*Desalination and Water Treatment*
www.deswater.com

doi:10.5004/dwt.2019.24385

164 (2019) 86–97
October

# Monitoring effluent quality of wastewater treatment plant by clustering based artificial neural network method

Elnaz Sharghi[a,*], Vahid Nourani[a,b], Atefeh AliAshrafi[a], Hüseyin Gökçekuş[b]

[a]*Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran, Tel. +98-914-417-8419, email: sharghi@tabrizu.ac.ir, (E. Sharghi), Tel. +98-914-403-0332, email: nourani@tabrizu.ac.ir (V. Nourani), Tel. +98-914-307-5722, email: atefehashrafi.a@gmail.com (A.A. Ashrafi)*

[b]*Faculty of Civil and Environmental Engineering, Near East University, Nicosia, Turkish Republic of Northern Cyprus, Tel. +90-548-853-4960, email: huseyin.gokcekus@neu.edu.tr (H. Gökçekuş)*

Received 21 January 2019; Accepted 10 May 2019

**ABSTRACT**

The artificial neural network (ANN), as a data-driven approach, is a powerful tool for forecasting effluent quality of wastewater treatment. However, selecting appropriate input variables is a major challenge in developing ANN models. Recent studies in various fields have highlighted the usefulness of different clustering methods in identifying appropriate input variables, which, however, has largely been unexplored in classifying wastewater quality parameters. This study was carried out to fill this knowledge gap. Three ANN models were developed with different clustering methods, to forecast effluent quality of Tabriz city's wastewater treatment plant. Model A used principal component analysis (PCA) for input selection, model B used those variables identified by non-linear mutual information (MI) measure. In model C, the self-organizing map (SOM) method was used as an artificial intelligence (AI)-based method to cluster data and impose the representative parameters of each cluster as inputs of ANN. Model C presented a more favorable and optimal ANN structure in comparison with models A and B and showed up to 8 % and 23% increment in determination coefficient (DC) efficiency criterion respectively. While the number of parameters involved in the wastewater treatment process are quite many, the proposed model by employing an AI-based clustering method could successfully predict the effluent quality using the minimum number of essential input parameters. Thus, this study highlights the superiority of the SOM technique in selecting dominant input variables for ANN modeling of WWTP efficiency performance, not only because of the enhanced performance of the model with respect to various indicators but also because such a superior result was achieved by an optimal ANN architecture.

*Keywords:* Wastewater treatment plant; Biochemical oxygen demand; Artificial neural networks; Clustering methods; Self organizing map; Tabriz wastewater treatment

## 1. Introduction

The population of the world has rapidly grown over the past few years,and the most visibly affected area of this population booming is access to adequate clean water. Regarding the scarcity of freshwater resources, it becomes important to increase the wastewater treatment performance for future uses of treated wastewater.

Water quality is a major concern around the world, and this concern is generally greatest when available water quantities are low, and maximum use must be made of the limited resources. Industrial and municipal wastewaters are major contamination sources of aquatic biota, according to the several thousand types of chemicals that release into the environment [1]. Improper operation of a wastewater treatment plant (WWTP) may bring about serious environmental and public health problems, and since discharging its effluent to a receiving water body can cause or spread various diseases to human beings and pose severe effect

*Corresponding author.

on the aquatic ecosystem, the importance of implementing effective monitoring and controlling techniques for wastewater systems is a well-known issue for water and environmental engineers. However, modeling a WWTP is a difficult task due to the complexity of the treatment processes [2]. The complex physical, biological and chemical processes involved in the wastewater treatment process exhibit nonlinear behaviors which are difficult to be described by classic linear mathematical models [3]. Therefore, reliable process monitoring of treatment systems may be achieved by developing robust non-linear methods that are capable of predicting the performance of WWTP based on the past observations of quality parameters.

Artificial neural networks (ANNs) as such non-linear models are computer techniques that attempt to simulate the functionality and decision – making processes of the human brain [4]. ANNs have been increasingly applied to various environmental modeling [5,6], and water quality problems [7,8]. In the field of WWTP modeling, ANNs have been successfully utilized for prediction of WWTP parameters [9–11], process control of WWTP [12–15], estimating output parameters and characteristics of WWTP [16,17]. Most of such studies need different input data, which is not easily accessible and makes the modeling process more expensive and time-consuming.

In recent WWTPs, advanced technologies and electronic sensors provide the opportunity of collecting a wide variety of variables. Besides, the dynamic behavior of WWTP becomes an obstacle to the effort of generating a simple relationship between the input and output parameters and predicting the effluent [18]. Therefore, the need for robust approaches that can extract the most meaningful and relevant information among "high-dimensional" data set is growing due to the huge amount of accessible data [18]. Furthermore, an ANN model which includes excessive information demands a large computational memory while the irrelevant variables make the learning process more complex and this issue could result in imprecise and mis-convergence [19].

In the light of aforementioned points, selection of the most dominant variables is a crucial issue in ANN-based model development. In this regard clustering as a data pre-processing approach can be used for dominant input selection for ANN or other data-driven methods, see [20,21]. In ANN-based modeling, clustering analysis is mostly implemented for classifying different data sets into relevant classes [22], and for optimizing the structure of the model by identifying the most pertinent inputs [19]. In the WWTP modeling content, various methods have been used in order to determine suitable inputs for forecasting effluent quality. Lou et al. [10] computed the correlation coefficient (CC) between parameters pf WWTP and sludge volume index in order to select the dominant inputs for developing ANN model and predicting the sludge volume index. Dogan et al. [11] employed sensitive analysis to investigate the relative importance of the input parameters of the ANN model that was developed for predicting effluent Biochemical Oxygen Demand (BOD). Pie et al. [23] calculated the CC values of different parameters of industrial WWTP with the aim of selecting important input parameters and predicting the effluent quality of WWTP. Wan et al. [21] applied adaptive-network-based fuzzy system (ANFIS) for predicting suspended solids (SS) and chemical oxygen demand (COD) of paper mill WWTP, in this way, fuzzy subtractive clustering was employed, meanwhile, PCA was applied to reduce the dimension of input variables.

Kohonen self-organizing map (SOM) as AI (artificial intelligence) based clustering method has superior visualization capability and the ability to highlight correlation patterns between the available data to classify and interpret the behavior of the process [24]. It operates as an effective tool to convert complex, nonlinear, statistical relationship between high dimensional data items into a simple, geometric relationship on a low-dimensional display so as to allow the number of clusters to be determined by inspection [25]. The SOM based classification is attractive, due to its topology preserving property and solving various problems that traditionally have been the domain of conventional statistical and operational research techniques [22]. SOM is trained using an unsupervised learning algorithm to identify hidden patterns in un-labeled input data. This unsupervised refers to the ability to produce a low-dimensional discretized representation of the input space. The lack of direction for the learning algorithm in unsupervised learning can sometimes be advantageous since it lets the algorithm to look back for patterns that have not been previously considered.

In comparison with other common clustering techniques such as K-means, SOM seems to have more robust result due to its higher accuracy, less computational time, and error rate [26]. Although successful applications of SOM have been reported in various hydro-environmental processes, see [22,25]. To the best of authors' knowledge, the capability of this robust method and its efficiency in increasing the accuracy of ANN models for predicting the effluent quality of WWTP had been overlooked. The main objective of the current study is to evaluate the application of the SOM clustering technique in identifying the dominant input variables to develop an ANN model for forecasting effluent BOD, as one of the most important parameters of Tabriz city WWTP, located at northwest Iran, to evaluate the performance of a real WWTP. Furthermore, mutual information (MI) as a non-linear measure is also used for dominant inputs selection of the ANN method. Unlike the SOM, the goal of MI as a supervised method is to find a mapping from inputs to outputs, so it requires the target parameter to be specified, whereas, unsupervised learning model (such as SOM) identifies the pattern class information heuristically [28]. Although linear CC has been widely used to find the strength of relationships between input and target variable in ANN based modeling of WWTP [10,11,23], wastewater treatment process is highly non-linear and consists of complex biochemical and chemical dynamic processes which make linear evaluation methods insufficient. Nourani et al. [29] criticized using linear input selection method (such as CC) in non-linear AI based modeling framework such as ANN. Finally, the obtained results by SOM and MI-based methods are compared with the results of multivariate statistical projection method of PCA which has been already applied as a data pre-processing method in WWTP context [30–32]. In fact, both the MI method and PCA are adopted as the preprocessing purpose before modeling WWTP by AI method. With the hypotheses that the application of preprocessing techniques would contribute to the increment of modeling accuracy. While the SOM investigates the relationship of wastewater parameters based on unsupervised and competitive learning approach, the MI method provides the chance of investigating this relation from supervised perspective and, while the SOM was utilized to approximate a nonlinear mapping between two data spaces, by PCA technique, linear relationship of the mentioned parameters was also examined. Hence, the clustering potential of the SOM have been explored and compared with different perspectives for deriving a comprehensive conclusion.

