conductivity is associated with major water quality parameters due to dilution effect of stream flow and can be used as indicator in determining the suitability of water for irrigation. Electrical conductivity is also considered to be a rapid and good measure of dissolved solids which reflects the pollution status as well as tropic level of the aquatic body [8]; [7]; [9]. Higher value of conductivity was due to presence of higher concentration of dissolved salts of cations such as calcium, magnesium and sulphate during the rains. The low conductivity might be responsible for the soft nature of the water and the significant changes in conductivity may be an indicator that a discharge or some other source of pollution has entered a stream [10].

2.3 Total Dissolved Solid has important consideration in determining its suitability for irrigation, drinking and industrial uses. For irrigation, water dissolved solid is a very important criterion due their gradual accumulation results in salinization of soil, thus, rendering the agriculture land non-productive [3].

2.4 Biochemical Oxygen Demand has limited the amount of 5 mg/l or less (5 days 20 °C) for drinking water source without conventional treatment but after disinfection [2]. The untreated discharge of municipal and domestic waste in water bodies increases the amount of organic content. Therefore the microbes present in water require more amount of oxygen for its degradation thus the BOD of water gets increased. BOD is the measure of the extent of pollutant in the water body.

2.5 Temperature is one of the most important factors for survival of aquatic life. The moderate change in temperature can seriously affect aquatic environment, including bacteria, algae, invertebrates and fish [3]; [7]. Verma and Singh [11] stated that, as the temperature increases, lesser amount of oxygen remains dissolved in the sample, hence BOD and COD increases. However, after a certain temperature all living organisms die.

2.6 Dissolved oxygen is an important indicator of water quality, ecological status, productivity and health of a reservoir. The range of dissolved oxygen recorded 4.8 mg/l – 8.2 mg/l shows the water to be of good quality and will support fish production (Mustapha 2008). The impact of low DO concentration or of anaerobic conditions is reflected in an unbalanced ecosystem, fish mortality, failure of eggs/larvae to survive, reduction in growth, odors and other aesthetic nuisances [2];[4].

2.6 pH value dependent on various phases of water treatment and water supply such as acid-base, neutralization, coagulation, sedimentation, corrosion control, etc. [3]. Due to acid rain or discharges, pH changes which in turn increases the bacterial degradation thereby depleting the DO and hence the BOD requirement exponentially rises [11].

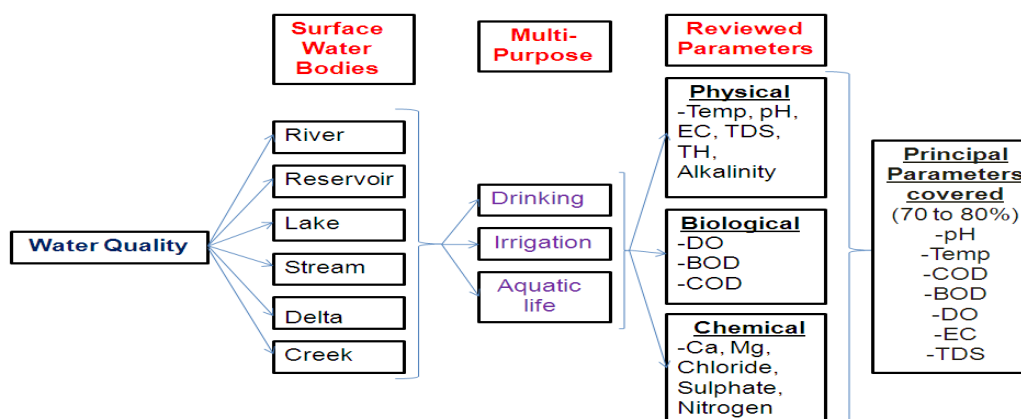


Fig.1. Water Quality parameters for surface water

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

Thus, from above discussion maximum researchers and Fig. 1 showed that the essential parameters water quality are COD, EC, TDS, BOD, Temperature, DO and pH. Those can be used for analysis and modeling purpose of reservoir water quality.

III. ANN ARCHITECTURE

Determination of appropriate network architecture is one of the most important, but also one of the most difficult, tasks in the model building process. The architecture of ANN is designed by the number of layers, number of neurons in each layer, weights between neurons, a transfer function that controls the generation of output in a neuron, learning laws that define the relative importance of weights for input to a neuron and number of training iterations [12].

3.1 Different Network and Training Functions

The backpropagation algorithm is used in layered feed-forward ANNs. The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on an output layer. There may be one or more intermediate hidden layers [11]. The BP algorithm uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute, and then the error (difference between actual and expected results) is calculated. Diamantopoulou [13] studied the cascade correlation algorithm and multi-layer feed-forward with hyperbolic-tangent transfer function. Various networks and training functions are listed see Table 1.

