International Journal of Advanced Technology in Civil Engineering

Volume 2 | Issue 1 Article 8

January 2013

WATER QUALITY PREDICTION IN DISTRIBUTION SYSTEM USING CASCADE FEED FORWARD NEURAL NETWORK

VINAYAK K PATKI
National Institute of Technology Karnataka (NITK), Surathkal, India, Patki.vnyk@rediffmail.com

S. SHRIHARI
NITK, Surathkal, Mangalore,Karnataka, India, S.SHRIHARI@GMAIL.COM

B. MANU
National Institute of Technology Karnataka (NITK), Surathkal, India, b.manu@nitk.ac.in

WATER QUALITY PREDICTION IN DISTRIBUTION SYSTEM USING CASCADE FEED FORWARD NEURAL NETWORK

VINAYAK K PATKI[1], S. SHRIHARI[2], B. MANU[3]

[1]Research scholar, National Institute of Technology Karnataka (NITK), Surathkal, India
[2]Professor, Department of civil Engineering, National Institute of Technology Karnataka (NITK), Surathkal, India
[3]Assistant Professor, Department of civil Engineering, National Institute of Technology Karnataka (NITK), Surathkal, India
E-mail: Patki.vnyk@rediffmail.com, s.shrihari@gmail.com, b.manu@nitk.ac.in

Abstract— Cascade feed forward ANN models have been developed by using pH, Alkalinity, Hardness, TS and MPN as the input variables to forecast water quality index (WQI) in the various zones of municipal distribution system. Different ANN models were developed using training data set and tested in order to determine optimum number of neurons in the hidden layer and best fitting transfer function. The study reveals that the predictions by logsigmoidal and pure linear transfer function are in good correlation with observed WQI as compared to tansigmoidal transfer function. It is also observed that the model performance changes considerably with change in hidden layer neurons. Hidden layer structure with seven neurons performs better, followed by hidden layer structure with four neurons and one neuron respectively.

Keywords*-component; Cascade ANN Network,Water quality index,transfer function,number of neurons*

I. INTRODUCTION (HEADING 1)

The artificial neural network (ANN), as its name implies, is a technique for the human brains problem solving process. Just as human apply knowledge gained from experience to know problems or situations; the structure of a neural network can be applied to powerful computations of complex non linear relationships [1]. The ANN method is regarded as a potentially useful tool for modeling complex non-linear system, whereas fuzzy logic (FL) and adaptive neuro-fuzzy interference system (ANFIS) are useful in cases wherein uncertainties and imprecision is involved [2]. However, a large number of factors affecting the quality have a complicated nonlinear relation with the linguistic variables; traditional data processing methods are no longer good enough for solving the problem [3]. Water distribution system plays a vital role in presenting a desirable life quality to the public. The welfare level of country is measured with the amount of water consumption for each person and the quality of the provided water [4]. The water quality varies temporally and spatially at source, treatment plant and in the distribution network. The water quality in the distribution system deteriorates due to pipe age, corrosion of pipe material, intrusion of contaminants through leakage and cross connections, leaching of pipe material, formation of biofilm in the pipes etc, and hence many uncertainties are involved till the water reaches to the users tap.

Main objective of present study is to develop best fitting ANN model for prediction of WQI in the municipal distribution system. In this study the Cascade Forward Back Propagation (CFBP) is used to forecast the variation in WQI with variation in water quality parameters, for the various zones in Solapur city. ANN models were developed by using pH, Alkalinity, Hardness, total solids (TS) and MPN as the input variables and WQI as the output variable. The ANN models are developed by using two years data set for training the model and one year data is used for testing the model performance. Different ANN models were developed using training data set and tested in order to determine optimum number of neurons in the hidden layer, best fitting transfer function.

II. METERIALS AND METHODS

A. Study Area and Water Quality Data

The municipal water distribution system of Solapur, India is taken as a case study for prediction and analysis of water quality in the distribution system. Fig. 1 shows the location sketch of three sources of water. The water quality at these three sources varies spatially and temporally. The water is distributed to Solapur city by dividing it into twenty nine zones. The water quality in the distribution system deteriorates due to pipe age, corrosion of pipe material, intrusion of contaminants through leakage and cross connections, leaching of pipe material, formation of biofilm in the pipes etc. The zone wise water quality data for years 2008, 2009 and 2010 is collected from Solapur Municipal Corporation, Solapur. Physico-chemical properties of water such as pH (0.09), dissolved oxygen (0.12), total alkalinity (0.01), total solids (0.13), total hardness (0.05) and most probable number (0.6) were used to get the WQI for various zones. Weight factors are given in parentheses.

