Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2020.26160

203 (2020) 80–90
November

Modeling of Bunus regional sewage treatment plant using machine learning approaches

Quoc Bao Pham[a,b], M.S. Gaya[c], S.I. Abba[d], R.A. Abdulkadir[c], Parvaneh Esmaili[e], Nguyen Thi Thuy Linh[f], Chetan Sharma[g], Anurag Malik[h], Dao Nguyen Khoi[i,*], Tran Duc Dung[j], Do Quang Linh[k]

[a]*Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam, Tel. +886978392744; email: phambaoquoc@tdtu.edu.vn (Q.B. Pham)*
[b]*Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam*
[c]*Department of Electrical Engineering, Kano University of Science and Technology Wudil, Kano-Nigeria, Tel. +2348023063442; email: muhdgayasani@gmail.com (M.S. Gaya), Tel. +2347038341284; email: rabiukk@gmail.com (R.A. Abdulkadir)*
[d]*Department of Physical Planning Development and Maintenance, Yusuf Maitama Sule University, Kano-Nigeria, Tel. +2348031334033; email: saniisaabba86@gmail.com (S.I. Abba)*
[e]*Department of Electrical and Electronic Engineering, Near East University, 99138 Nicosia-Northern Cyprus, Tel. +9054288225801; email: parvaneh.esmaili@neu.edu.tr (P. Esmaili)*
[f]*Thuyloi University, 175 Tay Son, Dong Da, Hanoi, Vietnam, Tel. +886978050973; email: linhntt@tlu.edu.vn (N.T. Thuy Linh)*
[g]*Civil Engineering Department, MITS Gwalior, M.P. 474006, India, Tel. +91-8755448704; email: chetan.cvl@gmail.com (C. Sharma)*
[h]*Department of Soil and Water Conservation Engineering, College of Technology, G.B. Pant University of Agriculture and Technology, Pantnagar-263145, Uttarakhand, India, Tel. +91-9548868511; email: anuragmalik_swce2014@rediffmail.com (A. Malik)*
[i]*Institute of Research and Development, Duy Tan University, Da Nang, Vietnam, Tel. +84 236 3 827 111; email: daonguyenkhoi@duytan.edu.vn (D.N. Khoi)*
[j]*Centre of Water Management and Climate Change, Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Vietnam, Tel. +84 902007 905; email: dungtranducvn@yahoo.com (T.D. Dung)*
[k]*NTT Institute of Hi-Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam, Tel. + 84 28 29 404 759; email: doqlinh@gmail.com (D.Q. Linh)*

Received 22 October 2019; Accepted 22 May 2020

ABSTRACT

Certain aspects of the dynamics of wastewater treatment plants appear to be chaotic, which makes modeling of the process of wastewater treatment plants extremely difficult. An appropriate model is key for the optimal operation of the plant. Conventional prediction techniques are not good enough to produce the desired results and determination of the suitable structure of using either fuzzy, artificial neural network or adaptive neuro-fuzzy interface system becomes cumbersome. This article proposed the application of advanced machine learning methodologies, for example, extreme learning machine (ELM), support vector machine (SVM) for modeling the Bunus regional sewage treatment plant. These advanced machine learning methods were also compared with conventional autoregressive integrated moving average (ARIMA). Observed data from the Bunus regional wastewater treatment plant was used for the modeling. The simulation results indicated that the ELM model performed better than the SVM and ARIMA models with a decrease in mean absolute percentage error by 19% and 29% than SVM and ARIMA models respectively. As the choice of input parameters often affects the modeling performance different combinations of input

* Corresponding author.

variables were selected. It was observed that influent biological oxygen demand, chemical oxygen demand, suspended solids, ammonium iron (NH_4) were able to model the process better than other input parameter combinations.

Keywords: Black-box models; Biochemical oxygen demand; Extreme learning machine; Wastewater treatment plant

1. Introduction

Biological oxygen demand (BOD) is one of the key parameters for determining the size and efficiency of wastewater treatment plants (WWTP) [1]. BOD is the amount of oxygen needed for the degradation of organic matter present in wastewater. BOD serves as an indicator of the clarification of water required before it is disposed into the receiving body [2]. The process of BOD oxidation is relatively slow and time-consuming over the production and analysis. The standard period of BOD is 5 d, in which about 60%–70% oxidation happens, while within ~95%–99% oxidation is attained in 20 d [3].

Demand for qualitative public health and environmental concerns have motivated research communities and practitioners to focus their attention on WWTP. Effective treatment of wastewater facilitates effective environmental management/protection process and eliminates major threats to public health [4]. BOD simulation is quite crucial in order to determine the approximate quantity of oxygen needed for the biological stabilization of organic matter present in the wastewater. Similarly, BOD serves as the indicator for evaluating the size and efficiency of WWTP, hence its prediction become indispensable. This can be achieved with the application of an appropriate forecasting tool. A good prediction tool/method leads to higher efficiency, which in turn reduces energy consumption and cuts operational/maintenance cost significantly, which improves the reliability and performance of the plant. Despite modeling of other wastewater quality indicators are paramount but several studies indicated that BOD and chemical oxygen demand (COD) is one of the most important wastewater quality indicators which determine the performance of the plant [5–12].

However, developing an accurate model for wastewater treatment plants can be extremely difficult due to the chaotic, dynamic nature of the wastewater treatment process. Different new forecasting strategies such as an artificial neural network (ANN), genetic algorithm, and fuzzy logic have been used for modeling the industrial WWTPs in Taiwan [13]. These advanced methods have helped to plan the control strategy in the successful management of the WWTP. Maleki et al. [14] predicted the influent parameters in WWTP using an autoregressive integrated moving average (ARIMA) and neural network auto-regression (NNAR) models. Although the ARIMA model showed acceptable performance, the results demonstrated that NNAR exhibited better performance. Similarly, the application of ANN and multiple linear regression (MLR) for the prediction of COD in New Nicosia WWTP using different influent parameters have been studied in Abba and Elkiran [15]. The results indicated the superiority of ANN over MLR in both the training and testing phases. The performance of ANN in prediction is found better than MLR in different fields also Sharma et al. [16]. In Verma et al. [17] the ability of five different machine learning approaches, including multilayer perceptron (MLP), *K*-nearest neighbor, support vector machine (SVM), random forest, and multivariate adaptive regression splines is employed to estimate the total suspended solids (TSS) in a WWTP using different input parameters. The results showed better performance of MLP than other models. Also, in Hamed et al. [18] ANN model is developed to predict BOD and TSS values measured at different places within a treatment plant. The results indicated that ANN emerged as a reliable model for predicting the performance of the treatment plant. In Civelekoglu et al. [8] application of ANN and adaptive neuro-fuzzy interface system (ANFIS) models to simulate the COD in WWTP has been studied. The ANFIS model was found more reliable for the estimation of plant performance. In other studies, ANN [19,20], and ANFIS [5,21] were used to develop modeling WWTP; however, most of these strategies have some limitations such as the determination of optimum model structure and the requirement for design expertise.

