Australian Journal of Basic and Applied Sciences, 6(1): 62-69, 2012
ISSN 1991-8178

Multiple Inputs Artificial Neural Network Model For The Prediction Of Wastewater Treatment Plant Performance

Mohammed S. Jami, Iman A.F. Husain, Nassereldeen A. Kabashi, Norhafiza Abdullah

Bioenvironmental Engineering Research Unit (BERU), Department of Biotechnology Engineering, International Islamic University Malaysia

Abstract: It is difficult to unveil the complicated interrelationships of wastewater parameters using linear models. A statistical modeling tool called artificial neural network (ANN) is used in this work to predict the performance of wastewater treatment plant (WWTP). Extensive influent and effluent parameters database containing measured data spanning over two years of period was used to develop and train ANN using ANN toolbox in commercially available software, MATLAB. The data were obtained from one of Sewage Treatment Plant in Malaysia. The input parameters for the ANN were BOD, SS, and COD of the influent, while the output parameters were combination of the effluent characteristics. The networks for single input-single output were compared with those of single input-multiple output. The ANN was developed for raw and screened data and the results were compared for both networks. It was found that the use of data screening is essential to come up with a better ANNs model. From the regression analysis, networks with one hidden layer and 20 neurons were found to be the best one for single input-single output approach. While the best network for the multiple inputs-single output approach was with BOD as outputs and 30 neurons. The multiple inputs- single output ANN models developed can be used in analyzing how wastewater parameters such as Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and Suspended Solids (SS) are affecting each other.

Key words: Artificial neural network, Wastewater treatment, Regression analysis.

INTRODUCTION

Artificial neural networks (ANNs) have been widely used for analyzing and simulating the nonlinearity of the data in the form of input-output. Alternatively, multivariate statistical methods were used for analyzing data. However, these statistical methods are only applicable when linear relationships are present between variables (Oliveira-Esquerre, *et al*., 2004). On the other hand, ANNs are good for analyzing incomplete data sets, fuzzy or incomplete information and for highly complex and ill defined problems, where humans usually make decisions on an intuitional basis (Sencan *et al*., 2005; Raduly, *et al*., 2007) .

Wastewater treatment influent parameters such as biochemical oxygen demand (BOD), suspended solids (SS), Ammonia Nitrogen (NH_3-N) and chemical oxygen demand (COD) are important factors that determine the volume and strength of pollution of municipal wastewater. As a result, their amounts in the wastewater affect the quality of effluent (Hamed *et al*., 2004). The connection of these parameters; both influent and effluent is always a concern to wastewater treatment plant (WWTP) process engineers in order for them to optimize the performance of the plant and ensure the safety of the environment for a proper plant operation. However, WWTP has variant inflow as well as fluctuating concentrations of the parameters. This dynamic behavior of WWTP becomes an obstacle to the effort of generating a relationship between the input and output parameters and predicting the effluent strength (Hong *et al*., 2003). Moreover, each WWTP in this world is unique because the characteristics of the influents of WWTP depend on the nature of the local lifestyle and are time-dependent. Likewise, the presence of bioorganic components fluctuate the performance of the plant. These factors contribute to difficulties in monitoring and controlling the process of WWTP. Such problems could lead to damage of the bioreactors, or even the environment if the incompletely treated effluent finds its way to the nature (Mjalli *et al*., 2007).

Removing pollutants that are incorporated in wastewater is the major objective of any WWTP in the world. The domestic and industrial wastewater have been characterized by high pollutants such as Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), Ammonia Nitrogen (NH_3-N), Phosphorus, heavy metals and pathogens. The pollutants pose a severe effect on human beings and aquatic ecosystem, if they were to be discharged into the water ways prior to treatment. To comply with the new stringent limit of wastewater effluent quality outlined by the law enforcement bodies became very demanding, due to the complex nature of the WWTP. Maintaining effluent quality of WWTP can be achieved by developing a model for predicting the plant performance based on past observations of important parameters of the plant (Mjalli *et al.*, 2007). In waste water treatment plants (WWTP), monitoring and controlling the influent

Corresponding Author: Mohammed S. Jami, Bioenvironmental Engineering Research Unit (BERU), Department of Biotechnology Engineering, International Islamic University Malaysia P.O. Box 10, 50728 Kuala Lumpur/Malaysia.
E-mail: saedi@iium.edu.my

parameters is a plant-dependent strategy that requires the understanding of the plant performance and the factors that affect the influent characteristics such as, time, season and the nature of the local people's life style. Moreover, the control of the waste water plant requires the monitoring of influent parameters such as biochemical oxygen demand (BOD), suspended solids (SS), ammonia nitrogen (NH_3-N) and chemical oxygen demand (COD).

