Open Access

Drinking Water
Engineering and Science
Discussions

Predicting turbidity and Aluminum in drinking water treatment plants using Hybrid Network (GA- ANN) and GEP

5 **Eng. Ruba Alsaeed**[1] ***, Bassam Alaji** [2] **, Mazen Ebrahim** [3] **.**

[1] Engineer at the Department of Sanitary and Environmental Engineering, Faculty of Civil Engineering, Damascus University, Damascus, Syria. s

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

10 3 Prof. Department of Engineering Management and Construction, Faculty of Civil Engineering, Damascus University, Damascus, Syria

Correspondence to: . Ruba Alsaeed (Rureta89@yahoo.com)

Abstract: Turbidity is the most important parameter needed to check the status of drinking water, as it is an
15 integrated parameter because its high values indicate high values of other parameters related to water quality. Coagulation and flocculation are the most essential processes for the removal of turbidity in drinking water treatment plants. Using alum coagulants increases the aluminum residuals in treated water, which have been linked to Alzheimer's disease pathogenesis.

In this paper, a hybrid algorithm (GA-ANN) used to predict the turbidity values in the drinking water purification
20 plant in Al Qusayr was used.

The models were constructed using raw water data: turbidity of raw water, pH, conductivity, temperature, and coagulant dose, to predict the turbidity values coming out of the plant.

Several models were built and fitness was detected for each model, the network with the highest fitness was selected, and then a hybrid prediction network was constructed.

25 The selected network was the most able to predict turbidity of the outlet with high accuracy with a correlation coefficient (0. 9940) and a root mean square error of 0.1078.

And 4 equations for determining the value of the residual aluminum was obtained using Gene expression method, and the best equation produced results with a very good accuracy, in this regard it can be referred to RMSE =0.02 R = 0.9 for the best model.

30

Drinking Water
Engineering and Science
Discussions

1. Introduction:

Due to the population growth around the globe, the scarcity of water resources has been a serious issue in the latest decades. (Haghiri et al 2018)

5 Surface water is one of the main sources of drinking water, but it is usually unsafe to be used without a treatment. There are many types of drinking water purification plants, according to the characters of the raw water. Still the primary goal of them is to produce safe water that does not contain pathogenic, microorganisms or toxic compounds.

Turbidity is an important indicator in drinking water quality. (Chenyu and Haiping 2020)

10 It is a lack in the water clarity, because of the presence of suspended or colloidal particles.

The excessive presence of colloids in raw water sources can cause many problems. It may disturb the operations of the water treatment plant (WTP) and can be very costly. (ABIDEEN, M 2016)

Water turbidity impedes the appropriate disinfection operation; it increases the needed dose of used disinfectant. Therefore, turbidity measurement is a basic test for assessing water quality, and it is an important operational

15 parameter to evaluate the stability of the drinking water purification plant's work.

Suspensions seek to settle at the bottom of the waterbed. However, fine particles with dimensions from 3-4 to 0.1 microns and colloidal particles with dimensions of 0.1-0.001 microns remain in the suspended state water.

Thus, the particles suitable for natural deposition technology are only those which their dimensions greater than 50-30 microns .

20 Table.1 shows the sedimentation period due to the gravity for different impurities in one meter of water with a temperature of 20℃, based on Stokes's Law:

Table 1. Classification of sedimentation period according to the magnitude of the particles

Time to participate	Diameter		Type
	mm	μ m	
1second	10	10^4	(Gravel)
10second	1	10^3	(Sand)
2 minute	10^{-1}	10^2	(Fine Sand)
2Hours	10^{-2}	10	(Clay)
8 Day	10^{-3}	1	(Bacteria)
2 years	10^{-4}	10^{-1}	(Colloids)
20 years	10^{-5}	10^{-2}	(Colloids)

25 Another important parameter of water quality for treated water is the amount of residual aluminum, especially when aluminum coagulants are used in the flocculation process [2].

The residual aluminum increases water turbidity, it also may have some health effects on consumers, aluminum hydroxide can build up on the inside of the pipes, reducing flow, and it has been suggested as a causal factor in Alzheimer's disease.

30 -The process of coagulation and flocculation of water impurities has been used for a long time. Coagulation is the process of increasing the sizes of dispersed particles by using materials that are able to standardize them. Although traditional modeling has been used to describe biological processes, it has been built by writing equations related to the speed of microorganism growth, substrate consumption, and product formation. Because

Drinking Water
Engineering and Science
Discussions
Open Access

microbiological reactions are non-linear and time dependent, with a rather complicated nature, such models have

35   many limitations.

