International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:15, No:8, 2021

An Approach for Coagulant Dosage Optimization Using Soft Jar Test: A Case Study of Bangkhen Water Treatment Plant

Ninlawat Phuangchoke, Waraporn Viyanon, Setta Sasananan

***Abstract*—**The most important process of the water treatment plant process is coagulation, which uses alum and poly aluminum chloride (PACL). Therefore, determining the dosage of alum and PACL is the most important factor to be prescribed. This research applies an artificial neural network (ANN), which uses the Levenberg–Marquardt algorithm to create a mathematical model (Soft Jar Test) for chemical dose prediction, as used for coagulation, such as alum and PACL, with input data consisting of turbidity, pH, alkalinity, conductivity, and, oxygen consumption (OC) of the Bangkhen Water Treatment Plant (BKWTP), under the authority of the Metropolitan Waterworks Authority of Thailand. The data were collected from 1 January 2019 to 31 December 2019 in order to cover the changing seasons of Thailand. The input data of ANN are divided into three groups: training set, test set, and validation set. The coefficient of determination and the mean absolute errors of the alum model are 0.73, 3.18 and the PACL model are 0.59, 3.21, respectively.

***Keywords*—**Soft jar test, jar test, water treatment plant process, artificial neural network.

I. Introduction

WATER supply is fed to Bangkok by three water treatment plants (WTP), namely BKWTP, Mahasawat Water Treatment Plant (MSWTP), and Samsen Water Treatment Plant (SSWTP). All three use raw water from the Chao Phraya River. In this research, BKWTP is set as a case study. With the largest production of 4,400,000 cubic meters per day, there are 18 clarifiers equipped in two production lines with different chemicals (i.e., Alum and PACL) [1]. The giant BKWTP is a huge challenge in terms of operational cost-effectiveness. Optimal chemical dosages are required for optimum cost-effectiveness.

As a guideline for optimal chemical dosages, a traditional jar test has been used for a number of years, although there are a lot of disadvantages. Apparently, the behavior of being offline (labor process) among equipped online sensors (e.g., pH meter, flow meter) causes suboptimal operation since the operator cannot receive information on time, resulting in the so-called bottle-neck problem. In order to alleviate the offline problem, the virtual version of the Jar Test is proposed with the help of ANN modeling and the so-called Soft Jar Test (SJT).

Ninlawat Phuangchoke is a graduate student, Asst Prof. Dr. Waraporn Viyanon is a faculty member in the department of Computer Science, Faculty of Science, and Dr. Setta Sasananan is a faculty member in the department of Civil and Environmental Engineering, Faculty of Engineering, at Srinakharinwirot University in Bangkok, Thailand (e-mail: ninlawat.phu@g.swu.ac.th, waraporn@g.swu.ac.th, setta@g.swu.ac.th).

The primary goals of this research mainly are: (i) to set up ANN models (SJT) for chemical dosage prediction (i.e., alum and PACL based on Jar Test results; and (ii) to evaluate the performance and limitations of JST.

II. Water Treatment Process: BKWTP Case Study

The BKWTP's water treatment process is shown in Fig. 1. As shown in the figure, BKWTP receives raw water from Chao Phraya River at the Samlae pumping station and is conveyed through the 18-kilometer-long canal. The water is pumped into the filtration plant through a filter by rough and fine screens and then chlorine is added to kill germs and algae and to adjust the pH; this process is called the retreatment and pH adjustment process. After that, the water flows into the clarification process, Alum and PACL are added for the coagulation process to destabilize colloid and generate small floc. To increase floc concentration, a coagulation aid is included to form a large floc and it is able to precipitate (sedimentation) into the bottom of the clarifier. The water coming out of the clarifier tank is controlled for turbidity at no more than four Nephelometric Turbidity Units (NTU). The filtration process consists of two layers: coal and sand. The turbidity of the filtered water at this point is not more than 1 NTU. Finally, chlorine is added again to meet sanitation hygiene standards before pumping water to the public.

