Environmental Health Engineering and Management Journal 2023, 10(3), 293–300
http://ehemj.com
doi 10.34172/EHEM.2023.33

Environmental Health Engineering and Management Journal

Original Article

Environmental Health Engineering and Management

Open Access Publish Free

Modeling of wastewater treatment plant in Hama city using regression and regression trees

Heba Bodaka[1*], Nahed Farhoud[1], Eyad Hlali[2]

[1]Department of Environmental Engineering Technologies, Aleppo University, Aleppo, Syria
[2]Department of Computer Engineering, Aleppo University, Aleppo, Syria

Abstract

Background: Modeling of wastewater treatment plants is necessary to predict their later works. In this research, three methods were compared to predict some parameters at the outlet of wastewater treatment plant in Hama city in Syria.

Methods: In this paper, three methods (linear regression, power regression, and regression trees) to model wastewater treatment plant in Hama city were compared to predict the parameters at the outlet of the plant ($cBOD_{5out}$, COD_{out}, TSS_{out}) in terms of the parameters at the inlet of the plant (Q_{in}, $cBOD_{5in}$, COD_{in}, TSS_{in}).

Results: When predicting $cBOD_{5out}$, the values of RMSE of the test data set were 4.4105, 4.3875, and 3.8418; when predicting COD_{out}, the values of RMSE of the test data set were 6.9325, 6.8003, and 5.3232; and when predicting TSS_{out}, the values of root mean squared error (RMSE) of the test data set were 3.7781, 3.6936, and 3.2391 using linear regression, power regression, and regression trees (RTs), respectively.

Conclusion: According to the results, the RTs outperforms in predicting $cBOD_{5out}$, COD_{out}, and TSS_{out} because this method achieved the least RMSE of the test data set.

Keywords: Linear models, Decision trees, Water purification, Syria

Citation: Bodaka H, Farhoud N, Hlali E. Modeling of wastewater treatment plant in Hama city using regression and regression trees. Environmental Health Engineering and Management Journal 2023; 10(3): 293–300. doi: 10.34172/EHEM.2023.33.

Article History:
Received: 5 September 2022
Accepted: 29 January 2023
ePublished: 27 July 2023

***Correspondence to:**
Heba Bodaka,
Email: hebabodaka7@gmail.com

Introduction

Wastewater treatment is an important measure that should be taken seriously for the betterment of our society and future (1). The use of treated wastewater in the irrigation of agricultural lands is necessary due to the lack of water and the change of climate (2). Wastewater treatment is a set of processes that improve the quality of wastewater and reduce its harmful impact on humans (3). Pollutants are removed from wastewater through treatment (4). Modeling of wastewater treatment plant is a challenging work due to the complexity of treatment processes (5). Effluent treatment plant modeling is fundamental to estimate process and work of the plant. Some of the important treatment variables cannot be measured online. For example, biological oxygen demand (BOD_5) needs an incubation of 5 days, so this is hard to be obtained (1). Lately, computer modeling techniques have been applied in several ecological topics (6) due to the significance of sewage water process and its act in decreasing the ecological pollution (7). Various approaches have been used to model and control the operation of wastewater treatment plants, such as expert systems (8), knowledge-based systems, neural networks (9), mixed approaches (10), and machine learning methods (11,12). There are some main variables such as chemical oxygen demand (COD), BOD, and total suspended solids (TSS) used to evaluate the performance of a sewage water treatment plant (13,14).

Predicting BOD_5 and COD rather than measuring them may be an environmentally and economically safe method due to the time required to measure them and the required measurement procedures that involve the use of many dangerous chemicals (15). Many researchers have applied multilinear regression (MLR), power regression, and regression trees (RTs) models to predict various water quality parameters. Abba et al predicted the effluent COD of Nicosia wastewater treatment plant using MLR, and they got the value of root mean squared error (RMSE) equal to 0.0121, the predictor parameters were BOD, COD, pH, conductivity, total phosphates (T-P), total nitrogen (T-N), suspended solid (SS) and TSS (16). Ewaid et al predicted the water quality index of the

Tigris River in Baghdad using power regression, and they concluded that this method can be used to predict with acceptable accuracy (17). Baki et al created models with linear regression and power regression to estimate BOD at the entrance of the sewage treatment plant in Antalya. The input discharge (Q) was one of the input parameters of the model, and the RMSE values were 54.0593 and 52.1535 for linear and power regression, respectively, so the power model outperformed (18). Granata et al used RT to predict BOD_5, COD, and TSS, and it has achieved durability, accuracy, as well as great generalizability. The RMSE values for predicting BOD, COD, and TSS were 103, 2395, and 3486, respectively. The predictor parameters were residential area percentage, commercial area percentage, drainage area, industrial area percentage, institutional area percentage, freeway percentage, open space area percentage, impervious area percentage, runoff and precipitation depth (9). Faraji-Khiavi et al used the binary logistic regression in order to determine the effect of demographic, economic, social, and disease status on the use of health services during the COVID-19 pandemic by the elderly in Iran, through a study conducted in 21 public health centers in Sirjan, southern Iran, and the results showed demographic, social, and economic disparities in the use of health services among the elderly (19).

