Scientific
Research

Simulation of Low TDS and Biological Units of Fajr Industrial Wastewater Treatment Plant Using Artificial Neural Network and Principal Component Analysis Hybrid Method

Naser Mehrdadi[1], Hamed Hasanlou[1], Mohammad Taghi Jafarzadeh[2], Hamidreza Hasanlou[3], Hamid Abdolabadi[1]

[1]Department of Environmental Engineering, Faculty of Environment, University of Tehran, Tehran, Iran
[2]Department of Environmental Preservation, National Petrochemical Company, Tehran, Iran
[3]Department of Environmental Science, Faculty of Science, Zanjan University, Zanjan, Iran
Email: hasanlou.hamed@ut.ac.ir

ABSTRACT

Being familiar with characteristics of industrial town effluents from various wastewater treatment units, which have high qualitative and quantitative variations and more uncertainties compared to urban wastewaters, plays very effective role in governing them. With regard to environmental issues, proper operation of wastewater treatment plants is of particular importance that in the case of inappropriate utilization, they will cause serious problems. Processes that exist in environmental systems mostly have two major characteristics: they are dependent on many variables; and there are complex relationships between its components which make them very difficult to analyze. In order to achieve a better and efficient control over the operation of an industrial wastewater treatment plant (WWTP), powerful mathematical tool can be used that is based on recorded data from some basic parameters of wastewater during a period of treatment plant operation. In this study, the treatment plant was divided into two main subsystems including: Low TDS (**Total Dissolved Solids**) treatment unit and Biological unit (extended aeration). The multilayer perceptron feed forward neural network with a hidden layer and stop training method was used to predict quality parameters of the industrial effluent. Data of this study are related to the Fajr Industrial Wastewater Treatment Plant, located in Mahshahr—Iran that qualitative and quantitative characteristics of its units were used for training, calibration and validation of the neural model. Also, Principal Component Analysis (PCA) technique was applied to improve performance of generated models of neural networks. The results of L-TDS unit showed good accuracy of the models in estimating qualitative profile of wastewater but results of biological unit did not have sufficient accuracy to being used. This **model** facilitates evaluating the performance of each treatment plant units through comparing the results of prediction model with the standard amount of outputs.

Keywords: Fajr Industrial Wastewater Treatment Plant; Simulation; Artificial Neural Network; PCA; Low TDS; Biological Unit

1. Introduction

An increasing awareness about the environmental issues has motivated researchers to explore for new technologies that would be capable of predicting wastewater treatment plant (WWTP) operation. Improper operation of a WWTP may bring about serious environmental and public health problems, as its effluent to a receiving water body can cause or spread various diseases to human beings. Operation of a WWTP is often affected by a variety of physical, chemical, and biological factors. In order to follow the treatment plant performance during

the operation, effluent measurements would not be sufficient. Predicting any of these parameters, depending upon the influent water quality, will help the operator to control the system and to take necessary precautions before any problem arisen. In the industry, the use of modern control strategies is a must due to increasingly stringent regulation of effluent quality [1]. Operational control of a biological WWTP is often a complicated task, which is caused by variation in raw wastewater compositions, strengths and flow rates of wastewater owing to the changing and complex nature of the treatment process. Modeling a WWTP is considered a difficult task due to