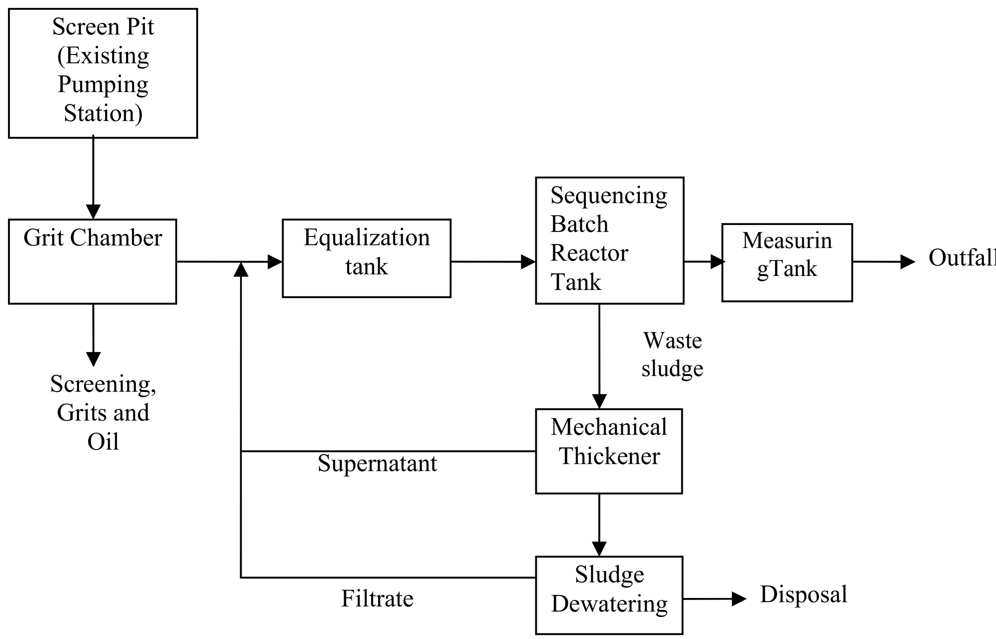

Fig. 1: The basic design of wastewater treatment plant at BTR

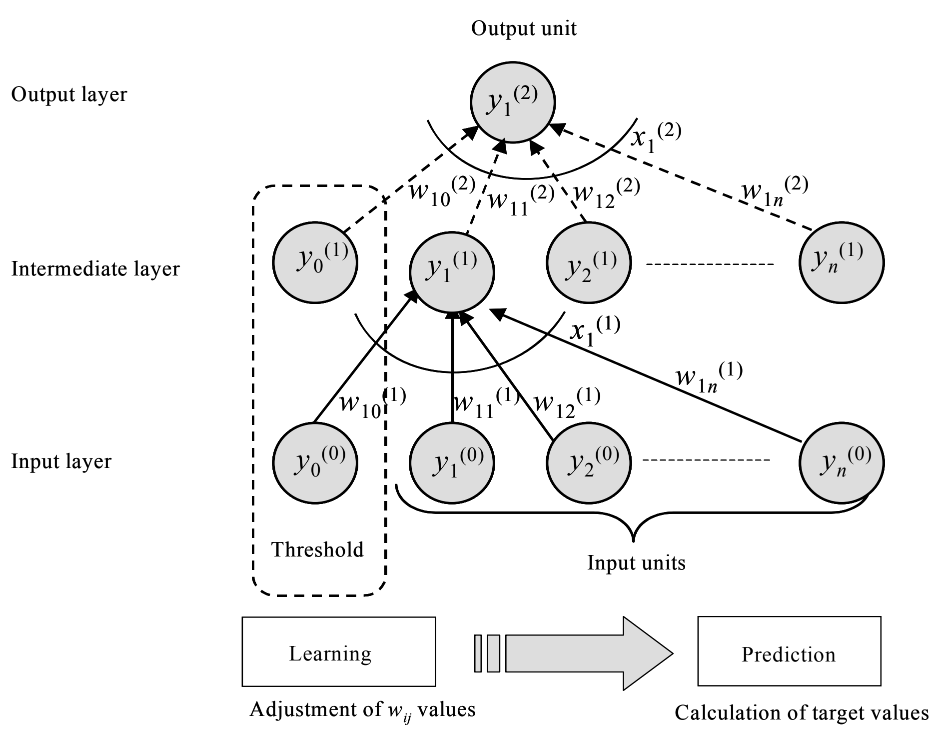

Fig. 2: Topology of the artificial neural net work model with an input layer, one hidden layer and an output laye.

Table 1: Regression and correlation results for raw and screened data of the single input-single output approach.

Network		R value & RMSE					
		Raw data			Screened data		
	neurons	10	20	30	10	20	30
BODin- BODout	R MSE	2.18203e-1 8.70493e-0	5.62677e-2 14.78775e-0	4.53120e-1 6.96598e-0	3.29477e-1 3.70349e-0	3.55638e-1 5.37020e-0	4.01611e-1 3.94935e-0
BODin- CODout	R MSE	2.31460e-1 99.04173e-0	2.2689e-1 112.3294e-0	-1.70396e-1 101069.43e-0	5.53426e-1 20.97258e-0	6.02204e-1 20.02581e-1	5.39250e-1 23.00198e-0
BODin- SSout	R MSE	2.37593e-1 14.30095e-0	2.55030e-1 14.12424e-0	1.11429e-1 33.77045e-0	4.75264e-1 4.58004e-0	3.76077e-1 5.20329e-0	1.67767e-1 11.03946e-0
CODin- CODout	R MSE	5.87835e-2 116.2917e-0	-6.77928e-2 372.0270e-0	1.73716e-1 215.343888e-0	1.50919e-1 37.06769e-0	-8.87101e-2 128.61582e-0	8.45436e-2 58.78197e-0
CODin- BODout	R MSE	1.65696e-1 9.42502e-0	1.47050e-1 10.89874e-0	4.44468e-1 7.34956e-0	3.05822e-1 4.39412e-0	2.22051e-1 5.42473e-0	1.25589e-1 11.43135e-0