In this paper, three methods, linear regression, power regression, and RTs, were used to model the wastewater treatment plant in Hama city, Syria. Then, the methods were compared and the one that achieves the least RMSE for the test data set was selected.

Materials and Methods

Hama wastewater treatment plant

The wastewater treatment plant in Hama city is located next to the village of Arza Al-Sharqiya, in the northern part of Hama city, 7 km from the city center, and it is situated on the banks of Orontes River where treated water is disposed of in the Orontes River, and it is on a total area of 80 000 m^2. The plant was put into service on June 3, 2005, and the amount of wastewater it receives is approximately 50 000 m^3/day from the living wastewater of Hama city. Its design life ends when its load values and incoming flow are higher than the loads it was designed to receive. Monthly data were collected for the parameters, flow (Q), carbonaceous biochemical oxygen demand ($cBOD_5$), COD, and TSS, at the entrance and the exit of the plant for the years from 2014 to the first month of 2020 measured by (m^3/day) for the flow and by (mg/L) for the remaining parameters from the wastewater treatment plant in Hama city.

Multi linear regression

Regression is an approach used to find the relationship between a dependent variable and a set of independent variables. It is used in many fields, including engineering, finance, business, medical, and others. There are several regression techniques that have been presented in the literature and it is used for research objectives (20).

Multi linear regression is based on the idea of a linear relation between the independent variable and the dependent variable (21). It uses the least squares method. Equation (1) shows the relationship of multi linear regression.

$$Y_i = b_0 + b_1 x_1 + b_2 x + \ldots + b_m x_m \quad (1)$$

where x_1, x_2, ..., x_m are the predictors, b_0 is the constant of the regression, and (b_1, b_2, ..., b_m) are the predictors coefficients of (x_1, x_2, ..., x_m) (22).

Similar to simple regression, the regression coefficients (b_0, b_1, b_2, ..., b_m) are calculated by reducing the sum of the differences (e_{yi}) between the actual values and the estimated values through the model as shown in Eq. (2) (23).

$$\sum_{i=1}^{n} e_{yi}^2 = \sum_{i=1}^{n} \left(Y_i - b_0 - bX_{1i} - b_2 X_{2i} - \ldots - b_m X_{mi}\right)^2 \quad (2)$$

Several authors used MLR to obtain a statistical model (24-26). Figure 1 shows a linear regression curve.

Regression methods have been widely used in predicting influent and effluent wastewater variables (27), and many researchers have used them extensively to assess quality parameters in civil flow, tanks, superficial water, as well as the plants of sewage treatment (28-33).

The regression equation is simple, and it takes much less time than other machine learning algorithms, but the majority of the real-world issues behave non-linearly behavior, so the linear regression does not fit neatly into their data, as a linear relationship is assumed between the input and output variables (20).

Power regression

It is a form of regression, in which the dependent variable is related to the independent variable raised to a power. The form of the power regression is as Eq. (3).

$$y = ax^b \quad (3)$$

Where x is the predictor variable and b is a constant. Power regression will not allow an independent variable of 0.

Figure 2 shows the shape of the power regression of $y = x^2$ (20).

The function of multi power regression is expressed as shown in Eq. (4).

$$y = b_0 * x_1^{b_1} * x_2^{b_2} * \ldots * x_m^{b_m} \quad (4)$$

Since y represents the dependent variable, while (x_1, x_2, ..., x_m) represent the independent variables, and (b_0, b_1, ...,

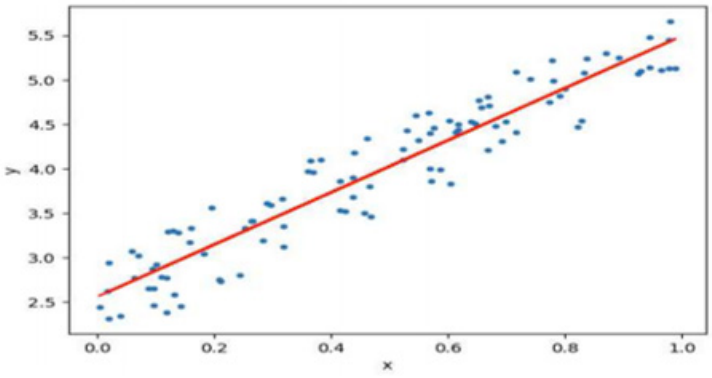

Figure 1. Linear regression curve. Adapted from Iqbal (20)