complexity of the treatment processes. A better control of a WWTP can be achieved through developing a robust mathematical tool for predicting plant performance based on past observations of certain parameters. The complex physical, biological and chemical processes involved in wastewater treatment may exhibit non-linear behaviors which are difficult to describe by linear mathematical models. The modeling process that is traditionally used in bioprocesses is based on balance equations together with rate equations for microbial growth, substratum consumption and formation of products, and since microbial reactions coupled with environmental interactions are nonlinear, time-variable and of a complex nature, traditional deterministic and empirical modeling has shown some limitations [2]. Nowadays, many studies based on intelligent methods were conducted in wastewater treatment, which are about predictions of WWTP output parameters. Among these, Artificial Neural Networks (ANNs) are complex, nonlinear analysis mathematical systems adapted to recognize patterns with regard to the structures and parameters of the networks chosen for each application. They comprise elements corresponding to some biological neuron characteristics, which are linked to each other through weighted connections, called "synaptic weights" [3,4]. The ANN modeling approach does not require a description of how the processes occur in either the micro or macro environments and requires only the knowledge of important factors that govern the process. Due to the complexity of the processes, they are difficult to be modeled and simulated using conventional mathematical modeling. However, ANN is used in many areas of science and engineering as a promising tool because of its simplicity in simulation, prediction and modeling of process performance [5]. Recent studies indicated that consideration of statistical principles in the ANN modeling process will improve its performance [6]. For example, Principal Components Analysis (PCA), which is a technique that shows an orthogonal variable transformation, can be used for pruning ANNs and improving nonlinear mapping. PCA is a useful statistical technique and is a common technique for finding patterns in data of high dimension. Most of the studies aimed to predict overall treatment plant performance or performance of a particular process, but none of them have considered the consecutive subsections (primary and secondary treatment units) in the whole WWTP. The Models for considering the main treatment units separately and estimating multiple parameters have not yet been sufficiently developed. It is clearly mentioned by Mingzhi *et al.* (2009) that in spite of some successful practical applications, there is still not an all-inclusive procedure or method to design such intelligent controllers by far because of its semi-empirical nature [7]. Studies are generally based on two or three parameters input and single output. In this study, wastewater treatment plant was divided into two main units as biological treatment and low TDS unit and the models were established for estimating pH, Chemical Oxygen Demand (COD), Total Dissolve Solids (TDS), turbidity, total nitrogen (Total N) and total phosphorous(Total P), depending on the unit. This study aimed to develop estimation models that can provide accurate prediction of the Fajr Petrochemical Treatment Plant operation. The best prediction performance was achieved when the data were preprocessed using PCA before feeding into back propagation [8].

Brief Description of Fajr Petrochemical Effluent Treatment Plant

The Effluent Treatment Plant (ETP) receives effluents from different petrochemical companies located in the petrochemical zone of Bandar-e Imam—Southern Iran. The plant serves for treatment of wastewater before discharging into the sea or using for irrigation.

The ETP receives three types of wastewater streams:

1) Low salinity wastewater from the Amirkabir, Booali Sina, and Fajr Petrochemical Companies. The flow of this stream is in the range of 144 - 421 m^3/h.

2) High salinity wastewater from the Khuzestan petrochemical company. This wastewater stream has a high solid content. The flow of this stream is in the range of 17 - 69 m^3/h.

3) The low salinity wastewater will be pretreated in an oil-water separation area mainly consisting of a surge basin, an API **(The name is derived from the fact that such separators are designed according to standards published by the American Petroleum Institute)** oil-water separator, an equalization basin and a **DAF (Dissolved Air Flotation)** unit.

In a second step, the oil-free wastewater will be treated in an activated sludge stage together with the supernatant from offsite septic tanks. The activated sludge stage consists of an aeration basin and two clarifiers. After biological treatment, the flow passes a chlorination tank before discharging into the sea. A filtration unit is also foreseen, which will be additionally employed, when the treated water will be used for irrigation purposes (**Figure 1**). The high salinity effluent of Khuzestan Petrochemical Company (TDS = 127 g/l) will be treated in aerated lagoons separate from the treatment of low salinity wastewater. In order to facilitate biological treatment, high salinity effluent is mixed with treated effluent from the oily and sanitary water treatment facilities (Q_{max} = 229 m^3/h, TDS = 1.7 g/l). Additional treatment units are foreseen for oil water separation and sludge treatment. Chemical dosing facilities provide the possibility of pH-adjustment, addition of polyelectrolyte for flotation unit and sludge dewatering, addition of emulsion breaking polymer, nutrients for biological systems, and chlorination.