B. Artificial Neural Network Models for Prediction

Neural networks have seen an explosion of interest over the last few years and are being successfully applied across an extraordinary range of problem domains, in areas as diverse as finance, medicine, engineering, geology, physics and biology. The excitement stems from the fact that these networks are attempts to model the capabilities of the human brain. From a statistical perspective neural networks are interesting because of their potential use in prediction and classification problems. Artificial neural networks (ANNs) are non-linear data driven self-adaptive approach as opposed to the traditional model based methods. They are powerful tools for modeling, especially when the underlying data relationship is unknown. ANNs can identify and learn correlated patterns between input data sets and corresponding target values. After training, ANNs can be used to predict the outcome of new independent input data. ANNs imitate the learning process of the human brain and can process problems involving non-linear and complex data even if the data is imprecise and noisy. Thus they are ideally suited for the modeling of agricultural data which is known to be complex and often non-linear.

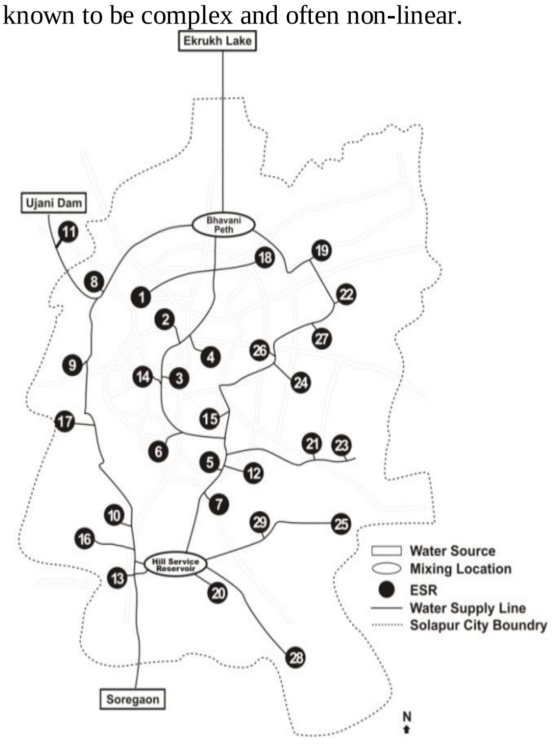


Fig.1 Location Sketch of Water Sources

1. Characteristics of Neural Network

- The NNs exhibit mapping capabilities, that is, they can map input patterns to their associated output patterns.
- The NNs learn by examples. Thus, NN architectures can be trained with known examples of a problem before they are tested for their "inference" capability on unknown instances of the problem. They can, therefore, identify new objects previously untrained.
- The NNs possess the capability to generalize. Thus, they can predict new outcomes from past trends.
- The NNs are robust systems and are fault tolerant. They can, therefore, recall frill patterns from incomplete, partial or noisy patterns.
- The NNs can process information in parallel, at high speed, and in a distributed manner.

***2 .**Basics of Artificial Neural Networks*

The terminology of artificial neural networks has developed from a biological model of the brain. A neural network consists of a set of connected cells. The neurons receive impulses from either input cells or other neurons and perform some kind of transformation of the input and transmit the outcome to other neurons or to output cells. The neural networks are built from layers of neurons connected so that one layer receives input from the preceding layer of neurons and passes the output on to the subsequent layer.

A neuron is a real function of the input vector (yj.....yk). The output is obtained as

$$f(x) = f a_i + [\sum_{n}^{k} W_{ij} \times Y_j] \tag{1}$$

Where f is a function, typically the sigmoid (logistic or tangent hyperbolic) function. A graphical presentation of neuron is shown in the Fig.2. Mathematically a Multi-Layer Preceptor network is a function consisting of compositions of weighted sums of the functions corresponding to the neurons. Feed-forward networks are especially useful in function approximation when a set of inputs and outputs is all that is known of the system, which is the situation in this study. Feed-forward networks have their neurons arranged in layers.

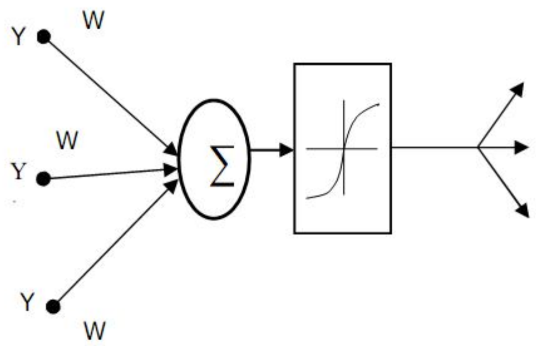


Fig.2 Structure of Neuron

*3. **Cascade Forward Back Propagation Algorithm***

The cascade back-propagation (CFBP) algorithm is the basis of a conceptual design for accelerating learning in artificial neural networks developed by Scott Fahlman at Carnegie Mellon in 1990.It is so named because it combines features of the back-propagation and cascade-correlation algorithms. Like other algorithms for learning in artificial neural networks, the CFBP algorithm (Fig.3) specifies an iterative process for adjusting the weights of synaptic connections by descent along the gradient of an error measure in the vector space of synaptic-connection weights. The error measure is usually a quadratic function of the differences between the actual and the correct outputs.CF models are similar to feed-forward networks, but include a weight connection from the input to each layer and from each layer to the successive layers.