A new and emerging black-box algorithm for single hidden layer feedforward networks (SLFNs) is the extreme learning machine (ELM) model, which was proposed in Huang et al. [22,23] to overcome the disadvantages of the traditional feed-forward backpropagation ANN (i.e., over-fitting, slow learning speed, and local minima). ELM has been widely adopted for the classification of municipal water samples [24], real-time BOD estimation [25], prediction of coagulant dosing [26]. In a recent study [27] ELM was found to be better than SVM to monitor water quality in the water treatment plants. Although few studies in the field of WWTP have made comparisons between ELM, SVM, and traditional ARIMA models, to the best of the authors' knowledge, no such study has been conducted at the Bunus regional sewage treatment plant (BRSTP). Therefore, the main aim of this study is to develop and compare the potential of a new data-driven approach, that is, ELM, and to compare it with the SVM and traditional ARIMA technique in predicting the effluent BOD_{eff} of BRSTP. The ELM is a new learning method so, to assess its performance it is compared with the widely adopted traditional ARIMA model and advanced SVM method. ARIMA and SVM represent two of the highly adopted learning methods in which ARIMA uses autoregressive properties for optimization of the model while SVM uses structural risk minimization (SRM) to find the global optima. Studies indicate that different machine learning methods are much better than ARIMA and the SVM method is found better than other machine learning methods in optimization and forecasting [28,29].

In hydro-environmental modeling demand for qualitative public health and environmental concerns have motivated research communities and practitioners to focus their attention on modeling wastewater treatment plants. Effective treatment of wastewater facilitates the effective environmental management/protection process and eliminates major threats to public health [4]. There are numerous challenges related to wastewater modeling that water resources engineers/researchers are facing. The prime purpose of the proposed models is to provide a consistent prediction using several models that may not be achievable due to the dynamic nature and non-stationarity of observed data. In addition, many existing systems are poorly documented, and modeling the system can provide a concise way to capture the process of the existing system. This information can then be used to facilitate maintaining the system or to assess the system with the goal of improving it by the decision-makers.

2. Methodology

2.1. Extreme learning machine

As an emerging black-box data-driven algorithm, the ELM comprises a SLFNs [22]. ELM was found to have several advantages over the traditional neural network [30]. For more detailed information on ELM, refer to [31,32].

The ELM structure is formed of a single hidden-layer feedforward neural network where the output weight matrix β is analytically characterized and the input weight matrix W is randomly selected. For a dataset with N arbitrary distinct samples (x_i, t_i) where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{in}]^T \in R^m$, the ELM structure is formulated as:

$$y_j = \sum_{i=1}^{\bar{N}} \beta_i g_i(x_i) = \sum_{i=1}^{\bar{N}} \beta_i g_i(w_i x_j + b_i), \quad j = 1, 2 \ldots n \tag{1}$$

where $\bar{N}$, $g_i(x_i)$, $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$, $\beta_{ii} = [\beta_{ii1}, \beta_{ii2}, \ldots, \beta_{iin}]^T$ and b_i are hidden nodes, activation function, the weight vector connecting the input nodes to the i^{th} hidden node, the weight vector connecting the output nodes with the i^{th} hidden node and the i^{th} hidden node threshold, respectively. The structure of the ELM network used in this study is presented in Fig. 1a.

2.2. Support vector machine

SVM [33] is another new artificial intelligence-based model comprised of both the classification and regression marching learning concepts. As a data-driven model, SVM has a promising ability to handle both classification and prediction problems. The SVM implements SRM which minimizes an upper bound to the generalization error instead of minimizing the training error. Based on this principle, the SVM achieves an optimum network structure. In addition, SVM is equivalent to solving a linear constrained quadratic programming problem so that the solution of the SVM is always unique and globally optimal [34]. The radial basis function kernel was applied in this study due to its robustness to simulate complex nonlinear functions [35]. In SVM, a linear regression was first fitted to the data and then the outputs go through a non-linear kernel to follow the non-linear pattern of the data (Fig. 1b). Given a set of training data $\{(x_i, d_i)\}_i^N$ (x_i is the input vector, d_i is the actual value and N is the total number of data patterns), the general SVM function is given as:

$$y = f(x) = w\phi(x_i) + b \tag{2}$$

where $\phi(x_i)$ indicates feature spaces, non-linearly mapped from input vector x [33].

2.3. Auto-regressive integrated moving average

The ARIMA model is one of the most widely used classical models for time series forecasting and provides a compatible approach to the problems of forecasting [36,37]. The pre-assumption of linear patterns is one of the disadvantages of the ARIMA time series. Ensuring the stationarity of the data is essential in any type of ARIMA modeling. The three main stages of the ARIMA model include parameter estimation, checking based on diagnostic processes, and model identification [38].

2.4. Model development and performance indicators

In model development, various steps are carried out, including data collection and pre-processing model design, model training, testing, and model execution. Prior to model development in any data-driven approach, the input selection is quite significant in different models (ELM, SVM, and ARIMA). For the purpose of this research, the data obtained from the BRSTP was randomly divided into 75% training and 25% testing. Although different methods are available for input selection, two methods were considered in this study (i) Pearson and Spearman correlation analysis to determine the strength and relations between inputs and outputs and (ii) auto-correlation function (ACF) and the partial autocorrelation function (PACF) (Fig. 2). More detail about ACF and PACF is provided in Appendix-A.

Subsequently, a set of three different models were derived on the basis of significant input variables. The normalization was carried out before the model training, as follows.

$$y = 0.5 + \left(0.5 \times \left(\frac{x - \bar{x}}{x_{\max} - x_{\min}}\right)\right) \tag{3}$$

Here, y, x, and $\bar{x}$ are normalized, measured, and mean of the measured data respectively. $x_{\max}$ and $x_{\min}$ represent the maximum and minimum value of the observed data.