2. Materials and Methods

2.1. Principal Component Analysis (PCA)

PCA is a projection-based statistical tool that is traditionally used for reduction of dimensionality [9]. The purpose of PCA is to identify linear correlations between random variables aiming at data dimensionality reduction. The distribution of the variables is explained by a few linear features called principal components or factors, which describe major trends in the data. Consider a p-dimensional data set X that consists of zero-mean elements, $X_1, X_2 \cdots X_p$. Then, the maximum number of PCs is p and the jth principal component is:

$$Y_j = h_{1j}X_1 + h_{2j}X_2 + \Lambda + h_{pj}X_p \quad (j=1,2,\Lambda,p) \quad (1)$$

where h_{ij} is the weight value that reflects the contribution of X_i to a PC Y_j. These equations can be described as:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ M \\ Y_p \end{bmatrix} = \begin{bmatrix} h_{11} & h_{21} & \Lambda & h_{p1} \\ h_{12} & h_{22} & \Lambda & h_{p2} \\ M & M & & M \\ h_{1p} & h & \Lambda & h_{pp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ M \\ X_p \end{bmatrix} \text{ or } Y = H'X \quad (2)$$

The application field of PCA is so wide that it includes process monitoring and fault diagnosis, the development of soft sensors and process optimization. As tried in this study, PCs that were extracted from data reduction through PCA can be used as inputs to the next step analysis. Recently, many studies have applied PCA to describe nonlinear processes by integrating PCA with other nonlinear modeling approaches such as ANNs [10]. In this research SPSS software (version 15th) was used to perform PCA. The general steps of Principle Component Analysis is shown in **Figure 2**.

2.2. ANN Model Development

The parameters involved in this study were selected from the most critical operational parameters in wastewater treatment in environmental engineering (**Table 1**). They are also critical wastewater quality parameters.
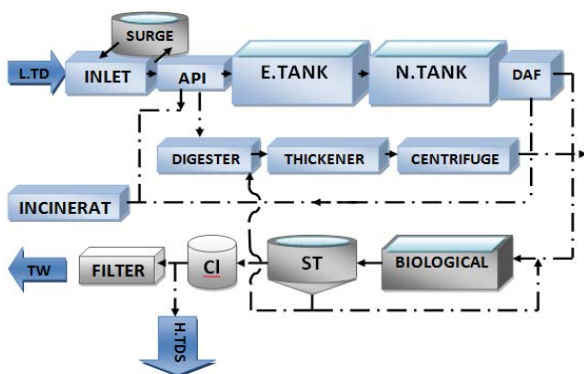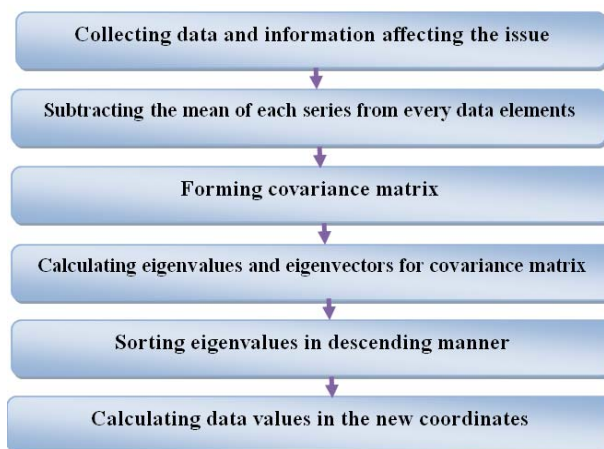

Figure 1. Flow diagram of Fajr ETP.

Figure 2. Steps of Principle Component Analysis.

Table 1. Maximum, average and minimum values of Fajr ETP parameters used in models.

