




Prediction of the Optimal Dosage of Poly Aluminum Chloride for Coagulation in Drinking Water Treatment using Artificial Neural Networks

Predicción de la Dosis Óptima de Policloruro de Aluminio para el Proceso de Coagulación en el Tratamiento de Agua Potable mediante Redes Neuronales Artificiales

 C. Izquierdo¹  B. Pezantes¹  E. Ayala¹

¹Grupo de investigación de inteligencia artificial y tecnologías de asistencia – GI IATA, Universidad Politécnica Salesiana, Cuenca, Ecuador

E-mail: cizquierdov@est.ups.edu.ec; bpezantes@est.ups.edu.ec; eayala@ups.edu.ec

Abstract

The addition of chemicals in drinking water treatment is usually a manual procedure performed by highly trained and experienced persons. To solve this problem, this study is based on the analysis of data collected from a raw water source located in Ecuador. Then, using the information on the physical-chemical parameters of the raw water such as pH, turbidity and color, the definition of the doses of Poly Aluminum Chloride (PAC), and the input and output variables of the dosage process are identified. Consequently, the implementation of an intelligent control system based on Artificial Neural Networks (ANN) is proposed in order to reduce the dependence on experienced people. These experiments start with data collection and analysis in order to establish the variables involved in the process. The proposed neural model has three hidden layers, and it uses adaptive gradient algorithms. An analysis of the results was performed using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). The PAC predictive model in the training phase gives a MAPE value of 0.0425 for the not adjusted values and 0.0262 for the adjusted numerical values. However, in the test phase the neural model achieves a MAPE of 0.057 for the not adjusted PAC values and 0.041 for the adjusted values. This alternative provides an efficient solution to solve dosing problems in drinking water treatment plants (DWTP), with reliable results according to RMSE and MAPE metrics.

Index terms— Drinking water, Dosing, DWTP, Coagulant chemicals, Artificial neural networks, Control system.

Resumen

La adición de las sustancias químicas en el tratamiento del agua potable comúnmente es un procedimiento manual realizado por personal altamente capacitado y experimentado. Esta resulta una tarea crítica debido a que requiere cierto nivel de experiencia para una correcta dosificación. Como posible solución, este estudio se basa en el análisis de datos recolectados de una fuente de agua cruda ubicada en Ecuador. Utilizando la información de los parámetros físicoquímicos del agua cruda, como el pH, turbidez y color, se identifican las dosis de Policloruro de Aluminio (PAC), y las variables de entrada y salida del proceso. En consecuencia, se propone la implementación de un sistema de control inteligente basado en Redes Neuronales Artificiales (RNA) con la finalidad de reducir la dependencia del personal experimentado. Para ello, se parte con la recolección y análisis de datos y así establecer las variables involucradas en el proceso. El modelo neuronal propuesto dispone de tres capas ocultas y utiliza algoritmos de gradiente adaptativo. El análisis de los resultados se realizó mediante el error porcentual absoluto medio (MAPE) y el error cuadrático medio (RMSE). El modelo predictivo de PAC en etapa de entrenamiento indica un valor MAPE de 0,0425 para los valores no ajustados y de 0,0262 para los valores numéricos ajustados. Sin embargo, en la etapa de prueba el modelo neuronal alcanza un MAPE de 0,057 para los valores de PAC no ajustados y de 0,041 para los ajustados. Esta alternativa brinda una solución eficiente a la hora de resolver problemas de dosificación en las plantas de tratamiento de agua potable (PTAP), teniendo resultados confiables según las métricas RMSE y MAPE.

Palabras clave— Agua potable, Dosificación, PTAP, Químicos coagulantes, Redes neuronales artificiales, Sistema de control.

Recibido: 11-01-2023, Aprobado tras revisión: 14-06-2023

Forma sugerida de citación: Izquierdo, C.; Pezantes, B.; Ayala, E. (2022). "Prediction of the Optimal Dosage of Poly Aluminum Chloride for Coagulation in Drinking Water Treatment using Artificial Neural Networks".