Coagulation is a complicated nonlinear process because there are many physical and chemical variables that influence this process (Kim, C. M., & Parnichkun, 2017)

Modeling a complex process such as in water treatment plants is not easy, due to the non-linear processes (physical, chemical, biological, and biochemical) (S.I. Abba et al , 2019)

40   Modeling is usually done in one of the following ways:

1- Numeric or deterministic methods.

2- Data-Driven methods: methods based on data.

Techniques based on data have gotten a lot of attention recently. Due to their ease of use in the field of process monitoring implementation, and fewer underlying requirements. (Haghiri et al 2018)

45   In recent years, a variety of artificial intelligence (AI) techniques, have been used in modeling complex nonlinear water treatment processes, Due to the fact they have many benefits in conventional data modeling, When it is required to explicitly and accurately determine an output related to an input and an unknown physical process. (Kim, C, & Parnichkun, M. 2017)

Modeling has been performed for Bansong drinking treatment plant, hybrid of k-means

50   clustering and adaptive neuro-fuzzy inference system. Raw water quality data were classified into four clusters according to its properties by a k-means clustering technique. Then the ANFIS model was built. The research indicated that k-means-ANFIS models can be used as a robust tool during rainy seasons which is the most challenging period of operation. RMSE = 0.0572 , R= 0.9. (Kim, C, & Parnichkun, M. 2017)

In another study, (J. Tomperi et al) Residual aluminum in drinking water was predicted using Multiple Linear

55   Regression (MLR) and Artificial Neural Network (ANN) models.

Variables that affected the amount of residual aluminum according to the study were: raw water temperature, raw water KMnO4 and PAC/KMnO4 (Poly-Aluminum Chloride/Potassium permanganate)-ratio. The study gave a good result for the two methods, but the Alum was not used in the plant that the study passed its data on. So it cannot be considered in such cases.

60   Another study done by (Daghbandan, A et al) Polyelectrolyte, pH, turbidity, polyaluminium chloride, temperature, and electrical conductivity were used as input parameters in a study used GMDH to predict Aluminum and turbidity in drinking water plant, the study gave great results, we added the turbidity of raw water into the inputs.

While the methods described above can be used to predict residual turbidity and aluminum, they have drawbacks

65   that should be considered.

Artificial neural networks have a low reproducibility due to the randomly given weight and bias. Therefore, it is useful to use hybrid methods to develop its performance.

This study adjusted the architecture of the ANN network with Genetic algorithms.

For aluminum residuals ANN, MLR methods were used in previous studies.

70   Many new studies gave a promising result when using GEP, so it is appropriate to use in many fields in modeling. Because of its great ability to model nonlinear relations. And that is why it was used to predict the residual aluminum.

Drinking Water
Engineering and Science
Discussions

2. Materials and Methods:

75 **2.1 Data:** Research was carried out in the water purification plant in Qusayr , Homs. Using a set of three years' data. The plant was established on Orontes River (Asi river) to supply Hama city with drinking water. Its capacity today is about 6000 cubic meters per day .

Figure.1 shows an aerial photo of Qusayr plant.

80 **Figure1. An aerial photo of the study area. © Google Earth**

The raw water from Orontes River enters the plant, and the amount of water drained for the station is controlled at the outlet.

All chemical additives are added before the main dispenser. Then, water is distributed into 4 circular

85 precipitations each with a diameter of 31 m, then to the sand filters, a number of 20 filters of the double-exposed type, which works on gravity and is equipped with a filter layer of up to 1.5 m.

Table2. Statical characteristics of water that was taken from Qusayr drinking treatment plant

Column type		input				output
Parameter	Tur- in	Conductivity	pH	T	Alum dose	tur out
	Mg/L			°C	Mg/L	Mg/L
Min	6.9	310	6.5	8.2	0	0.14
Max	30.4	416	8.53	21.4	11.16	4.18
Mean	14.63	346.40	7.43	15.53	3.68	1.29
Std. deviation	4.41	19.11	0.32	2.66	2.31	0.99

90 **2.2 Artificial Neural Network:**

Artificial neural networks (ANNs) are one of the machine learning methods, many researchers have analyzed the quality of water and have built various deep learning models to predict the different parameters related to water quality. (Venkata Prasad et al 2020)

ANN are computational techniques designed to simulate the way in which the human brain performs a specific

95 task, by parallel processing consisting of simple processing units. These units are neurons (or nodes), which have

a neural property, as they store scientific knowledge and experimental information to make it available for use by adjusting weights. (Ghatfan et al 2018)

Fig. 2 shows a simple scheme for the process of passing inputs, calculating weights, and obtaining outputs within an artificial neuron and its similarity to a biological nerve cell. As a person has input units that connect him to the outside world, known as the five senses, Likewise, neural networks need input and processing units that of which calculations are made to set weights and through which we get the appropriate response for each of the network inputs.

