

DOI: <http://dx.doi.org/10.21123/bsj.2022.6452>

Modeling Jar Test Results Using Gene Expression to Determine the Optimal Alum Dose in Drinking Water Treatment Plants

Ruba D. Alsaeed^{1*} 

Bassam Alaji¹ 

Mazen Ibrahim² 

¹Department of Sanitary and Environmental Engineering, Faculty of Civil Engineering, Damascus University, Damascus, Syria.

²Department of Engineering Management and Construction, Faculty of Civil Engineering, Damascus University, Damascus, Syria.

*Corresponding author: rureta89@yahoo.com

E-mails address: alajibassam@gmail.com , mazen7790@yahoo.com

Received 29/6/2021, Accepted 12/8/2021, Published Online First 20/3/2022, Published 1/10/2022



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Abstract:

Coagulation is the most important process in drinking water treatment. Alum coagulant increases the aluminum residuals, which have been linked in many studies to Alzheimer's disease. Therefore, it is very important to use it with the very optimal dose. In this paper, four sets of experiments were done to determine the relationship between raw water characteristics: turbidity, pH, alkalinity, temperature and optimum doses of alum $[Al_2(SO_4)_3 \cdot 14H_2O]$ to form a mathematical equation that could replace the need for jar test experiments. The experiments were performed under different conditions and under different seasonal circumstances. The optimal dose in every set was determined, and used to build a gene expression model (GEP). The models were constructed using data of the jar test experiments: turbidity, pH, alkalinity, and temperature, to predict the coagulant dose. The best GEP model gave very good results with a correlation coefficient (0.91) and a root mean square error of 1.8. Multi linear regression was used to be compared with the GEP results; it could not give good results due to the complex nonlinear relation of the process. Another round of experiments was done with high initial turbidity like the values that comes to the plant during floods and heavy rain. To give an equation for these extreme values, with studying the use of starch as a coagulant aid, the best GEP gave good results with a correlation coefficient of 0.92 and RMSE 5.1

Key words: Alum dose, Artificial intelligence, Coagulant, Gene expression, Multi linear regression, Turbidity.

Introduction:

The drinking water resources scarcity has been a serious issue for many decades¹. Surface water is one of the main drinking water sources, which is usually unsafe to use without a treatment².

There are different types of drinking water treatment plants, according to the characters of the resource water. The main objective of them all is to produce water that does not contain microorganisms or toxic compounds that are biologically and chemically safe for human consumption³.

Because of its ability to solubilize, pure water is not found in nature. It usually contains dissolved impurities like minerals, organic compounds and gases that alter the physical, chemical and biological characteristics of water⁴.

Biological, organic, and inorganic materials may be found in water. The majority of impurities

are colloids. The colloids could be classified as hydrophilic such as soap soluble starch, synthetic detergents, and hydrophobic such as clay particles, metal oxides. A hypothetical molecular structure of humic acid, is shown in Fig. 1. The behavior of colloids is controlled by principal phenomena like electrostatic forces, van der Waals forces, and Brownian motion. Most of these collides have a negative charge and their colloidal dispersions are stabilized due to electrostatic repulsion, which prevents particle aggregation and overcomes Vander Waals' forces⁵. These collides could not be removed by normal filtration or precipitation processes.

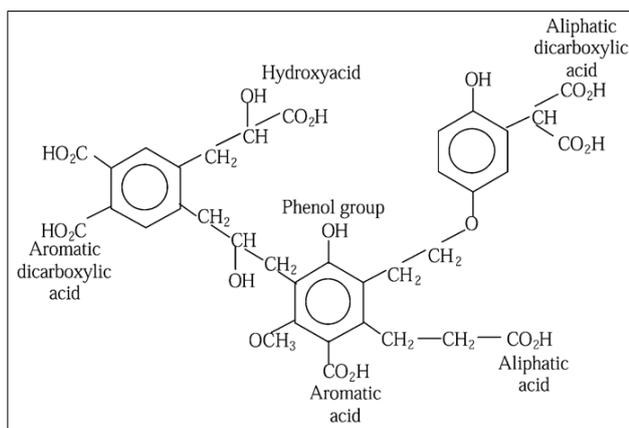


Figure 1. Hypothetical molecular structure of humic acid, showing important functional groups⁶

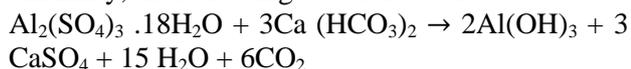
To remove these impurities from water, conventional treatment is frequently used. It is a combination of the following steps: Coagulation, Flocculation, Sedimentation and Filtration.

Coagulation and flocculation are the main and important processes of water purification. Coagulation is the process of increasing the sizes of particles by using materials called coagulants that can standardize these particles and help to make them bigger and more able to be settled⁷.

The history of coagulation is nearly three thousand years old. There are so many different coagulants in which alum is the most used. Alum is a coagulant that is widely used in water treatment plants to remove turbidity and reduce natural organic matter.

Earlier research suggested several mechanisms for particles destabilization of aluminum during coagulation. These mechanisms are double layer compression, neutralization, adsorption, and sweep flocculation. Due to the complexity of the coagulation process, these mechanisms may exist either by themselves or they may exist in combination⁸.

When alum is added to water containing alkalinity, the following reaction occurs:



After adding alum to the water, it dissociates immediately, resulting in aluminum ion surrounded by six water molecules. Then the aluminum ion starts reacting with the water, forming large complexes.

Regardless of the species formed, the complexes are massive precipitates that enmeshed numerous colloids, removing them through entanglement⁹

The turbidity levels vary during the year, as it may increase in the wet seasons, due to the heavy rain and soil erosion^{10,11}. This means the need for high levels of alum, but the excess use of aluminum sulphate "alum" can leave high aluminum residuals concentrations in the treated water. Such residuals should be controlled and minimized because they can cause problems in the distribution system. High aluminum concentrations in treated water are associated with many problems, like the formation of aluminum precipitates that lead to increasing turbidity. In addition, it has been suggested as one of the main factors in Alzheimer's disease¹.

Modeling water treatment plants is very important in the processing, designing and operation time of the treatment units. Models are valuable tools. Some plants introduce a fully automated operation. In these types of plants, good predictive models are needed that can provide online data about water quality, chemical dose, and filter operation time⁵.

In recent years, modelling and optimization have become increasingly important in most fields. Optimization helps with gaining a better understanding of the system. Earlier, traditional modelling has been used to describe biological processes; it has been made by giving equations related to the speed growth of microorganisms, the consumption of substrate, and product formation. Such models have many limitations according to the fact that reactions are non-linear and time dependent¹².

Modelling complex processes such as those in water treatment plants is not an easy thing, due to the non-linear processes happening.¹³

Recently too many efforts in modeling real-world problems have been made by using the artificial intelligence (AI) tools available. One of the most promising tools is gene expression programming, which is still young in environmental science applications. It could provide a nonlinear equation, that is good accurate, and easy to be used for giving the optimal dose of alum. Previous studies^{14,15,16,1,12} used AI techniques in drinking water modelling are described in Table.1 with the method used in modelling and the predicted values (the output of the Models).

In this study, a gene expression method was used, to model the optimal dose of alum in drinking water treatment plants. By modelling the results of different sets of experiments, studying the effect of the following parameters on the optimal dose, Turbidity, pH, the temperature and the alkalinity.

Table1. Previous studies used AI techniques in drinking water modelling

References	Method	Predicted Value	Parameters used
Heddami 2021 ¹⁴	ERT	coagulant dosage	Turbidity, pH, dissolved oxygen, electrical conductivity, and the water temperature
Heddami et al., 2011 ¹⁵	GRNN	coagulant dosage	TU, CO , TE , DO, UV , pH
A. J. León-Luque, C. L. Barajas.2018. ¹⁶	ANN	coagulant dosage	pH, Temperature, Alkalinity Turbidity, Conductivity
Kim, Chan Moon, and Manukid Parnichkun.2017. ¹²	Neuro-fuzzy	Turbidity coagulant dosage	Turbidity Temperature pH Conductivity
Haghiri, S et al . 2018. ¹	ANN	coagulant dosage	pH, Temperature, Alkalinity Turbidity

Materials and Methods:

Data Used

Research had took place in the water purification plant in Qusayr , Homs. The plant has a capacity of

6000 m³ per day. The statistical characteristics of the water in the plant during the study is describes in Table. 2.