CODin- SSout	R MSE	2.43896e-1 14.16660e-0	3.05934e-1 14.08984e-0	9.6815e-2 23.84508e-0	1.67314e-1 14.32691e-0	1.82710e-1 6.51484e-0	4.94775e-1 4.59753e-0
SSin- SSout	R MSE	2.43112e-1 23.76881e-0	1.88508e-1 140.9760e-0	2.33920e-1 22.41529e-0	3.87641e-1 4.88578e-0	-1.34887e-1 8.15365e-0	3.35314e-1 11.07111e-0
SSin- BODout	R MSE	-5.10172e-2 10.13807e-0	1.75257e-1 67.30890e-0	5.10105e-2 16.85928e-0	2.79885e-1 3.68421e-0	1.92904e-1 7.46435e-0	-3.08416e-2 14.54059e-0
SSin- CODout	R MSE	-1.76058e-1 61.03835e-0	1.31927e-1 117.2118e-0	1.93165e-1 306.05628e-0	-5.20707e-2 117.53354e-0	5.68348e-1 19.92609e-0	4.45934e-2 155.1662e-0

In order to optimize the performance of the treatment plant and ensure the fulfillment of the department of environments (DOE) effluent requirements, both influent and effluent parameters are monitored and controlled. However, the dynamic behavior of the WWTP complicates the process to generate relationships between input and output. Thus, artificial neural networks (ANN) provide a powerful tool for the analysis of the linear or non-linear relationships between variables. Some research works have been done concerning the application of ANN to water and wastewater treatments systems (Grishma and Chellam, 2003; Hanbay *et al*., 2008; Haykin, 1994; Hong *et al*., 2003).