100
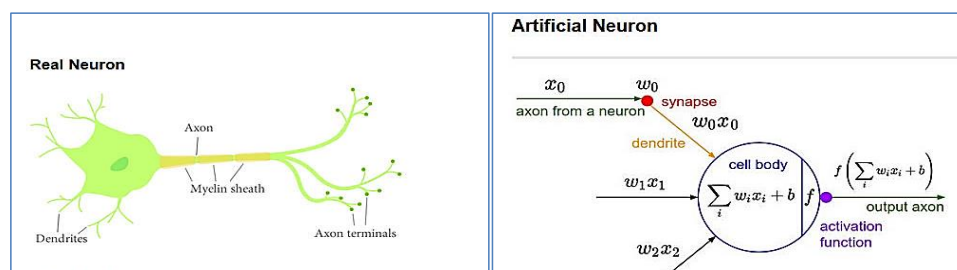
Figure 2: The similarities between a biological neuron and an artificial neuron.

105

- Basic components of an artificial neural network

It consists of the following basic components or at least some of them :(input layer - output layer - hidden layers - interconnections (weight)**. (WILMOT, et al 2005)

2.3 (GAs) Genetic Algorithms: (MELANIE, M. 1996)

Genetic algorithm is known as a smart algorithm that can be used to find and improve the solution of complex problems, it is one of the efficient research methods based on the principles of natural selection and genetics. The genetic algorithm is successfully applied to find an acceptable solution (near to ideal) in matters related to science, including medical and engineering sciences, because it has greatly reduced the time and effort required by program designers.

110

2.43 Gene Expression Modeling:

Ferreira et al. proposed the GEP algorithm in 2001.

It is a form of genetic algorithm that was first proposed based on Darwin's theory of evolution. And it works in the same way as a group in evolution abandons undesirable members and produces genetically modified offspring. No functional relationship is considered at the start of this method's operation.

115

This method optimizes the model's construction as well as its components.

Natural selection and genetic recombination are the ideas that underpin these adaptive machine methods.

These computer techniques produce entities (chromosomes representing potential solutions) that learn about the given problem and, given enough time to simulate, adapt to their environments (objective function and constraints).

120

As opposed to conventional modeling methods, GP has the advantage of not assuming any priori functional form of the solution. In a typical regression method, for example, the model structure is specified in advance (which is difficult to do in general) and the model coefficients are calculated.

125

Drinking Water
Engineering and Science
Discussions

GP uses a genetic algorithm as its basic search strategy, but GP varies from standard GA in that it usually works for parse trees rather than bit strings. A parse tree is constructed using a terminal set (the problem's variables)

130   and a function set (the simple operators that make up the function gene 2005).

The basic steps for this approach are the following:

1- Create a new mating pool from the current population by selecting solutions based on their fitness (as determined by the evaluation/objective function);

2- apply the genetic operators (crossover, mutation) to selected mating pairs from the mating pool, resulting

135   in offspring structures that can be incorporated into the new population

3- The current population will be replaced by a new population.

Each chromosome has at least a head and a tail, the head contains the function node and terminal node and the tail contains just the terminal node, the number of the tail gene is linked to the head gene number. As shown in the following equation:

140
$$t = h \times (n_{max} - 1) + 1$$

t is the number of the tail gene, h is the head gene numbers , n_{max} is the largest breach amount.

2.2 Mythology:

- For the prediction of turbidity a hybrid ANN- GA was obtained.

145   The data has been processed and outlier values were excluded, which are conflicting values, outliers and missing values, because it hinders the proper training of the neural network and greatly affect its performance.

First, a genetic algorithm was used to determine the input parameters used.

The Fitness function used in the comparison of input patterns was the inverse of the generalized regression network error (GRNN) resulting from the test set.

150   The generalized regression neural network (GRNN) is rapidly trained and sensitive to the explanatory inputs for the dependent variable adopted in the study.

Table 3. Parameters of genetic algorithms for selecting appropriate inputs for predicting turbidity

Network Parameters	Description
Search Method	Ga
Population Size	50
Generation Number	50
Crossover Rate	0.9
Mutation Rate	0.1

The parameters used at first as inputs were chosen according to their importance to the turbidity values of the

155   treated water. Turbidity of raw water, affects coagulation and rates of Settling, Conductivity because it is an online measure of water quality. Temperature because it affects reaction rates and rates of settling. pH because the pH determines the solubility. Alum dose Impacts turbidity by neutralizing negatively charged particles and instigating sweep flocculation. (Robert McArthur2015)

160 Table.4 shows the five best possible input patterns for ANN model to predict turbidity emission. We studied the first model and the sensitivity of the input elements on the proposed neural network model at the end of the research.

