

# Determination of the Optimal Dosage of Aluminum Sulfate in the Coagulation-Flocculation Process Using an Artificial Neural Network

A. J. León-Luque, C. L. Barajas, and C. A. Peña-Guzmán

**Abstract**—The process of coagulation and flocculation is one of the most important operations among the water purification process, but its effectiveness is affected due to the calculation of the coagulant dosage which is performed by the Jar tests or the use of the *Streaming Current Detector* (SCD), having as main disadvantage that it does not take into account the change of the physicochemical parameters of the water in real time and also the need to obtain an optimal operation point for the equipment. In this paper the optimal dosage of Aluminum Sulfate ( $\text{Al}_2(\text{SO}_4)_3 \cdot 18\text{H}_2\text{O}$ ) is determined using a model of Artificial Neural Network (ANN) that, when faced with real time variations of turbidity is able to calculate an indicated dose of coagulant, with the aim of achieve effective coagulation in the trial water and avoid excessive or insufficient presence of coagulant, minimize the need to make jars test continuously and reduce economic losses due to inadequate spending of coagulant.

The ANN created is able to calculate the dosage based on the value of initial turbidity of the fluid to be treated with a MSE 0 mg/L, achieving removal percentages greater than 93% for most cases.

**Index Terms**—Aluminum sulfate ( $\text{Al}_2(\text{SO}_4)_3 \cdot 18\text{H}_2\text{O}$ ), artificial neural networks, coagulation, flocculation, optimal dosage of coagulant.

## I. INTRODUCTION

The potabilization process aims to offer safe water that fulfill the requirements established in the local regulations. Inside these processes, one of the most used and important in superficial waters is the coagulation and flocculation, which aims to eliminate turbidity in order to remove the suspended particles in water using flocculants [1].

To reach high levels of reduction is necessary to look for an optimal dosage of coagulant, which generally is calculated by means of the “Jar Test,” working under a uniform condition of turbidity.

However the turbidity values change temporarily on the water, affecting the effectiveness of this process, due to each turbidity value may need a different coagulant dose [2].

Coagulation and flocculation processes has been object of study and research by different approaches. One of these has been the application of artificial intelligence, through different computational tools like diffuse systems and neural

networks [3].

Previous investigations have implemented Artificial Neural Networks satisfactorily in the calculation of the coagulant dosage in Drinking Water Treatment Plants; in the case of Canada it was applied in the plant of Sainte Fox in the year 1997, with a group of four ANN, each one with a behavior learnt for each specific station, minimizing the need of Jar Tests to determine a coagulant dose [2]. Other studies show the importance of the relation of some physicochemical parameters regarding the coagulant dose, this is the case of an optimal dose program predictor based in the logic of the RNA implemented in Salisbury, Australia, which has the turbidity, color and alkalinity of the water as parameters of entrance of the network, giving a value of coagulant dose for each admitted stage [4]. Incidentally they have made several studies about the advantages of using artificial intelligence in order to optimize the calculation of the dose of coagulant using operation of treatment plants data as base for the design of each predictive model [5].

This article shows the use of an Artificial Neuronal Network as an alternative to determine the optimal dose of coagulant for different concentrations of initial turbidity employing synthetic water, having the turbidity as the only intake variable simplifying the structure of the Artificial Neural Network.

## II. MATERIALS AND METHODOLOGY.

### A. Determination of the Most Representative Physicochemical Parameters for the Calculation of the Optimal Coagulant Dosage

The parameters selected were, in first instance, turbidity, because of its relationship with the colloidal particles or suspended material that affect the degree of transparency of the water; pH, since each coagulant has an optimal range of pH within which the action turns out to be more efficient; alkalinity, because if the water has a high degree of alkalinity the reaction of coagulant gets affected; Color, because of the contribution of color made by the colloidal particles; water conductivity, as an indirect measurement of ions or dissolved solids present in the water and temperature, since the lower the temperature the longer it takes for flocculation of particles to occur.