Model	Parameter	Max	Ave.	Min
(Low TDS Influent)	COD (mg/l)	3850.0	1018.3	180.0
	pH	12.0	6.9	6.2
	Oil (mg/l)	756.0	180.7	50.0
	Phenol (mg/l)	13.4	10.9	0.1
	TDS (mg/l)	2910.0	512.3	459.0
L.TDS	pH	8.4	7.2	6.5
	TDS (mg/l)	2810.0	1098.1	415.0
	Turbidity N.T.U	121.0	59.7	30.6
Biological unit	pH	7.9	7.3	6.9
	COD (mg/l)	438.0	198.9	50.0
	N (mg/l)	20.8	8.4	2.2
	P (mg/l)	6.4	1.6	0.5

In the model structure, the treatment plant was divided into two main subsystems including: Low TDS treatment unit and Biological unit (extended aeration). The first system composed of equalization tank, API, neutralization tank, DAF, extended aeration, clarifiers, disinfection, filtration, and treated effluent sump. The second subsystem model includes extended aeration and clarifiers. The procedure used to develop the ANN models is outlined in **Figure 3**.

2.3. Data Collection and Preprocessing

The total data set consists of wastewater parameters of fourteen industries that located in the petrochemical zone of Bandar-e Imam and was taken from the database of

Fajr ETP of Mahshahr. To construct the predictive models, 4 variables for the Low TDS unit model and biological unit model were chosen using engineering judgment regarding which ones might have an important effect on effluent quality prediction. The nomenclature and Schematic ANN model for Fajr ETP are presented in **Table 2** and **Figure 4**.
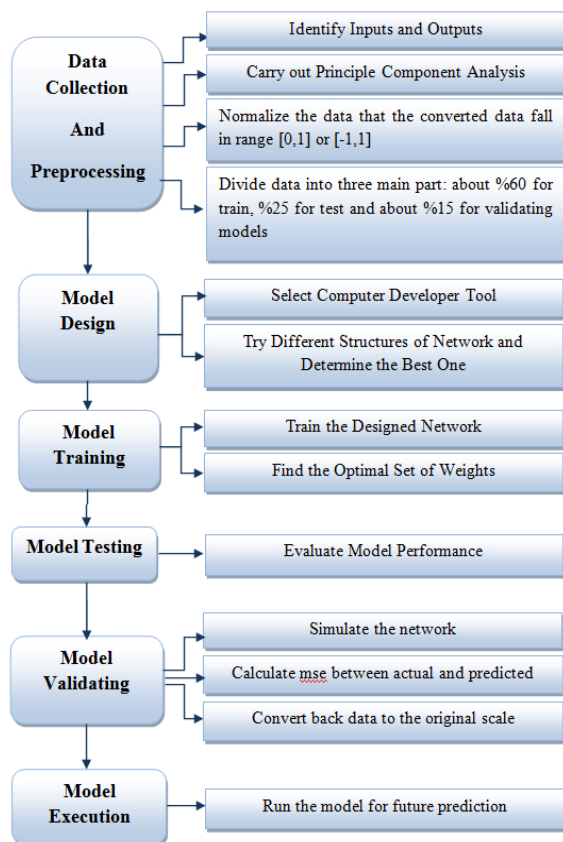

Figure 3. Model development process.

Table 2. Nomenclature.

Variable	Symbol
Influent pH of ETP	pH
Effluent pH of ETP	pH'
Influent COD of ETP (mg/l)	COD
Influent TDS of ETP (mg/l)	TDS
Effluent TDS of ETP (mg/l)	TDS'
Influent turbidity of ETP (N.T.U)	Turb
Effluent turbidity of ETP (N.T.U)	Turb'
Effluent turbidity of biological unit (N.T.U)	Turb`
Effluent nitrogen of biological unit (mg/l)	N
Effluent phosphor of biological unit (mg/l)	P

2.4. Model Structure Design

Neural Network Toolbox Neuro Solution 6 mathematical software was used to predict the effluent characteristics of Fajr ETP. The appropriate architecture of the neural network models was determined through several steps of training, testing and validating of the models. A three-layered feed-forward model with the Levenberg-Marquardt (LM) learning algorithm was used and trained with the error back propagation (**Figure 5**). The back propagation training with generalized delta learning rule is an iterative gradient algorithm designed to minimize the mean square error between the actual output of a multilayered feed-forward NN and a desired outputs.