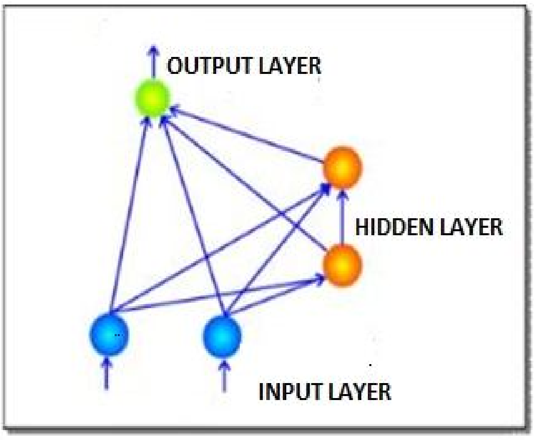


Fig.3 Cascade Forward Back Propagation Algorithm

There are two common criteria to stop training a network: (1) training cycles (epochs); and (2) desired errors. C.W. Dawson and R.L. Wilby (2001), suggested to typically applying 20,000 to 100,000 training cycles (epochs) to train the network when steepest descent method is used. The other criterion is to limit the difference between desired output and output calculated by the network .The training process may be brought to halt using either the worst error difference after complete presentation of all input output patterns, or the root mean square error summed over all patterns.

In practice, it is sometimes necessary to apply or compare both approaches to ensure the capability of the trained network in generalizing on the tested samples and application. The errors of tested samples is generally higher than the error of training sample as the network is trained to reduce the latter, not the former. However, the over-trained network would occasionally result in over fitting. Over fitting means the network can converge and yield a minimum or desired error in training samples but it cannot generalize well when validated with testing sample .The weights that produce the lowest error on the test sample would be used for the model.

C. Modelling Performance Criterion

In order to evaluate the prediction accuracy of ANN and multiple regression models four criterions were used for comparative evaluation of the performance of the model. The criterions employed are Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error and Coefficient of Correlation (Cc).

Mean Absolute Error (MAE)

MAE is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. Expressed in words, the MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. The mean absolute error is given by

$$MAE = 1/n \sum_{i=1}^{n} [observed - predicted] \quad (2)$$

Mean Square Error (MSE)

The mean squared error of an estimator is one of many ways to quantify the difference between values implied by an estimator and the true values of the quantity being estimated. It is the residual or error sum of squares divided by the number of degrees of freedom of the sum. This gives an estimate of the error or residual variance.

$$MSE = 1/n \sum_{1}^{n} (observed - predicted)^2 \quad (3)$$

Coefficient of Correlation (Cc)

It is a measure of the strength of the linear relationship between two variables. It is defined in terms of the (sample) covariance of the variables divided by their (sample) standard deviations

$$\mathrm{Cc} = \frac{\sum (x - x')(y - y')}{\sqrt{(x - x')(y - y')}} \quad (4)$$

Where, n= the number of data patterns in the dependent data set,x= the observed values, y= the predicted values , x'= mean of the observed values and y'= mean of the predicted values

III. RESULTS AND DISCUSSIONS

The ANN Architecture for WQI prediction is composed of one input layer with six input variables, one hidden layer in which number of neurons varied from one to ten and one output layer with one output variable. Tansigmoidal, Purelinear and Logsigmoidal transfer functions were used to construct the ANN model for various zones in the city. The observed WQI for twenty nine zones reveals that, for zone twenty two (90.64) and twenty eight (92.39) water quality is excellent, for zone four (55.37), twenty three (66.69) and twenty nine (62.14) water quality is medium, for zone two (35.85) water quality is bad and for remaining twenty three zones water quality is good (70 to 90).The average WQI is given in parentheses

The typical error analysis during training and testing for zone with bad, medium, good and excellent water quality is mentioned in the Tables 1 to 4. From Tables 1 to 4 it is observed that model performance varies considerably with change in transfer function and number of neurons in the hidden layer structure. The best fitting ANN model for each zone based on performance indices and transfer function is mentioned in Table 5. It can be observed from Table 5 that, out of twenty nine zones in the study area, for thirteen zones Logsigmoidal, for ten zones Purelinear and for remaining six zones Tansigmoidal transfer function performs better. Logsigmoidal transfer function performed better due to strong nonlinearity between input variables and output variable.