Table2. Statistical characteristics of water samples

Parameter	Tur- in mg/L	Alkaliniy	pH	T °C
Min	7.89	135	6.55	8.1
Max	56.35	195	8.53	21.4
Mean	17.06	162.24	7.4	15.5
Std. Deviation	8.2	12.78	0.3	2.7

The raw water turbidity varied between 6.89 and 60.35 during the time of the study, it is described in Fig.2

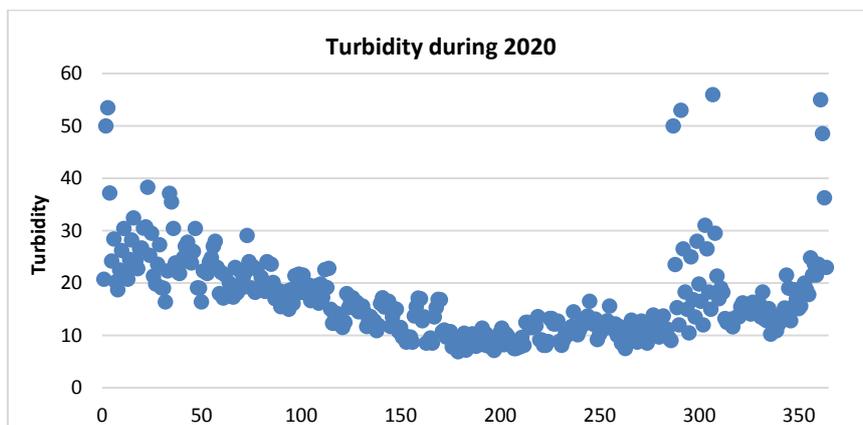


Figure 2. Turbidity interring the plant variation during the time of the study

As it is shown in Fig.2, the turbidity raises to about 40 NTU during the rainy seasons, with some extreme values near 60 NTU, and it is between 5 and 20 during the other seasons.

Jar Test Experiments:

A series of chemical experimental has been conducted using jar test experiments. The study focused on four aspects, the difference in the raw water turbidity, pH, temperature and alkalinity on the suitable alum dose. The next steps describe the way that each experiment had been conducted:

1- First of all, a Jar Test was performed for each initial turbidity level, (from 5 NTU to 60NTU ± 3) with a step of 5 NTU every time; 65 experiments were carried out overall. The alum dose was

changed between 5 and 30 mg/L. The turbidity values were as actual values as possible, and not a synthetic sample, because turbidity have so many different resources, and to avoid the mistake that would happen because of the wrong samples.

One level of turbidity does not occur during the time of the study, so synthetic samples were made, using kaolinite clay $Al_2Si_2O_5(OH)_4$.

At this part of the experiments, the pH was about 7.5 ± 0.4 , $T = 20 \pm 2^\circ C$, alkalinity = 140 ± 5 .

Fig. 3 shows the jar test apparatus used in the study.



Figure 3. Jar test used for the experiments

2- Rapid mixing was carried out at 150 rpm during two minute intervals to homogenize the distribution of coagulant. The slow mixing was done at a speed of 30 rpm during 10 minutes' intervals to ensure the joint of the small flocs formed in the coagulation, which were allowed to sediment for 30 minutes.
3-The second set of experiments, was to evaluate the effect of pH differences on each dose. Turbidity values were (10, 20, 30, and 40 ± 3), and the dose varied from 5 to 20 for the initial turbidity 10 , and from 10 to 30 with a step of 10 mg/L for other

turbidity values . pH was changed between 6 to 8.5 for each set of data for turbidity and alum, to study the effect of pH on changing the suitable dose. Then step was 2 repeated.

4-The temperature in the study area differs from one month to another during the year. Therefore, a set of experiments were studied for determining the optimal dose in each specific temperature.

This collection of experiments were done at different times of the year, the temperature of raw water was (9, 14, 18, 23 ±1), initial turbidity were (10, 20, 30, 40 ± 3) and the dose changed between 5 to 30 mg/L with a step of 5 mg/L. Then step 2 was repeated.

5-To evaluate the effect of the alkalinity changes, three different alkalinity levels were studied, 140 , 160 , 180 ± 5 , at six levels of turbidity (10 , 20, 30, 40, 50,60) , pH and temperature were constant, with different doses of alum. Then step 2 was repeated. Table.3 summarizes the steps of the experiments.

Table 3. The four sets of experiments details.

	Turbidity	Alum dose	pH	T	Alkalinity
Set 1	10 , 20 , 30, 40,50,60	5 ,10,15,20,25,30	7.5 ± 0.4	20 ±2	140 ± 5
Set 2	10, 20 ,30 ,40	5 ,10,15,20,25,30	6.5,7,7.5, 8,8.5	20 ±2	140 ± 5
Set 3	10, 20 ,30 ,40	5 ,10,15,20,25,30	7.5 ± 0.4	9 ,14,18 ,23 ± 1	140 ± 5
Set 4	10, 20 ,30 ,40 , 50 , 60	Optimal dose from set1	7.5 ± 0.4	20 ± 2	140 , 160 ,180 ± 5

After finishing the Jar Tests for each run in each set, the parameters for coagulation and flocculation were measured again (pH, turbidity, and aluminum residual) in order to determine the coagulant dosage that caused the highest percentage of turbidity removal, the optimal dose.

Multi Linear Regression: known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. MLR models are suitable to simulate relationship between many variables in different fields, based on least -squares fitting. It can determine the relationship between two or more parameters. As expressed in the following equation, where x is multiple components, y is the output

$$Y = a_0 + \sum_0^i a \times X_i$$

Y: dependent variable, X_i : explanatory variables, a_0 : constant , a :slope coefficient for each explanatory variable

Gene Expression: It is one of the AI models; GEP is a type of genetic algorithm that was first proposed in response to Darwin's theory of evolution. Which was proposed by Ferreira et al in 2001. It operates in the same way that a group abandons undesirable members and creates genetically engineered offspring in evolution. At the start of this method's process, no practical relationship is taken into account¹⁵.

GP's basic search strategy is a genetic algorithm, but it differs from standard GA in that it usually works with parse trees rather than bit strings. A terminal set (the problem's variables) and a function set are used to build a parse. Moreover, the GEP can solve problems in different fields with high performance. This method has been applied recently to recognize the manner of nonlinear systems. The first step of GEP operation, is determining the fitness function, which can be determined mathematically as the following equation:

$$F_i = \sum_{j=1}^{C_i} (M - |C_{i,j} - T_j|)$$

Hence, M is the range of selection, $C_{i,j}$ is the value returned by the individual chromosome i for fitness case j (out of C_t fitness cases), T_j is the target value for fitness case j¹⁰

Figure 4 shows the expression tree (ET) for a mathematical expression of the following example $(a + a \times b) - (a + \sqrt{b})$.

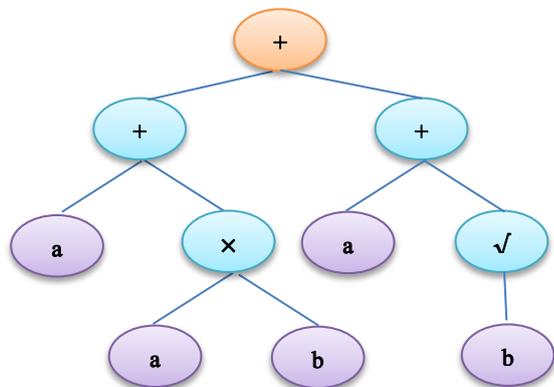


Figure 4. Expression tree diagram of the example

Fig.5 shows the steps of GEP models, its steps, first the randomly creation the initial population generation. Then, the chromosomes are expressed and excluded the tree expression for fitness evaluation. The individual is then selected according to their fitness to reproduce with the modification; these individuals are subject to the same development. This process is going in a repetition loop several times until a good solution is found.