Table 4. Different Pattern Used in the Study

ID	Best Network Features Mask	Fitness
1	Turbidity(out)=F(Turbidity in, Conductivity ,pH, T ,Alum dose)	70.55
2	Turbidity(out)=F(Turbidity in, T ,Alum dose)	70.12
3	Turbidity(out)=F(Turbidity in, Conductivity ,Alum dose)	69.41
4	Turbidity(out)=F(Turbidity in, Conductivity , pH , ,Alum dose)	69.33
5	Turbidity(out)=F(Turbidity in, pH ,Alum dose)	69.01

165 From Table.4, it was found that the best input patterns for the neural network are the presence of (Turbidity in, Conductivity, pH, Temperature, and Alum doses) as inputs, with close results in the absence of conductivity and pH.

After determining the inputs, a hybrid method using a combination of neural networks and genetic algorithms were used, the genetic algorithms were used to get the optimal parameters of the neural model (number of neurons and the activation functions) and to accelerate the convergence process by giving appropriate initial 170 weights to the network before training, which makes the network train more efficiently, and as well as greater generalization capacity.

The work was done using Matlab 2012a program language to create a feedforward network, with two layers in all used networks, one hidden layer and an Output Layer containing a single neuron (One Node) representing the predicted value of turbidity of the water coming out of the station, and the bias was considered as a weight of an 175 additional constant income, with assuming the number of neurons ranged between (10) and (50) neurons with a transmission step of (5) neurons, In the hidden and output layer the activation functions were: (tansig, logsig, purelin).

- (GA) network randomly generates members of the community for each scenario, which was specified as 100 chromosomes, where each chromosome is a component of real values of weights (Real Value Encoding), and 180 then determines the best performance of the ANN (less MSE) through the selection method for the chromosome (Roulette wheel Selection), where the parents are chosen in this method based on the fitness function, then the crossover process takes place based on the report of to two points according to the ratio of (85%), and with the probability of a mutation of the amount. (2%). Table .5 shows the parameters of (ANN-GA) used in the study.

Table 5. the parameters of (ANN-GA) used in the study

Network Parameters	Description
Neural Networks	Feed- Forward –Propagation
Neural Networks Architecture	5-n-a
Nu. Neurons	[10:5:50]
Transfer function	Tang; lodsig; purelin
Search Method	GA
Population size	100
Fitness function	MSE

Probability of crossover	0.85
Probability of mutation	0.02
Crossover technique	Two point
Selection method	Roulette wheel selection

185

To improve the initial weights of the network, the first step is the random generation of the primitive population consisting of a set of chromosomes, , then the fitness function is evaluated for each individual ,and the best individuals are selected, genetic algorithms steps are done (crosses and mutations) on them, and repeating the evaluation process of the new generation members, after obtaining the ideal weights, comes the step of training

190 the artificial neural network using the reverse propagation algorithm (LM) and calculating the mean square error, comparing it with the permissible error, this process is done when the number of generations is over or the permissible error is reached and then the best architecture and best RMSE evaluation are presented.

-The number of data used in the study was (300) views, (70%) of the observations were randomly selected as a training group, (15%) as an investigation and calibration group, and (15%) as a testing group, and the number of

195 iterations was (500) and number of retraining times = 5 times, in order to obtain the best architecture of the neural network, and the training results for the different networks are shown in Fig.3 .
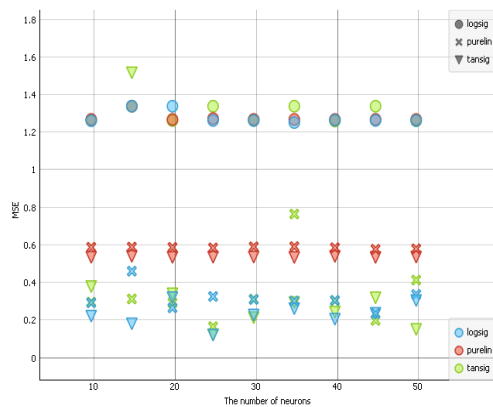

Figure 3. MSE Results of various networks

From the above figure, we note that the best architecture for the neural network is (5-25-1), according to the

200 lowest MSE value. The next step was training the best network with LM algorithm the, which shows good network performance in the training phase.