Revista Técnica "energía". No. 20, Issue I, Pp. 93-99

ISSN On-line: 2602-8492 - ISSN Impreso: 1390-5074

Doi: <https://doi.org/10.37116/revistaenergia.v20.n1.2023.562>

© 2023 Operador Nacional de Electricidad, CENACE



1. INTRODUCTION

Currently there are a lot of industrial processes that demand a specific control method for its proper operation. Within this group of procedures requiring modernization and implementation of control systems, are processes related to water treatment, specifically drinking water treatment, which involves transforming raw water from natural sources into drinking water within the parameters established under specific standards for human consumption. A conventional drinking water treatment process consists of sequential steps and the most important one, is coagulation since it ensures that the dosed quantities of coagulating chemicals are in accordance with the properties of the raw water such as color, turbidity, pH and alkalinity [1]. Obtaining the doses of these chemicals reacts to a non-linear response done by experts. It is performed by jar testing, which is not adaptive to changes in real time and needs a considerable amount of time for its execution [1]. This is an issue since there is an immediate and constant dependence on qualified and experienced operators.

To ensure good quality of treated water, operators must adjust the amounts of coagulant chemicals at certain time intervals or climatic conditions where the water has parameters outside the usual range. Excessive amounts of coagulant chemicals correspond to increased treatment costs and public health problems. An under dosage corresponds to a failure in the flocculation of the water and increases the frequency of maintenance of DWTP increasing the cost of production. Moreover, an implementation of algorithms based on artificial intelligence and machine learning, which have been investigated and implemented in treatment plants around the world, is proposed.

For instance, [2] uses the potential provided by artificial neural networks, supporting vector machines, and gene expression programming to approximate the model of trihalomethane formation generated by chlorine water disinfection processes. They obtained as a result three models that capture the complex nonlinear behavior of the collected data. They also indicated excellent predictive and generalization capability. Furthermore, it demonstrated that these types of models, which commonly need a large amount of data, apply to a smaller amount of data. In another research, [3] artificial neural networks (ANN) are utilized to model the PAC dose. This method responds well when obtaining the appropriate dose in real-time when a storm brings high turbidity in raw water. In fact, they defined the input variables using Pearson's correlation and validated their model using the mean square error obtained.

In addition, [4] developed a model where the type of coagulant to be used is set by decision trees and the dose was estimated by ANN, allowing to calculate from the raw water parameters (pH, turbidity, and temperature), the amount and type of coagulant to be used (PAC, PASS and PSO-M). Moreover, [5] developed a model to predict

turbidity and color of treated water at the outlet of the Rossdale WTP located in Edmonton, Alberta in Canada. In 2009, [6] determined that the coagulant dose cannot be settled under traditional mathematical models, because it depends on several factors. Stating that the prediction of coagulant by neural network provides high accuracy and faster convergence speed and can be used to predict in real time online.

These types of neural models are seen as standard estimators of nonlinear relationships and their predictive and generalization capabilities let them have successful applications in different fields of knowledge [7].

On account of the above-mentioned research, the objective of this study is to build a neural model that adequately adapts to the relationship between raw water quality and the doses of chemicals needed for treatment. Initially, the data obtained involves a dosing history over a period of 14 months. The correlation between raw water quality and coagulant dosage was found. We will have to find a middle ground in the learning of our model in which we are not underfitting and not overfitting. For this problem, the input data set for training should be subdivided into two: one for training and one for the test that the model will not know beforehand. This division is usually made of 80% for training and 20%. The Test set should have diverse samples and enough samples to be able to check the results once the model has been trained. We proceeded with the training and validation process of the neural model by using the adaptive gradient algorithm and the analysis of the results using MAPE and RMSE. The results for the training set were an RMSE value below 2.82 and for the MAPE, a value of less than 0.045. On the other hand, for the test, set a RMSE value below 3.3 and 0.06 for the MAPE. It has been observed that the RMSE metric does not predict whether the estimation model is ideal or not. On the contrary, MAPE offers a better way to determine the accuracy of the model. The higher model accuracy is achieved when the value of MAPE is lower. Proving that the system had the ability to get information from dosing background and be able to estimate PAC doses for different raw water qualities.