Table 1: Error Analysis for Zone Two (Avg.WQI-35.85) with Bad Water Quality Using CFBP Algorithm (60% Training Dataset)

Transfer Function	**Data**	**Error Analysis**	**No. of Neurons**									
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Tansigmoidal	**Training**	**MSE**	849.842	34.1726	**121.349**	45.9801	173.29	4.07986	85.655	5.7541	20.54	65.842
		MAE	17.9053	4.09675	**4.03983**	3.08209	6.6067	1.26369	4.7829	1.3403	3.9492	4.3462
		RMSE	29.1521	5.84573	**11.0159**	6.78086	13.164	2.01987	9.255	2.3988	4.5321	8.11431
		CC	-0.1302	0.93377	**0.80797**	0.92877	0.6133	0.9927	0.9067	0.9925	0.9638	0.95508
		MRE	42.1014	10.9645	**11.5108**	9.32694	13.365	3.51693	12.518	3.3287	10.796	8.80556
	Testing	**MSE**	17.8618	13.9593	**3.27986**	16.118	21.108	21.1085	13.82	20.367	10.207	18.5313
		MAE	3.87235	3.38209	**1.53506**	2.98026	3.6527	3.65271	2.4542	4.2389	2.8306	3.01257
		RMSE	4.22632	3.73622	**1.81104**	4.01473	4.5944	4.5944	3.7175	4.513	3.1948	4.3048
		CC	0.20032	0.20683	**0.69002**	0.04294	-0.0515	-0.0515	0.0369	0.6609	0.3951	-0.0681
		MRE	12.0184	10.6637	**4.93328**	9.41615	11.64	11.6401	7.9996	13.129	8.8639	9.72173
Purelinear	**Training**	**MSE**	95.5168	241.739	303.833	97.89	144.13	263.016	129.09	94.645	121.32	163.752
		MAE	7.67685	8.53369	10.5557	7.023	9.2436	12.1759	9.3312	8.4011	7.4998	9.75016
		RMSE	9.77327	15.5479	17.4308	9.89394	12.005	16.2178	11.362	9.7285	11.014	12.7966
		CC	0.81203	0.3462	0.43615	0.80966	0.7941	0.5787	0.748	0.8181	0.7586	0.71454
		MRE	20.4611	18.6879	27.7078	17.127	26.722	34.9175	25.606	23.854	18.26	27.5658
	Testing	**MSE**	80.3569	160.735	472.984	61.4785	1778.1	45.2154	169.18	133.2	1008.6	3103.25
		MAE	5.51645	11.9699	20.7722	6.87601	41.599	4.62776	11.194	9.3108	30.91	55.4136
		RMSE	8.9642	12.6781	21.7482	7.84082	42.167	6.72424	13.007	11.541	31.759	55.7069
		CC	-0.1906	-0.0836	-0.0362	-0.0074	0.5387	-0.0876	-0.011	-0.135	0.3226	-0.3514
		MRE	18.1617	38.5519	66.3799	21.5081	131.51	14.3805	36.043	30.267	97.9	176.666
Logsigmoidal	**Training**	**MSE**	826.947	742.78	743.055	694.473	694.47	831.096	874.93	742.78	675.18	733.002
		MAE	26.7545	26.7545	26.7766	25.2727	25.273	26.7545	28.953	26.755	24.336	26.5502
		RMSE	28.7567	27.254	27.259	26.3529	26.353	28.8287	29.579	27.254	25.984	27.074
		CC	0.4689	-3E-16	0.21445	0.73447	0.4552	-0.0489	0.9475	-3E-16	-0.13	0.60869
		MRE	80.8046	77.7673	77.8078	76.1739	76.174	81.0165	84.669	77.767	75.018	77.6775
	Testing	**MSE**	1042.94	828.097	828.097	828.097	828.1	888.577	828.1	828.1	828.1	828.097
		MAE	31.008	28.6857	28.6857	28.6857	28.686	29.7031	28.686	28.686	28.686	28.6857
		RMSE	32.2946	28.7767	28.7767	28.7767	28.777	29.809	28.777	28.777	28.777	28.7767
		CC	-0.1594	4.4E-15	4.4E-15	4.4E-15	4E-15	-0.4474	4E-15	4E-15	4E-15	4.4E-15
		MRE	99.1929	91.5537	91.5537	91.5537	91.554	94.8292	91.554	91.554	91.554	91.5537

Table 2: Error Analysis for Zone Four (Avg.WQI-55.37) with medium Water Quality Using CFBP Algorithm (90% Training Dataset)