Results and Discussions:

I-1-First set of experiments, the turbidity changed between 5 and 60 \mp 3, and the alum dose was between 5 and 30, pH was set about 7.5 \mp 0.2, and temperature at 20 \mp 0.2, Alkalinity 140 \mp 3.

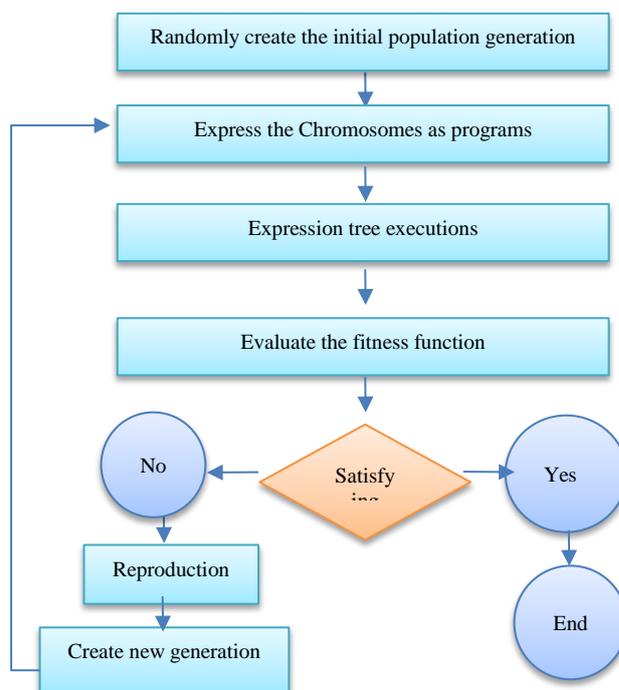


Figure 5. Flowchart of Gene Expression Programming.

After the rapid mix, slow mix, and precipitation. At the turbidity 30 NTU, both doses 15 and 20 mg/L were able to give acceptable final turbidity, but when the dose increased to 25 mg/L, the residual increased. It is obvious from results of the experiments that the percentage of turbidity removal was increasing with the increase of the alum dose until a certain value. Then it becomes useless to continue increasing the dose. Moreover, the percentage of removing decreased. Because the overdose causes charge reversal "destabilization". To form a precipitate that will attract suspensions at low concentrations of turbidity, an excessive amount of coagulant must be added, whereas at higher concentrations, coagulation occurs at a lower chemical dose because the suspensions provide the nuclei that aid sedimentation. The results are illustrated in Fig. 6.

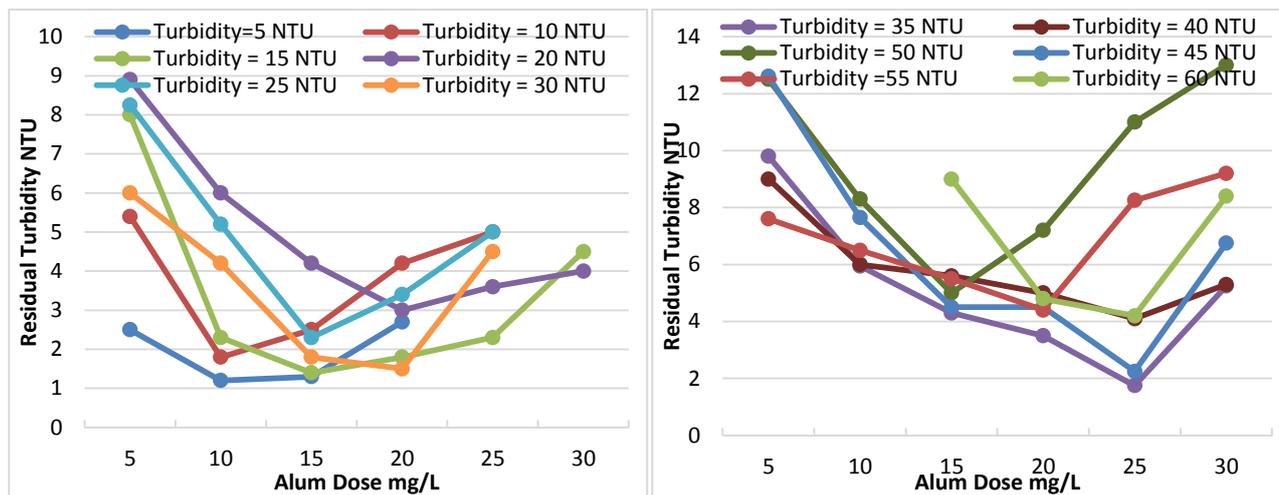


Figure 6. Residual turbidity for different alum dose and different initial turbidity

2- At the second set of experiments, turbidity was set at (20 , 30 , 40) \pm 2 NTU , and pH was changed between 6-8.5 for 6 different doses of

Alum (5, 10, 15, 20, 25, 30) mg/L, Alkalinity 140 \pm 3 , and the temperature at 20 \pm 2 . The results are listed in Table.4, and Fig7.

Table4. Turbidity removal at different alum doses

tur	Alum dose mg/L	pH					
		6	6.5	7	7.5	8	8.5
20	5	4.1	4.8	5.1	7	9	13
	10	3.5	4.2	5	5.4	8.8	12.3
	15	3	3.45	4	5	5.6	8
	20	1.9	2.2	2.5	3.1	5.6	7.5
	25	3	3.5	3.8	4.5	5.1	6.8
	30	3.1	3.6	4	5	5.6	8
30	5	5	5.03	5.51	7	9.23	16.5
	10	3.5	4.2	5.41	5.4	8.8	12.53
	15	3.41	3.4	4.23	4.9	5.2	9
	20	2.31	2.3	2.5	3.1	5.1	7.91
	25	3.1	3.5	3.8	4.5	5.51	7.3
	30	3.2	3.6	4.1	5	5.7	8.23
40	5	3.8	4.3	5.1	9	11	14
	10	3.5	4	4.8	5.4	8.8	13
	15	2.8	3.2	4	5	5.6	9
	20	2	2.9	3	4	5	8.5
	25	2.8	3.1	3.8	4.5	5.1	8
	30	3.55	3.6	4	5	5.5	9