A. The Conventional Jar Test

Currently, at the BKWTP, the optimum dosage of alum and PACL is obtained by performing a traditional laboratory jar test. Fig. 2 illustrates the jar test equipment used in the method that has been in place since 1979. The objective of this procedure is to simulate three key processes: (1) coagulation; (2) flocculation; and (3) sedimentation. All these processes are physically demonstrated in one-liter containers with a varied set of agitation [2]. Optimal chemical dosage can be obtained by manually changing chemical doses and considering residual turbidity. In other words, it can be accounted as trial-error by an expert. In BKWTP, Jar testing is performed twice a day at approximately 8:00 a.m. and 4:00 p.m.

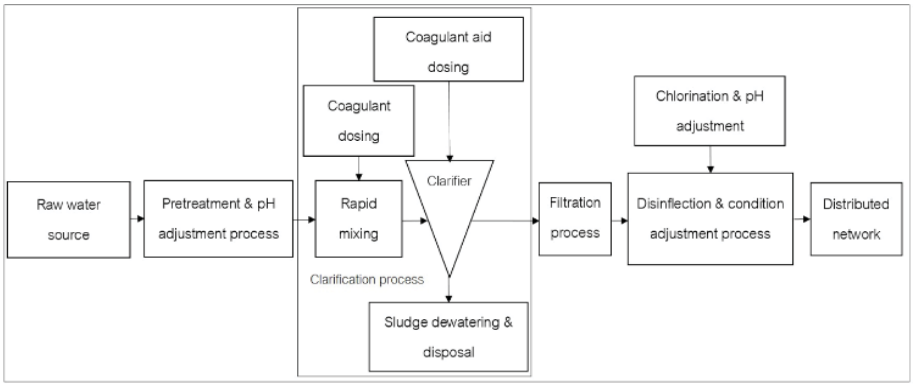


Fig. 1 Water treatment process [1]

The Jar test is considered a labor-consuming process and requires a skilled operator to promptly respond to any changes. Obviously, the jar test did not respond to any changes in the raw water on a real-time basis. On the other hand, when the Jar test is offline, there is a time lag as the test takes approximately 16-35 minutes, depending on the stirring speed of the water.

Fig. 2 Jar test equipment

B. Artificial Neural Network

ANN consists of a large number of processors connected to each other, and called neurons, as they are similar to biological neurons in the brain. The neurons are connected together and transmit signals from one neuron to another with weight in the neurons. Weight is responsible for long-term memory [3]. ANN learns by repeatedly adjusting the weight to make decisions close to humans' brains. ANN structure as shown in Fig. 3 contains components as follows:

(1) Input layer: The first layer of ANN consists of input neurons. This layer is the starting point for bringing input data into the system to the next layer of data processing.
(2) Hidden layer: The middle layers of ANN lie between the input and output layers, receiving data from the previous layer, and processing the data.
(3) Output layer: The last layer of ANN contains the output neurons, depending on the task of how many outputs there are. Usually, the output neurons are equal to the number of output variables used in the predictive model.

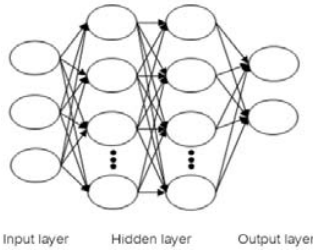


Fig. 3 Neural network architecture 3 layers

Our study proposes the SJT model using feed-forward back-propagation neural networks. When the neural network receives data and goes into a hidden layer, each input is multiplied by its weight value and processed by an activation function such as the sigmoid or tanh function. After that, the output comes out, which has calculated the error value to be an adapted weight in the training network to reduce error in the next iteration.

C. Data Collection & Filtration

The data collected in BKWTP used in this research cover the period from January 1, 2019 to December 31, 2019. A one-year span of data is recorded to cover the effects of seasonal changes in Thailand (i.e., summer, winter, and rainy seasons). All data concerned with the Jar test are categorized into two groups: (i) raw water quality; and (ii) chemical dosages. Raw water quality data include pH, alkalinity, turbidity, conductivity, and OC and the chemicals used, including alum and PACL. Some of the data can be collected by the sensors and some of them have to be collected manually. All parameters used are described in Table I, while the data collection schedule is illustrated in Table II. Due to the offline process of BKWTP, it must be fully upgraded to factory automation, and therefore most of the sensors are online. The sensors are often clogged and disrupted by a large amount of water production per day and sensor calibrations are regularly required. Mostly, data error arises from the data collected during the calibration period. In order to alleviate data error, the data filtration threshold was set according to interviews with the operators, as shown in previous works [1] and illustrated in Table III.