In this study past data spanning two years obtained from Bandar Tun Razak Sewage Treatment Plant (BTR STP), one of the wastewater treatment plants managed by Indah Water Konsortium (IWK), Malaysia's national sewerage company were used to develop the artificial neural network. The data were analyzed before and after screening. Moreover, two approaches were applied, the first with single input-single output while the second is multiple inputs-single output.

Method:

Data Collection:

Data that were obtained from Bandar Tun Razak STP contains values of WWTP parameters measured over the span of almost two years' readings of each parameter (104 data set). This is preferable because it includes all seasonal variations in the parameter that might influence the pattern of the data. Moreover, the more the data set, the more reliable is the designed model.

Plant Description:

Bandar Tun Razak Sewage Treatment Plant of Indah Water Konsortium Sdn Bhd (IWK) is a relatively new sewage treatment plant (STP) that employs Sequential Batch Reactor (SBR), one of the most advanced systems available (Figure 1). The brand-new STP receives and treats sewage from part of Kuala Lumpur. The design capacity of the facility is 25,000 m^3 per day and equivalent to 100,000 populations. It covers approximately 1000 ha from the vicinity developments. Currently the flow is around 11,700 m^3 per day which equivalent to 52,000 populations. SBR system is capable to remove not only carbonaceous compounds (BOD and COD test) and suspended solids (SS) but also nitrogen compounds (NH_3-N, Nitrate etc) in raw sewage.

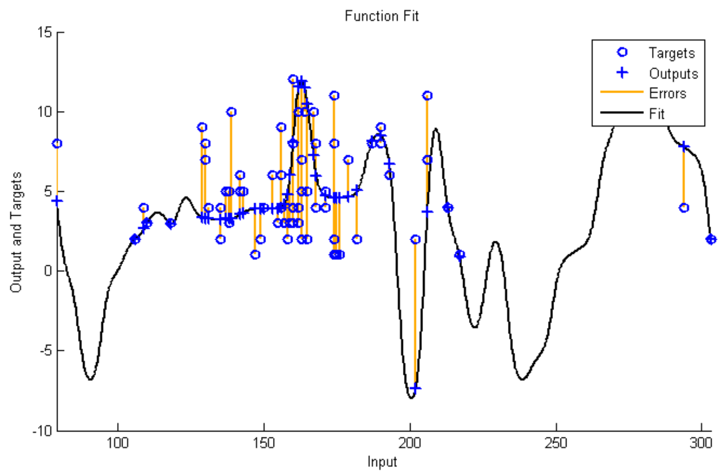

Fig. 3: Fitting graph for the raw data network with BODin-BODout and 30 neurons.

***Artificial Neural Network Modeling**:*

An artificial neural net work can be defined as a distributed computational system composed of a number of individual processing elements operating largely in parallel, interconnected according to some specific topology (architecture), and having the capability to self-modify connection strengths during the processing of element parameters (learning) (Iwata *et al*., 2007). Figure 2 shows the topology of the feed-forward ANN model used in this work. The ANN basically consists of an input layer, one hidden or intermediate layer and an output layer.

The neurons of one layer are connected to the neurons of another layer with connection weight, but they are not connected to neurons of the same layer. Each output value from a node, Y_i is weighted by the value W_i and the input to the next node, X_i is obtained by summing the product Y_iW_i. The output from this neuron is obtained by applying a non-linear activation function to X_i. All other neurons are also connected to the neurons of another layer (not shown in the figure). The connection weights between neurons are optimized using the known input and target values through an iterative process and error-minimization technique, so that the network produces outputs close or equal to the known target values.