The following paper is organized as follows: an overview of the water treatment process, determination of the variables, data analysis, construction and training of the neural model, analysis of results and conclusions.

2. METHODOLOGY

2.1. Determination of Process Variables

The conventional drinking water treatment process consists of 6 stages where the predominant process is dosing. This defines the success of the following stages. The process diagram of the drinking water treatment process is shown in Fig. 1.



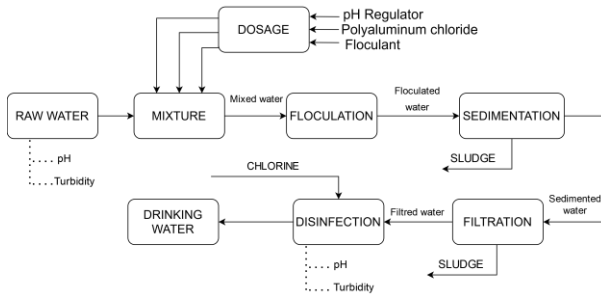


Figure 1: Conventional drinking water treatment process

The measurement of organoleptic and physical-chemical parameters is performed on the raw water. These measured values help in acquiring the amount of chemicals to be dosed for water coagulation. The amounts of chemicals can be tested by a jar test in the DWTP laboratory before dosing. Based upon the dosing process carried out at the DWTP, from the studies previously mentioned and the experience, the minimum and necessary amount of input variables involved in the process was defined. The input variables of the control system are pH, turbidity, and color of the raw water. These were selected through expert knowledge and were determined to be the minimum necessary to estimate the doses of chemicals considering the limited instrumentation existing in the DWTP. These variables can be quantified by means of sensors located at the inlet of the DWTP. The output variables are directly related to the amount of chemicals to be dosed in parts per million (ppm). Therefore, the control system in a generic way was shaped by the mentioned process variables and is structured in Fig. 2.

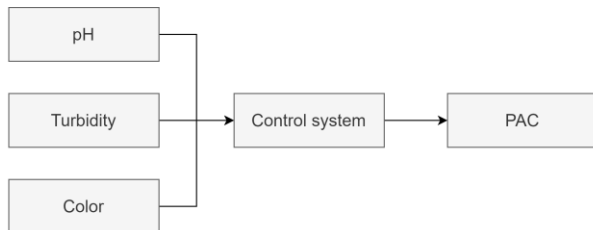


Figure 2: Control diagram for dosage

2.2. Data Collection and Analysis

The data obtained correspond to a water source that supplies one of the DWTP in Ecuador. This database contains the input parameters of the control system, which are the quantifiable characteristics of the water that took one year and two months of data collection. It started on October 1, 2017 and ended on December 31, 2018 with a total of 438 data points available. The raw water parameters of the raw water source are shown in Fig. 3 - Fig. 5.