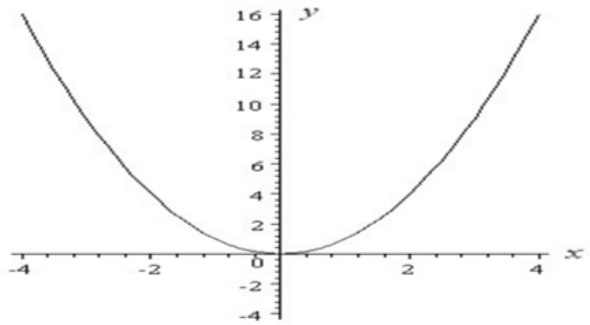

Figure 2. The shape of the power regression of the equation $y=x^2$. Adapted from Iqbal (20)

b_m) represent the regression coefficients (18).

The majority of real-world issues are linear over a very short period. As power regression is very close to the real issues. In addition, power regression techniques are used to get superior results with minimal error, but it becomes more difficult to understand and appliance than other regression models because the power of the variable increases the difficulty of the power regression technique (20).

Application of power regression

The applications of power regression are several, such as weather forecasting, ecological review, and physiotherapy (20).

Regression trees

RTs are a form of decision trees (34) that are widely used classic machine learning algorithm due to its intuitive and obvious model (35), and they are considered as continuous class decision trees (9).

RTs are hierarchical structures made up of nodes, branches, and leaves that represent the division of a field into a number of fields that can be very close to the target and with sufficient precision (36). Figure 3 shows the structure of a typical RT.

Tree leaves are numbers that represent the mean cases that reach the leaf. The tree is more complicated and bigger than the regression model (37).

The process of constructing RTs consists of a repeated process that divides the data into branches or sections. In the beginning, all the data in the training set are collected in a single partition. After that, the algorithm starts by dividing the data into the first two branches using each possible binary division of each domain. At each stage, the algorithm chooses the division that reduces the sum of squared differences from the average of the two isolated divisions (9).

The process is repeated until the number of records that reaching each node is as previously defined by the researcher, and thus, it becomes a leaf (9). In contrast to regression, RT has a great ability to capture complicated and high-level interactions between input variables (38).

The RT method was used for transfer debris (39,40), deluge estimation (41), average annual deluge estimation (42), depth cleaning estimation (43,44), and prediction of the amount of sediment in rivers (45).

Models' performance evaluation criteria

The evaluation criteria used in this research is the RMSE, and it is calculated using Eq. (5).

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}\left(\mathrm{K_i}-V_i\right)^2}{\mathrm{n}}} \quad (5)$$

Where K_i is the real value, V_i is the estimated value, and n is the number of records (33,46).

Results

Data splitting

The data have been divided into the training data set that is used in the creation of the model, and it has the largest percentage (65%), and the test data set that is not used to create the model, but the model's performance is evaluated by RMSE value when the test dataset is applied to the model, and it has the ratio of 35%. In this study, computer with Intel Core I3 1.2 GHz, 4 GB RAM, and MATLAB (Matrix Laboratory) software (R2016a) was used to get the results.

The first method (MLR)