## 2. Materials and methods

### 2.1. Study area and data

The Tabriz WWTP has been oriented at a distance of four kilometers west of Tabriz in the Qaramalek village lands (Fig. 1) and on the south side of the Ajichay River in an area of 72 ha. The capacity of Tabriz WWTP is about 765 $m^3/d$. It comprises fine and coarse screens, grit chambers, a pretreatment unit, anaerobic tanks, anoxic tanks, activated sludge aeration tanks, secondary sedimentation tanks, and a disinfection unit. Influent wastewater initially passes through fine and coarse screens and then enters to grit removal phase. A cylindrical-shaped grit chamber is capable of collecting particles greater than 20 mm. In the grit chamber, the inorganic matters settle with a hydraulic retention time of 3.8 min. After passing the pretreatment processes, the wastewater enters the anaerobic tank, anoxic tank and aeration tank, and secondary clarifier, respectively. The effluent is disinfected by the end of the process (Fig. 2). The capacity of this WWTP has been planned for wastewater from a population of six million by the year 2025.

The Tabriz WWTP was examined for five years (from 2013 to 2017) regarding BOD removal. Determining BOD values after five days ($BOD_5$) has been adopted as a compromise between a short test-period and the detection of a practically complete biological breakdown of organic materials. With domestic effluents at 20°C, a complete degradation (= 100% BOD) is achieved only after 21 d; however, after only 5 d, 70% of the biologically convertible substances are broken down,and it could be considered as a target parameter. Since the change in DO concentration has been measured after five days in Tabriz WWTP, and due to the Markov nature of treatment system, it is hypothesized that the BOD of the effluent at the current and previous time steps have a direct impact on the quality and the consumed oxygen after five days. Thus, $BOD_{t-1}$, $BOD_{t-2}$ are introduced to the model in order to examine this hypothesis. The assessment variables of the system include the temperature (T) of the ambient air and wastewater, flow rate (Q), pH, Electrical conductivity (EC), settleable solids (SS), total suspended solids (TSS), volatile suspended solids (VSS), total dissolved solids (TDS), mixed liquor suspended solids (MLSS), chemical oxygen demand (COD), all in the influent and BOD at the effluent as target.

The statistical features of the parameters, for the gathered data, from 2013 to 2017 are tabulated in Table 1.

Considering Table 1, it could be concluded that the pH was quite steady over 7.19–8.5. The temperature of the wastewater follows the seasonal temperature of Tabriz city, high in the summer and low in the winter. The highest standard deviation (90.99) belongs to EC which is an indicator of its fluctu-

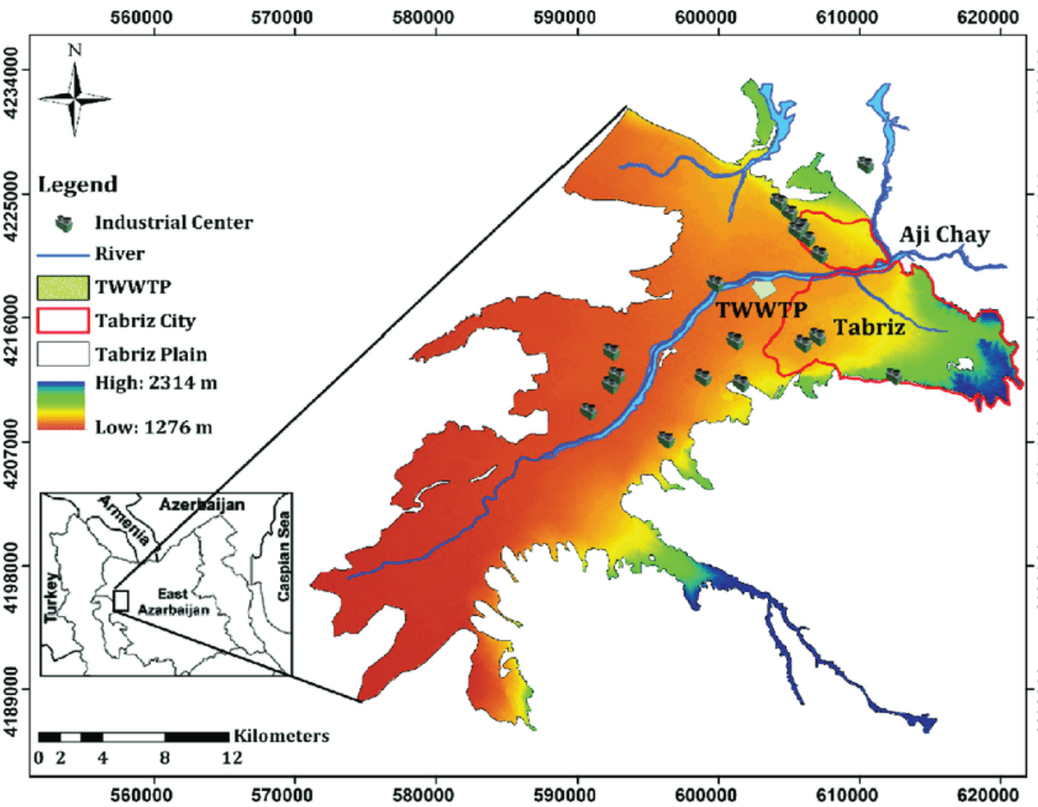



Fig. 1. Wastewater treatment location in Tabriz.

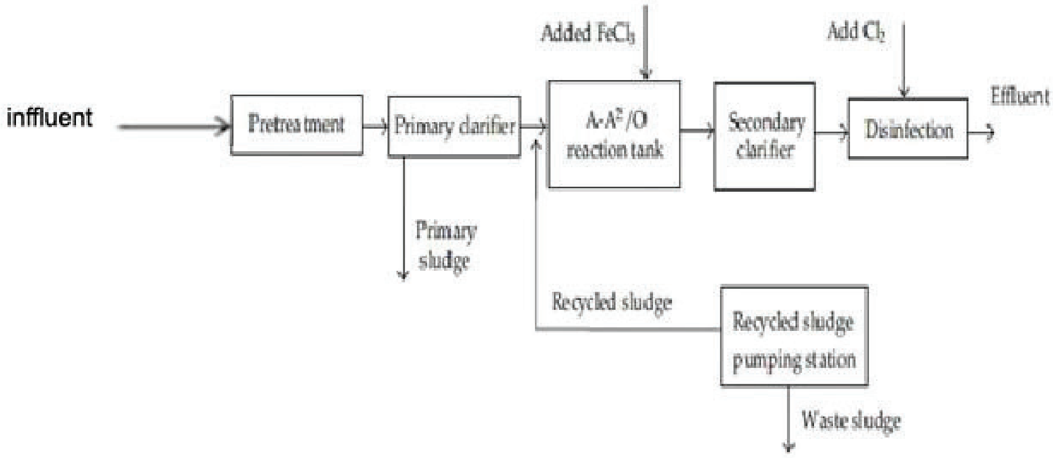



Fig. 2. Wastewater treatment process of Tabriz city WWTP.