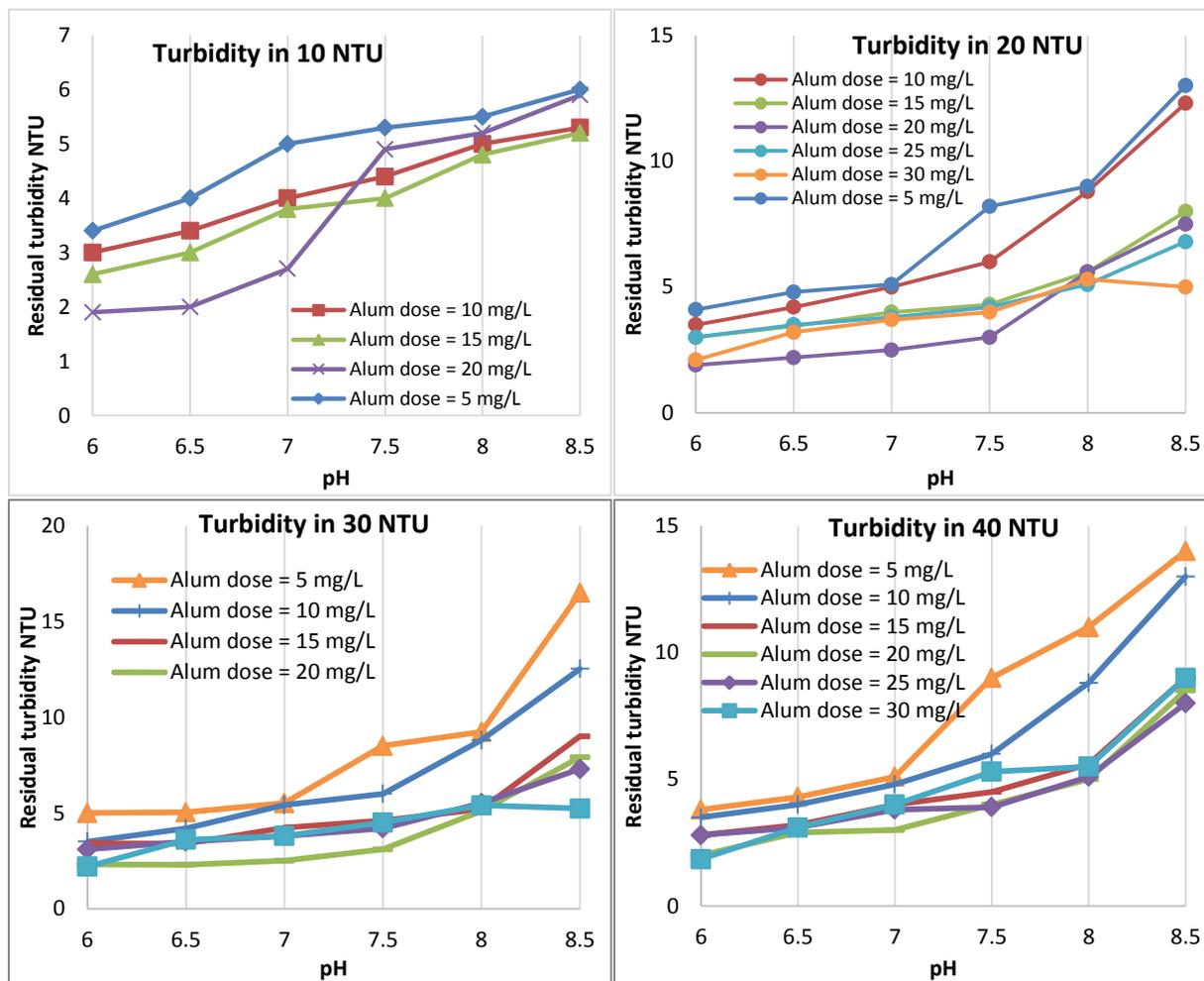


Figure 7. Residual turbidity for different pH level at different initial turbidity

At set 1 the optimal dose for initial turbidity 30 was 15 mg/L, whereas, with the drop of pH into 6.5 in set 2, the required dose for reaching an acceptable level of residual dropped to 10 mg/L. Likewise, for initial turbidity 10 NTU, when pH changed from 6 to 8 for the same dose, the effectivity varied from 66% to 32%.

When the pH value of the treated water is reduced, the coagulation efficiency becomes greater and the residual turbidity is lower because the coagulant pH value decreases, and this causes a lower required dose at many cases. At higher pH, the optimum alum dosage increases due to the decreased positive charge of the adsorbed species.

3- The third set of experiments, were to study the effect of the temperature, as temperature is an important parameter for water treatment. At this set of experiments, pH, and alkalinity were constant as possible at, 7.5 ± 0.4 , 140 ± 5 . The experiments were done at different times of the year, and the temperature were $9, 14, 18, 23 \pm 1$, studying the doses (5, 10, 15, 20, 25, 30) mg/L. The results are described in both Table.5 and Fig. 8.

Table5. Turbidity removal at different alum doses

Turbidity	T	5	10	15	20	25	30
10	9	9.3	7.6	5.73	5.3	4.3	3.23
	14	8.6	5.3	4.3	4.1	3.62	2.8
	18	7.14	4.18	3.8	2.72	1.82	1.4
	23	6.5	3.9	2.41	1.63	1.2	0.9
20	9	10	7	5.3	4.8	3.62	2.8
	14	8	4.3	3.8	2.1	1.82	1
	18	7.5	5.5	4.9	4.3	3.2	1.2
	23	5.8	3.72	1.5	0.89	0.84	0.72
30	9	11	10	8.9	7.6	5	9
	14	13	10	8.3	5	6.62	5.8
	18	10	7.3	5.3	2	1.5	1
	23	8	6.5	5	2.3	1.5	0.8
40	9	17	14	11.9	10.6	6	5.6
	14	15	12	10	5.3	4.9	8.8
	18	12	10.3	8.3	5	4.5	4
	23	8	5.5	5.3	4.5	3.8	4

The optimal dose has increased with the temperature decrease, which means they need more consumption of alum during the cold seasons. The optimal dose raised for raw water turbidity 10 NTU

from 10 to 20 when the temperature dropped from 23 to 9 and from 15 to 25 when the temperature for the raw water was 40 NTU, the alum dose raised about 0.7 mg/L for each 1 °C drop in the temperature. Low temperature affects coagulation and flocculation processes by altering coagulant

solubility, increasing water viscosity and retarding the kinetics of hydrolysis reactions. Higher coagulant dosage, the addition of flocculation aids, longer flocculation and sedimentation times; are required at lower temperatures.

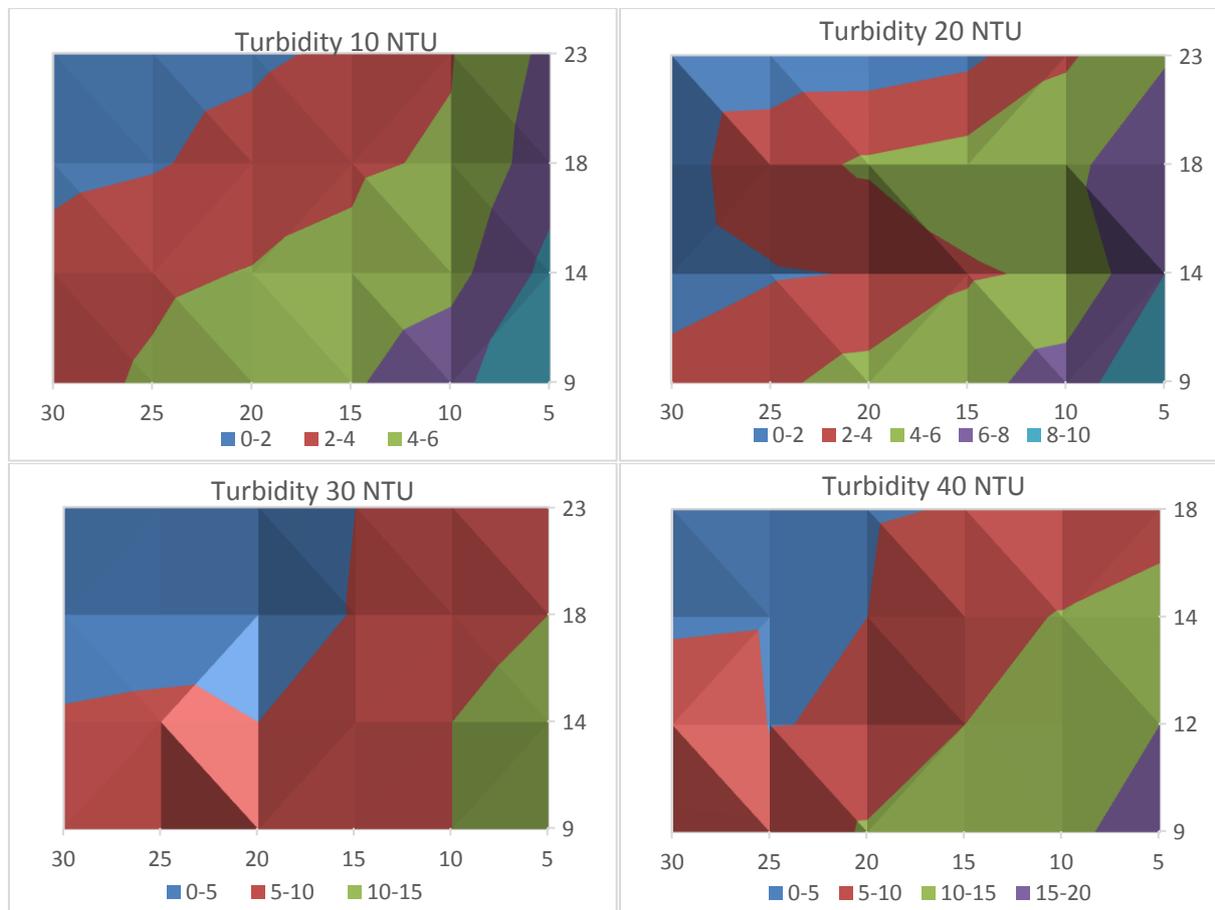


Figure 8. Residual turbidity for different alum dose at different temperatures

4- The fourth set of experiments were to study the optimal dose at different alkalinity levels, 140,160,180. pH, temperature were constant at, 7.5

$\pm 0.4, 20 \pm 2^\circ\text{C}$. For initial turbidity 10 ,20, 30, 40, 50,60 NTU , for the optimal dose taken from set1. And the percentage of removing is shown in Fig. 9.