Transfer Function	Data	Error Analysis	No. of Neurons									
			1	2	3	4	5	6	7	8	9	10
Tansigmoidal	Training	MSE	994.6084	872.3003	**29.27042**	40.1146	1215.91	308.1757	51.81491	120.6352	1735.943	30.71938
		MAE	20.17101	18.59667	**3.272194**	5.045894	23.76786	8.324909	4.689144	7.0542	36.59087	3.656897
		RMSE	31.53741	29.53473	**5.410215**	6.333608	34.86991	17.55493	7.198257	10.98341	41.66465	5.542507
		CC	0.471989	0.620964	**0.981866**	0.973476	0.395643	0.840764	0.967444	0.953548	-0.35559	0.981992
		MRE	28.52564	50.19113	**7.581011**	11.1636	62.98813	20.84619	10.47233	13.72618	88.10848	8.245503
	Testing	MSE	1418.587	12.17333	**1.377845**	13.91369	1418.587	3.900719	6.788159	12.17333	24.70979	311.9096
		MAE	24.13333	3.266667	**0.958967**	3.296933	24.13333	1.742267	2.2842	3.266667	4.3936	11.15683
		RMSE	37.66413	3.48903	**1.173816**	3.730107	37.66413	1.975024	2.60541	3.48903	4.970895	17.66096
		CC	0.5	0.999121	**0.999972**	0.996638	0.5	0.998912	0.998911	0.99621	0.999997	0.846817
		MRE	75.74717	7.106816	**3.120291**	7.927514	75.74717	5.420533	4.703062	7.106816	13.64744	36.17416
Purelinear	Training	MSE	114.0306	112.6279	107.4229	122.963	134.1102	111.4636	188.7271	125.0295	117.4051	104.3797
		MAE	9.546228	9.237231	8.8729	9.690634	9.608294	8.663369	11.86805	9.181756	8.770484	8.793878
		RMSE	10.67851	10.61263	10.3645	11.08887	11.5806	10.55763	13.7378	11.18166	10.83536	10.21664
		CC	0.925152	0.926239	0.928855	0.918418	0.914447	0.925989	0.907491	0.919347	0.934136	0.933079
		MRE	22.10361	23.3201	21.64163	24.12945	26.03193	21.51433	25.48923	25.16756	23.91276	21.72485
	Testing	MSE	230.7542	203.3599	228.8891	119.8796	120.2479	207.558	234.5248	163.6484	194.7959	88.96345
		MAE	11.53677	8.5609	11.78873	8.9283	8.365533	10.3766	12.24357	11.39863	12.32213	7.052833
		RMSE	15.19059	14.26043	15.12908	10.94895	10.96576	14.40687	15.3142	12.79251	13.95693	9.432044
		CC	0.937381	0.911811	0.942927	0.940004	0.958684	0.933248	0.955403	0.913152	0.985679	0.950601
		MRE	33.83521	27.78868	34.49244	27.30468	24.65046	29.23777	30.11304	27.35003	26.17726	22.98695
Logsigmoidal	Training	MSE	898.3425	755.4466	469.4076	907.3072	503.2506	473.6395	744.9153	894.7361	477.8978	685.6246
		MAE	25.3651	24.44688	16.77849	26.55313	18.7704	17.06122	25.94063	27.45109	17.60296	22.97604
		RMSE	29.97236	27.48539	21.66582	30.12154	22.43325	21.76326	27.29314	29.91214	21.86087	26.18443
		CC	0.328249	0.461831	0.959889	0.286002	0.965681	0.963535	0.326509	0.104425	0.960554	0.535475
		MRE	66.2139	61.47497	50.6218	67.08371	53.16182	51.01971	61.01866	66.21864	51.65929	59.45903
	Testing	MSE	661.46	927.69	661.4327	927.69	661.46	661.1159	927.69	928.9049	659.1214	661.3355
		MAE	22.53333	30.43333	22.5306	30.43333	22.53333	22.48153	30.43333	30.45187	22.27877	22.52083
		RMSE	25.71886	30.458	25.71833	30.458	25.71886	25.71217	30.458	30.47794	25.67336	25.71644
		CC	0.435421	8.37E-17	0.435421	8.37E-17	0.435421	0.478321	8.37E-17	-0.17015	0.432131	0.432314
		MRE	70.48401	79.22295	70.48099	79.22295	70.48401	70.43355	79.22295	79.30574	70.20292	70.47018

Table 3: Error Analysis for Zone Three (Avg.WQI-77.95) with Good Water Quality Using CFBP Algorithm (90% Training Dataset)