Each layer is fully connected to the previous layer, and has no other connection. After completing the training procedure of the neural network, the weights of MLP are frozen and ready for use in the testing mode [11].

To achieve proper models configuration, functions with relatively low error level, hyperbolic tansig, logarithmic sigmoid (logsig) and purelin were selected and compared with the same training, testing and validating data. Finally, tangent sigmoid (tansig) activation function and linear transfer function (purelin) were used in hidden layer and in output layer, respectively. For a total of 151 series of data, 90 series (60% of total data) were used in training, 31 (20% of total data) in testing and 30 (20% of total data) in validating procedure.

2.5. Model Training, Testing and Validating

The purpose of the training is to capture the relationship between historical data of model inputs and corresponding
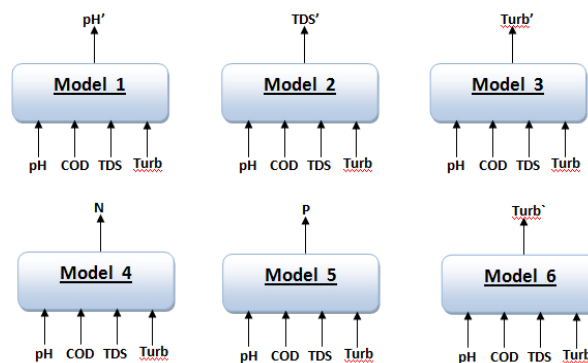

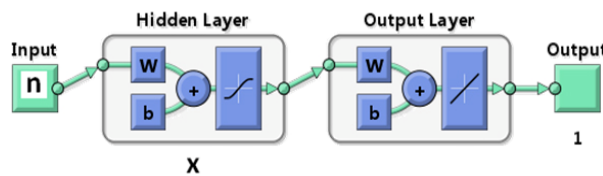**Figure 4. Schematic ANN models for Fajr ETP.**

Figure 5. The general structure of a back propagation ANN used in this work.

outputs. The back propagation is commenced by presenting the training data to the network at the input layer. The input signal flows through the network, producing an output signal, which is a function of the values of the connection weights, the transfer function and the network geometry. The learning process enables the network to find a set of weights that will produce the best possible input/output mapping. The output signal produced is then compared with the desired output signal with the aid of an error (mean squared error) function [12].

2.6. Model Execution

Once the training, testing and validating completed, we can run the models and obtain the predicted values.

3. Results and Discussions

The R value and **RMSE (Rout Mean Square Error)** error indicate how "close" one data series is to another.

In our case, the data series are the Target (actual) output values and the corresponding predicted output values generated by the model. R values range from –1.0 to +1.0. A larger (absolute value) R value indicates a higher correlation. The R values for the model on the training and test sets are close to each other, which mean the model generalizes well and is likely to make accurate predictions.

Results obtained from neural models are indicated in **Table 3**. MAE (Mean Absolute Error), determination coefficients (R^2) and RMS error are as shown in the table. R^2 is the proportion of the total variation in predicted values that can be accounted for by the linear relationship with measured values. R^2 near to 0 indicate that the data points do not fall in a well-defined straight-line shape, and the slope near to 1 in the regression equation represent the equivalence of measured and predicted values [13,14]. Using a random selection method, 60% of all data records were assigned to the training set and 20% for testing, while the remaining 20% were relegated to the validation set.

Table 3. Statistical parameters of models.