Three different models were formulated based on chosen different sets of input parameters as given below:

$$\text{M1} = \text{BOD}_{\text{eff}} = f\left(\text{BOD}_{\text{inf}}, \text{COD}_{\text{inf}}\right) \tag{4}$$

$$\text{M2} = \text{BOD}_{\text{eff}} = f\left(\text{BOD}_{\text{inf}}, \text{COD}_{\text{inf}}\text{SS}_{\text{inf}}\right) \tag{5}$$

$$\text{M3} = \text{BOD}_{\text{eff}} = f\left(\text{BOD}_{\text{inf}}, \text{COD}_{\text{inf}}, \text{SS}_{\text{inf}}, \text{NH}_{4\text{inf}}\right) \tag{6}$$

where SS, subscripts 'eff' and 'inf' represents suspended solids, effluent and influent, respectively.

However, it is essential to briefly explain the other influents parameter for example; COD is one of the most

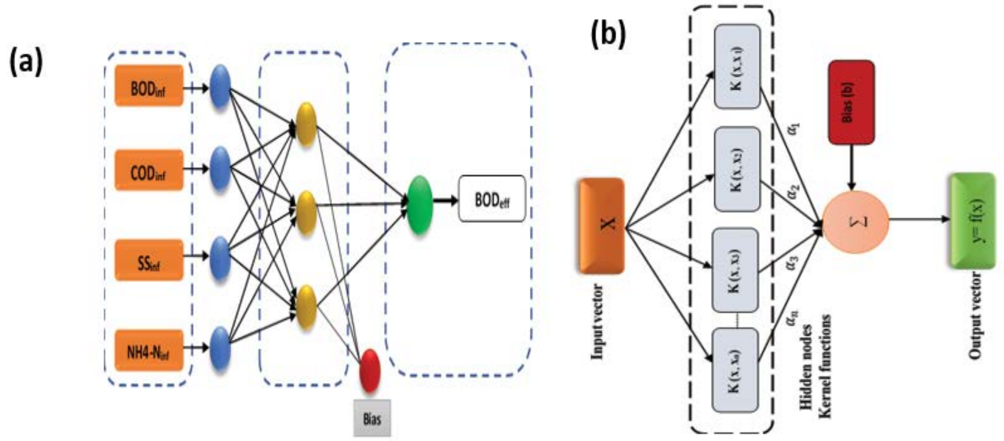

Fig. 1. (a) Topological structure of the extreme learning machine network and (b) conceptual architecture of the SVM algorithm.

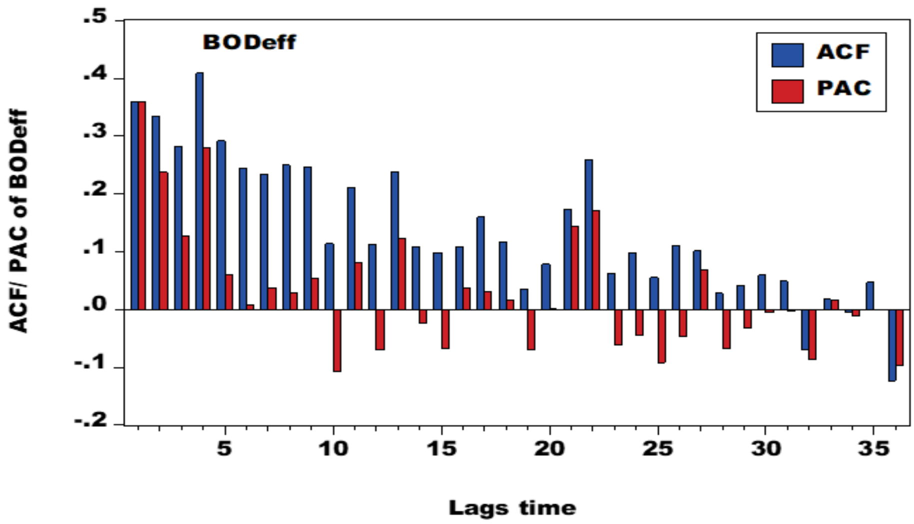

Fig. 2. Autocorrelation and partial autocorrelation functions of BOD_{eff}.

important parameters of water quality assessment employed for estimating the organic pollution of water. The COD is widely used as a measure of the susceptibility to oxidation of the organic and inorganic materials present in the water bodies. NH_4–N is an important parameter for water quality assessment; generally, the presence of nitrogen in wastewater indicates the presence of organic matter in it. Nitrogen is essential to the growth of Protista and plants, and such is known as nutrient or biostimulant. Sewage normally contained 99.9% of water and 0.1% of solids. Measuring suspended solids (SS) in water is used for control of various treatment processes and for the examination of wastewater quality. The level of suspended solids (or total suspended solids) in water and wastewater affects the quality of the water and how it can be used.

The parameters of the ARIMA model were optimized using different trial-and-error procedures. The model was fitted to the training data and the trained model was subsequently used to find the testing error using values of the testing data [37,39]. The parameters of the SVM model and the number of hidden neurons in ELM were also optimized to obtain the best architecture of SVM and ELM models. The optimal ELM model was selected following the process of Yaseen et al. [33,34]. The optimum SVM model was obtained by adopting different combinations of kernel function (γ) and the regularization constant parameter (C).

The performance of the models can be assessed through different statistical measures, that is, the Nash–Sutcliffe coefficient, root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) [40–43]. The predicting performance of the ELM, SVM, and ARIMA models, was evaluated using RMSE, MAPE, and mean squared error (MSE) in this study. RMSE, MAPE, and MSE are defined as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(\text{BOD}_{\text{obs},i} - \text{BOD}_{\text{pre},i}\right)^2}{N}} \quad (7)$$

$$\text{MAPE} = \frac{1}{N}\left[\sum_{i=1}^{N}\left|\frac{\text{BOD}_{\text{obs},i} - \text{BOD}_{\text{pre},i}}{\text{BOD}_{\text{obs},i}}\right|\right] \quad (8)$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\text{BOD}_{\text{obs},i} - \text{BOD}_{\text{pre},i}\right)^2 \tag{9}$$

where N, $\text{BOD}_{\text{obs},i}$, and $\text{BOD}_{\text{pre},i}$ are sample size, observed, and predicted values, respectively.

2.5. Bunus regional sewage treatment plant

The plant is located in the northeast of Kuala Lumpur, Malaysia, and has a capacity of 87,000 m^3/d. The BRSTP is the largest and modern treatment plant of Kula Lumpur which covers the catchment area of 70km^2 which includes three major towns, for example, Ampang Ulu Klang, Gombak, and Bunus. The BRSTP is based on the process of step-feed activated sludge process and uses different parameters, for example, BOD, COD, suspended solids, and nitrogen.