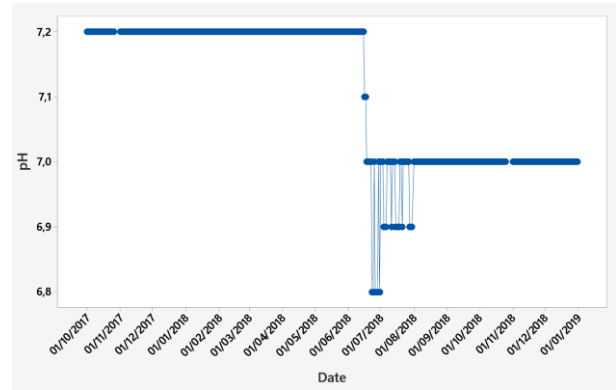


Figure 3: Raw water pH

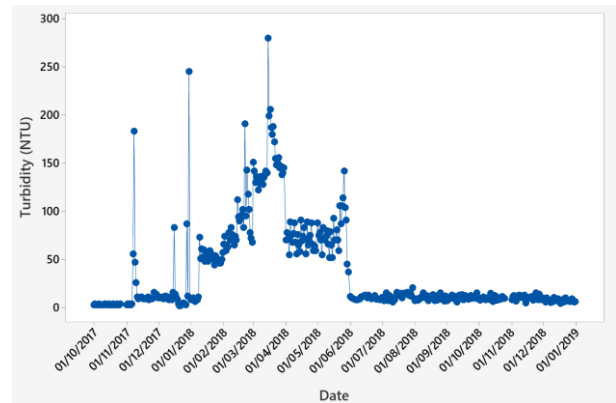


Figure 4: Raw water turbidity.

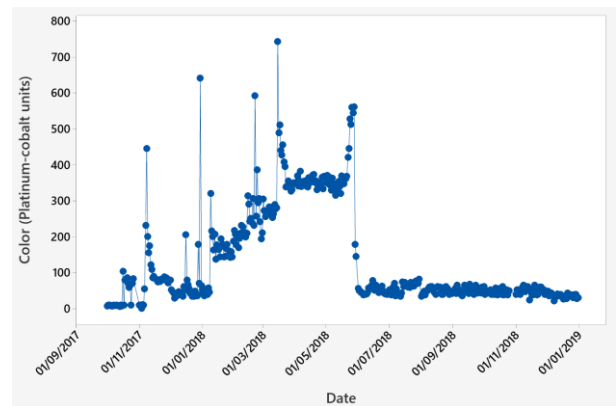


Figure 5: Raw water color

Fig. 6 shows the PAC dosages. The graph shows the dosage in normalized values. A linear scale between 0 and 1 was used to normalize the data using the following equation.

$$D_{normalized} = \frac{D - Min}{Max - Min} \quad (1)$$

Where D is the PAC dose, Max is the maximum PAC dose value, and Min is the minimum PAC dose value.

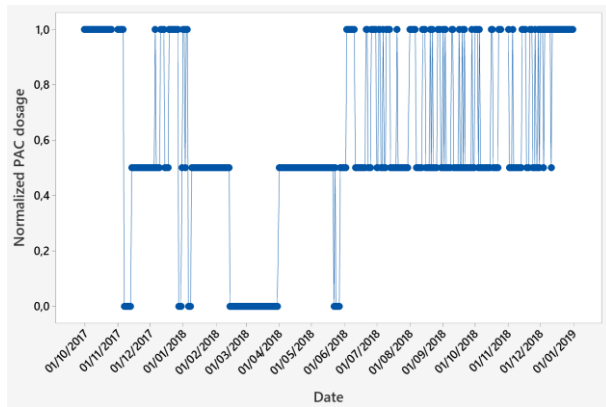


Figure 6: PAC Dosage

Pearson's correlation was used to determine the statistical relationship between the process variables. This is shown in Table 1.

Table 1: Pearson's Correlation Between each Input and Output

Pearson's correlation coefficient (r)				
	PAC	Color	Turbidity	pH
PAC	1	-0.63012	-0.70006	-0.32401
Color		1	0.89924	0.49197
Turbidity			1	0.46451
pH				1

The relationship between the PAC output and the Color, Turbidity and pH inputs is a negative association, i.e. as the raw water parameters increase their value for the amount of PAC doses decreases. It can also be said that this correlation between PAC doses and the parameters: Color and Turbidity is higher compared to the association between PAC and water pH. The color and turbidity of the water have a relatively high positive correlation, that means that as the color of the raw water increases its value so does the turbidity. Additionally, color and turbidity are positively related to pH but their relationship value is low. Finally, the coefficient of determination between the variables was obtained, which is illustrated in the following table:

Table 2: Determination Between each Input and Output

Pearson's correlation coefficient (r)				
	PAC	Color	Turbidity	pH
PAC	1	0.39706	0.49009	0.10498
Color		1	0.80864	0.24204
Turbidity			1	0.21577
pH				1

It is proved that the indicator of determination between the PAC, the variable to be predicted, and the Color, Turbidity and pH inputs is found to be less than

50% in each case. As a result, the model is a poor fit to its data. Consequently, the model belongs to a non-linear system and demands an intelligent control system that is suitable for the data.