Fig.4 illustrates the architecture of the neural model (5-25-1) which refers to the use of an input layer with five inputs (Tur-in, Conductivity, pH, T, Alum dose), with one hidden layer containing 25 neurons, and an output layer containing a single neuron representing the output turbidity. In addition to the use of the logsig circuit in

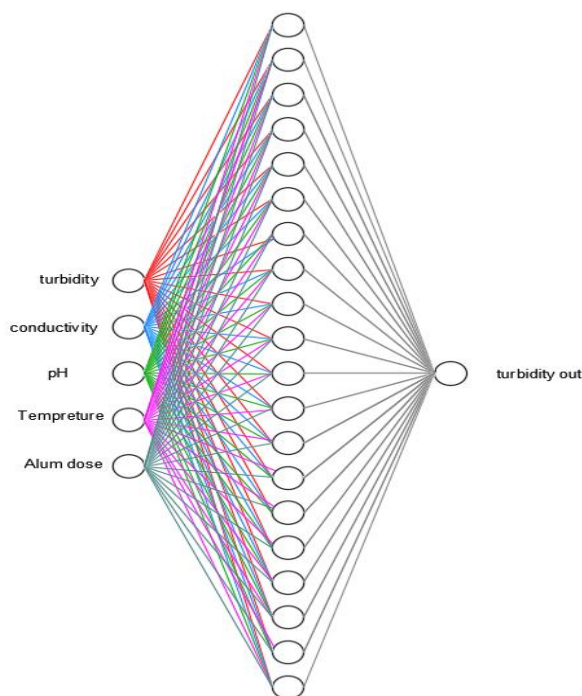205 the hidden layer, and the (tansig) function in the output layer.

Open Access
Drinking Water
Engineering and Science
Discussions

Figure 4. Architecture of the Neural Model

-For evaluating aluminum:

210 **Table. Staticall characteristics of water used for aluminum prediction**

Column type	Input					output
Parameter	Tur- in	Conductivity	pH	T	Alum dose	Resuidul aluminum
	Mg/L			°C	Mg/L	Mg/L
Min	6.9	310	6.5	8.2	0	0
Max	30.4	416	8.53	21.4	11.16	0.22
Mean	14.63	346.40	7.43	15.53	3.68	0.11
Std. deviation	4.41	19.11	0.32	2.66	2.31	0.12

In this part of the study a multivariate linear regression method (MLR) and a gene expression method were used to model residual aluminum in a drinking water treatment plants

215 The MLR model was used to be
 compared with the GEP result of this study.

MLR models are suitable to simulate the relationships between dependent and independent variables in many fields and studies, based on least squares fitting

It can be used to determine the relationships between two or more parameters.

220 As expressed in the following equation, where x is a multiple component, y is the output

$$Y = m_0 + \sum_0^i m \times X_i$$

9

Gene Expression models:

In the first stage of model building, data was entered, and a set of functions was selected to create primary chromosomes

225 For this, 70% of the data available from the drinking water purification plant in Quseir was entered to train the model.

The remaining data (30% of the data) building were used for validation.

In the next stage, the fit function was chosen, where the RMSE was chosen to evaluate the efficiency of the model.

230
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{n-1}}$$

Table6. Values of GEP Control Parameters

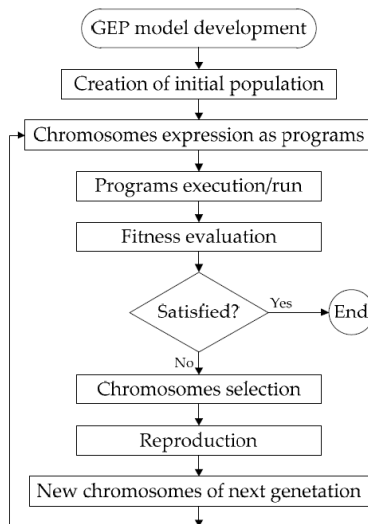| Number of chromosomes	40	Mutation rate	0.00138
Head long	8	Exchange Rates	0.003
Number of genes in the Chromosome	2	Inversion rate	0.0056

Figure 5. GEP flow Chart Liu, L, & Wang, M. (2019).

235

3. **Results and Discussion:**

The best architecture for the neural network was (5-25-1), it was trained with LM algorithm.

The statistical constants are shown in Table .6, we can note the good convergence between the real values and the values resulting from the neural model, hence the neural network (5-25-1) was able to predict the maximum

240 and minimum values in the series

Table .6 shows the results of the optimal network.

Drinking Water
Engineering and Science
Discussions

Table 6. Statistical results of best neural network architecture.