### B. Evaluation of the Optimal Dosage of Aluminum Sulfate ( $\text{Al}_2(\text{SO}_4)_3 \cdot 18\text{H}_2\text{O}$ ) for Different Scenarios by Jar Test

#### 1) Preparation of samples

Sixteen samples were prepared under different levels of initial turbidity that was within the range of 10-1000 NTU,

Manuscript received May 5, 2015; revised July 20, 2015.

A. J. León-Luque and C. L. Barajas G. are with the Faculty of Environmental Engineer, Santo Tomás Univeristy, Bogotá, Colombia (email: andrea.leon@usantotomas.edu.co, claudia.barajas@usantotomas.edu.co).

C. A. Peña-Guzmán is with the Department of Environmental Engineer, Faculty of Engineer, Santo Tomás Univeristy and Universidad Autónoma de Colombia, Bogotá, Colombia (e-mail: carlos.pena@usantotomas.edu.co, carpeguz@gmail.com).

which is a characteristic range of the rivers in Colombia. The samples were defined as follows: 10, 28, 46, 64, 82, 100, 180, 260, 340, 420, 500, 590, 670, 760, 840, and 1000 NTU. These levels were obtained by adding different samples a certain volume of artificial water that was prepared with different proportions of water, sand, clay and powdered refreshment drink (made up of sweetener, colorant and flavoring), in order to simulate suspended solids, color and organic composition usually found in water that is submitted to purification.

When the samples were prepared, we proceeded to setup the Jar Test with 10 beakers with 500 ml of water. Aluminum Sulfate  $Al_2(SO_4)_3 \cdot 18H_2O$  in a granular form was used as coagulant, and it was applied in samples in liquid form, this is why its preparation was necessary at a concentration of 5000 mg/L, in which 1ml applied on a sample of 500 ml represents a concentration of 10 mg /L.

## 2) Jar test

First of all, a Jar Test was performed for each initial turbidity level defined above, meaning that 16 experiments were carried out overall. The first seven experiments (from 10NTU to 180NTU) were made with six samples or jar each and the nine remaining experiments (from 260NTU up to 1000 NTU) were performed with a setup of 10 samples for each test.

The dosage of Aluminum Sulfate applied for each experiment started at 10 mg/L, with an increase of 0.5 mg/L between each sample. However starting from the experiment number 9, corresponding to 340 NTU, it was necessary to increase the dosage because this was insufficient for that level of turbidity and therefore did not showed good results of removal.

TABLE I: DOSAGE OF  $Al_2(SO_4)_3 \cdot 18H_2O$  FOR EACH EXPERIMENT

Initial Turbidity (NTU)	Dosage $Al_2(SO_4)_3 \cdot 18H_2O$ (mg/L)
10-180	10, 15, 20, 25, 30 y 35
260-340	10, 15, 20, 25, 30, 35, 40, 45, 50, 55
420-590	35, 40, 45, 50, 55, 60, 65, 70, 75, 80
670	40, 45, 50, 55, 60, 65, 70, 75, 80, 85
760	50, 55, 60, 65, 70, 75, 80, 85, 90, 95
840-1000	55, 60, 65, 70, 75, 80, 85, 90, 95, 100

After preparing the samples, a Jar Test was carried out for each level of initial turbidity defined above, where a certain amount of coagulant was added to each beaker. For the first seven experiments, which range from the 10 NTU to 180 NTU, six samples were prepared for each test and the concentration of Aluminum Sulfate were added from 10 mg/L to 35 mg/L, with a variation of 0.5 mg/L between each sample. For the nine remaining experiments with turbidity levels ranging from 260 to 1000 NTU, 10 samples setups were put in place for each level of turbidity.

Then, rapid mixing was carried out at 70 rpm during one minute intervals in order to homogenize the distribution of coagulant throughout the beaker and trigger the particle agglutination process. The slow mixing was done at a speed of 30 rpm during 15 minutes intervals to ensure the joint of the small flocs formed in the coagulation, which were

allowed to sediment for 20 minutes.

After finishing the Jar Tests with each of the samples, the reference parameters for coagulation and flocculation were measured again (pH, conductivity, turbidity, temperature, alkalinity and color) in order to determine the coagulant dosage that caused the highest percentage of turbidity removal, in other words the concentration of alum that showed the lowest level of turbidity in a sample of each of the 16 experiments.