2.3. Structure of the Neural Model

An Artificial Neural Network (ANN) represents a computer model based on the application of theoretical neurophysiology that replicates the way in which the human nervous system communicates and propagates. In [8] McCulloch & Pitts developed the first computer model that captures this work.

A Multilayer Perceptron (MLP) consists of a network architecture composed of one or several hidden layers, an interconnected system that examines information in a parallel but non-linear way, giving the ability to solve non-linear problems. The hidden layer makes a connection between each input and each output of the neural network, forming a fully connected or "dense" model. The information that reaches the input layer generates an activation pattern that in turn is an input signal applied to the neurons of the hidden layer [9]. If there is more than one hidden layer, the output signal of the first hidden layer is the input of the next hidden layer and so on until the output layer is accomplished. The matrix-expressed implementation of this algorithm called forward propagation is shown as follows:

$$\sum_{l=1}^n Z_l = W^{T[l]} A^{[l-1]} + B^{[l]} \quad (2)$$

$$A^{[l]} = f(Z)$$

Being l the number of layers of the MLP, f is the transfer activation function, $W^{[l]}$ is the current synaptic weights, $A^{[l-1]}$ represents the output of the previous layer, B is the bias vector and $A^{[l]}$ represents the output of the current layer.

An MLP is regularly trained with stochastic gradient descent methods [10], a technique in which the ANN parameters are updated at each iteration and the error is propagated backwards, updating the synaptic weights and decreasing the error in prediction.

In the present case study, the inputs of the ANN are the physical-chemical characteristics of the raw water: Turbidity, Color and pH. The output is the amount in parts per million of PAC.

Recordings of data corresponding to 438 dosages of the chemical agent PAC, with their respective input parameters, were used. The outcomes suggest that 80% of the data should be used for the training set, and the remaining 20% for the test set. Through the training process of the neural network, the best hyperparameters suit the model and allow the best accuracy in the prediction of the output were determined. Which are shown in the following table:



Table 3: Hyperparameters of the neural model

Dense Sequential Deep ANN			
Layer	Number of neurons	Number of parameters	Activation function
Input	3	-	-
Hidden 1	50	200	Sigmoid
Hidden 2	200	10200	Sigmoid
Hidden 3	50	10050	Sigmoid
Output	1	51	Linear
Total parameters: 20501 Trainable parameters: 20501 Untrainable parameters: 0 Learning rate: 0.1 Number of training epochs: 3000 Optimizer: Adaptive Gradient Algorithm (Adagrad) Root Mean Square Error (RMSE): 2.54			

The design of the deep ANN for PAC dosing is illustrated in the Fig. 7.

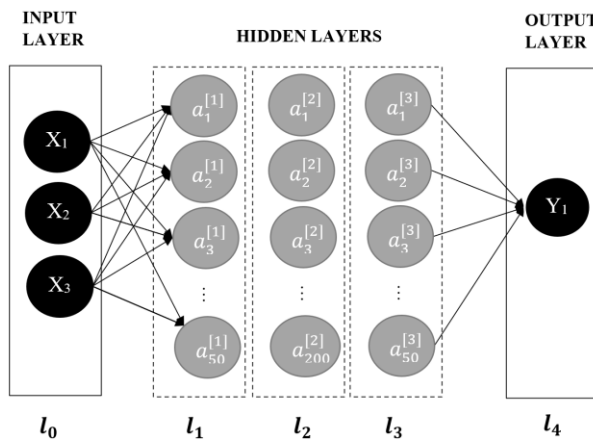


Figure 7: ANN Architecture

3. RESULTS AND DISCUSSIONS

Once the artificial neural model has been trained, a validation was performed using real vs. predicted data. For this, the data from the training set was initially employed, later the data for the test set was used, which represented new data for the neural model. The trained model has the ability to predict the PAC dose with the test data and with new data in real time.