	Number of Data	RMSE	R	NSE
All Data	300	0.1078	0.9940	0.9880
Training Set	210	0.0553	0.9985	0.9969
Validation Set	45	0.1760	0.9842	0.9674
Testing Set	45	0.1815	0.9807	0.960

245 Fig. 6 shows the linear compatibility between the model outputs and the real values of the model ANN (5-25-1), for the three groups (training, validation and testing, where the scatter plot of the values in Fig.5 shows the high accuracy of the model for the three groups. And for the total data set.
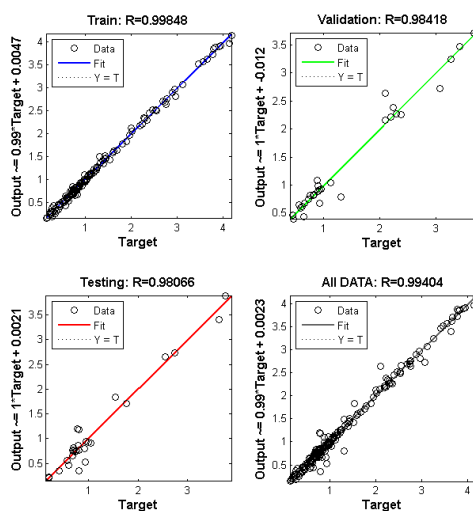
Figure 6. Correlation plots of process models

250

By examining the frequency errors shown in Fig.7, a high frequency of small errors below the values of 0.10 can be observed, which is an indicator of the compatibility between the true values of turbidity and the values predicted using the model (5-25-1). The hybrid network was useful in giving the optimal structural of the network. Comparing to ANN networks
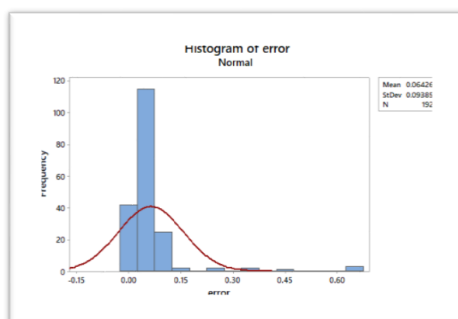
255

Figure 7. Frequency distribution of errors.

A sensitivity analysis was also conducted for the effect of the absence of any network input (Tur-out, Alum dose, T, pH, Conductivity, (Tur-in) on its performance, as the results in Table 7 indicate that the most influential

260 component is the coagulant dose, followed by (income turbidity, temperature, conductivity, PH), respectively.

Sensitivity analysis is the process of recalculating the outputs of the mathematical model under different variables to determine the importance of these variables in relation to the mathematical model. This helps to better understand the relationship between the inputs and outputs and arrange the variables according to their importance in relation to the model. There are a large number of ways to perform sensitivity analysis, the most important of

265 which are: at-a-time (OAT) and depends on removing one of the input variables and keeping the rest of the other variables constant and studying the effect of this variable on the output of the model and repeating that on the rest of the inputs, and the sensitivity is determined by observing changes in the output in many ways, such as determining the percentage of error resulting from removing the variable and its comparison with the total error.

Table. 7 Sensitivity analysis of the inputs of the model.

Input column name	Importance, %
Alum dose	49.07
Tur- in	21.12
T	16.88
Conductivity	9.54
PH	3.4

270

And to make it easier for the operator to handle the use of the neural model (5-25-1) The proposal: a graphical user interface (GUI) using Matlab 2014a language was created, with the ability to export it with an .exe extension to work on all Windows platforms. The GUI interface is shown in figure8.
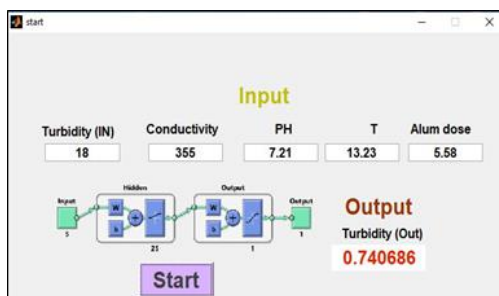
275

Figure 8. An interactive graphical user interface

- The results for Aluminum evaluating

MLR model**:**

280   The MLR model was built using Spss 21

Dependent: aluminum. Independent: Turbidity, Conductivity, pH, Temperature

Standard regression method: Enter

Developed MLR Aluminum simulation model was given as the following equation.