Table 1 Various Networks and Training Functions used in ANN

| Network Type | Brief Explanation | Training Functions | Brief Explanation |
|--------------|-----------------------------------|--------------------|-----------------------------------------------------------------|
| FFNN/ FFBP | Feedforward Back Propagation | Trainbfg | BFGS Quasi-Newton backpropagation |
| CCFF/ CF | Cascade Correlation Feed Forward | Trainbr | Bayesian Regularization backpropagation |
| MLP | Multilayer Perceptron | Traincgb | Conjugate Gradient backpropagation with Powell-Bale restarts |
| RBF | Radial Bias Function | traincgf | Conjugate Gradient backpropagation with Fletcher-Reeves updates |
| GRNN | General Regression Neural Network | Traincgp | Conjugate Gradient backpropagation with Polak-Ribiere updates |
| HNN | Hopfield Neural Network | Traingda | Gradient descent with adaptive learning rate backpropagation |
| FFTD | Feed forward Time Delay | Traingdm | Gradient descent with momentum backpropagation |
| RNN | Recurrent Neural Network | Traingd | Gradient – descent backpropagation |
| GUI | Graphical User Interference | Trainlm | Levenberg – Marquardt backpropagation |

Sentu and Regulwar [14] studied Multilayer Perceptron (MLP), Generalized Feed Forward (GFF), Time-Lag Recurrent Network (TLRN) and Radial Basis network (RBN) with a single training algorithm i.e. Levenberg-Marquardt and founded that Time-lag Recurrent Network and Radial Basis network (RBN) with Levenberg-Marquardt is able to forecast the reservoir inflow up to one month in advance with reasonable prediction accuracy. To overcome the shortcomings of traditional BP algorithm as being slow to converge and easy to reach extreme minimum value, the model adopts LM (Levenberg-Marquardt) algorithm to achieve a higher speed and a lower error rate.

3.2 Transfer Function

Hamid [15] used two training algorithms (i.e. Levenberg–Marquardt and Momentum) to train the network with two different transfer functions (i.e. *Sigmoid* and *Tanh*) to obtain the best results with respect to non-linearity of this phenomenon, see Table 2 and Fig. 2.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

Table 2 Various Transfer /Activation Function

| Transfer functions | Brief Explanation |
|--------------------|-----------------------------------------------------------------------------------|
| <i>tansig</i> | Hyperbolic tangent sigmoid transfer function / Tan Function / Hyperbolic function |
| <i>logsig</i> | Log-sigmoid transfer function /sigmoid function |
| <i>purelin</i> | Linear Transfer Function |

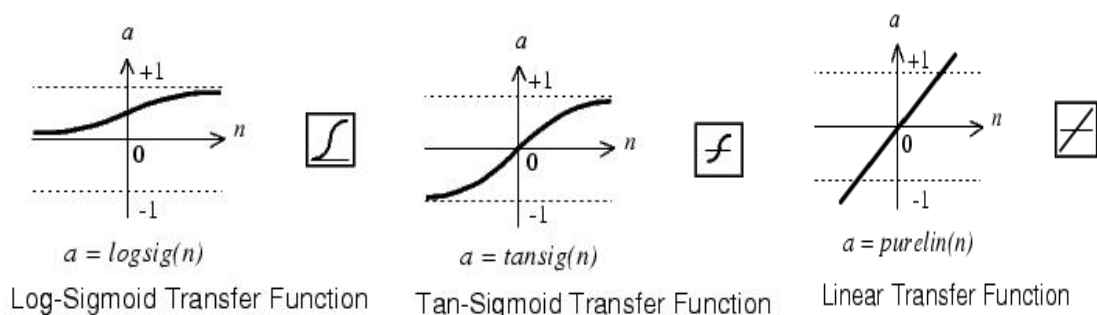


Fig. 2 Activation Functions “Logsig” “Tansig” and “Purlin”

Logistic and Tangent sigmoid nonlinear function is

$$\text{Logsig } \phi(x) = \frac{1}{1 + e^{-x}} \qquad \text{Tansig } \phi(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$$

3.3 Division of Data

It is common practice to split available data into three subsets; a training set, a testing set and a validation set, which is very data intensive. The same applies to cases where network geometry or internal parameters are optimized by trial and error. When limited data are available, it might be difficult to assemble a representative validation set. Division of dataset can be as training and testing of 73-30, 80-20, 90-10 respectively.