Model	Parameter	Target	Data	PCA-ANN R^2	PCA-ANN RMSE	PCA-ANN MAE	ANN R^2	ANN RMSE	ANN MAE	Records
Model 1	COD	pH								
	pH		All	0.86			0.82			151
	TDS		Train	0.92	0.118	0.089	0.9	0.162	0.124	90
	Turb		Test	0.85	0.113	0.077	0.86	0.150	0.111	31
Model 2	COD	TDS	Data	R^2	RMSE	MAE	R^2	RMSE	MAE	Rec.
	pH		All	0.82			0.77			151
	TDS		Train	0.88	133.82	94.10	0.86	143.94	105.64	90
	Turb		Test	0.67	116.76	91.54	0.60	134.21	107.03	31
Model 3	COD	Turb.	Data	R^2	RMSE	MAE	R^2	RMSE	MAE	Rec.
	pH		All	0.6			0.47			151
	TDS		Train	0.86	19.77	14.12	0.78	21.44	17.86	90
	Turb		Test	0.27	20.56	14.07	0.18	19.10	16.11	31
Model 4	COD	N	Data	R^2			R^2			Rec.
	pH		All	0.22			0.19			151
	TDS		Train	0.51			0.32			90
	Turb		Test	0.39			0.1			31
Model 5	COD	P	Data	R^2			R^2			Rec.
	pH		All	0.28			0.24			151
	TDS		Train	0.45			0.42			90
	Turb		Test	0.07			0.06			31
Model 6	COD	Turb.'	Data	R^2			R^2			Rec.
	pH		All	0.39			0.3			151
	TDS		Train	0.53			0.43			90
	Turb		Test	0.23			0.25			31

As described above, network training is done using the standard back propagation algorithm. The sigmoidal function is used as the transfer function in hidden layer and purelin transfer function in output layer due to its suitable application. The best results were obtained for the ANN composed of nine (9) neurons in hidden layer. In order to improve models performance, the PCA technique was applied for data preprocessing. PCA seeks relevant directions for the input data that maximize variance [15].

The combined use of PCA and ANN has been shown to provide prediction results that have statistical parameters significantly superior to those obtained using ANN alone. Application results for ANN and PCA-ANN structure are presented in **Figure 6**.
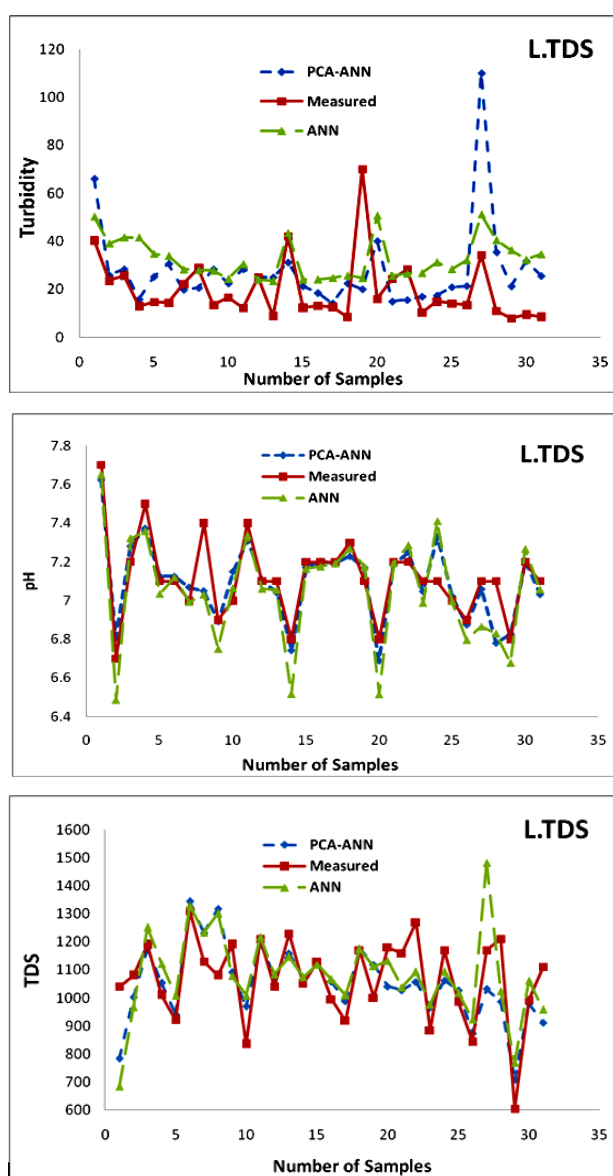