One of the main features of ANNs methodology is to enhance the networks' speed. Solution can be generated very quickly for most problems, if a number of conditions are followed. First, the network configuration should not be too large, so that the number of connections whose weights have to be calculated is not large. Second, the training examples set should not be too long since the smaller the set, the more times every example will go through the network and the faster the solution will be obtained. Finally, the composition of training examples must be homogenous. In other words, the more the examples are, the faster the learning process. Given these conditions, data screening process is carried out on raw experimental data to eliminate all out-of-range values the presence of which might be due to transcription or transposition errors, improper input of data, and experimental errors or human errors. The data screening was accomplished by removing values that are not within the range of ± 3σ and was completed using Microsoft Excel. The components of input layers consist of three parameters, BOD influent (BOD_{in}), COD influent (COD_{in}) and SS influent (SS_{in}). Moreover, the output parameters in this modeling are BOD effluent (BOD_{out}), COD effluent (COD_{out}) and SS effluent (SS_{out}).
Data is subdivided into three groups; training, validating, testing, in the ratio of 70%: 15%: 15%, respectively. The data sequence of the inputs was obtained after normalization of the data. The fitting graph for multiple inputs-single output were generated.

Determining the Number of Hidden Layers:

A single hidden layer with different count of neurons (i.e. 10, 20, or 30) has been selected for this study. Every node in the ANNs is linked with each other by a weighing factor (W). Back propagation network (BPN) algorithm developed by D. E. Rumerlhart (Rumerlhart *et al*., 1986) is favored for such application. The BPN is a very useful learning model for ANNs. The procedure of the BPN is the error at the output layer that propagates backward to the input layer through the hidden layer in the network to obtain the final desired outputs. The gradient descent method is utilized to calculate the weight of the network and adjusts the weight of interconnections to minimize the output error. The value of weight of interconnective neuron can be expressed as

$$W_{ij}^{m} = W_{ij}^{m-1} + W_{ij}^{m} = W_{ij}^{m-1} + \eta \cdot \delta_j^n \cdot A_i^{n-1}$$

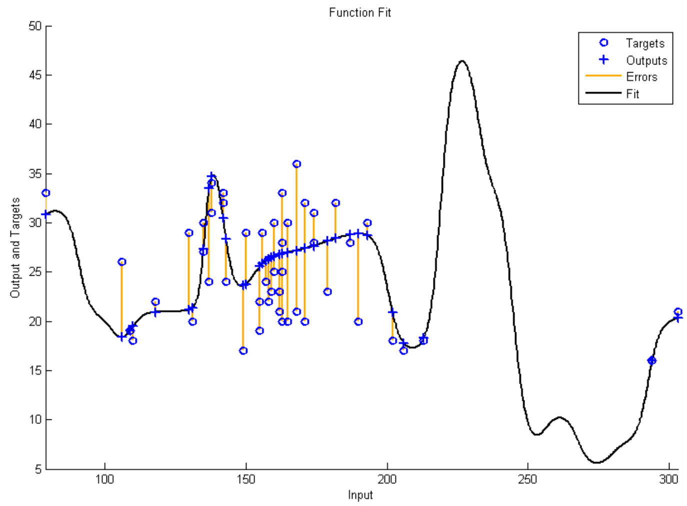

Fig. 4: Fitting graph for the screened data network with BODin-CODout and 20 neurons.

Where W_{ij} is the connective weight
η is the learning rate
δ^n_j is the error signal
A_i^{n-1} is the output value of sublayer related to the connective weight (W_{ij})
The BPN was trained to simulate and predict the output based on 56 and 30 sets of operational record.

Training of the Network:

Training of the network is performed by BPN. The main steps in the calculation of BPN are as follows:

- Step 1: Initialize all weights to small random values within the range.
- Step 2: Give the input vectors and output vectors.
- Step 3: Compute the output values in a feed forward
- Step 4: Use the values computed by the final layer unit and the corresponding target value to compute the delta quantities.
- Step 6: Update all weights.
- Step 7: Return to step 2 and repeat for each pattern until the iteration is reached.
- Step 8: Stop the procedure of training as the iteration is reached.

Two different activation functions were used during the feed forward of the training pattern, non linear sigmoid transfer function in the hidden layer and the linear transfer function in the input and output layer.
The equation for the sigmoid transfer function is

$$\mathrm{logsig(x)} = \frac{1}{1+e^{-x}}$$

Where x is the gain

Model Verification:

For model validation, the best model is selected according to the RMSE and R value. The RMSE for different assays are done for each parameter data set and different number of hidden neurons. The minimum value, the mean value and the standard deviation of the RMSE of these assays are registered. RMSE is calculated as follows

$$RMSE = \sqrt{\sum_{k=1}^{n} (y_k - \hat{y}_k)^2 / n^2}$$

Where $\hat{y}_k$ is the observed value, y_k is the predicted value.