Transfer Function	Data	Error Analysis	No. of Neurons									
			1	2	3	4	5	6	7	8	9	10
Tansigmoidal	Training	MSE	477.57	25.078	22.865	127.947	2.6776	591.46	3.2597	6.482789	6.473908	8.565341
		MAE	11.855	1.8868	3.1477	4.49137	1.0914	13.6515	1.1818	1.136228	1.218181	1.767669
		RMSE	21.853	5.0078	4.7818	11.3114	1.6363	24.3199	1.8055	2.546132	2.544388	2.92666
		CC	0.715	0.978	0.983	0.928	0.997	0.581	0.997	0.994	0.994	0.991
		MRE	14.833	2.8636	4.4888	6.97004	1.4358	36.3388	1.6837	1.915486	1.948948	2.64755
	Testing	MSE	33.209	2.5129	2.0332	2.568	4.7418	3.61863	2.0821	60.43425	53.72868	60.72692
		MAE	1102.9	6.3146	4.1341	6.59461	22.485	13.0944	4.335	3652.298	2886.771	3687.758
		RMSE	20.373	2.0497	1.3916	2.41807	3.7335	3.14837	1.8771	60.34017	50.62937	60.69567
		CC	0.773	0.705	0.958	0.987	0.45	-0.12	0.998	0.81	0.775	0.999
		MRE	23.163	2.2993	1.569	2.68541	4.1744	3.49501	2.0915	193.9762	161.5995	194.9385
Purelinear	Training	MSE	214.94	212.39	741.78	**205.861**	207.4	206.208	204.97	205.5121	410.2931	205.4047
		MAE	8.3436	8.2416	16.324	**7.58103**	7.8877	7.5399	7.0773	7.294128	15.74056	7.333844
		RMSE	14.661	14.574	27.236	**14.3479**	14.401	14.3599	14.317	14.33569	20.25569	14.33195
		CC	0.919	0.958	0.505	**0.964**	0.944	0.971	0.963	0.967	0.559	0.966
		MRE	22.748	22.625	42.84	**21.7859**	22.183	21.7525	21.212	21.46269	30.92674	21.49342
	Testing	MSE	3.5391	7.5528	25.514	**3.52525**	6.5902	1.29007	1.2092	3390.748	4122.92	3798.385
		MAE	1.6417	2.6365	4.9044	**1.61673**	2.429	1.00107	0.8321	58.0924	64.2	61.50983
		RMSE	1.8812	2.7482	5.0511	**1.87757**	2.5671	1.13581	1.0996	58.23013	64.20997	61.63104
		CC	0.974	0.981	-0.931	**0.999**	0.929	0.958	0.966	0.978	0	0.402
		MRE	1.8085	2.9075	5.4204	**1.8015**	2.6825	1.11908	0.933	186.5492	206.1617	197.8323
Logsigmoidal	Training	MSE	214.94	212.39	741.78	205.861	207.4	206.208	717.46	205.5121	410.2931	205.4047
		MAE	8.3436	8.2416	16.324	7.58103	7.8877	7.5399	18.063	7.294128	15.74056	7.333844
		RMSE	14.661	14.574	27.236	14.3479	14.401	14.3599	26.785	14.33569	20.25569	14.33195
		CC	0.919	0.958	0.505	0.964	0.944	0.971	0.963	0.967	0.559	0.966
		MRE	22.748	22.625	42.84	21.7859	22.183	21.7525	40.808	21.46269	30.92674	21.49342
	Testing	MSE	3.5391	7.5528	25.514	3.52525	6.5902	1.29007	1.2092	3390.748	4122.92	3798.385
		MAE	1.6417	2.6365	4.9044	1.61673	2.429	1.00107	0.8321	58.0924	64.2	61.50983
		RMSE	1.8812	2.7482	5.0511	1.87757	2.5671	1.13581	1.0996	58.23013	64.20997	61.63104
		CC	0.974	0.981	-0.931	0.929	0.999	0.958	0.966	0.978	0	0.402
		MRE	1.8085	2.9075	5.4204	1.8015	2.6825	1.11908	0.933	186.5492	206.1617	197.8323

Table 4: Error Analysis for Zone Twenty Two (Avg.WQI-90.64) with Excellent Water Quality Using CFBP Algorithm (90% Training Dataset)