Figure 6. Comparison of the ETP data with PCA-ANN structure model and ANN model.

Highly fitting predictions were obtained in the pH and TDS values compared turbidity. Model 1, Model 2 and Model 3 gave good correlations for pH, TDS and turbidity prediction in PCA-ANN structure than ANN structure. **Figure 6** presents a graphical representation of the measured and predicted data of pH, TDS and turbidity, using the best modeling structure, *i.e.* PCA-ANN with three layers and 9 neurons in hidden layer. Considering the results from models 4, 5, and 6 that were generated in order for predicting amounts of N, P, and Effluent turbidity of biological unit and its values are shown in **Table 3**, levels of R^2 in all these models were low and in general, it can not be used for predicting these parameters. This could be due to low number of available data for modeling, complexity of treatment biological processes, and large changes in the characteristics of the wastewater entering the treatment plant that by increasing the number of parameters recorded for training and creating models, we can partially increase the model accuracy.

In general, models 4, 5 and 6 cannot be used for prediction but as is clear from the results, using PCA technique has increased the accuracy of models compared to the case in which ANN was used alone, and it can be used for simulating the wastewater treatment process.

4. Conclusions

With regards to the fact that recorded data during plant operation were used for training and generating neural models, more data can enter into the model over the time and to increase the model accuracy, it can be calibrated with more data. To improve model accuracy, PCA technique was used that increased PCA-ANN hybrid structure of every 6 models.

Given the wide range of changes in effluent quality and small number of recorded data, accuracy of data in models 4, 5 and 6 was less than the other three. Considering that effluents of 14 industrial units located in Special Economic Zone of Mahshahr were treated in Fajr Industrial Wastewater Treatment Plant, and there was lots of changes in quality of input effluents, the models 1 to 3 have an acceptable performance. But if more data are fed to models, the performance will be improved and we have more accurate outputs that facilitates evaluating the performance of each treatment plant units through comparing the results of prediction model with the standard amount of outputs.

5. Acknowledgements

The authors gratefully appreciate the assistance of all technicians in Fajr ETP for providing the data set, National Petrochemical Company (NPC) of Iran for their support, Dr. Roohollah Noori for valuable guidance and

Mrs. Neda Pournasir for her support.

REFERENCES

[1] D. S. Lee and J. M. Park, "Neural Network Modeling for On-Line Estimation of Nutrient Dynamics in a Sequentially-Operated Batch Reactor," *Journal of Biotechnology*, Vol. 75, No. 2-3, 1999, p. 229. http://www.postech.ac.kr/ce/great/publication/interjour/interjour34.pdf doi:10.1016/S0168-1656(99)00171-6

[2] M. F. Homada and I. A. Al-Ghusian, "Integrated Wastewater Treatment Plant Performance Evaluation Using Artificial Neural Networks," *Water Science and Technology*, Vol. 40, No. 7, 1999, pp. 55-65. http://www.sciencedirect.com/science/article/pii/S0273122399005843

[3] F. I. M. Argoud and M. F. De Azevedo, "An Effective System for Automated Detection of Epileptiform Events in Long-Term EEG Based on Context Information," *Medical and Biological Engineering and Computing*, Vol. 44, No. 6, 2006, pp. 459-470. doi:10.1007/s11517-006-0056-y

[4] G. Tezel, E. Yel and R. K. Sinan, "Artificial Neural Network (ANN) Model for Domestic Wastewater Treatment Plant Control," *Balwois* 2010, Ohrid, 2010, p. 25.