Residual aluminum = 0.915 – 0.001×Turbidity – 0.001× T – 0.001×conductivity -0.094×pH +0.024 × dose

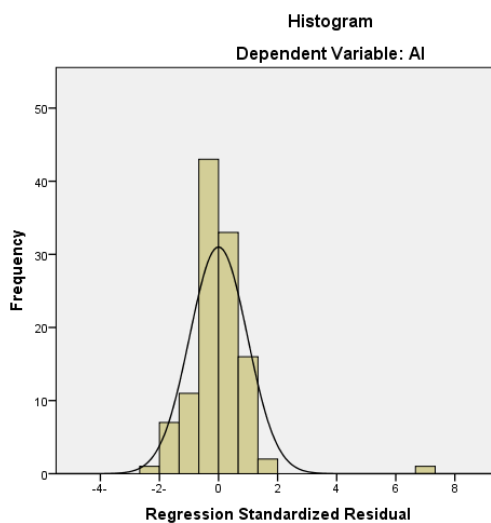285   With an accuracy of: R= 0.7 and RMSE = 0.2, and the histogram of the errors is shown in figure 9.

Figure 9. the frequency of the errors

As it is obvious that it wasn't satisfying enough to predict aluminum by MLR model, the residuals values were

290   very high compared to the values of the real aluminum values.

Gene Expression model:

GEP were used to make an equation that could determine the values of aluminum values in the outlet.

The input parameters were selected manually

First models were built using the inputs:

Drinking Water
Engineering and Science
Discussions

295   Turbidity, conductivity, Temperature, Alum dose, pH
Then different models were built deleting one or two of the input parameters. The bsest model was Model 1 with

the following parameter: Turbidity, conductivity, Temperature, Alum Dose, pH.

Table 8. The accuracy assessment table of gene-expression models.

Model number	Inputs	RMSE	R
1	Turbidity, conductivity, Temperature, Alum Dose , pH	0.02	0.9
2	Turbidity, Temperature, Alum Dose , pH	0.035	0.7
3	Turbidity, Alum Dose , pH	0.03	0.8
4	Turbidity, Alum Dose	0.025	0.7

300

As it's shown, minimizing the inputs did not give better results for the models. mentioning that during the

modeling. The

Figure10. Observed Vs Predicted Aluminum values for model 1

305

And as a result, 2 sub gene-expression trees were developed, all the sub gene-expression trees
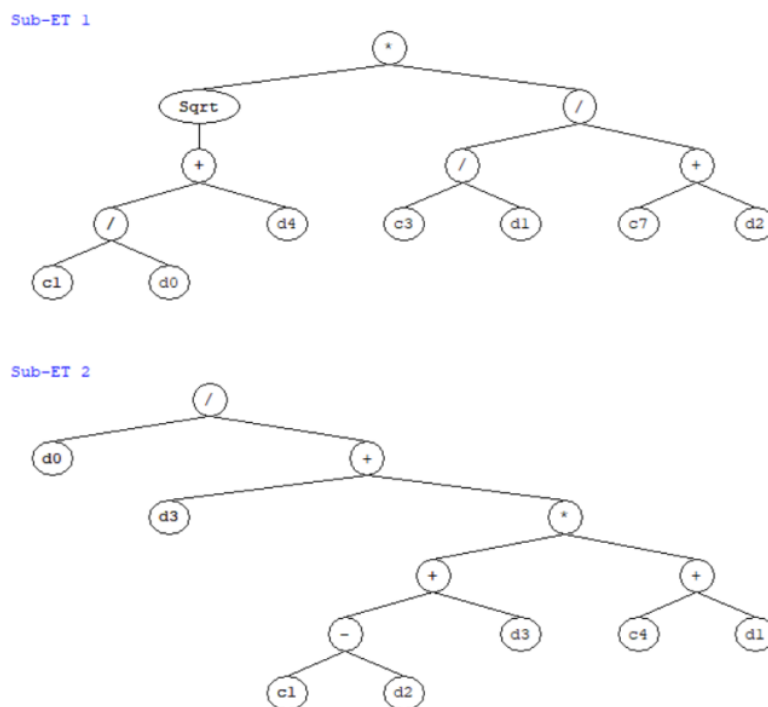were drawn in Figures

310

Figure 10. The sub gene-expression tree

The written equation for model one was as followed:

$$Reseduel\ Aluminum = \frac{\sqrt{\frac{T}{22.1339} + (Dose - 0.49006)}}{\frac{Cconductivity}{47.3113} \times (-4.31506 + pH)}$$

315

$$+ \frac{Turbidity}{((conductivity - 100.607) \times ((t - pH) - 4.55))) + Conductivity}$$

320   As was shown, models with the parameters mentioned in model1 is able to predict the residual aluminum with a
very geed accuracy.
Comparing to the previous studies, it is obvious that GEP method is very good at simulating the process of
drinking water, and it should be used to model and predict different parameters related to drinking water.

325   **Conclusion:**
Modeling methods that are based on data offer valuable tools for process modeling and controlling process in
drinking water treatment plants and providing a good alternative to the conventional methods based on numeric
simulation.