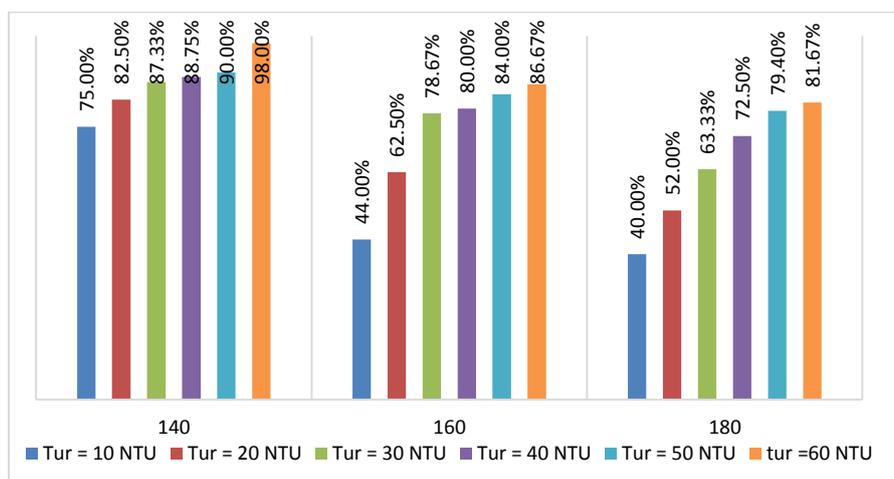


Figure 9. Turbidity removing percentage for different alkalinity levels

The percentage of removing turbidity had dropped from 75% to 40 % when the alkalinity increased from 140 to 180 for the initial turbidity 10 NTU. The same thing happened for the other initial turbidity levels, with different percentages. Low

water turbidity was more affected by the difference of alkalinity than the high ones.

The optimal dose from each run in every set was used to build up a model; two methods were used m MLR and GEP.

The data used for the model shown in Fig.10

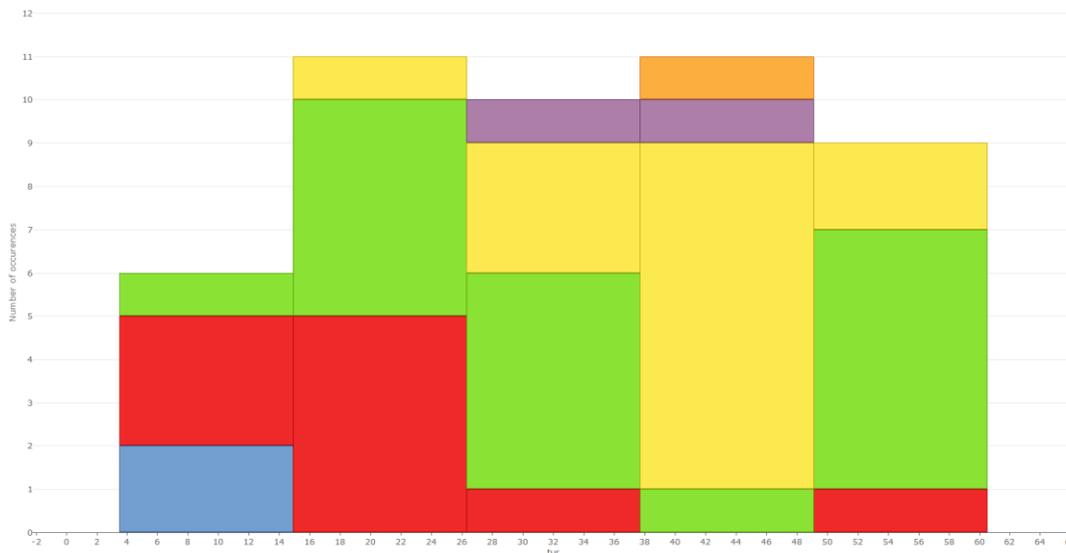


Figure 10. Optimal alum dose frequency for each initial turbidity used at the four sets of experiments

The performance of various models was evaluated using the statistical indices:

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (T_i - O_i)^2}{n}}$$

- Correlation Coefficient: R

$$R = \frac{\sum_{i=1}^n (P_{obs} - \overline{P_{obs}})(P_{pre} - \overline{P_{pre}})}{\sqrt{\sum_{i=1}^n (P_{obs} - \overline{P_{obs}})^2 \times \sum_{i=1}^n (P_{pre} - \overline{P_{pre}})^2}}$$

Multi Linear Regression Model (MLR):

MLR model was built with the use of SPSS 25 program.

Dependent variable: alum dose. Independent variable: Turbidity, alkalinity, pH, Temperature. Regression method: Enter. The developed MLR model was as the equation shown next .

$$\text{Alum dose} = -35.584 + 0.186 * \text{Turbidity} + 3.881 * \text{pH} - 0.555 * \text{T} + 0.192 * \text{Alkalinity}$$

$$R = 0.802, R \text{ square} = 0.643$$

The errors histogram is shown in Fig.11.

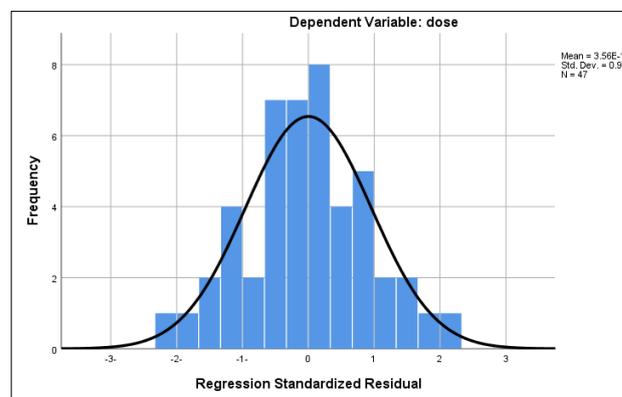


Figure 11. The errors frequency of the MLR model

MLR model was good; but not satisfying enough. Due to the complex non – liner relations.

Gene Expression Models:

First, 70% of the data were used to train the model. The remaining data were used for validation. Fitness function used to evaluate was RMSE.

There are various of parameters related to the Gene expression models; the most important ones are the Number of chromosomes, mutation and functions.

In this study two different numbers of chromosomes, and mutation rates were used.

The Values of GEP Parameters are shown in Table. 6.

Table 6. Values of GEP Control Parameters

	GEP 1-1	GEP 1-2	GEP 2-1	GEP 2-2
Function set	+, -, ×, /, √, exp, ln, log, 10 [^]	+, -, ×, /, √, exp, ln, 10 [^] , 3RT	+, -, ×, /, √, exp, ln, log, 10 [^]	+, -, ×, /, √, exp, ln, 10 [^] , 3RT
Number of chromosomes	600	600	700	700
Constants per gene	10	10	10	10
Function Fitness	RMSE	RMSE	RMSE	RMSE
Linking Function	+	+	+	+
mutation	0.00138	0.01	0.00138	0.01
Head size	8	7	8	7
Tail size				

GEP 2-2 gave the best results, as the increase in the number of chromosomes helps to improve the model most of the times, mutation rate was set as 0.00138, increasing it does not give a better solution in this case.

The same functions, function fitness, and linking function were used in the four models.

The results for the best model are shown in Table.7.