Table 1
Wastewater characteristics of Tabriz WWTP

| Parameter (Units) | Maximum | Minimum | Average | Standard deviation |
|---|---|---|---|---|
| $Q_{in}$ ($m^3$/day) | 173247 | 32245 | 145750.7 | 3.49 |
| $PH_{in}$ (°C) | 8.57 | 7.19 | 7.99 | 0.17 |
| $BOD_{in}$ (mg/L) | 350 | 210 | 290.59 | 20.71 |
| $COD_{in}$ (mg/L) | 600 | 348 | 478 | 33.11 |
| $SS_{in}$ (mg/L) | 1.2 | 0 | 0.45 | 0.143 |
| $TSS_{in}$ (mg/L) | 384 | 214 | 304.36 | 20.26 |
| $VSS_{in}$ (mg/L) | 324 | 113 | 208.025 | 22.57 |
| $TDS_{in}$ (ppm) | 1520 | 852 | 1212.4 | 76.58 |
| $T_{in}$ (C) | 26.7 | 12.2 | 20.83 | 2.99 |
| MLSS (mg/L) | 2910 | 1113 | 2204.02 | 142.9 |
| $EC_{in}$(µSiemens/cm) | 1860 | 1334 | 1594.66 | 90.99 |
| $BOD_{out}$ (mg/L) | 44 | 13 | 23.05 | 3.02 |

ation over time. The ratio of BOD/COD is about 0.60 which is in the normal range of municipal wastewater, emphasizing that it is biodegradable wastewater. Due to the probability of bulking occurrence in activate sludge process, the MLSS is unstable,and it ranges from 1113 to more than 2000 mg/L.

Correlation coefficients between $BOD_{eff}$ and other 13 parameters were calculated in order to analyze the effect of each variable on the effluent BOD. It is obvious from Table 2 that CCs between $BOD_{eff}$ and $BOD_{in}$, COD, TSS are greater than 0.20 which indicates that these parameters, in comparison with other parameters, have a stronger correlation with effluent BOD. It worth mentioning that CC deals with the only linear relationship while WWT process is a Markov and nonlinear phenomenon. As seen in Table 2, if CC was used as a measure to select dominant inputs, both $BOD_{in}$ and $COD_{in}$ were selected mathematically, but it is clear physically that these two are dependent variables and it doesn't seem logical to impose both as the input of the model in the next step. Therefore, CC is an inadequate tool for explaining the most pertinent variables with the output.

## 2.2. Feed-forward neural network

ANN as a "black box tool" follows human brains pattern in order to approximate the nonlinear relationship between inputs and outputs of any process [33]. ANN is capable of providing a framework for mapping the input set and the output set of variables. The feed-forward neural network (FFNN) and back propagation (BP) algorithm are widely applied in engineering problems. Three-layer FFNN

Table 2 correlation coefficients between BOD and water parameters

| Water parameter | $BOD_{in}$ | $COD_{in}$ | TSS | $BOD_{t-1}$ | $BOD_{t-2}$ | $pH_{in}$ | SS | $VSS_{in}$ | MLSS | $EC_{in}$ | $TDS_{in}$ | T | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $BOD_{eff}$ | 0.33 | 0.29 | 0.25 | 0.25 | 0.21 | 0.12 | 0.16 | 0.052 | 0.043 | 0.031 | 0.018 | 0.013 | 0.007 |

which is trained by the BP algorithm seems to be adequate for prediction purposes in the engineering problems [34]. BP algorithm needs no prior knowledge of weak learner and is a commonly applied method for model training due to its high level of versatility, flexibility, and accuracy. The term "feed-forward" refers to the fact that information flows only from the input layer to the hidden neurons and finally to the output layer. In FFNN models the following equation can be used for target value determination [35]:

$$\hat{y}_k = f_o\left[\sum_{j=1}^{M_N} w_{kj}, f_h\left(\sum_{i=1}^{N_N} w_{ji}x_i + w_{jo}\right) + w_{ko}\right] \tag{1}$$

where $w_{ji}$ denotes a weight in the hidden layer which connects the *ith* neuron in the input layer and the *jth* neuron in the hidden layer, *jth* hidden neuron's bias is indicated by $w_{jo}$, $f_h$ is the activation function of the hidden neuron, $w_{kj}$ is a weight in the output layer connecting the *jth* neuron in the hidden layer and the *kth* neuron in the output layer, $w_{ko}$ is the bias for the *kth* output neuron, $f_o$ is the activation function for the output neuron, $x_i$ is *ith* input variable for the input layer, $\hat{y}_k$ and $y$ represent the calculated and observed output variables, respectively. $N_N$ and $M_N$ are the number of neurons in the input and hidden layers, respectively. The weights are different in the hidden and output layers, and their values can be changed during the process of the network training. In this study, determination coefficient (DC) [Eq. (2)] and root mean square error (RMSE) [Eq. (3)], which rank from 0 to 1, were used as the efficiency criteria in order to assess the performance of the ANN modeling [36]. The smaller RMSE denotes a more convenience result, and the DC which is close to 1 denotes a better fit of data. While the value of RMSE depends on the units of the predicted variable, DC is dimensionless and can be used to compare models by different units. Also, the RMSE values can be used to distinguish model performance in a calibration period with that of a validation period as well as to compare the individual model performance to that of other predictive models. Legates and McCabe [37] indicated that these two criteria are capable of evaluating hydro-environmental models sufficiently.

$$DC = 1 - \frac{\sum_{i=1}^{n}\left(I_i - \hat{I}_I\right)^2}{\sum_{i=1}^{n}\left(I_i - \hat{I}_i\right)^2} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(I_i - \hat{I}_i\right)^2}{n}} \tag{3}$$

where $\bar{I}_i$ indicates the average value over n data, $\hat{I}_i$ and $I_i$ are stand for the observed data and predicted data, respectively. In order to obtain convergence within a reasonable number of cycles, and adjusting values to the same scale, the input and output data should be normalized and scaled to the range of 0–1 by Eq. (4) [38].

$$x_{ni} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

where $x_i$ is the initial value, $x_{max}$ and $x_{min}$ are the maximum and minimum of the initial values, and $x_{ni}$ is the scaled value.

To develop the ANN model for predicting effluent BOD, the data set was divided into 70% training data set and the rest 30% was proportioned into validating and testing data sets, 15% for validating and 15% for the testing purposes. As described above, network training was carried out using the standard back propagation algorithm. The sigmoidal function was used as the transfer function in both the hidden and output layers due to its suitable application, especially, for continuous-value input/output pairs. The optimal hidden layer was determined by varying the total number of nodes from 1 to 10. The stop criteria are based on the RMSE for the validation set instead of that for the training set to ensure model generalization.

In the case of using the trial-error method in current work for predicting BOD of WWTP within 13 input variables, $2^{13}$–1 trials would be needed for finding appropriate input combination which was a time-consuming process. All the mentioned 13 input variables do not affect the output parameter equally and do not contribute to the informative data set. Therefore, using the selected pertinent variables as the model inputs not only simplifies the structure of the model but also yields more convincing results [22]. In the presented study PCA as a linear statistical method, and MI as a supervised non-linear method, and SOM as an unsupervised non-linear method were applied and compared in order to evaluate the best approach for appropriate input selection.

### *2.3. Principal component analysis (PCA)*

During the last two decades, multivariate statistical techniques have become increasingly popular in many fields. PCA analysis and developments based on PCA, such as principal component regression (PCR) and projection to latent structures (PLS), have been applied successfully to model various industrial processes [39], including WWT modeling [40].

The basic idea behind PCA is that the colinear nature of data is utilized to reduce the dimensionality of the measurement space by introducing a number of pseudo-variables (principal components). These pseudo-variables describe the main mechanisms that drive the process and normally are fewer than the number of measured variables. PCA is one of the multivariate statistical methods which can use to reduce the complexity of input variables when we have a huge volume of information and we want to have a better interpretation of variables [41].