A MLR model that relates $cBOD_{5out}$ in terms of Q_{in},

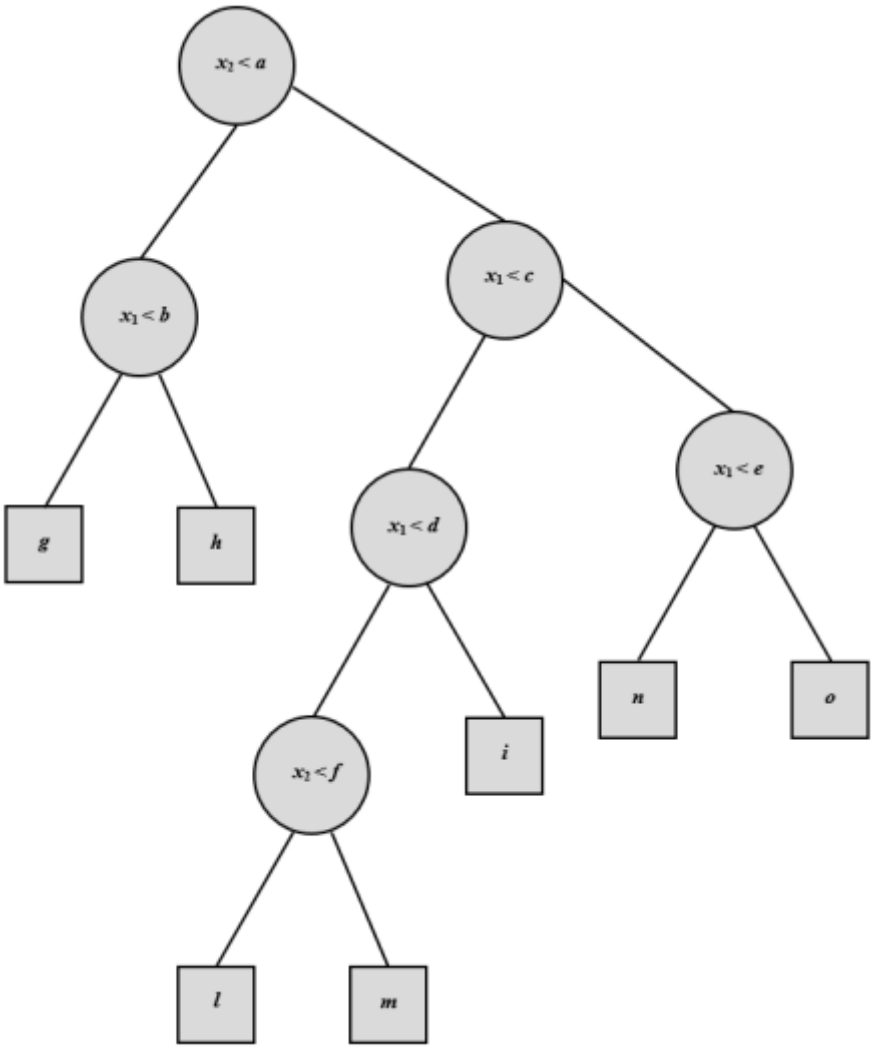

Figure 3. The structure of a typical regression tree. Adapted from Granata et al (9)

$cBOD_{5in}$, COD_{in}, and TSS_{in} was found, and the result was as Eq. (5).

$$cBOD5_{out} = 9.7227 - 0.000048227Q_{in} - 0.012666cBOD_{5in} + 0.010203COD_{in} + 0.0071636\ TSS_{in} \quad (5)$$

MLR model that relates COD_{out} in terms of Q_{in}, $cBOD_{5in}$, COD_{in}, and TSS_{in} was found, and the result was as Eq. (6).

$$COD_{out} = 38.892 - 0.000033158\ Q_{in} - 0.048097cBOD_{5in} + 0.032239COD_{in} + 0.044383\ TSS_{in} \quad (6)$$

MLR model that relates TSS_{out} in terms of Q_{in}, $cBOD_{5in}$, COD_{in}, and TSS_{in} was found, and the result was as Eq. (7).

$$TSS_{out} = -0.99199 - 0.000045951\ Q_{in} - 0.018757\ cBOD_{5in} + 0.021291COD_{in} + 0.026576\ TSS_{in} \quad (7)$$

The RMSE values of the train and test data set were also calculated, and the results were as shown in Table 1.

According to Table 1, the RMSE value when predicting COD_{out} for the testing data set, was 6.9325, which is higher than the RMSE value (0.0121) achieved in another study (16). This discrepancy is because of the differences in the predictor parameters between the two research, since the predictor parameters in the study by Abba and Elkiran were BOD, COD, pH, conductivity, T-P, T-N, SS and TSS (16), while the predictor parameters in the present research were Q_{in}, $cBOD_{5in}$, COD_{in}, and TSS_{in}.

The second method (power regression)

Power regression model that relates $cBOD_{5out}$ in terms of Q_{in}, $cBOD_{5in}$, COD_{in}, and TSS_{in}, was found, and the result was as Eq. (8).

$$cBOD5_{out} = 5.8498 * (Q_{in})^{-0.1828} * (cBOD_{5in})^{-0.25868} * (COD_{in})^{0.5471} * (TSS_{in})^{0.10978} \quad (8)$$

Power regression model that relates COD_{out} in terms of Q_{in}, $cBOD_{5in}$, COD_{in}, and TSS_{in} was found, and the result was as Eq. (9).

$$COD_{out} = 10.7930*(Q_{in})^{-0.041335}*(cBOD_{5in})^{-0.30643}* (COD_{in})^{0.36039}*(TSS_{in})^{0.26771} \quad (9)$$

Power model that relates TSS_{out} in terms of Q_{in}, $cBOD_{5in}$, COD_{in}, and TSS_{in}, was found, and the result was as Eq. (10).

$$TSS_{out} = 0.0194*(Q_{in})^{-0.099645}*(cBOD_{5in})^{-0.47954}* (COD_{in})^{0.90184}*(TSS_{in})^{0.77636} \quad (10)$$

The RMSE values of the train and test data set were also calculated, and the results are shown in Table 2.