[5] S. N. Prakash, A. Manikandan, L. Govindarajan and V. Vijayagopal, "Prediction of Biosorption Efficiency for the Removal of Copper (II) Using Artificial Neural Networks," *Journal of Hazardous Materials*, Vol. 152, No. 3, 2008, pp. 1268-1275. doi:10.1016/j.jhazmat.2007.08.015

[6] A. Kardam, K. R. Raj and J. K. Arora, "Artificial Neural Network Modeling for Sorption of Cadmium from Aqueous System by Shelled Moringa Oleifera Seed Powder as an Agricultural Waste," *Journal of Water Resource and Protection*, Vol. 2, No. 4, 2010, pp. 339-344. doi:10.4236/jwarp.2010.24039

[7] H. Mingzhi, Y. Ma, W. Jinquan and W. Yan, "Simulation of a Paper Mill Wastewater Treatment Using a Fuzzy neural Network," *Expert Systems with Applications*, Vol. 36, No. 3, 2009, pp. 5064-5070.

[8] K. P. Oliveira-Esquerre, M. Mori and R. E. Bruns, "Simulation of an Industrial Wastewater Treatment Plant Using Artificial Neural Networks and Principal Components Analysis," *Brazilian Journal of Chemical Engineering*, Vol. 19, No. 4, 2002, pp. 365-370. doi:10.1590/S0104-66322002000400002

[9] H. Zhao, *et al.*, "Modeling Nutrient Dynamics in Sequencing Batch Reactor," *Journal of Environmental Engineering*, Vol. 123, No. 4, 1997, p. 863. doi:10.1061/(ASCE)0733-9372(1997)123:4(311)

[10] D.-J. Choi and H. Y. Park, "A Hybrid Artificial Neural Network as a Software Sensor for Optimal Control of a Wastewater Treatment Process," *Water Research*, Vol. 35, No. 16, 2001, pp. 3959-3967. doi:10.1016/S0043-1354(01)00134-8

[11] S. Hayken, "Neural Networks: A Comprehensive Foundation," Prentice-Hall International, Inc., Englewood Cliffs, 1994.

[12] M. Vyas, B. Modhera, V. Vyas and A. K. Sharma, "Performance Forecasting of Common Effluent Treatment Plant Parameters by Artificial Neural Network," *ARPN Journal of Engineering and Applied Sciences*, Vol. 6, No. 1, 2011, pp. 38-42. http://www.arpnjournals.com/jeas/research_papers/rp_2011/jeas_0111_439.pdf

[13] R. Noori, M. A. Abdoli, A. Ameri and M. Jalili-Ghazizade, "Prediction of Municipal Solid Waste Generation with Combination of Support Vector Machine and Principal Component Analysis: A Case Study of Mashhad," *Environmental Progress & Sustainable Energy*, Vol. 28, No. 2, 2009, pp. 249-258.

[14] R. Noori, M. A. Abdoli, A. Ameri, M. Jalili-Ghazizade and R. Samifard, "Comparison of ANN and PCA Based Multivariate Linear Regression Applied to Predict the Weekly Municipal Solid Waste Generation in Tehran," *Iranian Journal of Public Health*, Vol. 38, No. 1, 2009, pp. 74-84.

[15] R. Noori, A. R. Karbassi, K. H. Ashrafi, M. Ardestani and N. Mehrdadi, "Development and Application of Reduced-Order Neural Network Model Based on Proper Orthogonal Decomposition for BOD5 Monitoring: Active and Online Prediction," *Environmental Progress & Sustainable Energy*. doi: 10.1002/ep.10611