PCs specified by the equation as:

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \ldots + a_{in}X_n \tag{5}$$

In Eq. (5), $Z_i$ represents specific PCs, $a_i$ is related to eigen vector and $X_i$ are also input variables [42]. This information achieved by solving Eq. (12):

$$|R - I\lambda| = 0 \tag{6}$$

where $I$ is unit matrix, $R$ is variance-covariance matrix and λ is eigen value. From these eigen values, we can achieve the eigen vectors. Details for mastering the art of PCA can be found in [43].

In this study, the PCA was employed to classify relevant variables of wastewater system, and to express their "interrelation patterns" on the effluent BOD. Existence of a strong correlation between the variables that are included in the study is essential for a good factor analysis. Kaiser-Meyer-Olkin (KMO) test that measures the property of our data for factor analysis can also measure the sampling adequacy for each variable in the model. Hence, this measure for sample adequacy was used in order to verify the applicability of PCA as:

$$KMO_j = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} u_{ij}^2} \quad (7)$$

where *rij* is the correlation matrix and *u* is the partial covariance matrix. MATLAB software was used for this purpose.

PCA as a popular tool for dimensionality reduction is widely used to identify essential information and convert high-dimensional data into a lower-dimensional space for providing more comprehensible data. However, PCA is focused on finding orthogonal projections of the dataset that contains the highest possible variance in order to find hidden linear correlations between variables of the dataset. In other words, if some of the variables in the dataset are linearly correlated, PCA can find directions that represent the data, but if the data are not linearly correlated, PCA would be inadequate. To address this problem, supervised non-linear MI and unsupervised non-linear were also used in the present study.

### 2.4. Mutual information (MI)

MI, in information theory, is described as a measure that specifies the "stochastic dependency" between two random variables without considering their relations nature [44]. In other word, MI quantifies the amount of information between random variables.

Let $X$ be a system which takes on $x_1$, $x_2$,…..,$x_M$ values. The Shannon entropy ($H$) of this system is defined as [45]:

$$H(X) = -\sum_{i=1}^{M_x} p(x_i) \log p(x_i) \quad (8)$$

where $M_x$ and $p(x_i)$ indicate the number of possible states and the probability of each possible values, respectively.

For two separate systems of $X$ and $Y$, the joint entropy is:

$$H(X, Y) = -\sum_{i=1}^{M_x} \sum_{j=1}^{M_Y} P(x_i) \log p(x_i) \quad (9)$$

where $p(x_i, y_j)$ is the joint probability, $X$ is in state $x_i$ and $Y$ is in state $y_j$, $M_X$ and $M_Y$ are numbers of states that could be different [46]. Mutual information $I$ ($X$, $Y$) between systems $X$ and $Y$ is described as:

$$I\,(X, Y) = H\,(X) + H\,(Y) - H\,(X, Y)\,, \geq 0 \quad (10)$$

Overall equation from (8) to (10) yields:

$$I(X,Y) = \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} P(x_i, y_j) \log\left(\frac{p(X_i, y_j)}{p(X_i,)p(y_j)}\right) \quad (11)$$

In Eq. (10), $p(x_i)$ and $p(y_i)$ are the probability distributions of $X$ (input variable) and $Y$ (output variable) respectively, and $p(x_i, y_i)$ is the joint probability distribution of $X$ and $Y$. Since the underlying relationship between variables is described by these distributions, their precise determination is necessary for MI evaluation. However the exact distribution of variables is not clear in pragmatic matters. Therefore, estimation is required for computing entropy H. Add to that, MI, as a supervised method, requires knowledge of the desired output values (which is also referred to the supervisory signal) and the goal of supervised method is providing best approximation of the relationship between inputs and output data. Readers for mastering in the entropy and MI knowledge and excessive details are referred to Cover and Tomas [47].

The minimum redundancy and maximum relevance (MRMR) as an input selection algorithm is widely utilized to evaluate the MI and conclude whether the given candidate parameter should be included in input variables. This algorithm chooses a subset of variables (S) which has minimum redundancy and maximum relevance with output layer [48]. This criterion for system X and Y is described as [49]:

$$J_{MRMR}(X_M) = I\ (X_n;\ Y) - \frac{1}{|S|} \sum_{j \in S} I(X_n \mid X_j) \quad (12)$$

### 2.5. Self-organizing map (SOM)

In the context of the unsupervised neural network, SOM is capable of classifying pattern of data without any prior knowledge about the process in which the data is generated. Therefore, by maintaining the topology structure of the data, SOM converts intricate nonlinear relationships among "high-dimensional" data into a comprehensible geometric relationship on a "low dimensional display" [50]. SOM can group the items with common similarities and plot the similarities on 1 or 2-dimensional plots [22]. Therefore, SOMs carry out dimension reduction and similarity demonstration. The SOM contains an input layer and an output layer (which is known as Kohonen layer) (Fig. 3). The objective is to find a suitable set of prototype vectors for each unit so that the network models the distribution of the input data in the output space [51].

At each irritation, the initial weight is appointed randomly. The distance among inputs and weight neurons calculated by putting input vector ($x$) through the network. The most common criterion to compute the distance is *Euclidean distance* as [50]:

$$\|x - w\| = \sqrt{\sum_{i=1}^{n} (x_i - w_i)^2} \quad (13)$$

The neuron which possesses the most similar weight to the input is called the best matching unit (BMU). Among training processes, the winner neuron or BMU and its neighboring neurons are given time in order to learn by changing the weights gradually, to further reduce the distance between the weights and the input vector [50]:

$$W(t+1) = w(t) + \alpha\,(t)\,h_l n\,(x - w(t)) \quad (14)$$

where $t$ is time and $\alpha$ is the learning rate which is ranging in [0 1], $h_{ln}$ is Gaussian function that is widely used as the neighborhood function and is centered in winner neuron whose position is indicated by $l$ and $n$.

$$h_{\ln} = \exp\left(-\frac{\left\| l - n^2 \right\|}{2\sigma(t)^2}\right) \quad (15)$$

The distance between $l$ and $n$ on the SOM network is indicated by $l$–$n$; and σ is the neighborhood radius. The training process is repeated until convergence. After the development of the SOM network, the relevant information is shown in the homogeneous groups on the map unit.

## 3. Results and discussion

In this study, three ANN-based models were developed in order to evaluate the process of wastewater treatment. In the following, obtained results by each of the models are presented and discussed.

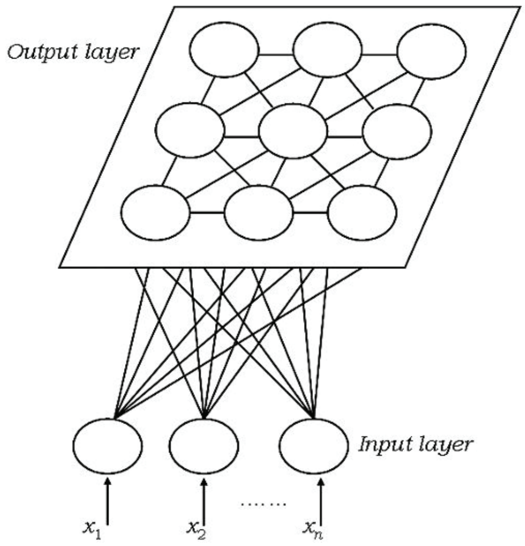



Fig. 3. SOM clustering scheme.

In model A, in order to apply PCA, it is necessary to evaluate the KMO test which measures sampling adequacy for each variable. The KMO value was calculated using Eq. (7), and the value of 0.73 indicated that the data is quite suited for PCA analysis. After standardization of the input variables for PCA application, variance-covariance symmetrical matrix R was formed from order 13 (equivalent to the number of potential input variables). After solving Eq. (6), 13 eigen values and for every eigen value 13 eigen vectors were obtained and by using them, 13 PCs were formed from input variables. The characteristics of the computed PCs are presented in Table 3 and the variance percentage of each component is tabulated in Table 4.