The third method (RTs)

The RTs were adjusted by specifying a minimum number of observations in each terminal leaf that called in MATLAB software by 'MinLeafSize' ranging from 1 to 20, the values of RMSE were calculated for the train and test data set, and the results are shown in Table 3.

According to Table 3, the RMSE values when predicting $cBOD_{5out}$, COD_{out}, and TSS_{out} for the testing data set using RTs were 3.8418, 5.3232, and 3.2391, which are lower than the RMSE values (103, 2395, and 3486) achieved in another study (9). This discrepancy is because of the differences in the predictor parameters that were residential area percentage, commercial area percentage, drainage area, industrial area percentage, institutional area percentage, freeway percentage, open space area percentage, impervious area percentage, runoff and precipitation depth in the study of Granata et al (9), while the predictor parameters in the present research were Q_{in}, $cBOD_{5in}$, COD_{in}, and TSS_{in}.

By comparing the tree methods used to model the plant in this research, in terms of RMSE for the test data, the results are summarized in Table 4.

According to Table 4 and Figure 4, the results in the three methods used in this research were convergent.

Discussion

As shown in Table 4, power regression models outperform linear regression models when predicting all parameters, and the reason for this is the power regression models are much close to the real reality representation as stated in the references (18,20).

According to Table 3, RTs were experimented with a number of records in the terminal leaf ranging from 1 to 20, and then, RMSE values were calculated for the train and test data set at each number of records in the terminal leaf. The number of records in the terminal leaf which achieved the least RMSE value for the test data set was selected. They were 7, 13, 9 achieved RMSE values for the test data set equal to 3.8418, 5.3232, and 3.2391 when predicting $cBOD_{5out}$, COD_{out}, and TSS_{out}, respectively. It was revealed that the performance of the RTs model was reliable and has high generalizability, which is consistent with the results of the present study (9).

Also, as shown in Table 4, the RTs method is superior in parameters prediction that are $cBOD_{5out}$, COD_{out}, and TSS_{out} because it achieved the least RMSE values for the test data set, which were 3.84, 5.32, and 3.24, when predicting $cBOD_{5out}$, COD_{out}, and TSS_{out}, respectively, which is consistent with the results of the study of Suchetana et al (38). They reported that the RTs have a different capability to capture complicated interactions between input variables more than regression.

In the present research, RMSE values of 3.8418, 5.3232, and 3.2391 achieved when predicting BOD_5, COD, and TSS using RT, which are less than the RMSE values (103, 2395, and 3486) achieved in the study of Granata et al (9). This discrepancy is due to the differences between the predictor parameters (residential area percentage, commercial area percentage, drainage area, industrial area percentage, institutional area percentage, freeway percentage, open space area percentage, impervious area percentage, runoff and precipitation depth) used in the study by Granata et al (9), and the predictor parameters (Q_{in}, $cBOD_{5in}$, COD_{in}, and TSS_{in}) used in the present

Table 1. RMSE values for predicting $cBOD_{5out}$, COD_{out}, and TSS_{out} by MLR method

The parameter to be predicted	RMSE training data set	RMSE testing data set
$cBOD_{5out}$	2.2166	4.4105
COD_{out}	3.5929	6.9325
TSS_{out}	2.3567	3.7781

Table 2. RMSE values when predicting $cBOD_{5out}$, COD_{out}, and TSS_{out} using power regression method

The parameter to be predicted	RMSE training data set	RMSE testing data set
$cBOD_{5out}$	2.2558	4.3875
COD_{out}	3.5861	6.8003
TSS_{out}	2.3683	3.6936

Table 3. RMSE values for predicting $cBOD_{5out}$, COD_{out}, and TSS_{out} by RT method

The parameter to be predicted	Min leaf size	RMSE training data set	RMSE testing data set
$cBOD_{5out}$	7	1.7841	3.8418
COD_{out}	13	4.05376	5.3232
TSS_{out}	9	2.2075	3.2391

Table 4. The RMSE values of the test data when predicting $cBOD_{5out}$, COD_{out}, and TSS_{out} using the three methods MLR, power regression, and RT

Modeling method	$cBOD_{5out}$	COD_{out}	TSS_{out}
Linear regression	4.4105	6.9325	3.7781
Power regression	4.3875	6.8003	3.6936
RT	3.8418	5.3232	3.2391