Simulation analysis by comparing predicted values versus observed values were performed for the model verification. A good agreement between the observed and the predicted data confirms the validity of the methodology developed.

Table 2: The R and MSE values for the multiple inputs—single output approach

NO of neurons Network	R value & RMSE Raw data 10	20	30	Screened data 10	20	30
inputs-BODout	7.57295e-2 3.35977e-1	6.54542e-2 3.259032-1	6.75799e-2 4.92340e-1	1.11839e-1 5.29738e-1	1.00870e-1 4.42281e-1	3.46948e-1 4.00950e-2
inputs-CODout	5.12232e-2 4.3173e-1	5.10658e-2 3.63832e-1	5.09801e-2 3.32268e-1	6.46825e-2 3.48635e-1	2.52537e-1 3.35991e-1	5.26220e-2 6.17708e-1
inputs-SSout	4.15562e-2 2.85145e-1	4.19057e-2 3.81616e-1	1.04606e-1 1.32424e-1	6.10977e-2 3.96357e-1	9.96014e-2 1.36641e-1	1.58717e-1 1.95877e-1

RESULTS AND DISCUSSION

Overall, 54 networks were generated for the raw and screened data and 6 networks for multiple inputs-single output approach. Data from Table 1 shows *R* and RMSE of each ANN generated with variation in the number of hidden neurons (i.e. 10, 20, and 30). Regression, *R* values measure the correlation between outputs and targets. An *R* value of 1 means a close relationship, 0 a random relationship while the RMSE is the average squared difference between outputs and targets, the lower values are the better. Moreover, zero RMSE means no error.

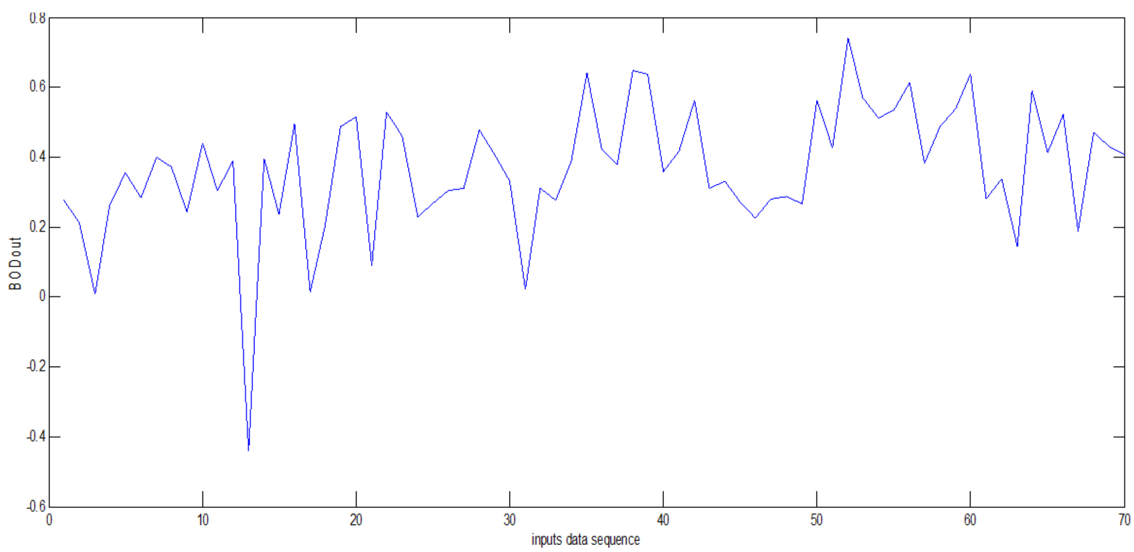

Fig. 5: Fitting graph for th enetwork with multiplt inputs-BODout and 30 neurons.