This plan aims to control environmental pollution by improving the quality of effluents. The raw sewage passes into a primary settler which allows the solids to settle at the bottom of the clarifier by the action of gravitational sedimentation [43]. Fig. 3 shows the influents and effluents process. To have a perfect understanding of wastewater management, it is necessary to understand the hydro-environmental processes that govern the wastewater pattern and its phenomenon. The past few decades have witnessed several studies on the wastewater phenomenon as a result of the interest in studying both regional and global patterns of hydro environmental changes. The wastewater pattern is modeled using two primary approaches: (i) mathematical or physically-based models such as models that deploy partial differential equations, and (ii) artificial intelligence (AI) models such as soft computing methods [13]. However, several studies need to be done and especially on the wastewater variables to select from the physical or artificial intelligence (AI) based models, the physical models still displayed various weaknesses in contrast, while AI based models which are associated with nonlinear optimization algorithm emerged to replace linear mathematical optimization process.

3. Results and discussion

3.1. Modeling performance of ELM, SVM and ARIMA model

Both descriptive statistical analysis and correlation matrix were used to explore the type, degree, and extent of the relationship between influent and effluent parameters. Spearman correlation coefficient (R) was calculated to measure the degree of the linear relationship between two variables. The significance of R is checked at a 5% significance level (Appendix-B). A negative R-value indicates an inverse relationship between two variables while vice-versa for positive value [5]. Hence, a weak R-value (close to zero) depicts that the application of conventional techniques is not useful in modeling such complex interactions and there is a great need to introduce more robust tools. As defined in Eqs. (4)–(6), three different models considering different input combinations were trained using ELM, SVM, and ARIMA models in order to forecast the effluent BOD_{eff}.

Table 1 shows that M3, having four input variables (BOD_{inf}, COD_{inf}, SS_{inf}, $\text{NH}_{4\text{inf}}$), produces the best result. A closer assessment of the outcome revealed that M3 has the lowest RMSE, MSE, and MAPE in both training and testing phases (Table 1) among the 3 models (M1, M2, and M3). The result also indicated that the ELM model % error was reduced by ~5% on the addition of input variables in the testing phase. No clear trend in the performance of SVM and ARIMA was found by increasing the number of input variables. It can also be noted that the values of RMSE, MSE, and MAPE are lowest for the ELM model which indicates better performance of ELM than the other two models. It can be seen that the performance of the models considering the ELM model can be related to M3 > M2 > M1. It can be noted that M1 with two input combinations (BOD_{inf}, COD_{inf}) has the best performance indicator considering the SVM model. The values of RMSE, MSE, and MAPE were found lowest for the M1 model than M2 and M3. The results also indicated that the M1 model exhibits 4% and 1% lesser values of MAPE in the testing phase in comparison to M2 and M3. This indicates that MAPE is the lowest in the M1 model. A conventional linear model like ARIMA is usually employed as the reference method for evaluating another non-linear black-box model. Overall, it can be seen from Table 1 that M1 is the best model among the three different models, which can be justified by the values of the performance indicators (RMSE, MSE, and MAPE). The MAPE in the testing phase for the M1 model is lower than ~2% and 4% from M2 and M3. The hierarchical performance of the ARIMA models is M1 > M2 > M3.

The predicted value of BOD_{eff} using different models is also compared with observed BOD_{eff} and shown as box plots in Fig. 4, which demonstrates the similarity of predicted

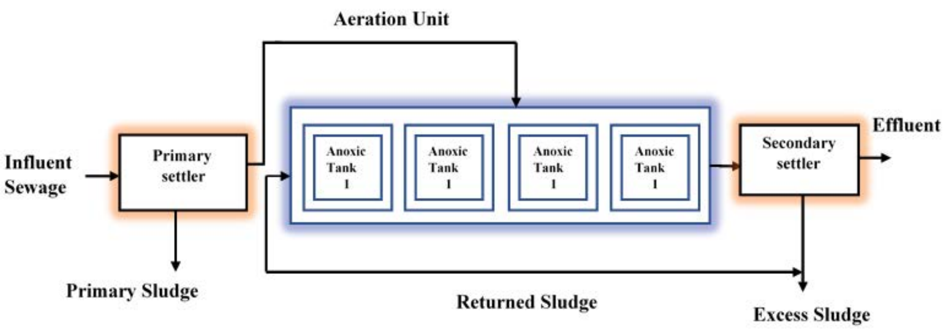

Fig. 3. Schematic diagram of Bunus regional sewage treatment plant layout.

Table 1
Performance indicators of ELM, SVM, and ARIMA models

Techniques	Models	Training			Testing		
		RMSE	MSE	MAPE	RMSE	MSE	MAPE
	M1	0.2194	0.0481	0.0724	0.0766	0.0059	0.4825
ELM	M2	0.1738	0.0302	0.0712	0.0880	0.0077	0.4745
	M3	0.1099	0.0121	0.0711	0.0712	0.0051	0.4276
	M1	0.2287	0.0503	2.4999	0.0892	0.0080	0.6180
SVM	M2	0.2277	0.0519	4.9408	0.0953	0.0091	0.6558
	M3	0.2265	0.0523	5.1062	0.1006	0.0101	0.6280
	M1	0.2271	0.0516	0.0933	0.0936	0.0094	0.7219
ARIMA	M2	0.2315	0.0536	0.1065	0.0738	0.0055	0.7244
	M3	0.2268	0.0514	0.0968	0.0752	0.0056	0.7451

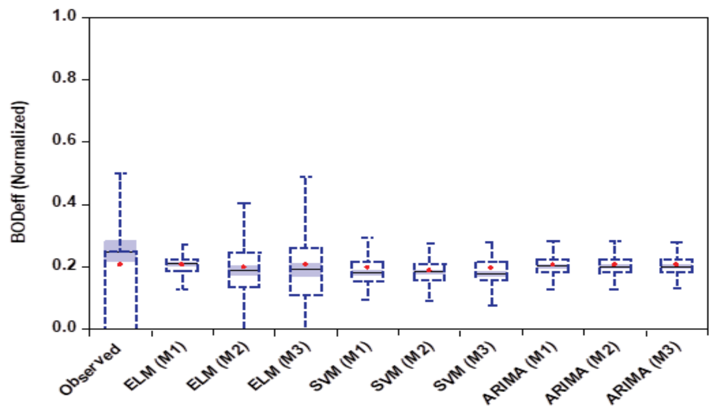

Fig. 4. Box plot of the observed BOD_{eff} compared with the forecasted models of ELM, SVM, and ARIMA.

output with the observed values using different models. It can be observed that the spread of predicted values using SVM and ARIMA models is almost similar irrespective of the number of input parameters. However, considering ELM, the spread is getting similar to the observed output by increasing the number of inputs. The median of the model outputs is more or less similar to the observed data irrespective of the model used and the number of input parameters. Overall, the ELM (M3) ranked the best model among all the models.