D.Data Analysis

All of the filtrated data were statistically analyzed and shown in Table IV. All the statistical parameters are within the expected range since errors are filtrated out. A plot of raw water turbidity, as graphically shown in Fig. 4, illustrated a change in raw water turbidity throughout the year. The highest turbidity was 126 NTU during the rainy season in September. The standard deviation and average turbidity were 14.86 and 27.42, respectively, indicating that there was a significant change in the turbidity of the raw water. In these situations, it is necessary to check the accurate predictions of the chemicals used in the plant more often, which is a limitation of the traditional jar test.

TABLE I
ALL PARAMETERS FOR THIS RESEARCH

Order	Parameter	Online/Offline	Description [4]
1	pH	Online	Measure acid/base of water.
2	Alkalinity (mg/L)	Online	Measure resistance changing aid of water.
3	Raw water/ dosed water Turbidity (NTU)	Online/Offline	Measure of the purity of the water, which may be visible to the naked eye. Therefore, turbidity is an important measure of water quality.
4	Conductivity (µs/cm)	Online	Measure electrical conductivity of the water.
5	OC (mg/L)	Online	Measure Biological oxygen demand (BOD_5), which is the amount of oxygen that bacteria and other microorganisms consume in a water sample during the period of 5 days at a temperature of 20°C to degrade the water contents aerobically.
6	Alum dosage	Offline	The mean of the most and least amount of alum used from the jar test divided by 2.
7	PACL dosage	Offline	The mean of the most and least amount of PACL used from the jar test divided by 2.

TABLE II
SCHEDULE TIME FOR MEASURE PARAMETERS AT BKWTP

Parameter	Time					
	0:00	4.00	8.00	12.00	16.00	20.00
Turbidity	X	X	X	X	X	X
pH	X	X	X	X	X	X
Conductivity	X	X	X	X	X	X
Alkalinity	X	X	X	X	X	X
OC	X	X	X	X	X	X

TABLE III
FILTER DATA AT BKWTP

Parameter	Criteria
Turbidity (NTU)	> 15
pH	6.8-8.5
Conductivity (µs/cm)	> 90
Alkalinity (mg/L)	35-125
OC (mg/L)	≤ 5.5

TABLE IV
STATISTIC OF RAW WATER PARAMETERS

Parameter	Max	Min	Mean	Median	Range	STD
Turbidity	126	15	27.42	24	111	14.86
Alkalinity	115	69	97.14	97	46	7.19
pH	8.2	7.25	7.7	7.7	0.95	0.16
Conductivity	3500	201	361.16	322	3299	237.1
OC	5.42	1.62	3.67	3.62	3.8	0.74
Alum	52.5	12.5	26.45	27.5	40	7.11
PACL	29	0	11.23	13	29	6.61

III. DATA RATIONALIZATION

Figs. 5 and 6 show the correlation coefficient of water quality parameters with the content of alum and PACL, respectively. It shows that the relationship of water quality parameters with the content of alum and PACL was non-linear. We can make a preliminary analysis that if the turbidity is high, a large amount of alum and PACL is required. On the other hand, alkalinity and pH are inversely proportional to the amount of alum and PACL.