Table 2: Regression and correlation results for raw and screened data of the single input-single output approachIt can be observed from the table that the networks with higher neurons number have a less MSE and higher *R* value. Moreover, the data screening resulted in a higher *R* value for most of the networks. The *R* value has been improved up to 70% for some of the networks. Moreover, for the raw data the network with BODin-BODout and 30 neurons resulted in the lowest RMSE and highest *R* value which are 6.97 and 45%, respectively. While for the screened data, the network with BODin-CODout with 20 neurons corresponded to the highest *R* of 60% and lowest RMSE of 2.003. It can be noticed that screening the data resulted in a better network with higher *R* value and less RMSE. This can be related to the fact that data screening accelerates the process by making prior selection of the patterns or learning examples, so that only necessary patterns are chosen. Moreover, the data used for training, or the training examples, is shorter, thus lessening the time for ANNs to do forecasting. Data screening also results in a homogenous new set of data which leads to well-trained and accurate networks that are more constructive for analysis. The fitting graphs for both networks are shown in Figures 3 and 4.

The fitting graphs show that the pattern of the data from training done using past observation (past data) is useful for prediction of new values. When prediction is carried out, new set of input is entered into the network;

it is fitted into the graph to produce the predicted values. To evaluate the efficiency of the network in terms of prediction, the real values of the entered input is also included in the network. The error of the prediction is the difference between the fitted (output) values and the real (target) values and this is shown in the graph as the yellow lines.

Table 2 shows the *R* and RMSE of simulation done for the multiple inputs-single output approach. The table shows that the networks for the screened data have resulted in up to 80% improvement in the *R* and RMSE values. Moreover, higher neuron number gave better results as was concluded for the previous approach. The best network for the raw data was with SS as output and 30 neurons. However, the *R* value was very low (10%). While for the screened data the best network was with BOD as output and 30 neurons which has an *R* and RMSE values of 34% and 0.04, respectively. Moreover, the data screening resulted in 70 % improvement of the *R* and RMSE values. Figure 5 shows the fitting graph of the selected network.

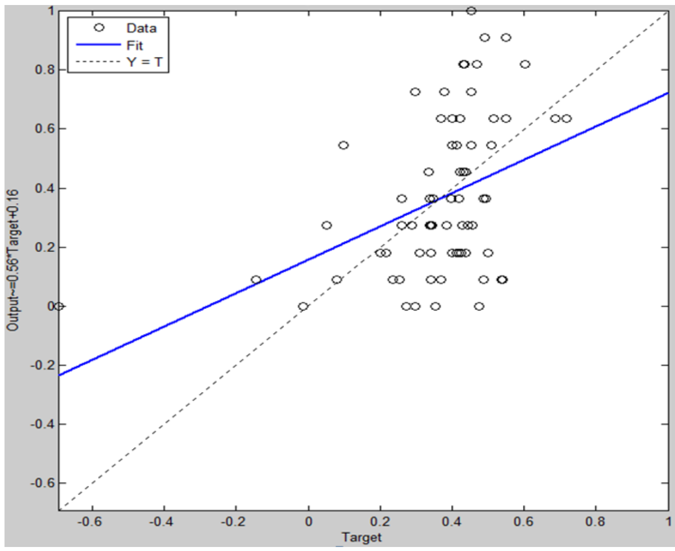

Fig. 6: The regression Analysis of the network with multiple inputs-BODout with 30 neurons for the screened data

From the regression analysis as shown in Figure 6, it is obvious that networks using screened data with one hidden layer and 30 neurons turns out to be the best one that gave better results in terms of *R* values and RMSE and thus was selected for the subsequent modeling analysis in this paper which is prediction. Moreover, the values of the RMSE which resulted from the training of the simulation of the data showed that there is a quite significant deviation between the real and predicted values. Moreover, using multiple input-single output approach resulted in a lower RMSE and showed to be better than the single input-single output approach. This is because in multiple inputs-single output approach, all the influent parameters are analyzed together accounting for the complexity of the wastewater treatment plant and the interactions among these parameters. However, in the single input-single output approach, one parameter is used while ignoring other parameters which also have an effect on the effluent characteristics.