Drinking Water
Engineering and Science
Discussions
Open Access

330   Many articles studied modeling methods to predict the coagulant dose, thus not enough had studied the prediction of turbidity and the aluminum residuals.

Articles that discussed turbidity have used k-means clustering and adaptive neuro-fuzzy inference system (Kim, C. M., & Parnichkun, 2017) and GMDH (Daghbandan, A et al)

In this research GA was used to obtain the optimal structure of an artificial intelligence network used to predict turbidity of the treated water, as it is a great parameter to evaluate the stability of the plant process.

335   In this way, GAs with a specific encoding scheme is first presented to evolutionary design of the ANN neural networks used in predicting turbidity.

The parameter selected was similar to some studies before (Kim, C. M., & Parnichkun, 2017) with adding the alum dose as an input.

The inputs parameters have a significant effect on the models accuracy, and the most affecting parameter on

340   turbidity were, alum dose, raw water turbidity, temperature, conductivity and pH, respectively. And GUI inference were obtained for the ease of use from different operators.

And the study gave a good prediction models that were created using gene expression method with a few important variables, for predicting residual aluminum.

As contract of another article (J. Tomperi et al), the MLR could not give a good equation in this case.

345   The residual aluminum in drinking water plants can be predicted with very good accuracy using resulted nonlinear equation. Differences between the results can be noticed when using deferent inputs, and the best inputs were alum dose, raw water turbidity, temperature, conductivity and pH with a very close results when the absence of conductivity.

The created prediction models could be improved by using different variable selected or using different

350   modeling methods. Yet, the results of this study were promising and models obtained can be used in an early warning to give information about the water treatment plant

355

Data availability

The datasets of the current study are available from the authors upon reasonable request.

360

References:

ABIDEEN, M. B. Z. . OPTIMISATION OF COAGULATION PROCESS IN WATER TREATMENT PLANT USING
STATISTICAL APPROACH (Doctoral dissertation, Universiti Teknologi Malaysia). 2016

365

Daghbandan, A., Khalatbari, S., & Abbasi, M. M. . Applying GMDH-type neural network for modeling and prediction of
turbidity and free residual aluminium in drinking water. Desalination Water Treat, 140, 118-131. 2019
 doi: 10.5004/dwt.2019.23357

370   Haghiri, S., Daghighi, A., & Moharramzadeh, S. . Optimum coagulant forecasting by modeling jar test experiments using
ANNs. Drinking Water Engineering and Science, 11(1), 1-8. doi.org/10.5194/dwes-11-1-2018. 2018

Drinking Water
Engineering and Science
Discussions
Open Access

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. Applied Water Science, 7(7), 375   3885-3902. DOI 10.1007/s13201-017-0541-5. 2017

Chenyu Song, Haiping Zhang . Study on turbidity prediction method of reservoirs based on long short term. Ecological Modelling 432;. 2020

380   Ghatfan Abd Alkarim Ammar , Amer Qousai AL_Darwis , Alaa Ali Slieman . GA-ANN model for Interpolation the Missing Data of daily Rainfall within AL-Ghab plain;;tishreen University magazine. 2019

Liu, L. W., & Wang, Y. M. . Modelling reservoir turbidity using Landsat 8 satellite imagery by gene expression programming. Water, 11(7), 1479. 2019
385

MELANIE, M. An Introduction to Genetic Algorithms. A Bradford Book the MIT Press Cambridge, Massachusetts, London, England, Fifth printing, 1999, 162.

Robert Harris McArthur (2015). DEVELOPMENT OF ARTIFICIAL NEURAL NETWORKS FOR THE OTIMIZATION
390   OF DRINKING WATERTREATMENT, Department of Civil Engineering,University of Toronto.2015

S.I. Abba, Rabiu Aliyu Abdulkadir, M.S. Gaya, M.A. Saleh, Parveneh Esmaili, Mustapha Bala Jibril. Neuro-fuzzy ensemble techniques for the prediction of turbidity in water treatment plant.2019.
DOI:10.1109/NigeriaComputConf45974.2019.8949629
395

Prasad, V. V. D., Venkataramana, L. Y., Kumar, P. S., Prasannamedha, G., Soumya, K., & Poornema, A. J.). Water quality analysis in a lake using deep learning methodology: prediction and validation. INTERNATIONAL JOURNAL OF
. doi.org/10.1080/03067319.2020.1801665.ENVIRONMENTAL ANALYTICAL CHEMISTRY.2020

400   WILMOT, C.G; MEI, B. Neural Network Modeling of Highway Construction Costs. Journal of Construction Engineering and Management ASCE, July (2005).