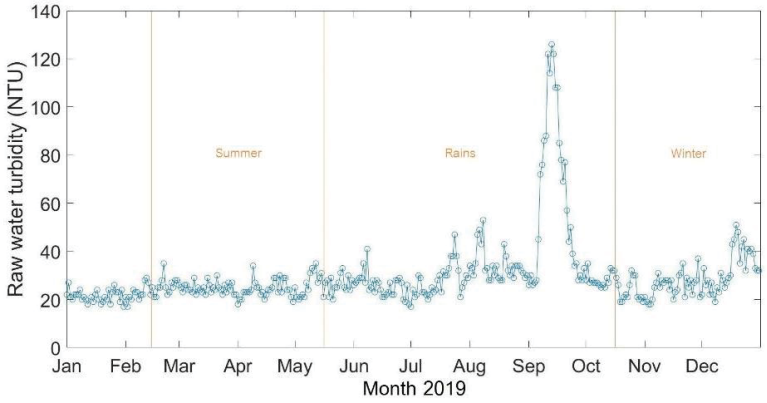


Fig. 4 Daily raw water turbidity

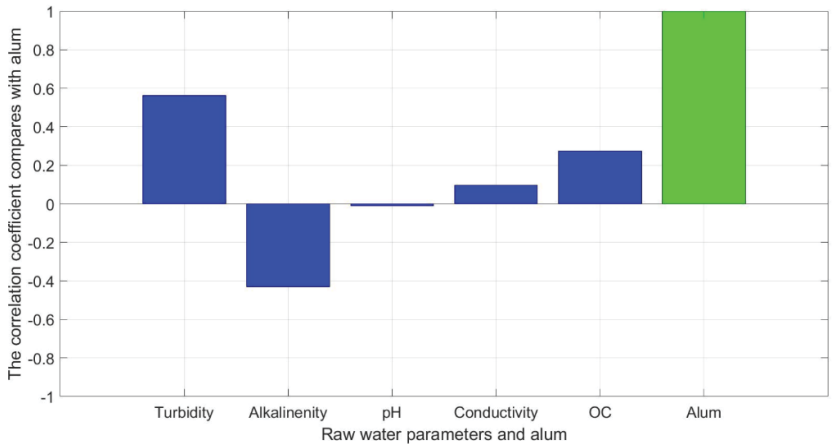


Fig. 5 Correlation coefficient with the alum

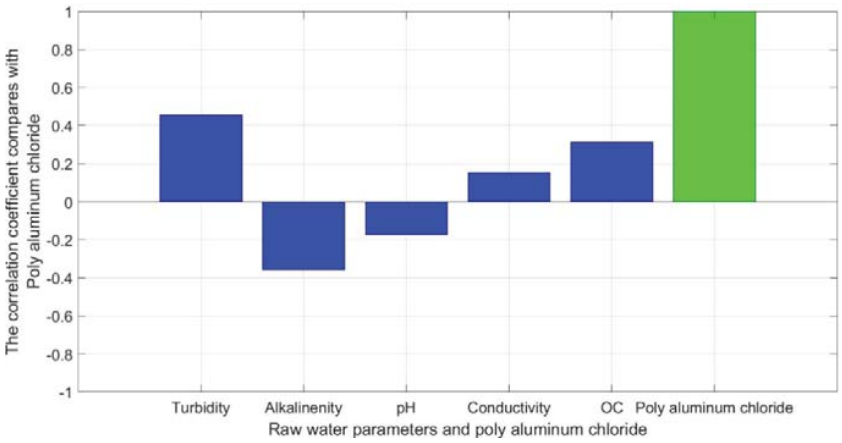


Fig. 6 Correlation coefficient with the PACL

TABLE V
INPUTS (I) AND OUTPUTS (O) FOR DEVELOPING ANN MODELS

ANN model	Raw water					Alum	PACL
	pH	Alkalinity	Turbidity	Conductivity	OC		
1	I	I	I	I	I	O	-
2	I	I	I	I	I	-	O

IV. SJT MODELLING

The modeling process of SJT consists of data collection and preparation, data partitioning, model architectural optimization, and model performance evaluation as shown in Fig. 7.

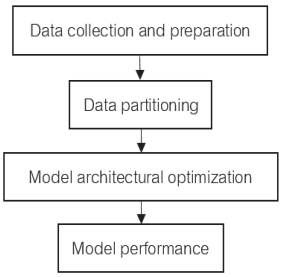


Fig. 7 Build SJT

A. Data Collection and Preparation

The process of data collection and preparation İS described in Section V, Operational Data and Section VI, Data Filtration.