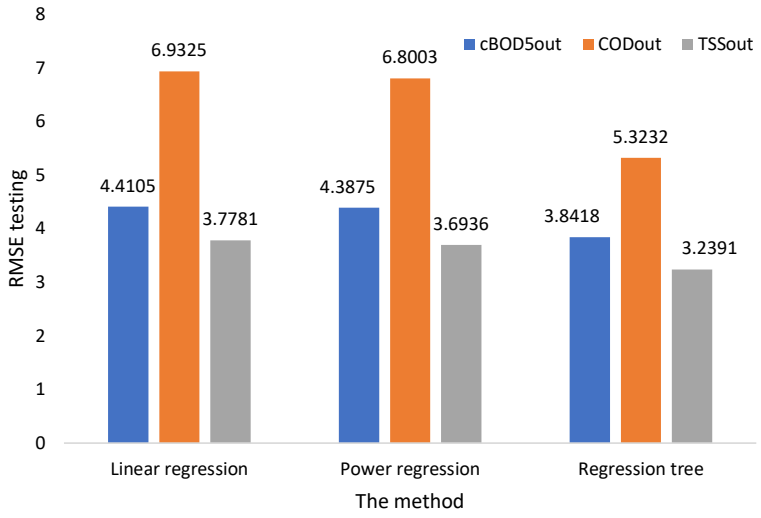

Figure 4. Comparison of MLR, power regression, and RT for predicting $cBOD_{5out}$, COD_{out}, and TSS_{out}

research.

The RMSE achieved in the study by Abba and Elkiran (16) when predicting COD_{out} using MLR was 0.0121, which is less than the RMSE value achieved in the present research (6.9325). This discrepancy is due to the differences between the predictor parameters (BOD, COD, pH, conductivity, T-P, T-N, SS, TSS) used in the study by Abba and Elkiran (16), and the predictor parameters (Q_{in}, $cBOD_{5in}$, COD_{in}, and TSS_{in}) used in the present research.

Conclusion

In this paper, MLR, power regression, and RTs were used to predict $cBOD_{5out}$, COD_{out}, and TSS_{out} at the outlet of the wastewater treatment plant in Hama city in terms of parameters Q_{in}, $cBOD_{5in}$, COD_{in}, and TSS_{in} at the inlet of the plant, and then, to compare the three methods to predict each parameter and select the method that achieves the least RMSE value for the test data set. The results showed the superiority of RTs method over the MLR and power regression methods. The best method to predict $cBOD_{5out}$, COD_{out}, and TSS_{out} is RT because it achieved the least RMSE value for the test data set, and this can be explained by the ability of the RTs method to provide good generalization in handling nonlinear relationships between the parameters to be predicted and the predictive parameters, and the parameters of wastewater treatment plants during the treatment make complex and nonlinear action.

Acknowledgments

The authors would like to thank the wastewater treatment plant in Hama city in Syria for providing support for this research.

Authors' contribution

Conceptualization: Eyad Hlali.
Data curation: Nahed Farhoud.
Formal analysis: Heba Bodaka.
Funding acquisition: Heba Bodaka.
Investigation: Heba Bodaka.
Methodology: Nahed Farhoud.
Project administration: Nahed Farhoud.
Resources: Nahed Farhoud.
Software: Eyad Hlali.
Supervision: Nahed Farhoud.
Validation: Eyad Hlali.
Visualization: Heba Bodaka.
Writing–original draft: Heba Bodaka.
Writing–review & editing: Eyad Hlali.

Competing interests

The authors declare that they have no competing interests.

Ethical issues

The authors hereby certify that all data collected during the study are as stated in this manuscript, and no data from the study has been or will be published elsewhere separately.

References

1. Vijayan A, Mohan GS. Prediction of effluent treatment plant performance in a diary industry using artificial neural network technique. J Civil Environ Eng. 2016;6(6):254. doi: 10.4172/2165-784x.1000254.
2. Samandari M, Movahedian Attar H, Ebrahimpour K, Mohammadi F, Ghodsi S. Measurement of ampicillin and penicillin G antibiotics in wastewater treatment plants during the COVID-19 pandemic: a case study in Isfahan. Environ Health Eng Manag. 2022;9(3):201-11. doi: 10.34172/ehem.2022.21.
3. Akuma DA, Hundie KB, Bullo TA. Performance improvement of textile wastewater treatment plant design by STOAT model simulation. Environ Health Eng Manag. 2022;9(3):213-21. doi: 10.34172/ehem.2022.22.
4. Sol D, Laca A, Laca A, Díaz M. Microplastics in wastewater