It should be noted that the neural model predicts decimal values due to the activation functions used in its neurons, which may be found in the database. The PAC dosage values are integer values (ppm). Hence, the next step has been used to approximate each datum to its immediate superior in order to get the predicted doses correctly. This does not affect the output results because the dosages are made through doses with specific steps (50, 60, 70). Therefore this approach helps us to stay in the practical range of dosages. However, for their

representation, both results were considered as an illustration. The performance and precision of the neural model with respect to the real data of the dosages were through the Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE), metrics expressed by the following equations:

$$RMSE = \sqrt{\frac{\sum (X_i - Y_i)^2}{N}}$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{X_i - Y_i}{X_i} \right|$$

Where X_i is the measured value, Y_i is the predicted value, and N represents the number of samples. The results obtained in the training and testing phase are shown in the Table 4.

Table 4: Accuracy of the neuronal model

	RMSE	MAPE
Training set (Not adjusted values)	2.549	0.0425
Training set (Adjusted values)	2.816	0.0262
Test set (Not adjusted values)	3.141	0.057
Test set (Adjusted values)	3.254	0.0401

Fig. 8 and 9 show the performance of the Deep ANN through the comparative graphs of the real data vs. the data predicted by the neural network in the training set. In the same way, the performance of the Deep ANN with the test set is shown through the Fig. 10 and 11.

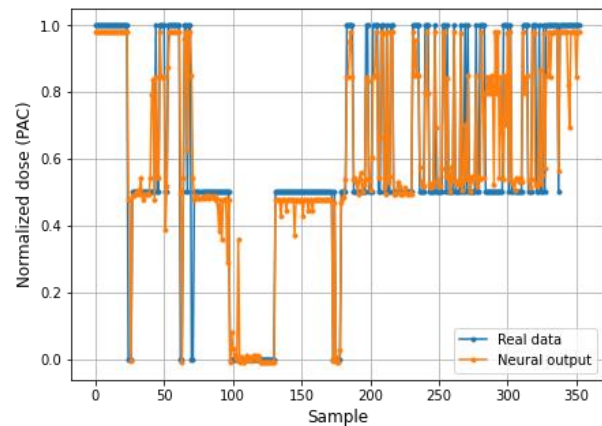


Figure 8: Training set (Not adjusted values)



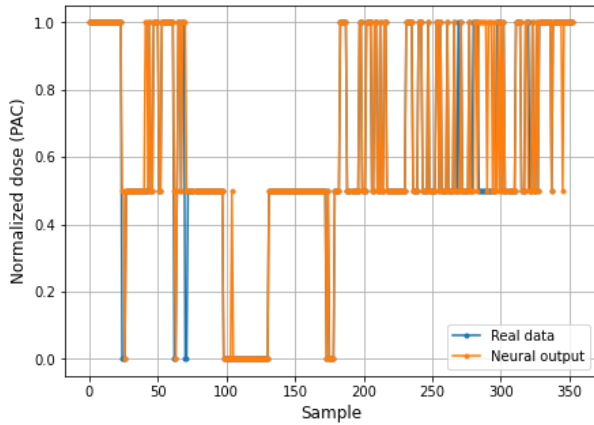


Figure 9: Training set (Adjusted values)