Since the timing of raw water parameter measurement and the Jar test are not the same, raw water parameter measurements were performed every 4 hours, while the Jar test was performed twice a day at 8:00 a.m. and 4:00 p.m. To build a model, we used the same frequency range, and thus, matched the data to the Jar test period at 8.00 a.m. and 4.00 p.m. This means that the Jar test results at 4.00 PM. were more influential than those of 8.00 a.m.

B. Data Partitioning

At this stage, the data were randomly divided into three groups: training set, test set, and validation set at ratios of 2/3, 1/6, and 1/6 of the data respectively. The training set is for pattern recognition or learning. The comparisons between candidate models are conducted through a testing set. To prevent overfitting, the validation set is used with early stopping criteria [6].

C. Model Architectural Optimization

Optimal model architecture (i.e., number of layers and

number of hidden nodes) can be found using a systematic trial-error approach [5]. The best candidate is the one that gives the largest coefficient of determination (R^2) and the smallest Mean Absolute Error (MAE). In this research, no more than two hidden layers are investigated. On the other hand, [7] the trial-error process is conducted through a systematic change via increment of the number of neural nodes in each hidden layer starting with five nodes and increasing by five nodes at a time through 100 nodes (i.e., a total of 20 values (5, 10, 15, ..., 100)). An example of a two-hidden layer neural network architecture is shown in Fig. 8. In this research, there are two ANN models: (i) an ANN model for prediction of the alum dosage and, (ii) an ANN model for PACL prediction. The inputs and output for each model are shown in Table V.

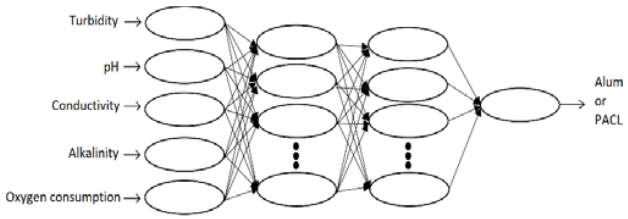


Fig. 8 An example architecture of two hidden layers

D.Model Architectural Optimization

All candidate ANN models are evaluated using the coefficient of determination (R^2) and Mean Absolute Error (MAE). R^2 is a measurement used to explain how much variability of one factor can be caused by its relationship to another related factor. This correlation, known as the "goodness of fit," is represented as a value between 0.0 and 1.0.

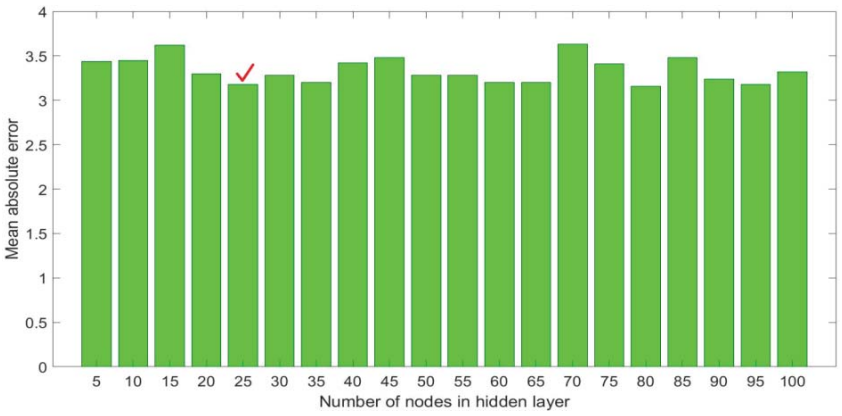
$$R^2 = 1 - \frac{RSS}{TSS} \tag{1}$$

where RSS = Sum of the square of residuals; TSS = Total sum of square.