Transfer Function	Data	Error Analysis	No. of Neurons									
			1	2	3	4	5	6	7	8	9	10
Tansigmoidal	Training	MAE	0.8806	0.35005	0.50208	17.5697	1.92135	0.91121	0.51517	0.53285	6.91292	0.29011
		MSE	1.66807	0.60734	0.80508	925.24	31.4811	14.9225	0.7967	0.70615	229.294	0.53046
		RMSE	1.29154	0.77932	0.89726	30.4178	5.6108	3.86297	0.89258	0.84033	15.1424	0.72833
		MRE	0.96244	0.38171	0.54159	19.2645	2.40701	0.99165	0.59274	0.60996	7.91931	0.31026
		CC	0.98	0.99	0.99	0.12	0.89	0.94	0.53	0.99	0.52	0.99
	Testing	MAE	2.20083	0.86823	2.5221	40.8667	3.2193	1.68483	0.89523	1.58273	4.2	7.24103
		MSE	4.94899	0.93099	8.76505	2402.25	10.7441	2.95029	0.836	3.25126	18.92	81.565
		RMSE	2.22463	0.96488	2.96058	49.0128	3.27782	1.71764	0.91433	1.80312	4.34971	9.03133
		MRE	2.34864	0.92879	2.7067	43.9192	3.43547	1.80044	0.9564	1.69206	4.49269	7.65208
		CC	-0.95	0.78	-0.31	0.832	-0.95	0.98	0.67	-0.29	.62	-0.99
Purelinear	Training	MAE	0.68082	0.51114	0.68625	0.61189	0.66017	0.67812	**0.67446**	0.70925	0.5452	0.58886
		MSE	1.13416	0.57159	1.56621	1.07923	1.11635	0.92352	**0.95383**	1.20089	0.88102	0.78238
		RMSE	1.06497	0.75604	1.25148	1.03886	1.05657	0.961	**0.97664**	1.09585	0.93863	0.88452
		MRE	0.77403	0.55396	0.80689	0.70821	0.75751	0.74712	**0.74505**	0.78094	0.63538	0.6574
		CC	0.99	0.99	0.99	0.99	0.99	0.99	**0.99**	0.94	0.99	0.99
	Testing	MAE	1.9316	1.07913	2.31603	2.20003	3.61483	1.24373	**1.5503**	3.14913	0.83913	1.28673
		MSE	4.6028	1.88848	6.28239	5.39471	13.5566	1.55104	**3.19714**	10.2289	0.89281	2.12996
		RMSE	2.14541	1.37422	2.50647	2.32265	3.68192	1.24541	**1.78806**	3.19826	0.94488	1.45944
		MRE	2.07139	1.13978	2.48106	2.35434	3.86258	1.32551	**1.66427**	3.36352	0.89003	1.38054
		CC	0.87	-0.98	0.59	0.81	0.87	-0.8	**0.99**	0.89	0.68	0.67
Logsigmoidal	Training	MAE	4.07879	0.59285	0.39521	0.57364	4.07878	0.55465	0.9446	0.58327	1.29896	0.5459
		MSE	48.3309	1.15377	0.48042	1.31821	48.3307	0.72111	3.8464	1.06618	12.8119	1.97862
		RMSE	6.95204	1.07414	0.69312	1.14813	6.95203	0.84918	1.96122	1.03256	3.57937	1.40664
		MRE	4.82852	0.63818	0.42097	0.62549	4.82852	0.60649	1.02209	0.63857	1.58957	0.5902
		CC	0.62	0.92	0.92	0.92	0.37	0.9	0.75	0.92	0.39	0.91
	Testing	MAE	4.2	2.3287	2.11937	1.3601	4.2	1.73857	2.2027	2.37393	0.72473	1.49623
		MSE	18.92	8.61367	5.00787	2.09051	18.92	3.92294	5.48635	6.32315	0.54475	2.55688
		RMSE	4.34971	2.9349	2.23783	1.44586	4.34971	1.98064	2.34229	2.51459	0.73807	1.59903
		MRE	4.49269	2.50242	2.26714	1.45351	4.49269	1.86571	2.35811	2.5341	0.77259	1.59855
		CC	0.62	-0.3	0.79	0.82	0.62	0.8	-0.92	-0.79	0.99	-0.71

Table 5: Zone Wise Best Fitting ANN Models

Zone	No of Neurons in Hidden Layer	Transfer Function	Cc	MSE	MAE
1	8	Logsigmoidal	0.966	33.610	3.560
2	5	Purelinear	0.874	1.33	0.839
3	4	Purelinear	0.999	1.212	0.997
4	10	Tansigmoidal	0.999	0.660	0.741
5	1	Logsigmoidal	0.997	1.095	0.880
6	2	Tansigmoidal	0.999	1.200	0.800
7	8	Logsigmoidal	0.994	11.080	2.530
8	4	Logsigmoidal	0.995	4.774	1.844
9	7	Tansigmoidal	0.980	22.935	3.203
10	3	Tansigmoidal	0.999	4.867	1.884
11	8	Logsigmoidal	0.999	4.420	1.112
12	2	Purelinear	0.999	2.980	1.478
13	3	Logsigmoidal	0.999	1.507	1.014
14	8	Logsigmoidal	0.999	2.233	1.200
15	7	Purelinear	0.998	5.266	1.985
16	7	Purelinear	0.999	1.820	1.082
17	4	Logsigmoidal	0.779	237.635	6.455
18	7	Logsigmoidal	0.999	2.595	1.253
19	1	Purelinear	0.998	20.570	2.111
20	4	Tansigmoidal	0.999	2.847	1.250
21	7	Purelinear	0.998	4.291	1.804
22	6	Logsigmoidal	0.993	1.759	1.308
23	9	Purelinear	0.999	15.804	3.513
24	7	Purelinear	0.999	2.178	1.190
25	4	Logsigmoidal	0.999	21.198	4.032
26	3	Logsigmoidal	0.987	147.978	4.942
27	10	Purelinear	0.996	5.668	2.093
28	9	Logsigmoidal	0.992	1.394	1.137
29	3	Tansigmoidal	0.999	3.521	1.780