The testing results also indicate that the MAPE using ELM reduced by 19% and 29% than SVM and ARIMA models, respectively. Generally, the non-linear black-box models (ELM and SVM) are considered to be superior to conventional time-series (ARIMA) models in terms of computational run-time and modeling efficiency, which is due to their promising ability to model highly complex and non-linear processes.

3.2. Modeling speed of ELM, SVM and ARIMA models

Apart from modeling performance, computational time is also an important aspect to choose a better model. The computational time taken by different modeling techniques was observed. The computational time is taken by ELM, SVM, and ARIMA models to run their respective best combinations of input–output parameters, that is, M1, M2, and M3 are indicated in Table 2. The best performing combination from M1, M2, and M3 for each modeling technique is also presented in Table 2. Both SVM and ARIMA showed the performance of the M1 model better than other models, while ELM showed an input parameter combination of the M3 model better than other combinations. It can be observed that the ELM M3 model showed the least RMSE, MAE, and MAPE than the M1 model of SVM and ARIMA. The ELM has also shown fast learning speed and lowest computational run time than other models. These results are also in close agreement with similar previous studies [30,32].

Another way to visualize the performance of different models is to compare the time series plots of observed and predicted values. The time series plots of predicted BOD_{eff} using best among M1, M2, and M3 input combination for each model with observed BOD_{eff} are shown in Fig. 5. It can be observed that the ELM (M3) model is satisfactorily able to predict low as well as high values of BOD_{eff}. The SVM (M1) model was able to predict only a few of the data satisfactorily. Most of the predicted values were found to be lying between 0.12 and 0.28 which indicates that it was not able to model the variability. The ARIMA (M1) model performance was the poorest among all three methods. Careful

Table 2
Comparison of the best model and the running time for ELM, SVM, and ARIMA

Techniques	Models	Training			Testing			Run-time
		RMSE	MSE	MAPE	RMSE	MSE	MAPE	
ELM	M3	0.1099	0.0121	0.0711	0.0712	0.0051	0.4276	5.26
SVM	M1	0.2287	0.0503	2.4999	0.0892	0.0080	0.6180	9.40
ARIMA	M1	0.2271	0.0516	0.0933	0.0936	0.094	0.7219	19.13

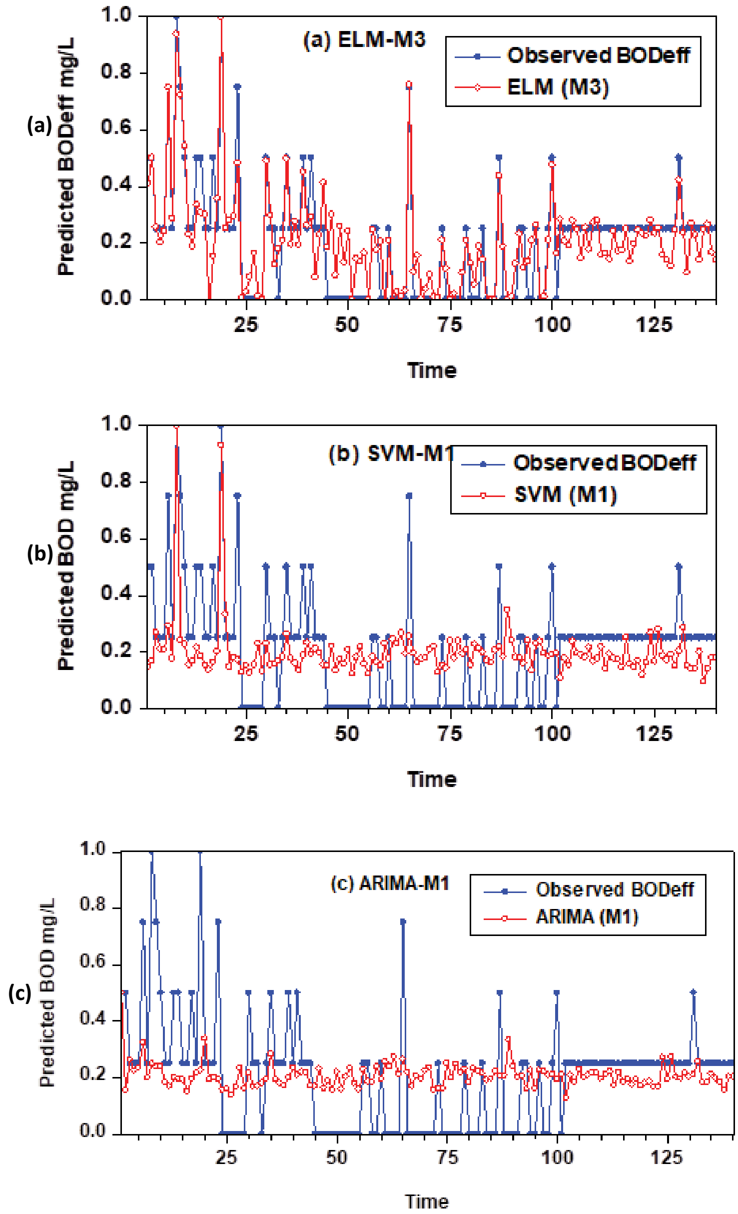

Fig. 5. Time series of observed vs. predicted BOD_{eff} considering models using (a) ELM, (b) SVM, and (c) ARIMA methods.

observation shows that the best sets of model inputs were not the same for each of the employed prediction techniques, signifying that the individual model type responds in a different way to different input parameters.

Fig. 6 depicts the bar plot of the performance indicators for the best model. It was noted that the smaller the values of RMSE, MSE, and MAPE, the more accurate the forecasting results [19]. Hence, the nonlinear black-box models demonstrated high forecasting ability in Bunus WWTPs, and can, therefore, be considered as a valuable and reliable forecasting tool for the WWTPs. The error outcomes could be justified by considering the studies conducted by Guo et al. [44].