As shown in Table 4, over 90% of the variation within the data cloud be explained by the former two PCs, so the importance of the thirteen variables for the former two $PC_s$ was based on the following order according to Table 3:

$PC_1$: 11 > 7 > 10 > 4 > 3 > 5 > 9 > 13 > 12 > 6 > 2 > 8 > 1

$PC_2$: 7 > 10 > 4 > 5 > 3 > 9 > 12 > 13 > 6 > 2 > 8 > 1 > 11

The projections of the process vectors into the space of the first two loading vectors are shown in Fig. 4 which reflects the relative importance of parameters based on their vector value. It could be hypothesized that parameters 10(TDS), 11(EC), 7(MLSS), 4(COD) are important variables. For investigating the exactness of this hypothesis, the t-test, a statistical hypothesis test in which the test statistic follows a Student's t-distribution was applied. In another word, the t-test method was adopted in order to determine whether the hypothesized variables' mean is statistically in meaningful distance from all variables' mean or not, and how large is the difference between the mean of selected parameters and the mean of whole variables in PCA1 and PCA2. Since calculated t-value (2.37) is less than the t- table value at an alpha level of 0.05 (3.11), it could be concluded that the difference is very subtle and null hypothesis (the sample mean equals to the proposed population mean) is acceptable. Hence, it could be acclaimed, with 95% confidence,

Table 3
The composition of the principal components for model A

| Number of Parameter | Parameter | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 | PC11 | PC12 | PC13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $Q_{in}$ | –6.0E-16 | –1/24E-15 | –6E-16 | 2E-15 | –2E-15 | 2E-14 | 1E-14 | –7E-15 | –5E-16 | 4E-15 | 1E-14 | –2E-13 | –1 |
| 2 | $PH_{in}$ | 0.001 | 0.002 | 0.000 | –0.002 | 0.003 | –0.004 | 0.001 | –0.009 | 0.008 | –0.001 | –0.003 | –1.000 | 2E-13 |
| 3 | $BOD_{in}$ | 0.067 | 0.075 | –0.017 | –0.320 | 0.321 | 0.076 | –0.865 | –0.038 | –0.169 | 0.007 | 0.006 | 0.000 | –9E-15 |
| 4 | $COD_{in}$ | 0.110 | 0.126 | –0.032 | –0.519 | 0.553 | 0.442 | 0.440 | 0.000 | 0.077 | –0.008 | –0.023 | 0.003 | 8E-15 |
| 5 | $TSS_{in}$ | 0.065 | 0.078 | –0.020 | –0.321 | 0.259 | –0.885 | 0.151 | 0.104 | –0.034 | 0.020 | 0.029 | 0.004 | –1E-14 |
| 6 | $T_{in}$ | 0.002 | 0.005 | 0.004 | –0.008 | 0.010 | –0.071 | 0.109 | –0.930 | –0.342 | –0.006 | 0.013 | 0.006 | 5E-15 |
| 7 | $MLSS_{in}$ | 0.457 | 0.778 | 0.384 | 0.172 | –0.089 | 0.006 | –0.001 | 0.005 | –0.003 | 0.000 | 0.001 | 0.001 | –1E-15 |
| 8 | $SS_{in}$ | 0 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 4E-13 |
| 9 | $VSS_{in}$ | 0.0467 | 0.057 | –0.029 | –0.692 | –0.716 | 0.050 | 0.004 | –0.004 | 0.000 | –0.001 | –0.001 | 0.000 | 5E-16 |
| 10 | $TDS_{in}$ | 0.3523 | 0.205 | –0.903 | 0.129 | –0.049 | 0.002 | –0.002 | –0.004 | 0.000 | –0.001 | 0.000 | 0.001 | 2E-16 |
| 11 | $EC_{in}$ | 0.8019 | –0.567 | 0.186 | 0.011 | –0.010 | –0.001 | 0.002 | 0.000 | –0.001 | 0.000 | 0.000 | 0.000 | 1E-16 |
| 12 | $BOD_{t-1}$ | 0.0048 | 0.006 | 0.001 | –0.018 | 0.015 | –0.068 | –0.085 | –0.173 | 0.463 | –0.822 | –0.259 | 0.008 | –1E-14 |
| 13 | $BOD_{t-2}$ | 0.0048 | 0.006 | 0.001 | –0.015 | 0.012 | –0.059 | –0.091 | –0.217 | 0.543 | 0.551 | –0.585 | 0.008 | –1E-14 |

Table 4
Percentage variance of the 13 principal components of model B

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 | PC11 | PC12 | PC13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variance explained percent | 68.78 | 22.15 | 6.45 | 1.59 | 0.86 | 0.07 | 0.03 | 0.0018 | 0.0013 | 0.0001 | 0.0001 | 0.0 | 0.0 |
| Variance cumulative percent (%) | 68.78 | 90.93 | 97.38 | 98.97 | 99.83 | 99.90 | 99.93 | 99.97 | 99.99 | 99.99 | 99.99 | 100 | 100 |

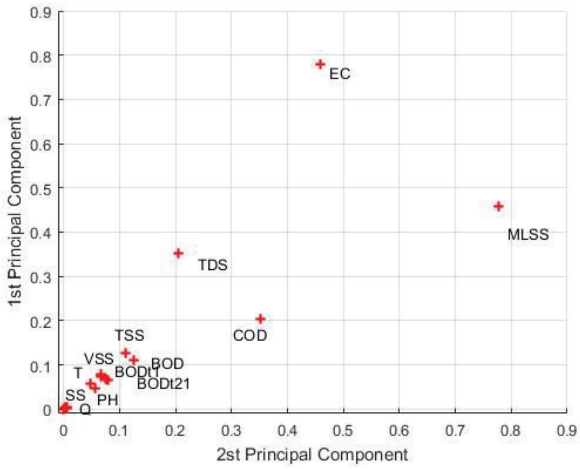



Fig. 4. The characteristics of PC1 and PC2.

Table 5
MI and scores

| Output parameter | Input variable | MI | MRMR |
|---|---|---|---|
| $BOD_{eff}$ | $BOD_{in}$ | 0.25 | 0.063 |
| | $VSS_{in}$ | 0.24 | 0.042 |
| | TSS | 0.24 | 0.012 |
| | MLSS | 0.22 | 0.036 |
| | $COD_{in}$ | 0.17 | 0.004 |
| | TDS | 0.17 | 0.004 |
| | EC | 0.12 | 0.0015 |
| | $BOD_{t-2}$ | 0.10 | 0 |
| | $PH_{in}$ | 0.9 | 0 |
| | T | 0.09 | 0.0001 |
| | $BOD_{t-1}$ | 0.045 | 0.0004 |
| | SS | 0 | 0 |
| | Q | 0 | 0 |

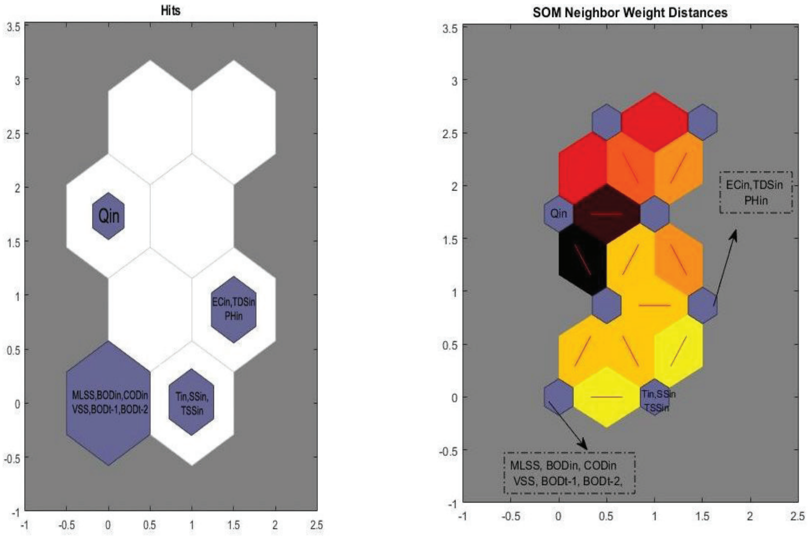
that the selected variables were representative of each PCA. As a result, MLSS, EC, TDS, COD were concluded as the most important parameters so that the input variables reduced from thirteen to four (MLSS, EC, TDS, COD) variables that should be introduced to the ANN model for predicting effluent BOD.