The number of neurons in the hidden layer affects the model performance.Fig.4 shows the performance of ANN model during training and testing for the typical zone two, four, fourteen and twenty two. From Fig.4 it is observed that the model performance changes considerably with variation in number of neurons in the hidden layer. In hidden layer number of neurons are varied from 1-10. From Table 5 it is observed that hidden layer structure with seven neurons performs better, followed by hidden layer structure with four neurons and three neurons respectively. The zone wise best fitting hidden layer structure changes due to zone wise change in statistical values (mean, standard deviation, variance etc.) for various water quality parameter viz. pH, alkalinity, hardness, DO, total solids and MPN.

From Fig. 4 it can be observed that during training and testing of ANN models, ANN models with different hidden layer structures shows high coefficient of correlation (Cc) nearer to one many times, in such cases best fitting model is to be selected based on mean absolute error (MAE) and mean relative error(MRE).From Table 5 it is also observed that the ANN models shows very high degree of correlation between observed and predicted values, almost for all twenty nine zones.

IV. CONCLUSIONS

The studies on prediction of water quality index (WQI) in the distribution system for Solapur city has been carried out by using ANN models. Performance of ANN models were tested by using modeling performance criterions. The study reveals that model performance changes considerably with change of transfer function and hidden layer neuron structure. Predictions by logsigmoidal and purelinear transfer function are in good correlation with observed WQI as compared to tansigmoidal transfer function. Out of twenty nine zones in the study area for thirteen zones Logsigmoidal, for ten zones Purelinear and for remaining six zones Tansigmoidal transfer function performs better. Hidden layer structure with seven neurons performed better, followed by hidden layer structure with four neurons and three neurons respectively.

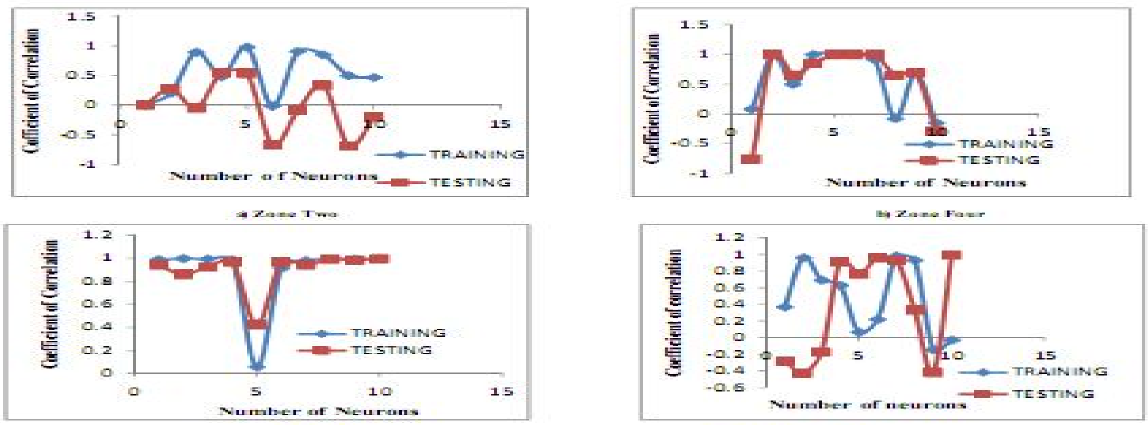
REFERENCES

[1] Kuo, Y.M., Liu, C.W, Lin, K.H. (2004). "Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan". Water Research, (38), 148-158.

[2] Yan, H., Zou, Z. and Wang, H. (2010). "Adaptive neuro fuzzy interference system classification of water quality status". Journal of Environmental science, 22(12), 1891-1896.

[3] Singh, K., Bansant, A., Malik, A. and Jain G. (2009). "Artificial neural network modeling of the river water quality – A case study". Ecological Modeling, (220), 888-895.

[4] Kumar, N.V., Mathew, S. and Swaminathan, G. (2009). "Fuzzy information processing for assessment of ground water quality". International Journal of Soft Computing 4(1), 1-9.



a) Zone Two
b) Zone Four
c) Zone Fourteen
d) Zone Twenty two

Fig.4 : ANN Model Performances vs. No.of Neurons in Hidden Layer for Zone a) Two, b) Four, c) Fourteen and d)Twenty two