A comparison of the results is also presented and examined using a Taylor two dimensional diagram [45,46] and provided in Fig. 7. The Taylor plot is used to evaluate the performance of three different models using correlation and

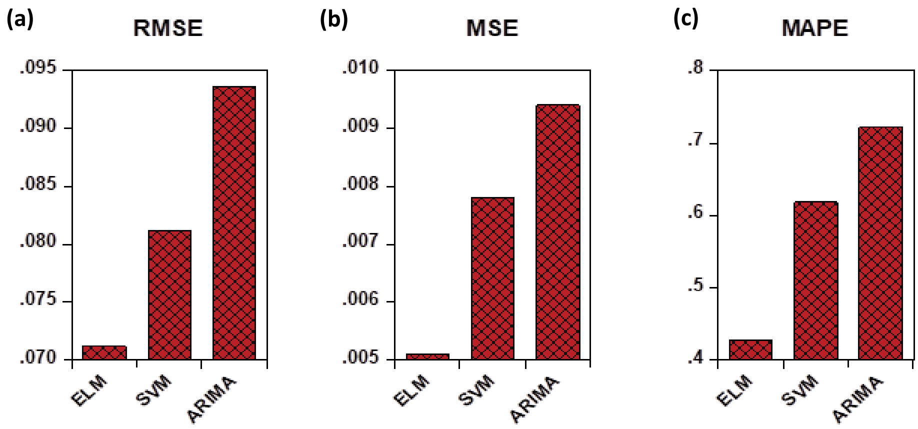

Fig. 6. Bar plots of the performance of the best models for ELM, SVM, and ARIMA methods (a) RMSE, (b) MSE, and (c) MAPE.

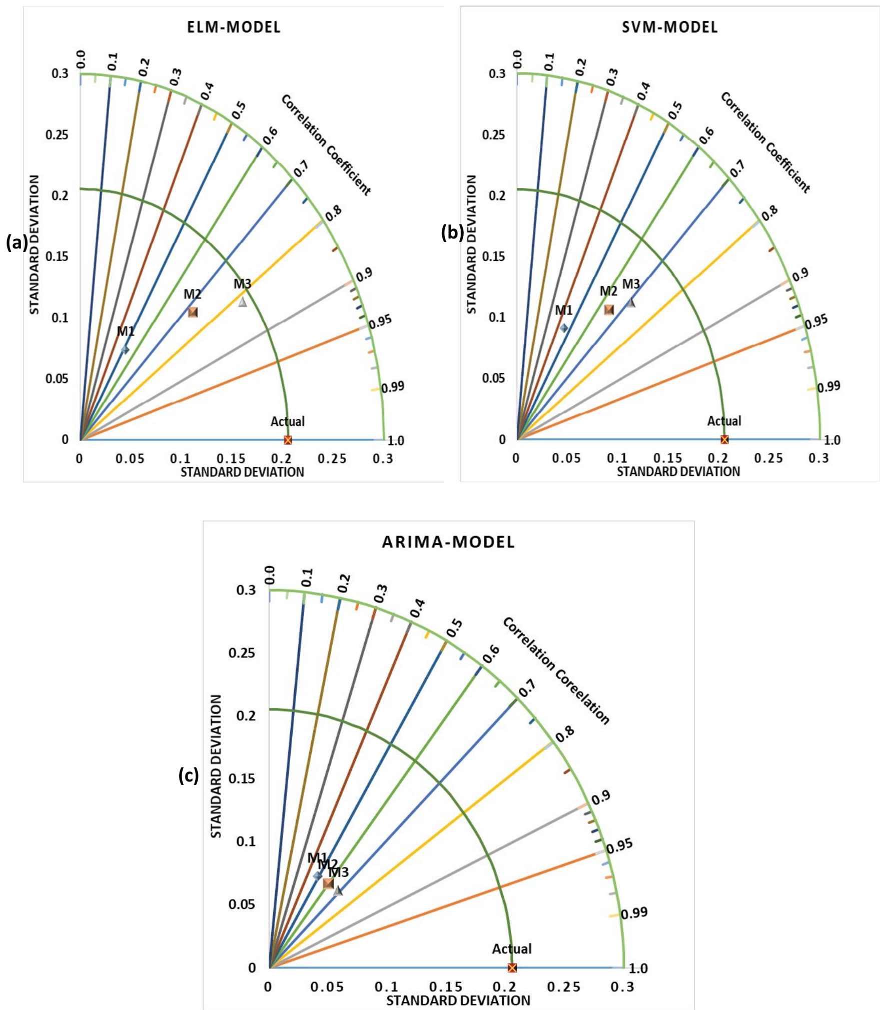

Fig. 7. Taylor diagram for (a) ELM, (b) SVM, and (c) ARIMA models.

standard deviation; in other words, the models (ELM, SVM, and ARIMA) can be visualized in terms of predictive skills and how close the predicted values from observed BOD_{eff}. It can be seen that the ELM (M3) model performed better than all other models with a correlation coefficient greater than 0.8. The standard deviation of predicted BOD_{eff} is ~0.2 which is similar to observed. It was also shown in Yaseen et al. [31] that ELM ranked highest in performance than other models, for example, M5 tree, support vector regression, and multivariate adaptive regression splines models. It can also be noted that the M3 model used in SVM for the prediction of BOD_{eff} is having better correlation, similar spread with observed BOD_{eff} than M1 and M2 models. So, it can be said that, for the prediction of BOD_{eff} in BRSTP, input parameters used in M1 and M2 are not sufficient to predict BOD_{eff} effectively, so input parameters used in the M3 model could be a better choice.

For data intelligence algorithms such as ELM and SVM, the optimization process is only employed when it is necessary. In the case of Bunus sewage treatment plant, both the model and parameters employed require no optimization techniques owing to the consideration of the previous studies conducted in similar case studies using a mathematical optimization approach. It is stated that the ELM was recently developed as a new learning approach whose major advantage is in its ability to map the internal features without the need to iteratively tune the parameters of the hidden neuron as required in a traditional ANN model [23]. The input and hidden neuron weights are computed randomly in the ELM from several pre-assigned neurons without having to pass through all the neurons in the model [22]. Furthermore, the generalization capability of the ELM is acceptable, and it requires less computation time [13]. Hence these models can still be feasible in various modeling processes without been subjected to optimization processes.