and drinking water treatment plants: occurrence and removal of microfibres. Appl Sci. 2021;11(21):10109. doi: 10.3390/app112110109.
5. Djeddo m, Aouatef H, Loukam I. Wastewater Treatment Plant Performances Modelling Using Artificial Neural Networks 2018.
6. Maier HR, Dandy GC. Neural network based modelling of environmental variables: a systematic approach. Math Comput Model. 2001;33(6-7):669-82. doi: 10.1016/s0895-7177(00)00271-5.
7. Khodadadi M, Mesdaghinia A, Nasseri S, Ghaneian MT, Ehrampoush MH, Hadi M. Prediction of the waste stabilization pond performance using linear multiple regression and multi-layer perceptron neural network: a case study of Birjand, Iran. Environ Health Eng Manag. 2016;3(2):81-9. doi: 10.15171/ehemj.2016.05.
8. Baeza J, Gabriel D, Lafuente J. An expert supervisory system for a pilot WWTP. Environ Model Softw. 1999;14(5):383-90. doi: 10.1016/s1364-8152(98)00101-7.
9. Granata F, Papirio S, Esposito G, Gargano R, de Marinis G. Machine learning algorithms for the forecasting of wastewater quality indicators. Water. 2017;9(2):105. doi: 10.3390/w9020105.
10. Sànchez M, Cortés U, Lafuente J, Rodriguez-Roda I, Poch M. DAI-DEPUR: an integrated and distributed architecture for wastewater treatment plants supervision. Artif Intell Eng. 1996;10(3):275-85. doi: 10.1016/0954-1810(96)00004-0.
11. Sànchez M, Cortés U, Béjar J, de Gracia J, Lafuente J, Poch M. Concept formation in WWTP by means of classification techniques: a compared study. Appl Intell (Dordr). 1997;7(2):147-65. doi: 10.1023/A:1008202113300.
12. Atanasova N, Kompare B. Modelling of wastewater treatment plant with decision and regression trees. In: Proc. of the Workshop in Binding Environmental Sciences and Artificial Intelligence, ECAI; 2002; Lyon.
13. Tumer AE, Edebali S. An artificial neural network model for wastewater treatment plant of Konya. Int J Intell Syst Appl Eng. 2015;3(4):131-5. doi: 10.18201/ijisae.65358.
14. Hamada M, Adel Zaqoot H, Abu Jreiban A. Application of artificial neural networks for the prediction of Gaza wastewater treatment plant performance-Gaza strip. J Appl Res Water Wastewater. 2018;5(1):399-406. doi: 10.22126/arww.2018.874.
15. Al-Asheh S, Mjalli FS, Alfadala HE. Forecasting influent-effluent wastewater treatment plant using time series analysis and artificial neural network techniques. Chem Prod Process Model. 2007;2(3):1-23. doi: 10.2202/1934-2659.1063.
16. Abba SI, Elkiran G. Effluent prediction of chemical oxygen demand from the astewater treatment plant using artificial neural network application. Procedia Comput Sci. 2017;120:156-63. doi: 10.1016/j.procs.2017.11.223.
17. Ewaid SH, Abed SA, Kadhum SA. Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis. Environ Technol Innov. 2018;11:390-8. doi: 10.1016/j.eti.2018.06.013.
18. Baki OS, Aras E, Özel Akdemir Ü, Yılmaz BA. Biochemical oxygen demand prediction in wastewater treatment plant by using different regression analysis models. Desalin Water Treat. 2019;157:79-89. doi: 10.5004/dwt.2019.24158.
19. Faraji-Khiavi F, Jalilian H, Heydari S, Sadeghi R, Saduqi M, Razavinasab SA, et al. Utilization of health services among the elderly in Iran during the COVID-19 outbreak: a cross-sectional study. Health Sci Rep. 2022;5(5):e839. doi: 10.1002/hsr2.839.
20. Iqbal MA. Application of Regression Techniques with their Advantages and Disadvantages. Elektron Magazine; 2020. p. 11-7.
21. Mutombo NM, Numbi BP. Development of a linear regression model based on the most influential predictors for a research office cooling load. Energies. 2022;15(14):5097. doi: 10.3390/en15145097.
22. Gaya MS, Abba SI, Abdu AM, Tukur AI, Saleh MA, Esmaili P, et al. Estimation of water quality index using artificial intelligence approaches and multi-linear regression. IAES Int J Artif Intell. 2020;9(1):126-34. doi: 10.11591/ijai.v9.i1.pp126-134.
23. Dogan E, Ates A, Yilmaz EC, Eren B. Application of artificial neural networks to estimate wastewater treatment plant inlet biochemical oxygen demand. Environ Prog. 2008;27(4):439-46. doi: 10.1002/ep.10295.
24. Ghasemi J, Saaidpour S. Quantitative structure-property relationship study of n-octanol-water partition coefficients of some of diverse drugs using multiple linear regression. Anal Chim Acta. 2007;604(2):99-106. doi: 10.1016/j.aca.2007.10.004.
25. Chenini I, Khemiri S. Evaluation of ground water quality using multiple linear regression and structural equation modeling. Int J Environ Sci Technol. 2009;6(3):509-19. doi: 10.1007/bf03326090.
26. Saleem A, Dandigi MN, Vijay Kumar K. Correlation-regression model for physico-chemical quality of groundwater in the South Indian city of Gulbarga. Afr J Environ Sci Technol. 2012;6(9):353-64. doi: 10.5897/ajest12.047.
27. Heddam S, Lamda H, Filali S. Predicting effluent biochemical oxygen demand in a wastewater treatment plant using generalized regression neural network based approach: a comparative study. Environ Process. 2016;3(1):153-65. doi: 10.1007/s40710-016-0129-3.
28. Tümer AE, Edebalı S. Prediction of wastewater treatment plant performance using multilinear regression and artificial neural networks. In: 2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA). Madrid, Spain: IEEE; 2015. p. 1-5. doi: 10.1109/inista.2015.7276742.
29. Qiuhua L, Lihai S, Tingjing G, Lei Z, Teng O, Guojia H, et al. Use of principal component scores in multiple linear regression models for simulation of chlorophyll-a and phytoplankton abundance at a karst deep reservoir, southwest of China. Acta Ecol Sin. 2014;34(1):72-8. doi: 10.1016/j.chnaes.2013.11.009.
30. Uyak V, Ozdemir K, Toroz I. Multiple linear regression modeling of disinfection by-products formation in Istanbul drinking water reservoirs. Sci Total Environ. 2007;378(3):269-80. doi: 10.1016/j.scitotenv.2007.02.041.
31. Maniquiz MC, Lee S, Kim LH. Multiple linear regression models of urban runoff pollutant load and event mean concentration considering rainfall variables. J Environ Sci. 2010;22(6):946-52. doi: 10.1016/s1001-0742(09)60203-5.
32. Basant N, Gupta S, Malik A, Singh KP. Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water—a case study. Chemometr Intell Lab Syst. 2010;104(2):172-80. doi: 10.1016/j.chemolab.2010.08.005.