In model B, the MI measure was applied to determine the most pertinent inputs among 13 parameters ($Q_{in}$, $SS_{in}$, $PH_{in}$, $T_{in}$, $TSS_{in}$, $TDS_{in}$, $VSS_{in}$, $EC_{in}$, $BOD_{in}$, MLSS, $BOD_{t-1}$, $BOD_{t-2}$, $COD_{in}$). MI score between each of these 13 parameters and model output (BOD) was calculated using Eq. (11). Similarly, MRMR score was calculated using Eq. (12) in order to classify parameters based on the minimum redundancy and maximum relevance with the output. The results are presented in Table 5 which indicate that the $BOD_{in}$ has the highest MI and MRMR with $BOD_{eff}$ (output parameter).The t-test, was used in this model same as the model A to examine whether the hypothesized variables' mean is statistically in meaningful distance from all variables' mean. The calculated t-value (2.28) is less than the t- table value at an alpha level of 0.05 (3.11), it could be concluded that the difference is ignorable. Thus, it could be concluded, with 95% confidence, the selected four variables (MLSS, $BOD_{in}$, TSS and VSS) which possess the higher magnitude of MI and MRMR score could be the representative of all other parameters.

In model C, the SOM clustering approach was employed for recognizing homogenous variables of the wastewater treatment system. Euclidean distance criterion was applied in order to select the best representative of each cluster and centroid variables. The size of the Kohonen layer must be adequate enough to provide the appropriate number of clusters that could cover all information represented by the data set (52). Due to the intuitive capability, the Gaussian function in combination with Euclidian distance can provide a helpful result so that the Gaussian kernel was used to train the topology of the SOM that was orientated on a hexagonal network. Hits map of SOM indicates the number of variables and their positions (Fig. 5a).

Four data classes were obtained using the SOM technique (Fig. 5 and Table 6). Neighbor weight distances are shown in Fig. 5a. Neurons distance are designated with the colors in every region where the darker color indicates the larger distance.

According to the results, the first cluster contains VSS, $BOD_{t-2}$, $BOD_{t-1}$, MLSS and $COD_{in}$, $BOD_{in}$. It should be noted that the biological treatment system of the Tabriz WWTP is the activated sludge process, therefore, it is essential to maintain the yield coefficient in a specific rate since it links the rate of BOD, as the substrate utilization, to the rate of MLSS and VSS as the amount of microbial community. In another words, the amount of biomass produced during cell synthesis are highly related to the amount of substrate degraded. Thus, it can be concluded that this cluster represents the biological component of the treatment process. Also, industrial wastewater in Tabriz city is obligated to experience a pretreatment process before discharging to the municipal sewer system, so that the COD and BOD values are so close and denoting them in one group seems logical. Clustering $BOD_{t-2}$, $BOD_{t-1}$ parameters with other biological components is also an indicator of existed connection between these parameters with other quality characteristics and the exactness of time series hypothesis. The second cluster comprises T, SS, and TSS which can be determined by physical senses such as touch and sight. Therefore this cluster concludes physical condition. The3th




(a) (b)

Fig. 5. The 2-dimensional SOM clustering of wastewater treatments (a) SOM hits showing the number of cells (b) SOM neighbor weight distances plan.

Table 6
Clusters and members' distribution

| Cluster number | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Cluster members | MLSS, $BOD_{in,}$ $VSS_{in}$, $COD_{in}$, $BOD_{t-2}$, $BOD_{t-1}$, | $Ti_n$ ,$SS_{in}$, $TSS_{in}$ | $EC_{in}$, $TDS_{in}$, $PH_{in}$ | $Q_{in}$ |

Table 7
Performance of ANN for three models

| Model | ANN structure | Iteration number (epoch) | Determination coefficient (DC) | | | Root mean square error RMSE (normalized) | | |
|---|---|---|---|---|---|---|---|---|
| | | | Training | Validation | Test | Training | Validation | Test |
| A(PCA) | 4-5-1 | 200 | 0.64 | 0.62 | 0.54 | 0.083 | 0.071 | 0.095 |
| B(MI) | 4-10-1 | 500 | 0.48 | 0.47 | 0.48 | 0.099 | 0.097 | 0.099 |
| C(SOM) | 3-5-1 | 200 | 0.74 | 0.70 | 0.67 | 0.046 | 0.050 | 0.054 |

cluster includes EC, TDS, and pH, these three parameters are indicators of chemical component and salinity level. The EC and TDS member are highly correlated and usually expressed by a simple equation: TDS = k EC (in 25°C). EC is closely related to total dissolved solids in which suspended undissolved solids are not included. Also, it is clear from Fig. 5b that SOM was capable of discerning Q as outlying data and insulating it from other clusters which show the ineffectiveness of this parameter. Election of just one member from each cluster, while all cluster members follow the same pattern and play the same role in the ANN-based prediction model, might be effective from variable and noise diminishing perspective. Hence, the clustering result of the SOM model, which is based on the unsupervised and competitive learning algorithm, leads to variables reduction.

Accordingly, based on the Euclidean distance measure the pertinent member from each cluster extracted and introduced to ANN for predicting effluent BOD. T as representative of physical characteristics of influent and PH as representative of chemical characteristics of influent and $BOD_{in}$ as representative of biological characteristics were determined as the representatives of clusters. Therefore, by employing the SOM and identifying dominant data, sufficient data are selected from the data set and insignificant data eliminated. It can be concluded that, due to covering all physical, chemical and biological characteristic, results seem to be the best representative of inputs which fit the prediction purpose.

There are two essential parts that should be involved in model testing: accuracy and generalization performance. Testing the model's potential in predicting the target for the given data set that was used for model training is accuracy performance, and testing the model's potential in predicting the target for the data set that wasn't used in training stage is generalization performance [10]. These ANN modeling performance results are presented in Table 7 and denote the