4. Conclusion

Recent studies reveal that the Black-box models could be effective tools for forecasting complex and nonlinear interactions. The primary goal of this research was to explore the applicability of ELM algorithms for predicting effluent BOD in the BRSTP. In this study, ELM, SVM, and ARIMA models were used for modeling the daily effluent BOD in BRSTP. The effectiveness of ELM was examined and compared with SVM and ARIMA models and the results were evaluated in terms of widely-used performance indices. It was found that the non-linear black-box models (ELM and SVM) can potentially improve the modeling of BOD_{eff} in comparison to the linear models, for example, ARIMA. The combination of input parameters used in the M3 model was found the best choice among all three models. The ELM (M3) model showed a reduction in MAPE by 19% and 29% than SVM and ARIMA models, respectively, which is a significant improvement in BOD_{eff} modeling. It was also found that the addition of SS_{inf} was not able to significantly improve the model performance. The ELM can be a better choice to predict BOD_{eff} of Bunus WWTP than other widely adopted methods. Other advanced machine learning methods, for example, relevance vector machines, random forest, gradient boosting, etc can also be tested to model the process of WWTPs. The quality and quantity of the observed data highly affect the performance of the developed models, so it expected to get better results with larger datasets also.

References

[1] H.Z. Abyaneh, Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters, J. Environ. Health Sci. Eng., 12 (2014) 1–8.

[2] 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. Sustainable Energy, 27 (2008) 439–446.

[3] J.K. Jain, Wastewater Engineering: Including Air Pollution, Laxmi Publications, India, 2014.

[4] S. Das, K. Radhakrishnan, Multicriteria decision making model of wastewater reuse: a stakeholders perspective in the context of India, Desal. Water Treat., 163 (2019) 17–25.

[5] V. Nourani, G. Elkiran, S.I. Abba, Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach, Water Sci. Technol., 78 (2018) 2064–2076.

[6] Z.Q. Huang, J.X. Luo, X.J. Li, Y.J. Zhou, Prediction of Effluent Parameters of Wastewater Treatment Plant Based on Improved Least Square Support Vector Machine with PSO, First International Conference on Information Science and Engineering, Nanjing, China, 2009, pp. 4058–4061.

[7] F. Granata, S. Papirio, G. Esposito, R. Gargano, G. de Marinis, Machine learning algorithms for the forecasting of wastewater quality indicators, Water, 9 (2017) 1–12.

[8] G. Civelekoglu, N.O. Yigit, E. Diamadopoulos, M. Kitis, Modelling of COD removal in a biological wastewater treatment plant using adaptive neuro-fuzzy inference system and artificial neural network, Water Sci. Technol., 60 (2009) 1475–1487.

[9] D. Ribeiro, A. Sanfins, O. Belo, Wastewater Treatment Plant Performance Prediction with Support Vector Machines, P. Perner, Ed., Advances in Data Mining. Applications and Theoretical Aspects, Vol. 7987, 13th Industrial Conference, ICDM 2013, New York, NY, USA, 2013, pp. 99–111.

[10] S. Pakrou, N. Mehrdadi, A. Baghvand, ANN modeling to predict the COD and efficiency of waste pollutant removal from municipal wastewater treatment plants, Curr. World Environ., 10 (2015) 873–881.

[11] M.S. Nasr, M.A.E. Moustafa, H.A.E. Seif, G. El Kobrosy, Application of artificial neural network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT, Alexandria Eng. J., 51 (2012) 37–43.

[12] M.S. Gaya, N.A. Wahab, Y.M. Sam, A.N. Anuar, S.I. Samsuddin, ANFIS modelling of carbon removal in domestic wastewater treatment plant, Appl. Mech. Mater., 372 (2013) 597–601.

[13] W.C. Chen, N.-B. Chang, W.K. Shieh, Advanced hybrid fuzzy-neural controller for industrial wastewater treatment, J. Environ. Eng., 11 (2001) 1048–1059.

[14] A. Maleki, S. Nasseri, M.S. Aminabad, M. Hadi, Comparison of ARIMA and NNAR models for forecasting water treatment plant's influent characteristics, J. Civ. Eng., 22 (2018) 3233–3245.

[15] S.I. Abba, G. Elkiran, Effluent prediction of chemical oxygen demand from the wastewater treatment plant using artificial neural network application, Procedia Comput. Sci., 120 (2017) 156–163.

[16] C. Sharma, C.S.P. Ojha, A.K. Shukla, Q.B. Pham, N.T.T. Linh, C.M. Fai, H.H. Loc, T.D. Dung, Modified approach to reduce GCM bias in downscaled precipitation: a study in Ganga River Basin, Water, 11 (2019) 1–31.

[17] A. Verma, X. Wei, A. Kusiak, Predicting the total suspended solids in wastewater: a data-mining approach, Eng. Appl. Artif. Intell., 26 (2013) 1366–1372.

[18] M.M. Hamed, M.G. Khalafallah, E.A. Hassanien, Prediction of wastewater treatment plant performance using artificial neural networks, Environ. Model Software, 19 (2004) 919–928.

[19] M.S. Gaya, M.U. Zango, L.A. Yusuf, M. Mustapha, B. Muhammad, A. Sani, A. Tijjani, N.A. Wahab, M.T.M. Khairi, Estimation of turbidity in water treatment plant using

hammerstein-wiener and neural network technique, Indones. J. Electr. Eng. Comput. Sci., 5 (2017) 666–672.

[20] E. Sharghi, V. Nourani, A.A. Ashrafi, H. Gökçekuş, Monitoring effluent quality of wastewater treatment plant by clustering based artificial neural network method, Desal. Water Treat., 164 (2019) 86–97.

[21] M.S. Gaya, N. Abdul Wahab, Y.M. Sam, S.I. Samsudin, ANFIS modelling of carbon and nitrogen removal in domestic wastewater treatment plant, J. Teknol., 67 (2014) 29–34.

[22] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing, 70 (2006) 489–501.

[23] G. Huang, G.-B. Huang, S. Song, K. You, Trends in extreme learning machines: a review, Neural Networks, 61 (2015) 32–48.

[24] L.J. Zhao, X.K. Diao, D.C. Yuan, W. Tang, Enhanced classification based on probabilistic extreme learning machine in wastewater treatment process, Procedia Eng., 15 (2011) 5563–5567.

[25] P. Yu, J. Cao, V. Jegatheesan, X. Du, A real-time BOD estimation method in wastewater treatment process based on an optimized extreme learning machine, Appl. Sci., 9 (2019) 1–12.