Table 7. The accuracy table of gene-expression models.

The number of the Model	Inputs used		RMSE	R
1	Turbidity, Temperature, pH, alkalinity	Train	1.79	0.91
		Test	1.8	0.9

RMSE in training was higher than in the testing phase.

RMSE, RAE, MAE, and RRSE values were greater in training relative to testing. As the error increases, the performance of the model declines.

3 sub gene-expression trees were developed for models 2-2. The sub gene-expression trees, are shown in Fig.12

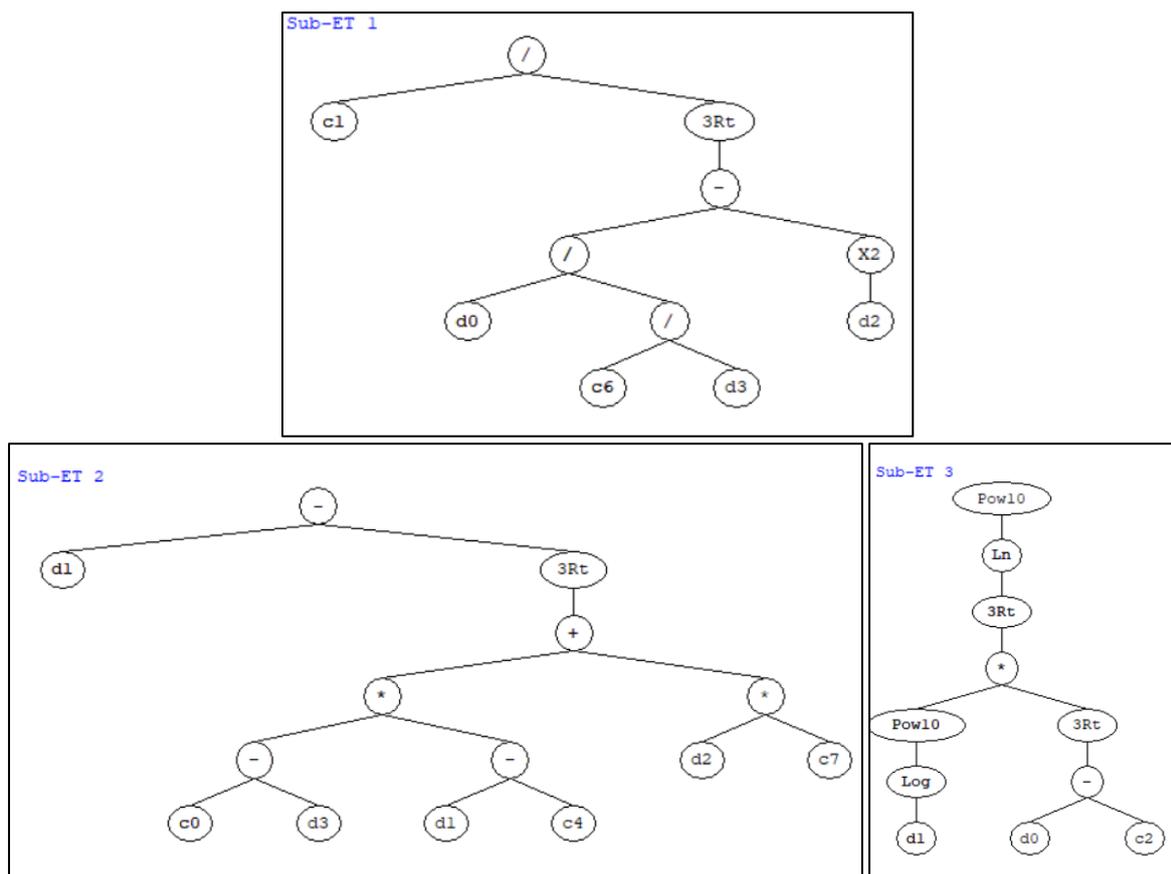


Figure 12. sub gene-expression tree for model 2-2

The constants and the parameter used in the equation showed in Fig. 12 are listed in Table.8.

Table 8. Values of parameters used in the equations

d_0	d_1	d_2	d_3
Turbidity	pH	T	Alkalinity
G1C1		6.9373149815363	
G1C6		7.08694101217425	
G2C7		7.16874839723298	
G2C0		92.054081850715	
G2C4		5.18388183250294	
G3C2		4.99437995203283	

Figure 13 represents a Comparing chart of part of the data with the results gained from the 2 models, MLR and GEP, as shown the GEP model was way ahead, and it can be used as a trusted modeling method in this case.

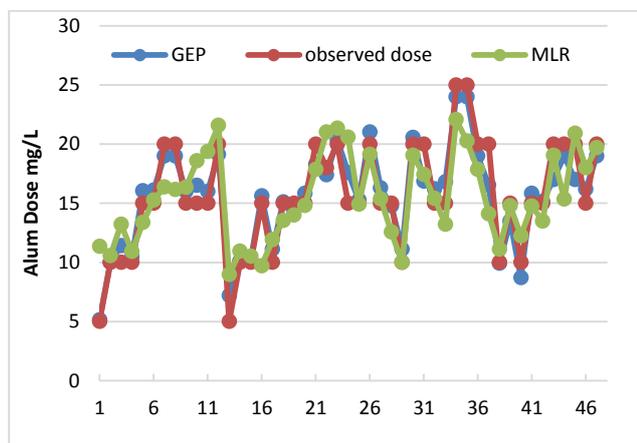


Figure 13. Comparing the observed data with the results of the two models

II- Another model was built for the high concentrations during fluids or other sudden circumstances; the same sets of experiments were conducted. First, different doses were used at different turbidity levels; pH was 7.5 and T 20°C . The results are represented in the following Table.9.

Table 9. Residual turgidity at different alum levels for different initial turbidity

tur	dose	tur out NTU	Removal percentage
	10	15	0.8
75	20	4.5	0.94
	30	10.5	0.86
	40	18	0.76
	10	40	0.6
100	20	30	0.7
	30	10	0.9
	40	4	0.96
	10	46.25	0.63
125	20	31.25	0.75
	30	17.5	0.8
	40	3.75	0.97
	50	5	0.96
	10	30	0.8
150	20	21	0.86
	30	9	0.94
	40	7.5	0.95
	50	3	0.98
	20	26.25	0.85
175	30	17.5	0.9
	40	8.75	0.95
	50	3.5	0.98
	20	30	0.7
200	30	14	0.8
	40	4.9	0.98
	50	2	0.99
	20	38.25	0.83
225	30	31.5	0.9
	40	11.25	0.9
	50	2.25	0.99
	60	4.5	0.98
250	20	25	0.9
	30	22.5	0.91
	40	12.5	0.95
	50	5	0.98
	60	7.5	0.98

Then the pH was adjusted between 6.5 and 8.5. The results are represented in Fig.14 pH has more effect on the optimal dose for the low concentrations than the high ones.

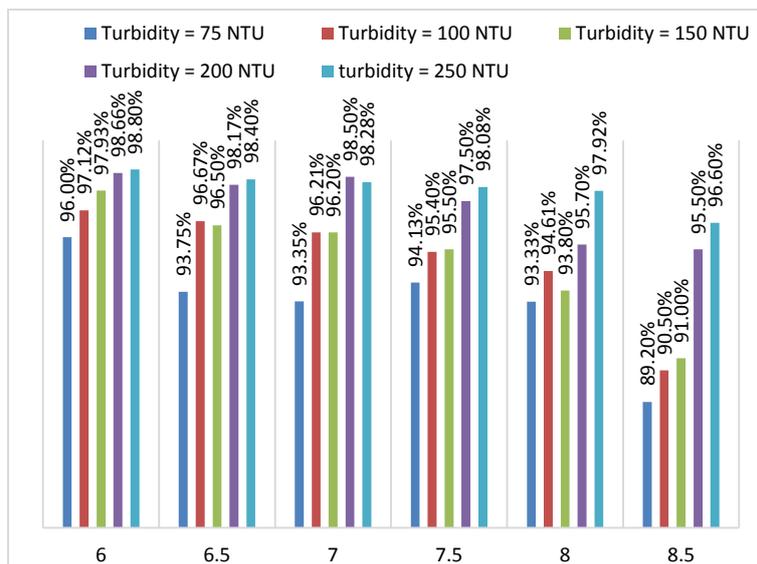


Figure 14. Turbidity removal for different pH levels at high turbidity concentrations

When the pH changed from 6.5 to 8.5. The removal of the optimal dose gained from set 1, changed from 96 to 89.2, for the initial turbidity 75 NTU; and from 98.8 to 96.6 % for the initial turbidity 200 NTU.