## 3) Determination of optimal coagulant dosage

Having defined an initial optimal dose, we proceeded to refine this dose, for this means we set it up using 10 different jars. In the jar positioned by number five, we added a dose exactly equal to the initial optimal dose obtained; for the four jars located to its left we added a lower dose lowering 0.1ml per jar at a time and, for the five jars by the right, we increased the dose in 0.1 ml per jar. This procedure was made tripled for each of the 16 levels of turbidity initially defined and it was performed in order to achieve the optimization of the coagulant dosage that would be entered in the creation of the Artificial Neural Network (ANN).

## C. Determination of the Optimal Dosage for Different Scenarios through Artificial Neural Network Taking into Account the Input Parameters Established

### 1) Relation of input parameters selected with coagulant dosage

With the results of the measurements of each of the selected parameters a correlation matrix was made, using MatLab, between the coagulant dosage for each level of turbidity against each physicochemical parameter, which allowed to establish the degree of correlation of each of the parameters compared to the behavior of the optimal coagulant dosage and in turn with the final turbidity level. This in order to determine if, for this case, all the monitored parameters should be defined in the ANN as input variables.

### 2) Statistical analysis

To eliminate the atypical data obtained in the realization of the Jar Test during the refining of coagulant dosage, those turbidities that were not related to the condition established by the authors were discarded. This condition was created taking into account the characteristics of a Gaussian distribution:

If  $|Xi - \bar{X}| > 1,1 * \sigma$  then  $Xi = \text{Outliers}$

If not, then  $Xi = Xi$

The value of 1.1 was established within the condition because this value allowed to define a range with which all values evaluated were not discarded and which in turn did not accept values distanced to the average.

With each optimal final turbidity and dose of coagulant obtained, the correspondent graph was made in order to analyze the results to determine if the data presented a linear behavior, which was determined by the use of a linear regression.

### 3) Linear regression of turbidity and optimal dose

This regression was performed using the following mathematical expression:

$$Y_i = B_0 + B_1 X_1 \quad (1)$$

where  $Y_i$  is the dependent variable and  $B_0, B_1$  are coefficients

that define the position and inclination of the line, also allows to see the influence of the explanatory variables ( $X_1, X_2, \dots, X_n$ ) on the dependent. Overall, the coefficient  $B_0$  is the cutoff of the line and the coefficient  $B_1$  represents the slope of the line [6].

Linear regression was performed with the values of coagulant dose applied and initial turbidity levels on each jar, using MATLAB's *regress* function, with which the values of the slope  $B_1$  and  $B_0$  cut off were obtained. With these values the relationship between optimal dose of applied coagulant ( $Y_i$ ) and the initial turbidity ( $X_i$ ) was calculated. The calculation of the regression allowed to obtain the trend line of the optimal dose of coagulant for each level of turbidity.

Subsequently was carried out the calculation of Mean Square Error (MSE) in order to observe how significant was the difference between real and simulated results. For the MSE the following formula was used [7]:

$$MSE = \sqrt{\sum_{i=1}^n (X_i - \bar{X}_i)^2} \quad (2)$$

Being the root of the sum of all estimated values minus simulated values squared. This formula was used to estimate the MSE, meaning the difference between simulated (trend line) and estimated (values of optimal dose).

#### 4) Creation of the artificial neural network (ANN)

After determining the associated between initial turbidity and optimal dose of coagulant and the function that represents this behavior, the ANN was calibrated by different levels of initial turbidity, optimal dose coagulant obtained in Jar Test and the optimal dose calculated by a linear regression.

Three cases were assigned for the training, for each case was established a type of network and a speed of propagation. For the first case was assigned a network with a propagation speed of 40, for the second case the same type of network of the first case was established, but with the difference that he was assigned a number of iterations (25 iterations), for the third case propagation speed was 0.5 and in all cases was calculated the MSE. For the training network the values obtained with the measurements of Jar test for each case above were used, regarding the ECM actual values were used, in other words the measured in Jar test and the simulated with linear regression.