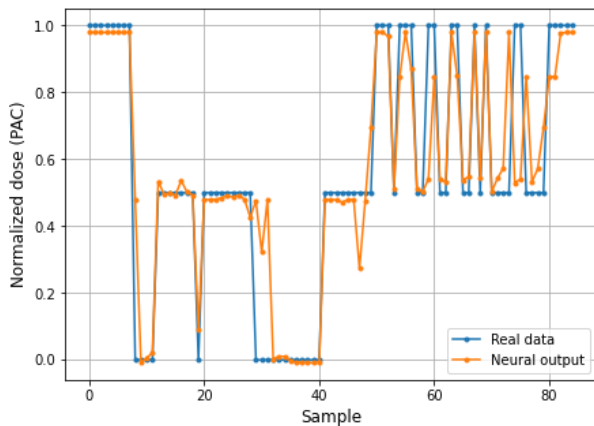


Figure 10: Test set (Not adjusted values)

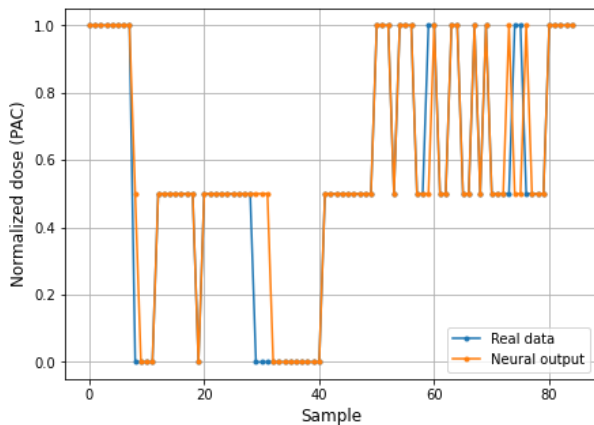


Figure 11: Test set (Adjusted values)

The main problem with the RMSE metric is that it does not predict whether the estimation model is good or not. MAPE offers a better way to determine model accuracy, while the value of MAPE is lower, then the model is more accurate.

As can be seen in Table 4, the PAC predictive model in the training phase achieves a MAPE value of 0.0425 for PAC values and 0.0262 for adjusted values. This

demonstrates that the predictions delivered by the neural model are quite accurate. However, in the testing phase, the neural model reaches a MAPE of 0.057 for PAC values and 0.041 for adjusted, which still displays the reliability and accuracy of the system.

4. CONCLUSIONS

In conclusion, this deep ANN gives an efficient solution to solve dosing problems that may occur in any DWTP. Within the DWTP, pH Regulator, PAC and flocculant are dosed, due to the focus of this project, a model with 3 inputs (pH, Turbidity, Color) - 1 output (PAC) was studied. Only the PAC will be used as output because it represents the minimum model that can be developed. To verify that the dosage is correct, pH, turbidity and color are measured at the outlet of the WTP (treated water) and check that the parameters are within the norm. The system automatically defines the appropriate dosage regardless of the parameters that are presented without the need to call or have the immediate help of a specialist. The various implementations of these systems have been studied within the state of the art in DWTPs in different parts of the world, and it has been possible to verify the success of this alternative in comparison to methods that do not use artificial intelligence.

This system can have a good reception in companies dedicated to the treatment of drinking water leading to answer the need of a possible booming market, which coincides with the implementation of automatic and intelligent systems in our country.

According to the metrics used, in the evaluation of the unadjusted training set, an RMSE of 2.549 and MAPE of 0.0425 were obtained. For the adjusted training data, an RMSE of 2.816 and a MAPE of 0.0262 were also acquired. As demonstrated in both cases of the training test, the MAPE is below 0.045, which reveals that the model is quite accurate.

Furthermore, for the unadjusted test set an RMSE value of 3.141 and a MAPE of 0.057 was obtained. For the adjusted data set an RMSE of 3.254 and MAPE of 0.0401 were obtained. As can be seen in the testing phase, for both cases, MAPE of less than 0.06 is examined, which still shows the efficiency of the ANN.