The last factor tested was the temperature; it was obvious that the temperature has a significant effect on the optimal dose. The effectivity of the same dose, dropped between 5 to 10 % when the temperature dropped about 15°C. The differences in the affectivity of the optimal dose for each initial turbidity, in different temperatures is shown in Fig.15.

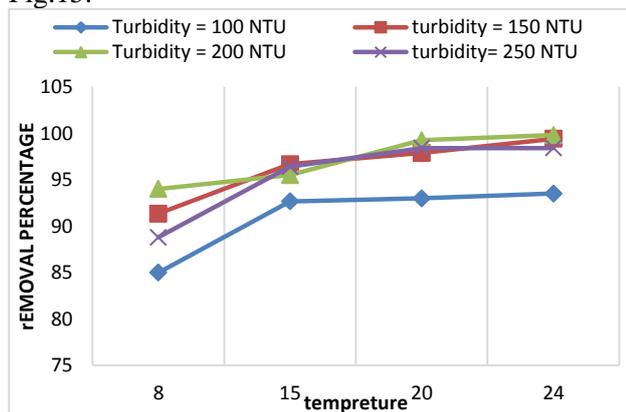


Figure 15. Turbidity removal at different temperatures for high turbidity concentrations

An addition set of experiments was added, to determine the required dose of starch when used as a coagulant aid, so the optimal dose and a much lower dose were tested for initial turbidity (100,

150, 200, 250) NTU, with adding starch doses (0.1 , 0.2, 0.3, 0.4) mg/L.

As a result of the experiments, it was obvious that an appropriate dose of starch can increase the required alum dose, and raises the affectivity of the optimal dose. The residual turbidity and the affectivity of removal are shown in Fig.16.

It was observed that the higher the starch dose is, the higher COD levels in the final water, so it is not recommended to use much higher concentrations.

-A nonlinear model was achieved using GEP with the same method used for the low turbidity levels, and the gained model were as the following equation:

$$\text{Alum Dose} = \frac{\text{Turbidity}}{T - 0.315147 \times \text{Turbidity}} + 8.10695 \times \text{pH} - 22.901728 - \text{Str} \times \text{pH}^2 - 3.94206 \times \text{Str} \times \text{pH} + \frac{1}{7} \times (8.2039 \times \text{pH} + 0.88341 \times \text{Turbidity} + 15.72247)$$

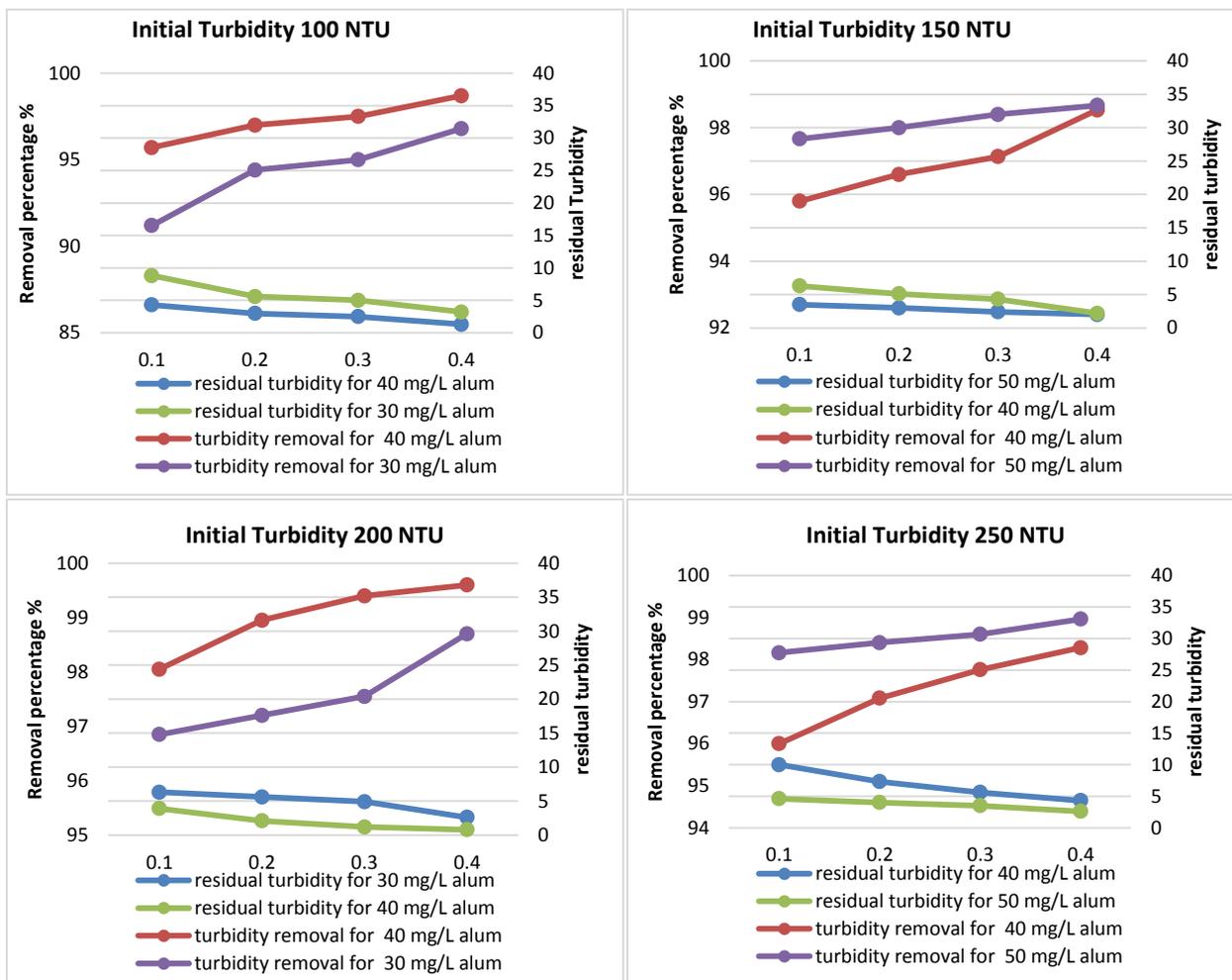


Figure 16. Residual Turbidity when using different starch doses as coagulant aid

And for the ease of use from the operators, the equation was made as a GUI inference was made, so the use of the equation becomes much easier. The inference is shown in Fig.17.

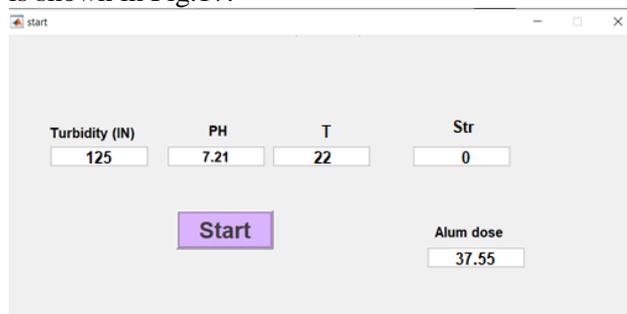


Figure 17. GUI inference for modeling the Alum dose at high initial turbidity

Conclusion:

Modelling the drinking water treatment plants is a very complicated process, and the most important thing is to determine the optimal dose of alum, when it is used as a coagulant, because it causes a lot of problems when it is not optimal. Like the reduction of pH of the water and the increasing of aluminum residuals.