For the validation of the ANN was performed the same procedure proposed in the calibration step, but the input vectors of this network were the same for the training and testing, in other words the values of initial turbidity and optimal dose obtained with the Jar tests. Just like the Network Training, were established three cases with the same conditions regarding speed of propagation of data and the number of iterations.

For the third network, Evaluation Network, were assigned the input vectors, in this case the initial turbidity values were random values in the range of 10-1000 NTU which were assigned with a determined value of coagulant dosage and as in the above two networks were established three cases for the network to choose the case that had better performance.

#### D. Verify the Effectiveness of the Implemented ANN

The verification of the proper operation of the ANN was performed by setting up the five Jars Tests compounds for

four samples with initial turbidities (random) 15, 90, 150, 500, and 950 NTU. These tests were performed triplicated for each level of initial turbidity mentioned and the doses used were the optimal dose values obtained with ANN.

TABLE II: DOSAGE OF  $Al_2(SO_4)_3 \cdot 18H_2O$  CALCULATED BY ANN

Initial Turbidity (NTU)	Dosage $Al_2(SO_4)_3 \cdot 18H_2O$ ANN
15	22 mg/L
90	27 mg/L
150	29 mg/L
500	55 mg/L
950	90 mg/L

### III. RESULTS

#### A. Relation of Parameters

Below, it shows a matrix of correlations between the different parameters that were monitored in Jar Tests, which was calculated in order to determine if it could be used as input variables for the ANN.

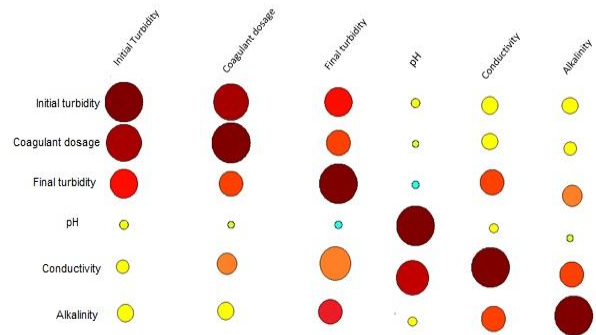


Fig. 1. Correlation matrix of parameters.

The sizes of the circle refers to the degree of correlation between the two variables, the larger and the darker its color is the value of the Pearson correlation coefficient is closer to one and therefore the variables presented have a strong linear relationship. According to the above representation, the variables of Initial Turbidity and Optimal Dosage for determining the optimal dose of coagulant were defined, this is the reason why parameters such as pH, conductivity and alkalinity, which presented a very low relationship versus coagulant dose and the initial turbidity, were not taken into account in the modeling of the ANN.

#### B. First Jar Test

It was found that the final turbidity value does not tend to followed a linear behavior but the final turbidity value for each initial stage can vary in a rank of 4.7 to 13.7 NTU, being values significantly small compared to the initial turbidity levels.

It was found that the turbidity removal level is significant because it had turbidity levels up to 1000 NTU and after the Jar Tests the highest final turbidity level was of 13.7 NTU. This could be confirmed with the removal percentage calculated for each experiment, which was made taking the data of the initial turbidity ( $I_t$ ) and the lower final turbidity value ( $F_t$ ) obtained in each proof.

It can be assumed that it exists a directly proportional relation between the initial turbidity and the removal percentage, where a higher initial turbidity a higher removal percent.

TABLE III: REMOVAL PERCENTAGE FOR EACH INITIAL TURBIDITY LEVEL

Initial Turbidity (NTU)	Final Turbidity (NTU)	Removal Percent (%)
10	7,1	29,0
28	6,1	78,2
46	4,7	89,8
64	7,5	88,3
82	8,2	90,0
100	5,6	94,4
180	12	93,3
260	11	95,8
340	13,7	96,0
420	8,7	97,9
500	7,1	98,5
590	13,2	97,8
670	9,1	98,6
760	9,7	98,7
840	11,7	98,6
1000	5	99,5

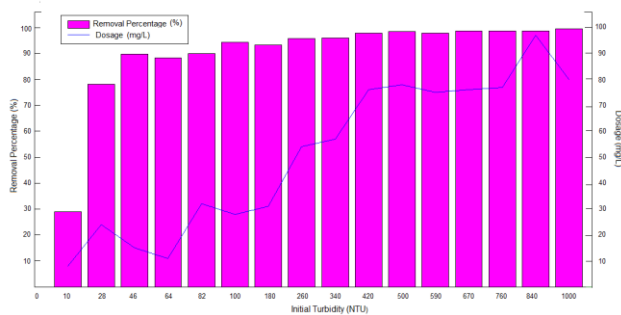


Fig. 2. First jar test removal percentage.