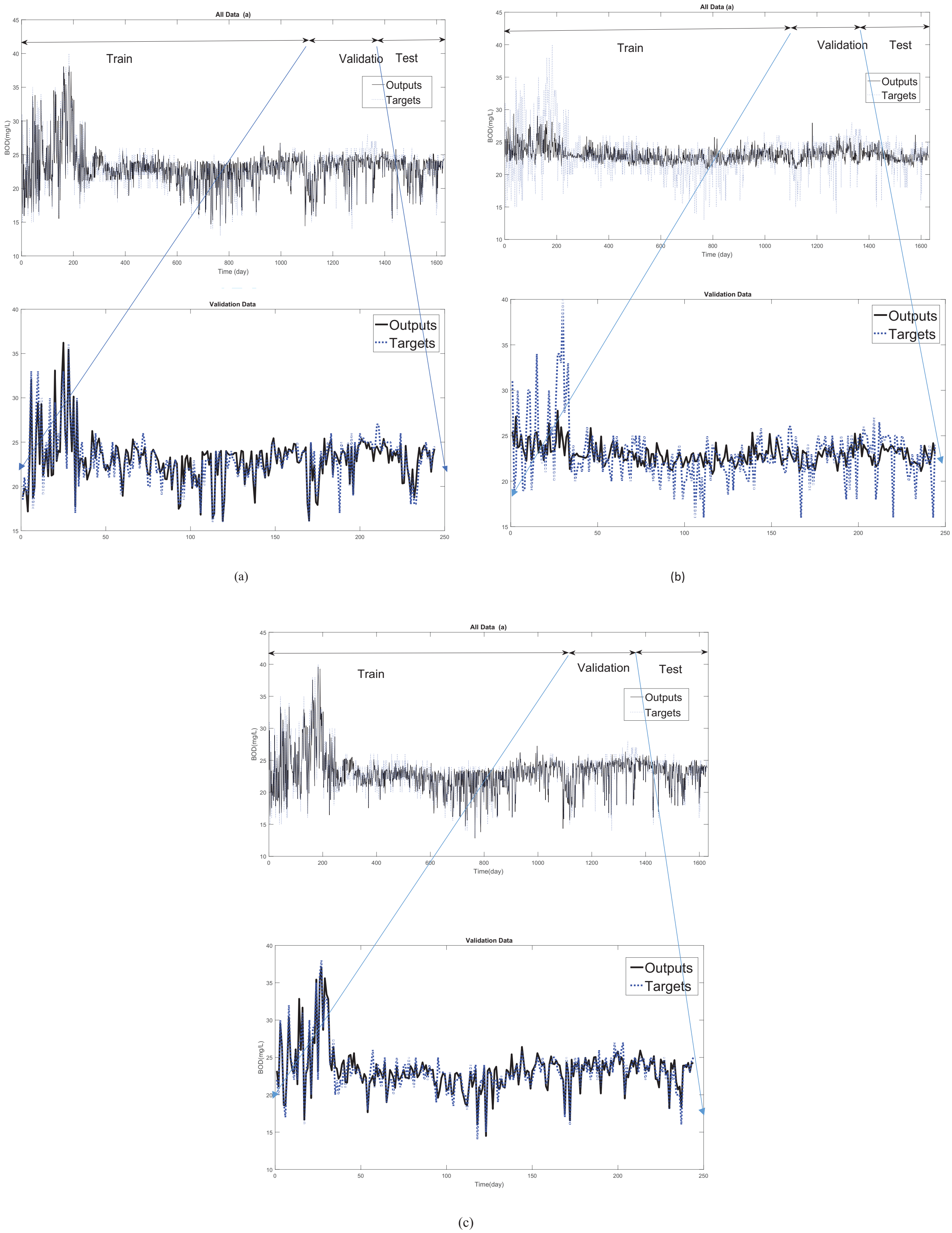



Fig. 6. Observed BOD vs. computed BOD in, a) Model A, b) Model B, c) Model C.

reliability of model C in which the input parameters have been clustered by SOM, since the DC values for verification steps increased to 8% and 23%, compared to the models A and B, respectively. It should be noted that none of PCA and MI methods are in the categories of the AI, however, SOM method is an AI-based clustering approach that in combination with AI-based modeling method (ANN) led to more precise effluent quality prediction.

Model C shows close values to the results of the RBF and MLP models that were developed recently by Hamada et al. [53], however, model C, owing to the SOM algorithm, has a simpler structure since it has used only three critical parameters, while in the mentioned studies five parameters have been utilized for predicting BOD. From the modeling perspective, clustering the data by SOM before developing the model is a two-phase process and in comparison with simple models may suffer from computational sufficiency. However, it should be mentioned that engineers who control and monitor the effluent quality of WWTP, faces a huge amount of operating variables. The developed model in current work would offer them an opportunity to minimize these parameters while preserving the crucial relationships of the data on the two-dimensional plane [54], hence, the data from various sources can be used efficiently. Furthermore, due to the capability of SOM in handling with noisy data set and unsupervised learning, there wouldn't be specific demand of processing knowledge. Therefore, the result confirmed that the proposed model would be sufficient and helpful for practical purposes.

Figs. 6a–c which depict observed versus predicted effluent BOD values for all three models indicate that models A and B underestimate the extremum values while almost all of the variations of the BOD are predicted by the model C. This model is capable of catching the peak values of effluent BOD, which emphasizes the superiority of SOM-ANN model and the best agreement between measured BOD and the model outputs.

## 4. Conclusions

This article investigated dominant input determination methods for ANN modeling of the WWTP. In this way, the FFNN was used to derive a convenient wastewater treatment model for predicting effluent BOD of WWTP. Model A utilized the PCA multivariate statistical method, in model B, non-linear MI measure was used for input selection. The comparison of models revealed that the model C which benefits SOM clustering method, outperformed others based on the performance evaluation measures, DC and RMSE. This model with three input variables (PH, T, $BOD_{in}$) led to better performance than the first model with, MLSS, EC, TDS, COD, inputs due to distinguishing parameters with maximum similarity and reducing them as well as covering all physical, chemical, biochemical characteristics of the wastewater. Furthermore, the obtained results show the merit of SOM in detecting the ineffective parameter and separating it from other clusters. Model B with four input variables (MLSS, $BOD_{in}$, VSS, and TSS) lacks precision because of neglecting physical parameters, which causes a negative effect on ANN's generalization and leads to unreliable results. Therefore, the multiple factors such as biochemical mechanism, the number of inputs, accuracy, etc. need to be considered for the choice of input variables.

Lack of available data and a wide range of involved parameters were the major difficulties and restriction of the current study. While, considering broad diversity of parameters (i.e., $NO_2^-$, $NO_3^-$, mixed liquor volatile suspended solids (MLVSS), total phosphate (TP), grease, dissolved oxygen (DO), etc.) can offer more reliable and comprehensive result of clustering WWTP parameters. For further studies, other clustering methods such as K-Means method which is a method of vector quantization could be applied in order to verify the SOM method. Add to that, application of an adaptive neuro-fuzzy inference system or adaptive network-(ANFIS), a kind of artificial neural network that is based on a fuzzy concept, could be compared to FFNN performance. More specifically, other parameters of WWTP may be models via the proposed method.

## References


[1] M.C. Jami, I.A. Husain, N.A. Abdullah, Multiple inputs artificial neural network model for the prediction of wastewater treatment plant performance, J. AJBAS., 6(1) (2012) 62–69.

[2] D. Hanby, I. Tukoglu, Y. Demir, Prediction of wastewater treatment plant performance based on wavelet packet decomposition and neural networks, Expert Syst. Appl., 34(2) (2008) 1038–1043.

[3] C.H. Wen, C.A. Vassiliadis, Applying hybrid artificial intelligence techniques in wastewater treatment, Eng. Appl. ArtifIntell., 11(6) (1998) 685–705.

[4] I. Zaheer, C.G. Bai, Application of artificial neural network for water quality management, LTI., 5(2) (2003) 10–15.

[5] A. Nabavi-Pelesaraei, R. Bayat, H. Hosseinzadeh-Bandbafha, H. Afrasyabi, K.W. Chau, Modeling of energy consumption and environmental life cycle assessment for incineration and landfill systems of municipal solid waste management-A case study in Tehran Metropolis of Iran, J. Clean Prod., 148 (2107) 427–440.

[6] Y. Seo, S. Kim, V.P. Singh, Comparison of different heuristic and decomposition techniques for river stage modeling, Environ. Monit. Assess., 190(7) (2018) 392.

[7] E. Olyaie, H. Banejad, K.W. Chau, A.M. Melesse, A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: A case study in United States, Environ. Monit. Assess., 187(4) (2015) 189.

[8] S. Shamshirband, E. Jafari Nodoushan, J.E. Adolf, A. Abdul Manaf, A. Mosavi, K.W. Chau, Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters, Eng. Appl. Comp. Fluid., 13(1) (2019) 91–101.

[9] M. Djeddou, B. Achour, The use of a neural network technique for the prediction of sludge volume index in municipal wastewater treatment plant, LARHYSS J., 24 (2015) 351–370.

[10] I. Lou, Y. Zhao, Sludge bulking prediction using principle component regression and artificial neural network, Math. Probl. Eng., (2012) Article ID 237693.

[11] E. Dogan, A. Ates, E.C. Yilmaz, B. Eren, Application of artificial neural networks to estimate wastewater treatment plant inlet biochemical oxygen demand, Environ. Prog. Sustain., 27(4) (2008) 439–446.

[12] A.E. Tumer, S. Edebali, An artificial neural network model for wastewater treatment plant of Konya, IJISAE., 3(4) (2015) 131–13.