Many articles have studied different modelling methods to predict coagulant dose, using artificial intelligence models, like ANN, and Fuzzy logic, but there is no specific equation to determine the alum dose, which would be much easier to use from the operators of the plants.

GEP is a promising tool for modeling treatment plants, and it is being used in recent studies in the environmental sciences. GEP gave a good predicting equation, that is flexible and able to model such complex relations with an accuracy of $R = 0.91$. The most affecting parameter on the GEP model, initial turbidity and temperature, with importance of 44.35 % and 28.03 %.

MLR could not give a good result. Whereas; GEP was a good and reliable method to determine the alum dose needed, in different circumstances.

At the rainy seasons, high turbidity levels occur, which is an extreme level, so it had to be studied individually. In order to build a specific model, another set of experiments were done and a good resulted model was gained using GEP, with the accuracy of $R=0.9$, $RMSE= 5.2$.

The study of using starch was used, to avoid consuming high concentrations of alum, it gave

good results in reducing the needed dose, but it should be noticed that it affect the COD continent in the treated water.

Authors' declaration:

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Damascus.

Authors' contributions statement:

R.A. developed the theory, did the experiments and analyzed the results.

B.A. and M.I. supervised and revised the findings of this work. All authors discussed the results and contributed to the final manuscript.

References:

1. Haghiri S, Daghighi A, Moharramzadeh S. Optimum coagulant forecasting by modeling jar test experiments using ANNs. *Drink. Water Eng Sci.* 2018; 11(1): 1-8. <http://dx.doi.org/10.5194/dwes-11-1-2018>
2. El-Chaghaby G, Rashad S, Moneem M A. Seasonal variation and correlation between the physical, chemical and microbiological parameters of Nile water in selected area in Egypt (Case study): physical, chemical and microbiological parameters of Nile water. *Baghdad Sci J*, 2020; 17(4): 1160-1160.
3. Song C, Zhang H. Study on turbidity prediction method of reservoirs based on long short term. *Ecol Modell.* 2020; 432. <https://doi.org/10.1016/j.ecolmodel.2020.109210>
4. Saritha V, Srinivas N, Srikanth Vuppala NV. Analysis and optimization of coagulation and flocculation process. *Appl Water Sci.* 2017;7(1): 451-460.
5. Al-Baidhani JH, Alameedee MA. Prediction of water treatment plant outlet turbidity using artificial neural network. *Int J Curr Eng Sci.* 2017; 7(4): 1559-1565.
6. Lanciné G D, Bamory K, Raymond L, Jean-Luc S, Christelle B, Jean B. Coagulation-Flocculation treatment of a tropical surface water with alum for dissolved organic matter (DOM) removal: Influence of alum dose and pH adjustment. *J Int Environ Appl Sci.* 2008; 3(4): 247-257.
7. Alsaeed R, Alaji B, Ebrahim M: Predicting turbidity and Aluminum in drinking water treatment plants using Hybrid Network (GA- ANN) and GEP, *Drink. Water Eng Sci Discuss.* in review, 2021:1-17. <https://doi.org/10.5194/dwes-2021-8>.
8. Dentel, Steven K., and James M. Gossett. Mechanisms of coagulation with aluminum salts. *J Am Water Works Ass.* 1988;80(4): 187-198.
9. Mackenzie L Davis. *Water and wastewater engineering: design principles and practice.* McGraw-Hill Education, 2010: 6-15.
10. Hassan F M, Naji H F, Al-Azaway E S. The study of some physical and chemical characteristics in drinking water treatment plant of Jurf Al-Sakar Subdestric in Babylon Governorate, Iraq. *Baghdad Sci J.* 2007; 4(3): 338-343.
11. Al-Kenzawi M A, Al-Haidary M J, Talib A H, Karomi M F. Environmental Study of Some Water Characteristics at Um-Al-Naaj Marsh, South of Iraq. *Baghdad Sci. J.* 2011; 8(1): 531-538.
12. Kim C M, Parnichkun M. Prediction of settled water turbidity and optimal coagulant dosage in drinking water treatment plant using a hybrid model of k-means clustering and adaptive neuro-fuzzy inference system. *Appl Water Sci.* 2020; 7(7): 3885-3902 . DOI 10.1007/s13201-017-0541-5 .
13. Abba S I, Rabiou Aliyu Abdulkadir, Gaya M S, Saleh M, Parvaneh Esmaili, Mustapha Bala Jibril. Neuro-fuzzy ensemble techniques for the prediction of turbidity in water treatment plant. 2nd International Conference of the IEEE Nigeria Computer Chapter. (Nigeria Comput Conf).IEEE.2019:1-6. DOI:10.1109/NigeriaComputConf45974.2019.8949629
14. Heddad S. Extremely Randomized Tree: A New Machines Learning Method for predicting Coagulant Dosage in Drinking Water Treatment Plant. *Water Engineering Modelling and Mathematic Tools.* Chennai: Elsevier, 2021: 475-490. <https://doi.org/10.1016/B978-0-12-820644-7.00013-X>.
15. Heddad S, Bermad A, Dechemi N. Applications of Radial Basis Function and Generalized Regression Neural Networks for Modelling of Coagulant dosage in a Drinking Water Treatment: Comparative Study. *ASCE J Environ Eng.* 2011; 137(12):1209-1214. DOI: 10.1061/(ASCE)EE.1943-7870.0000435
16. León-Luque A J, Barajas C L, Peña-Guzmán C A. Determination of the optimal dosage of aluminum sulfate in the coagulation-flocculation process using an artificial neural network. *Int J Environ Sci. Dev.* 2016; 7(5): 346.

تقدير الجرعة الأمثل من مخثر الشبة في محطات تنقية مياه الشرب باستخدام برمجة التعبير الجيني

مازن ابراهيم³

بسام العجي²

ربا دحام السعيد¹

^{1,2} كلية الهندسة المدنية، قسم الهندسة الصحية و البيئية، جامعة دمشق، دمشق، سوريا.
³ كلية الهندسة المدنية، قسم الإدارة الهندسية و التشييد، جامعة دمشق، دمشق، سوريا.

الخلاصة:

يعد التخثير من أهم العمليات في محطات تنقية مياه الشرب، ان استخدام مخثر الشبة يؤدي الى زيادة تركيز الألمنيوم المتبقي في المياه، والذي تم ربطه في العديد من الدراسات بمرض الزهايمر. لذلك فإنه من المهم استخدامه في الجرعة الأمثل. في هذه الدراسة، تم إجراء أربع مراحل من التجارب لتحديد تأثير مواصفات المياه الخام، مثل: العكارة، pH، القلوية والحرارة على الجرعة الأمثل من مخثر الشبة $[Al_2(SO_4)_3 \cdot 14H_2O]$ للحصول على علاقة رياضية تمكن من الاستغناء عن الحاجة لتجارب الجرعة. تم إجراء التجارب بظروف مختلفة وعلى مدى فصول السنة، وتم تحديد الجرعة الأمثل لكل دورة من التجارب لتشكيل نموذج يعتمد على التعابير الجينية. تم بناء النموذج عن طريق بيانات مواصفات المياه: العكارة، pH، الحرارة والقلوية للتنبؤ بقيمة جرعة الشبة الأمثل اللازمة، النموذج الذي تم الحصول عليه أعطى نتائج جيدة بمعامل ارتباط 0.91 ومربع جذر الأخطاء 1.8. تمت مقارنة النتائج مع نموذج انحدار خطي والذي لم يكن كافياً لإعطاء نتائج جيدة نظراً لطبيعة العلاقة الغير خطية المعقدة. سلاسل أخرى من التجارب تم القيام بها لتحديد الجرعة الأمثل اللازمة أثناء حدوث الفيضانات والتي تصل فيها قيم عكارة الى قيم عالية مع دراسة استخدام النشاء كمادة مساعدة للتخثير، تم الحصول على نموذج جيد للتنبؤ، بقيمة معامل 0.92 وجذر مربعات أخطاء 5.1.

الكلمات المفتاحية: الشبة، الذكاء الاصطناعي، التخثير، التعبير الجيني، الانحدار الخطي المتعدد، العكارة.