It can be observed that the turbidity removal percentage is more than 50% for most of the Jar tests, only in the first Jar Test did not reach an equal removal to that percentage (although it is important to take into account that the first sample has a 10NTU turbidity), by this, it can be deduce that

the removal percentage of turbidity expected for the network is going to be keep inside a rank of the 80 to 98%, that is considered satisfactory.

### C. Tripled of the Refine of the Optimal Dose

Different results were obtained in the final turbidity value compared to that obtained in the first Jar Test because the higher final turbidity tests for refine dose was 11 NTU while for the first test was 13.7 NTU.

It showed that the optimal dose for these experiments ranged in a few milliliters regarding the optimal dose defined in the first Jar Tests. Additionally it was shown that the value of coagulant dosage was maintained in the three repetitions of the refining. In terms of percentage removal Jar tests performed for the refining of the optimal coagulant dose, the lowest percentage of removal was 31% and the highest was 99.57%.

### D. Artificial Neural Network

The artificial neural network model is divided in three types of networks: Calibration network, Validation network and Simulation network [8].

The calibration network adopted a pattern of the behavior of the coagulant dose for each level of initial turbidity, which tends to increase as the value of initial turbidity increases. The validation network represents the values of coagulant dose defined by the Jar Tests compared with the simulated values by the Validation Network, showing that both variables had the same behavior, allowing to confirm that the network already has trained and has adapted the pattern of behavior that has the coagulant dose for each turbidity level. The simulation network showed the value of coagulant dose to apply for a random value of initial turbidity previously entered to the network, it was observed that the value of the coagulant dose was minor to the values of dose obtained with the Jar Test. It can be observed in the graphic of the simulation network that the network has a satisfactory adjust, showing that error level of the network is lowly significant. In Fig. 3 is shown the behavior of each of the networks previously mentioned.

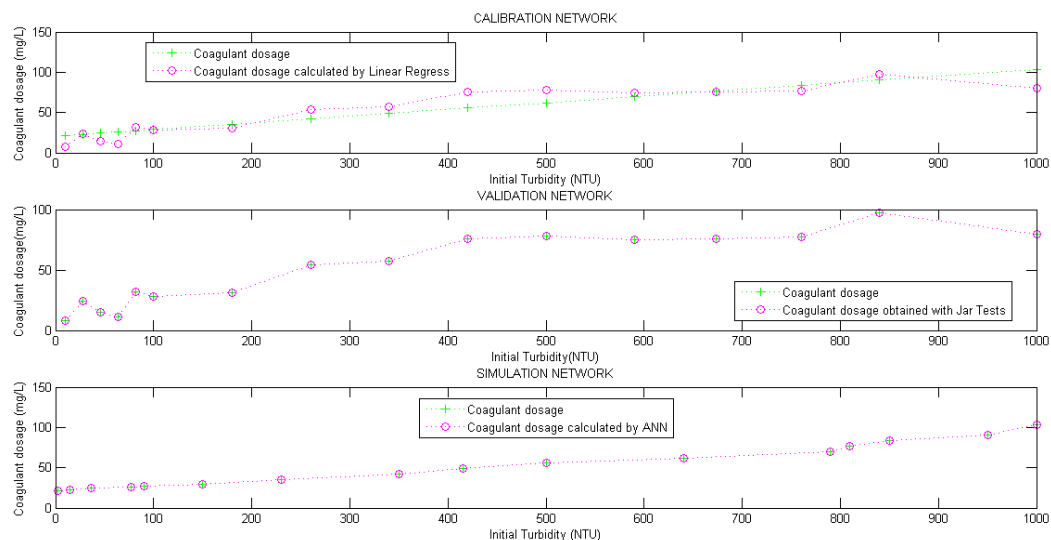


Fig. 3. Artificial neuronal network.