3.4 Data Pre-processing

In order to ensure that all variables receive equal attention during the training process, they should be standardized. In addition, the variables have to be scaled in such a way as to be commensurate with the limits of the activation functions used in the output layer. For example, as the outputs of the logistic transfer function are between 0 and 1, the data can be scaled in the range 0.1–0.9 or 0.1–0.8. Before using the data for ANN, the data has been pre-processed (normalized) by different tools in various studies because ANNs work best if the inputs are scaled to the same range of values [9]. Scaled data within range of 0.0 to 1.0 calculated by using below equation.

$$x_i = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})}$$

Areerachakul [16] and Najah [10] studied on normalized data within range of 0.1 to 0.9 by using below equation.

$$x_i = 0.1 + 0.8 \frac{(x - x_{\min})}{(x_{\max} - x_{\min})}$$

Normalized data (Filtered Values / Transformed Data) between the range -1 and 1 using the following equation

$$z_p = 2 \times \frac{(x_p - x_{\min})}{(x_{\max} - x_{\min})} - 1$$

Where z_p is the normalized or transformed data series, x_p is the original data series, x_{\min} , x_{\max} are the minimum and the maximum values of the original data series.

3.5 Type of Connection

FFNN, where nodes in one layer are only connected to nodes in the next layer, have been used for prediction and forecasting applications. Deciding the appropriate amount of training is therefore crucial to the success of neural networks' learning. The network needs first to be trained before interpreting new information. During training of the

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

network, data are processed through the network until it reaches the output layer (forward pass). In this layer, the output is compared to the measured values (the true output). The difference or error between both is processed back through the network (backward pass) updating the individual weights of the connections and the biases of the individual neurons.

3.6 Geometry

Network geometry determines the number of connection weights and how these are arranged. This is generally done by fixing the number of hidden layers and choosing the number of nodes in each of these. The number of hidden neurons is calculated by an empirical formula:

$$N_h = (n / k (m + p))$$

Where N_h = number of hidden neurons, n = number of training sets, m = number of input neurons, P = number of output neurons, and K = noise factor [11]. A neuron is a non-linear algebraic function, parameterized with boundary values. The numbers of neurons in layers can be determined by the number of variables and number of data patterns, as follows: a) The number of neurons in the input and output layer is equal to the number of input and output variables that have been used in network creation; b) The number of neurons in the hidden layer is equal to the number of training patterns used for model training; c) The number of neurons in the summation layer is equal to the number of output variables plus one [5].

3.7 Epoch Size

The epoch size is equal to the number of training samples presented to the network between weight updates. If the epoch size is equal to 1, the network is said to operate in on-line mode. If the epoch size is equal to the size of the training set, the network is said to operate in batch mode. In many applications, batch mode is the preferred option, as it forces the search to move in the direction of the true gradient at each weight update. Generally Epoch size takes 1000 iterations.

3.8 Performance Indicators

Performance of ANN was assessed based on error measures, namely, Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Index of Agreement (IA) and Percent of Prediction within a Factor of 1.1 (FA 1.1) see Table 3. The mean squared error (MSE) function is most commonly used, but other error functions have also been proposed. The advantages of using the MSE include that it is calculated easily, that it penalizes large errors, that its partial derivative with respect to weights can be calculated easily and that it lies close to the heart of the normal distribution. The coefficient of Determination (R^2), measures the strength and direction of the linear relation between two variables [16]

Table 3 Statistical metrics used in model performance evaluation (C_p and C_o are the predicted and observed concentrations, respectively)

| Measures | | Mathematical Expression |
|----------|------------------------------|---------------------------------------------------------------------------------------------------------------|
| R^2 | Determination of Coefficient | $R^2 = \frac{\sum(C_p - \bar{C}_o)(C_o - \bar{C}_o)}{\sqrt{\sum(C_p - \bar{C}_o)^2 \sum(C_o - \bar{C}_o)^2}}$ |
| RMSE | Root Mean Squared Error | $RMSE = \sqrt{\frac{1}{n} \sum (C_p - C_o)^2}$ |
| MAE | Mean Absolute Error | $MAE = \frac{1}{n} \sum C_p - C_o $ |

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

| | | |
|--------|----------------------------------------------|--------------------------------------------------------------------------------------------|
| IA | Index of Agreement | $IA = 1 - \frac{1}{n} \sum \left(\frac{ C_p - C_o }{ C_p - C_o + C_o - C_o } \right)^2$ |
| FA 1.1 | Percent of Prediction within a Factor of 1.1 | $0.9 < \frac{C_p}{C_o} < 1.1$ |