33. Zare Abyaneh H. Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. J Environ Health Sci Eng. 2014;12(1):40. doi: 10.1186/2052-336x-12-40.
34. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and Regression Trees. Boca Raton, FL: CRC Press; 1984.
35. Jiao SR, Song J, Liu B. A review of decision tree classification algorithms for continuous variables. J Phys Conf Ser. 2020;1651(1):012083. doi: 10.1088/1742-6596/1651/1/012083.
36. Cichosz P. Data Mining Algorithms: Explained Using R. United Kingdom: John Wiley & Sons; 2015.
37. Witten I, Frank E, Hall MA. Data Mining: Practical Machine Learning Tools and Techniques. USA: Morgan Kaufmann Publishers; 2011.
38. Suchetana B, Rajagopalan B, Silverstein J. Assessment of wastewater treatment facility compliance with decreasing ammonia discharge limits using a regression tree model. Sci Total Environ. 2017;598:249-57. doi: 10.1016/j.scitotenv.2017.03.236.
39. Bhattacharya B, Solomatine DP. Neural networks and M5 model trees in modelling water level–discharge relationship. Neurocomputing. 2005;63:381-96. doi: 10.1016/j.neucom.2004.04.016.
40. Najafzadeh M, Laucelli DB, Zahiri A. Application of model tree and evolutionary polynomial regression for evaluation of sediment transport in pipes. KSCE J Civ Eng. 2017;21(5):1956-63. doi: 10.1007/s12205-016-1784-7.
41. Solomatine DP, Xue Y. M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. J Hydrol Eng. 2004;9(6):491-501. doi: 10.1061/(asce)1084-0699(2004)9:6(491).
42. Singh KK, Pal M, Singh VP. Estimation of mean annual flood in indian catchments using backpropagation neural network and M5 model tree. Water Resour Manag. 2010;24(10):2007-19. doi: 10.1007/s11269-009-9535-x.
43. Etemad-Shahidi A, Ghaemi N. Model tree approach for prediction of pile groups scour due to waves. Ocean Eng. 2011;38(13):1522-7. doi: 10.1016/j.oceaneng.2011.07.012.
44. Ghaemi N, Etemad-Shahidi A, Ataie-Ashtiani B. Estimation of current-induced pile groups scour using a rule-based method. J Hydroinformatics. 2012;15(2):516-28. doi: 10.2166/hydro.2012.175.
45. Goyal MK. Modeling of sediment yield prediction using M5 model tree algorithm and wavelet regression. Water Resour Manag. 2014;28(7):1991-2003. doi: 10.1007/s11269-014-0590-6.
46. Vyas M, Modhera B, Vyas V, Sharma AK. Performance forecasting of common effluent treatment plant parameters by artificial neural network. ARPN J Eng Appl Sci. 2011;6(1):38-42.