Regarding the level of removal obtained with the model of RNA it was observed that its results are significantly higher

compared with the removal level obtained with the Jar Tests initially made (see Fig. 4).

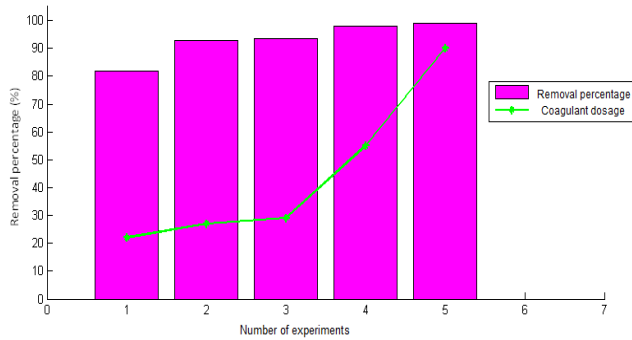


Fig. 4. ANN removal percentage.

#### IV. CONCLUSIONS

The MSE obtained with the ANN was 0 mg/L, which demonstrates that the use of this model for predicting optimal dose would be successful and in addition would decrease the need for consecutive Jar Tests.

The percentage of turbidity removal obtained with the RNA was greater than 50% for each case, showing that the model is highly efficient, with lower values of coagulant dose to those obtained in laboratory, optimizing the coagulation and flocculation process.

The effectiveness of the functioning of the ANN depends mainly on the quality and quantity of the data entered for its training and calibration, reason why it is important to define the input variables that really have prior influence on the estimation of coagulant dosage, since it will allow to obtain a network with a significantly small MSE and therefore calculate more accurate results, meaning that really optimal coagulant doses will be obtained.

#### REFERENCES

[1] D. Castaño, "Analysis of the influence of two raw materials coagulants on the residual aluminum in treated water," Industrial Chemistry, Technology School, Technological University of Pereira, Pereira, Colombia, 2011.

[2] G. Wu and S. Lo, "Effects of data normalization and inherent-factor on decision of optimal coagulant dosage in water treatment by artificial neuronal network," *Experts Systems with Applications*, vol. 37, pp. 4974-4983, 2010.

[3] S. Heddami, A. Bernard, and N. Dechemi, "ANFIS-based modelling for coagulant dosage in drinking water treatment plant: a case study," *Environmental Monitoring and Assessment*, vol. 184, 2012.

[4] P. Kulkarni and S. Chellam, "Disinfection by product formation following chlorination of drinking water: Artificial neural network models and changes in speciation with treatment," *Science of the Total Environment*, vol. 408, pp. 4202-4210, 2010.

[5] C. Villarreal and E. Caicedo, "Applied computational intelligence techniques to estimation models coagulant in water purification process," *Journal Faculty of Engineering University of ANTIOQUIA*, pp. 205-215, 2013.

[6] S. Weisberg, "Linear hypothesis: Regression (basics)," in *International Encyclopedia of the Social & Behavioral Sciences*, J. D. Wright, Ed., Oxford: Elsevier, pp. 153-156, 2014.

[7] S. Quintanilla, "Numerical methods applied to engineering," *Science & Development*, p. 10, Septiembre 2012.

[8] P. Ponce, "Artificial intelligence with engineering applications," Mexico: Alfaomega editor group, 2010.



**Andrea Juliana León Luque** was born in Valledupar, Colombia on July 3, 1991. She has a professional degree on environmental engineer, obtained at Santo Tomas University in Bogotá-Colombia by July, 2015.



**Claudia Lorena Barajas Garzón** was born in Bogotá, Colombia on March 8, 1993. She has a professional degree on environmental engineer in Santo Tomas University in Bogotá-Colombia obtained in July, 2015.

**Carlos Andres Peña Guzmán** was born in Bogotá, Colombia on March 1, 1983. He received the B.S. in environmental engineer and his M.S. in hydrosystems. Currently he is coursing on his Phd studies in water and sustainable development and is a professor in Universidad Santo Tomás and Universidad Autónoma de Colombia in Bogotá, Colombia.