From the result of the analysis and prediction, several discussions can be made. Data that has been screened thus becoming homogenous is more defined compared to non-homogenous data; consequently the correlation between parameters is more comprehensive and accurately linked, and lead to high *R* values. However after screening, the number of set of data is limited in contrast to raw data. This limitation of data can cause inefficiency of the networks; consequently the RMSE resulted from the prediction is relatively high for several networks. Moreover, not only the quantity of the data is important but also the quality. This includes the synchronization of data with time since the time gap between data used is not constant. Hence the ANNs is dynamic itself, such bias in data could affect the overall performance of the networks. To achieve optimum result, especially in unveiling the interrelationship between parameters, it is important to train the networks as many times as possible. A better computerized system should be used, if ANN is to be implemented in real wastewater treatment plant, for better optimization procedures.

Conclusion:

In this study, the data from an existing WWTP were used to train ANNs. It was found that screening of the data is essential to come up with better ANNs model. Moreover, using multiple inputs-single output approach results in a better ANN. ANNs developed in this study can be used to predict the performance of an existing

WWTP thus solving the problem of variations in parameters of WWTP which contributes to difficulties in monitoring and processing activities. ANNs is found to be applicable to analyze complex, non-linear, and dynamic data which makes ANNs a valid tool to study biological processes.

ACKNOWLEDGMENTS

The authors would like to thank the team of dedicated staff of Indah Water Konsortium at Bandar Tun Razak Sewage Treatment Plant for making available the data collected and for their kind support. This work was artially supported by IIUM Endowment Fund (Type B) and Fundamental Research Grant Scheme (FRGS). The authors would like to acknowledge the financial support with sincere gratitude.

REFERENCES

Grishma, R.S., S. Chellam, 2003. Predicting membrane fouling during municipal drinking water nanofiltration using artificial neural networks, Journal of Membrane Science, 2171(2): 69-86.

Hamed, M., M. Khalafallah, E. Hassanien, 2004. Prediction of Wastewater Plant Performance Using Artificial Neural Networks. Environmental Modeling and Software, 19(10): 919-928.

Hanbay, D., I. Turkoglu, I. Demir, 2008. Prediction of wastewater treatment plant performance based on wavelet packet decomposition and neural networks, Expert Systems with Applications: An International Journal, 34(2): 1038-1043.

Haykin, S., 1994, Neural Networks: A Comprehensive Foundation. Macmillan College Publishing Company, 696.

Hong, Y., M. Rosen, R. 2003. Bhamidimarri, Analysis of A municipal Wastewater Treatment Plant Using A neural Network-Based Pattern Analysis. Water Research, 37(7): 1608-1618.

Hong, Y.S.T., M.R. Rosen, R. Bhamidimarri, 2003. Analysis of a municipal wastewater treatment plant using a neural network-based pattern analysis, Water Research, 37(7): 1608-1618.

Iwata, M., M. Jami, S. Shiojiri, 2007. Artificial Neural Network Model to Predict Compression-Permeability Characteristics of Solid/Liquid Systems. Filtration Solutions, 7(4): 337-344.

Mjalli, F., S. Al-Asheh, H. Alfadala, 2007. Use of Artificial Neural Network Black-Box Modelling for the Prediction of Wastewater Treatment Plants Performance. Journal of Environemntal Management, 83(3): 329-338.

Oliveira-Esquerre, K., D. Seborg, M. Mori, R. Bruns, 2004. Application of Steady-State and Dynamic Modeling for the Prediction of BOD of an Areated Lagoon at A pulp and Paper Mill: Part II. Nonlinear Approaches. Chemical Engineering Journal, 105(1-3): 61-69.

Raduly, B., K. Gernaey, A. Capodaglio, P. Mikkelsen, M. Henze, 2007. Artificial Neural Networks for Rapid WWTP Performance Evaluation: Methodology and Case Study. Environmental Modeling and Software, *22*(8): 1208-1216.

Rumelhart, D.E., G.E. Hinton, J.L. McClelland, 1986. A general framework for parallel distributed processing. In: Parallel distributed processing: explorations in the microstructure of cognition, Cambridge, MA: The MIT Press, 1: 45–76.

Sencan, A., S. Kalogirou, 2005. A new Approach Using Artificial Neural Networks for Determination of the Thermodynamic Properties of Fluid Couples. Energy Conversion and Management, 46(15-16): 2405-2418.