IV. CONCLUSIONS

The physical chemical and Biological parameters such as factors COD , EC, TDS, BOD, Temperature, DO and pH are computed for drinking, irrigation, aquatic life purposes and it is clear that all parameters are equally important and these broad range parameters was to be studied with more details for water quality modeling of reservoir. Also, those parameters were selected due to their simple, fast and continuous measurement at water quality monitoring stations. Moreover, keeping in mind that ANNs are require less prior knowledge of the system under study, it is expected that it will be a more powerful tool in capturing interrelations between water quality variables. Different algorithms and transfer functions can be used for better output results for decision making of parameter such as Feedforward Neural network, Cascade correlation. Data division can be taken as 70-30 for training and testing of dataset. Data pre-processing of dataset can be done in the range of 0.0-1.0. Adopts LM (Levenberg-Marquardt) algorithm to achieve a higher speed and a lower error rate, Epoch size has to be taken as 1000. A selection of an appropriate number of neurons in the hidden layer is very important. The optimum size of networks is selected from the one which resulted in maximum coefficient of determination (R^2) for the training+ validation+ test set for goodness of fit , root mean square error (RMSE) and mean absolute error (MAE), index of agreement (IA) and percent of prediction within a factor of 1.1 (FA 1.1) measures the errors . Scatter plots and time series plots are used for visual comparison of the observed and predicted values. The new results and prediction of Network function and algorithm types in ANN for water quality are left to explore for better results.

REFERENCES

- Emamgholizadeh S., H. Kashi, I. Marofpoor and E. Zalaghi, "Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models", *International Journal of Environmental Science Technology*, Springer, Vol. 11, 2014, pp. 645-656.
- Alam M.J.B., Islam M.R., Muyen Z., Mamun M. and Islam S., "Water quality parameters along rivers", *International Journal Environmental Science Technology*. Vol. 4 (1), 2007, pp. 1735-1472.
- Choudhary R., Ratwani P. and Vishwakarma M., "Comparative study of Drinking Water Quality Parameters of three Manmade Reservoirs i.e. Kolar, Kaliasote and Kerwa Dam" *Current World Environment*, Vol. 6 (1), 2011, pp.145-149.
- Rankovic V., Radulovic J., Radojevic I., Ostojic A. and Comic L., "Neural network modeling of dissolved oxygen in the Gruza reservoir, Serbia", *Ecological Modeling ELSEVIER*, Vol. 221, 2010, pp. 1239-1244.
- Antanasijevic D., Pocajt V., Povrenovic D. Peric-Grujic A. and Ristic M., "Modeling of dissolved oxygen content using artificial neural networks: Danube River, North Serbia, case study", *Environmental Science Pollution Research Springer*, Vol. 20, 2013, pp. 9006-9013.
- Palani S., Liong S. and Tkalic P., "An ANN application for water quality forecasting", *Marine Pollution Bulletin, ELSEVIER*, Vol. 56, 2008, pp.1586-1597. DOI:10.1016/j.marpolbul.2008.05.021
- Gupta P., Vishwakarma M. and Rawtani P.M., "Assessment of water quality parameters of Kerwa Dam for drinking Suitability", *International Journal of Theoretical and Applied Sciences*, Vol. 1 (2), 2007, pp. 53-55.
- Mustapha M.K., "Assessment of the Water Quality of Oyun Reservoir, Offa, Nigeria, Using Selected Physico-Chemical Parameters", *Turkish Journal of Fisheries and Aquatic Sciences* Vol. 8, 2008, pp. 309-319.
- Heydari M., Olyae E., Mohebzadeh H. and Kisi O., "Development of a Neural Network Technique for Prediction of Water Quality Parameters in the Delaware River, Pennsylvania", *Middle-East Journal of Scientific Research*, Vol. 13 (10), 2013, pp. 1367-1376. DOI: 10.5829/idosi.mejrs.2013.13.10.1238
- Najah A., El-Shafie A. and Karim O.A., "Application of artificial neural networks for water quality prediction", *Neural Computational and Applied Science Springer*, Vol. 22 (1), 2013, pp. 180-201.
- Verma A.K. and Singh T.N., "Prediction of water quality from simple field parameters", *Environmental Earth Science, Springer*, Vol. 69, 2013, pp.821-829. DOI 10.1007/s12665-012-1967-6.
- Misaghi F. and Mohammadi K., "Estimating Water Quality Changes in the Zayandeh Rud River Using Artificial Neural Network Model", *The Canadian society for Engineering in Agricultural, food and biological systems*, 2003, pp. 03-17.
- Diamantopoulou M.J., Papamichai D.M. and Antonopoulos V.Z., "The Use of a Neural Network Technique for the Prediction of Water Quality Parameters", *Operational Research, An International Journal, ASCE*, Vol. 5 (1), 2005, pp. 115-125.
- Sentu D. and Regulwar D.G., "Inflow Prediction by Different Neural Network Architectures: A Case Study", *International Journal of Earth Sciences and Engineering*, Vol. 04 (6), 2011, pp. 225-230.
- Hamid Zare Abyaneh, "Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters." *Journal of Environmental Health Science and Engineering*, 2014, 12:40 doi: 10.1186/2052-336X-12-40
- Areerachakul S., "The Using Artificial Neural Network to Estimate of Chemical Oxygen Demand", *World Academy of Sciences, Engineering and Technology*, Vol. 79, 2013, pp. 455-461.