[26] X. Deng, C. Lin, Application of ELM to predict the coagulant dosing in water treatment plants, Water Sci. Technol. Water Supply, 17 (2017) 1053–1061.

[27] M. Djerioui, M. Bouamar, M. Ladjal, A. Zerguine, Chlorine soft sensor based on extreme learning machine for water quality monitoring, Arabian J. Sci. Eng., 44 (2019) 2033–2044.

[28] A.R.S. Parmezan, V.M.A. Souza, G.E.A.P.A. Batista, Evaluation of statistical and machine learning models for time series prediction: identifying the state-of-the-art and the best conditions for the use of each model, Inf. Sci., 484 (2019) 302–337.

[29] K. Kandananond, A comparison of various forecasting methods for autocorrelated time series, Int. J. Eng. Bus. Manage., 4 (2012) 1–6.

[30] R.C. Deo, M. Şahin, Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern Australia, Atmos. Res., 153 (2015) 512–525.

[31] Z.M. Yaseen, S.O. Sulaiman, R.C. Deo, K.-W. Chau, An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction, J. Hydrol., 569 (2019) 387–408.

[32] Z.M. Yaseen, O. Jaafar, R.C. Deo, O. Kisi, J. Adamowski, J. Quilty, A. El-Shafie, Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq, J. Hydrol., 542 (2016) 603–614.

[33] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.

[34] V.N. Vapnik, A.Y. Chervonenkis, On the method of ordered risk minimization I, Autom. Remote Control, 35 (1974) 1226–1235.

[35] J. Wang, H. Du, H. Liu, X. Yao, Z. Hu, B. Fan, Prediction of surface tension for common compounds based on novel methods using heuristic method and support vector machine, Talanta, 73 (2007) 147–156.

[36] D.Ö. Faruk, A hybrid neural network and ARIMA model for water quality time series prediction, Eng. Appl. Artif. Intell., 23 (2010) 586–594.

[37] S.I. Abba, M.S. Gaya, M.L. Yakubu, M.U. Zango, R.A. Abdulkadir, M.A. Saleh, A.N. Hamza, U. Abubakar, A.I. Tukur, N.A. Wahab, Modelling of Uncertain System: A comparison study of Linear and Non-Linear Approaches, IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Malaysia, 2019.

[38] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, 50 (2003) 159–175.

[39] M.E. Turan, M.A. Yurdusev, River flow estimation from upstream flow records by artificial intelligence methods, J. Hydrol., 369 (2009) 71–77.

[40] W.-c. Wang, K.-w. Chau, L. Qiu, Y.-bo. Chen, Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition, Environ. Res., 139 (2017) 46–54.

[41] V. Nourani, An emotional ANN (EANN) approach to modeling rainfall-runoff process, J. Hydrol., 544 (2017) 267–277.

[42] G. Elkiran, V. Nouren, S.I. Abba, J. Abdullahi, Artificial intelligence-based approaches for multi-station modelling of dissolve oxygen in river, Global J. Environ. Sci. Manage., 4 (2018) 439–450.

[43] N.S.A. Yasmin, M.S. Gaya, N.A. Wahab, Y.M. Sam, Estimation of pH and MLSS using neural network, Telkomnika, 15 (2017) 912–918.

[44] H. Guo, K. Jeong, J. Lim, J. Jo, Y.M. Kim, J.-p. Park, J.H. Kim, K.H. Cho, Prediction of effluent concentration in a wastewater treatment plant using machine learning models, J. Environ. Sci., 32 (2015) 90–101.

[45] K.E. Taylor, Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106 (2001) 7183–7192.

[46] M.A. Ghorbani, R.C. Deo, Z.M. Yaseen, M.H. Kashani, B. Mohammadi, Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran, Theor. Appl. Climatol., 133 (2018) 1119–1131.

Appendix A

Auto-correlation and partial auto-correlation function

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.359	0.359	18.411	0.000
		2	0.335	0.237	34.611	0.000
		3	0.282	0.128	46.132	0.000
		4	0.409	0.281	70.618	0.000
		5	0.292	0.061	83.157	0.000
		6	0.245	0.008	92.079	0.000
		7	0.234	0.038	100.28	0.000
		8	0.250	0.030	109.68	0.000
		9	0.246	0.054	118.85	0.000
		10	0.114	-0.107	120.84	0.000
		11	0.210	0.081	127.61	0.000
		12	0.113	-0.070	129.60	0.000
		13	0.239	0.123	138.55	0.000
		14	0.109	-0.024	140.43	0.000
		15	0.098	-0.068	141.96	0.000
		16	0.109	0.038	143.85	0.000
		17	0.160	0.031	147.97	0.000
		18	0.117	0.016	150.19	0.000
		19	0.036	-0.070	150.40	0.000
		20	0.079	0.001	151.43	0.000
		21	0.174	0.143	156.50	0.000
		22	0.259	0.172	167.81	0.000
		23	0.063	-0.061	168.49	0.000
		24	0.097	-0.045	170.10	0.000
		25	0.056	-0.092	170.64	0.000
		26	0.111	-0.047	172.77	0.000
		27	0.101	0.069	174.58	0.000
		28	0.028	-0.068	174.72	0.000
		29	0.041	-0.032	175.01	0.000
		30	0.060	-0.005	175.66	0.000
		31	0.049	-0.002	176.09	0.000
		32	-0.069	-0.086	176.96	0.000
		33	0.018	0.017	177.02	0.000
		34	-0.004	-0.011	177.02	0.000
		35	0.047	0.000	177.45	0.000
		36	-0.123	-0.096	180.35	0.000

Appendix B

Pearson and Spearman correlation and descriptive statistical analysis

Parameters	BOD_{inf}	COD_{inf}	SS_{inf}	$NH_4^-N_{inf}$	BOD_{eff}	Mean	Standard deviation	Minimum	Maximum
BOD_{inf}	1.0000					147.1571	29.5747	86.0000	257.0000
COD_{inf}	0.6046	1.0000				317.1071	63.0775	165.0000	566.0000
SS_{inf}	0.4711	0.5303	1.0000			140.1143	36.2140	35.0000	264.0000
$NH_4^-N_{inf}$	0.1573	0.0908	0.1250	1.0000		24.1357	3.3505	14.0000	36.0000
BOD_{eff}	0.5157	0.4071	0.1705	–0.0869	1.0000	1.8286	0.8218	1.0000	5.0000