According to the aforementioned values it is concluded that the Deep ANN model is correctly adapted to the nonlinear behavior describing the chemical dosing processes from the parameters of the raw water discharge and has the ability to gain knowledge from a dosing history. It is mentioned that in the proposed treatment process, only the PAC doses are predicted, the doses of pH regulator and flocculant are calculated in the field with the knowledge of the expert without affecting the operation of the DWTP.

A comparison of the different neural models for the prediction of coagulant chemicals can be proposed as a



future study. In addition, another system can be implemented where the outputs are all the coagulating chemicals. In this case only one output (PAC) is used as a starting point for future projects.

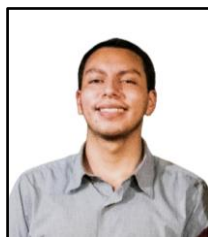
REFERENCES

- [1] J. A. Rojas, Potabilización del agua, Mexico, DF: ALFAOMEGA GRUPO EDITOR, 1999.
- [2] K. P. Singh and S. Gupta, Artificial intelligence based modeling for predicting the disinfection by-products in water, *Chemometrics and Intelligent Laboratory Systems*, vol. 114, pp. 122-131, 2012.
- [3] D. W. Guan and L. L. Shang, Effects of data normalization and inherent-factor on decision of optimal coagulant dosage in water treatment by artificial neural network, *Expert Systems with Applications*, vol. 37, n°7, pp. 4974-4983, 2010.
- [4] H. Bae, S. Kim and Y. Kim, Decision algorithm based on data mining for coagulant type and dosage in water treatment systems, *Water Science and Technology*, vol. 53, pp. 321-329, 2006.
- [5] Q. J. Zhang and S. J. Stanley, Real-time water treatment process control with artificial neural networks., *Journal of Environmental Engineering*, vol. 125, pp. 153-160, 1999.
- [6] Z. Song, Y. Zhao, X. Song and C. Liu, Research on prediction model of optimal coagulant dosage in water purifying plant based on neural network, 2009 ISECS International Colloquium on Computing, Communication, Control, and Management, vol. 4, pp. 258-261, 2009.
- [7] H. Yoon, S.-C. Jun, Y. Hyun, G.-O. Bae and Kang-Kun Lee, A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, *Journal of Hydrology*, vol. 396, pp. 128-138, 2011.
- [8] W. McCulloch S y W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The bulletin of mathematical biophysics*, vol. 5, n.º4, pp. 115–133, 1943.
- [9] A. Romero Herrera, "Aproximacion de funciones con redes neuronales y algoritmos evolutivos", Master's thesis, Universidad Autonoma del Estado de Mexico.
- [10] D. E. Rumelhart, G. E. Hinton y J. L. McClelland, "A general framework for parallel distributed processing", *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1, n.º 45-76, p. 26, 1986.



Cristopher Izquierdo Verdugo.-

He was born in Cuenca, Ecuador in 1998. He received his Mechatronic Engineer degree in Automation of Industrial Processes from the Salesian Polytechnic University in 2022; He did a student exchange through the ERASMUS+ program within the field of Artificial Intelligence at the University of Pisa, Italy in 2020; and his field of research is related to artificial intelligence and assistive technologies.



Braulio Pezantes Domínguez.-

He was born in Cuenca, Ecuador in 1998. He received his degree in Mechatronic Engineer in 2022 and currently, he is pursuing his Master studies in Electronics and Automation, both at the the Salesian Polytechnic University of Ecuador. His field of research is related to artificial intelligence, industrial automation and assistive technologies.



Edy Ayala Cruz.-

He was born in Cuenca in 1987. He received his degree in Electronic Technologist in 2009 and Electronic Engineer in 2011, both from the Salesian Polytechnic University of Ecuador; and his Master of Engineering Science (Electrical and Electronic) degree from Swinburne University of Technology in Australia, 2015. Currently, he is pursuing his Doctorate studies in Engineering at the University of Ferrara. His field of research is related to control systems in engineering and renewable energies.