[13] M. Vyas, B. Modhera, V. Vyas, A.k. Sharma, Performance forecasting of common effluent treatment plant parameters by artificial neural network, JEAS., 6(1) (2011) 38–42.

[14] D.J. Choi, H. Park, A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process, Water Res., 35(16) (2001) 3959–3967.
[15] M.M. Hamed, M.G. Khalafallah, E.A. Hassanien, Prediction of wastewater treatment plant performance using artificial neural networks, Environ. Model Softw., 19(10) (2004) 919–928.
[16] S.I. Abba, G. Elkiran, Effluent prediction of chemical oxygen demand from the WWTP using artificial neural network application, Procedia. Comput. Sci., 120 (2017) 156–163.
[17] H.G. Han, J.F. Qiao, Q.L. Chen, Model predictive control of dissolved oxygen concentration based on a self-organizing RBF neural network, Control. Eng. Pract., 20(4) (2012) 465–476.
[18] Y.S.T. Hong, M.R. Rosen, R. Bhamidimarri, Analysis of a municipal wastewater treatment plant using a neural network-based pattern analysis, Water Res., 37(7) (2003) 1608–1618.
[19] G.J. Bowden, G.C. Dandy, H.R. Maier, Input determination for neural network models in water resources applications. Part 1—background and methodology, J. Hydrol., 301(4) (2005) 75–92.
[20] S.E. Kim, I.S. Won, Artificial Neural Network ensemble modeling with conjunctive data clustering for water quality prediction in rivers, J. Hydro-Environ. Res., 9(3) (2015) 325–339.
[21] J. Wan, Prediction of effluent quality of a paper mill wastewater treatment using an adaptive network-based fuzzy inference system, Appl. Soft. Comput., 11(3) (2011) 3238–3246.
[22] V. Nourani, M. Mehrvand, A.H. Baghanam, Implication of SOM-ANN based clustering for multistation rainfall-runoff modeling, JUEE., 8(2) (2014) 198–210.
[23] T.Y. Pai, P.Y. Yang, S.C. Wang, M.H. Lo, C.F. Chiang, J.L. Kuo, Y.H. Chang, Predicting effluent from the wastewater treatment plant of industrial park based on fuzzy network and influent quality, Appl. Math. Model., 35(8) (2011) 3674–3684.
[24] Ö. Çinar, New tool for evaluation of performance of wastewater treatment plant: artificial neural network, Process Biochem., 40(9) (2005) 2980–2984.
[25] T. Kohonen, The self-organizing map, Neurocomputing., 21(1–3) (1998) 1–6.
[26] K.K. Jassar, K.S. Dhindsa, Comparative study and performance analysis of clustering algorithms, IJCA, 975 (2015) 8887.
[27] D. Aguado, T. Montoya, L. Borras, A. Seco, J. Ferrer, Using SOM and PCA for analyzing and interpreting data from a P-removal SBR, Eng. Appl. ArtifIntell., 21(6) (2008) 919–930.
[28] R. Sathya, A. Abraham, Comparison of supervised and unsupervised learning algorithms for pattern classification, IJARAI., 2(2) (2013) 34–38.
[29] V. Nourani, G. Andalib, D. Dąbrowska, Conjunction of wavelet transform and SOM-mutual information data pre-processing approach for AI-based Multi-Station nitrate modeling of watersheds, J. Hydrol., 548 (2017) 170–183.
[30] J. Qiao, H. Zhiqiang, L. Wenjing, Soft measurement modeling based on chaos theory for biochemical oxygen demand (BOD), Water., 8(12) (2016) 581.
[31] D.A. Cancilla, X. Fang, Evaluation and quality control of environmental analytical data from the Niagara River using multiple chemometric methods, Great Lakes Res., 22(2) (1996) 241–253.
[32] T.R. Holcomb, M. Morari, PLS/neural networks, Comput. Chem. Eng., 16(4) (1992) 393–411.
[33] A. Singh, M. Imtiyaz, R.K. Isaac, D.M. Denis, Comparison of soil and water assessment tool (SWAT) and multilayer perceptron (MLP) artificial neural network for predicting sediment yield in the Nagwa agricultural watershed in Jharkhand, India, Agr. Water Manage., 104 (2012) 113–120.
[34] K. Hornik, S. Maxwell, W. Halbert, Multilayer feed forward networks are universal approximators, Neural Netw., 2(5) (1989) 359–366.
[35] T.W. Kim, J.B. Valdés, Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks, J. Hydrol. Eng., 8(6) (2003) 319–328.
[36] A.R. Pendashteh, A. Fakhru'l-Razi, N. Chaibakhsh, L.C. Abdullah, S.S. Madaeni, Z.Z. Abidin, Modeling of membrane bioreactor treating hypersaline oily wastewater by artificial neural network, J. Hazard. Mater., 192(2) (2011) 568–575.
[37] D.R. Legates, J.G. McCabe, Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35(1) (1999) 233–241.
[38] G.B. Sahoo, C. Ray, Predicting flux decline in crossflow membranes using artificial neural networks and genetic algorithms, J Membr. Sci., 283 (2006) 147–157.
[39] C. Rosén, J. Röttorp, U. Jeppsson, Multivariate on-line monitoring: challenges and solutions for modern wastewater treatment operation, Water Sci. Technol., 47(2) (2003) 171–179.
[40] P. Teppola, S.P. Mujunen, P. Minkkinen, A combined approach of partial least squares and fuzzy c-means clustering for the monitoring of an activated- sludge waste-water treatment plant, ChemometrIntell. Lab. Syst., 41(1) (1998) 95–103.
[41] H. Çamdevýren, N. Demýr, A. Kanik, S. Keskýn, Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs, Ecol. Modell., 181(4) (2005) 581–589.
[42] J.C. Davis, Statistical and Data Analysis in Geology, 2nd ed, John Wiley and Sons. New York, 1986.
[43] H. Wackernagel, Multivariate Geostatistics: An Introduction With Applications, New York and London, 1995.
[44] R. Steuer, J. Kurths, C.O. Daub, J. Weise, J. Selbig, The mutual information detecting and evaluating dependencies between variables, Bioinformatics, 18(2) (2002) 231–240.
[45] C.E. Shannon, A note on the concept of entropy, Bell System Tech J., 27(3) (1948) 379–423.
[46] M.S. Babel, G.B. Badgujar, V.R. Shinde, Using the mutual information technique to select explanatory variables in artificial neural networks for rainfall forecasting, Meteorol. Appl., 22(3) (2015) 610–616.
[47] T.M. Cover, J.A. Thomas, Elements of information theory, John Wiley & Sons, 2012.
[48] R.K. Parviz, N. Mozayani, M.J. Motlagh, Mutual information-based input variable selection algorithm and wavelet neural network for time series prediction, International Conference on Artificial Neural Networks. Springer, Berlin, Heidelberg, 2008.
[49] G. Brown, A. Pocock, M.J. Zhao, M. Luján, Conditional likelihood maximisation: a unifying framework for information theoretic feature selection, J. Mach. Learn. Res., (2012) 27–66.
[50] T. Kohonen, Self-Organizing Maps, Springer, 1997.
[51] H.L. Garcia, I.M. González, Self-organizing map and clustering for wastewater treatment monitoring, Eng. Appl. ArtifIntell., 17(3) (2004) 215–225.
[52] T. Kohonen, S. Kaski, H. Lappalainen, Self-organized formation of various invariant-feature filters in the adaptive-subspace SOM, Neural. Comput., 9(6) (1997) 1321–1344.
[53] M. Hamada, A.Z. Hossam, A.J. Ahmed, Application of artificial neural networks for the prediction of Gaza wastewater treatment plant performance-Gaza strip, JARWW., 5(1) (2018) 399–406.
[54] A. Sharma, C.W. Omlin, Performance comparison of particle swarm optimization with traditional clustering algorithms used in self organizing map, Int. J. Comput. Int. Sys., 5(1) (2019) 1–12.