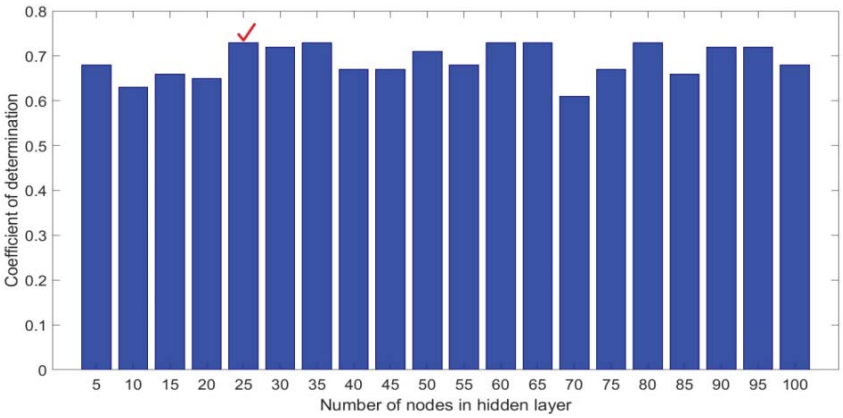
MAE measures the average magnitude of the errors in a set of predictions. It is the average over the test sample of the absolute differences between prediction and actual observation.

$$MAE = \frac{\sum_{i=1}^{n} | y_i - \hat{y}_i |}{n} \tag{2}$$

where $\hat{y}_i$ = Prediction value; y_i = True value; n = Total number of data points.

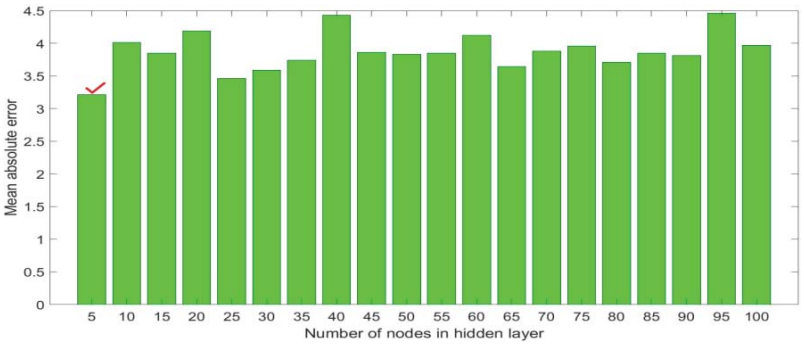


(a) MAE

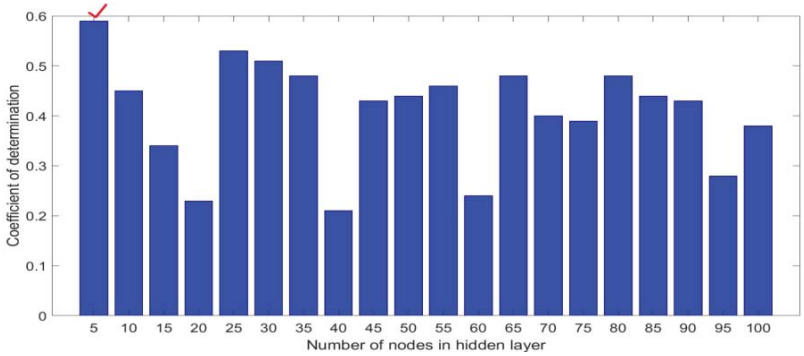


(b) R^2

Fig. 9 The alum prediction model with one hidden layer

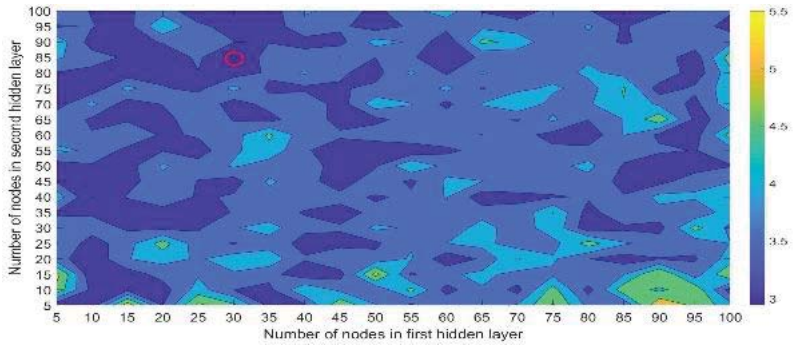


(a) MAE

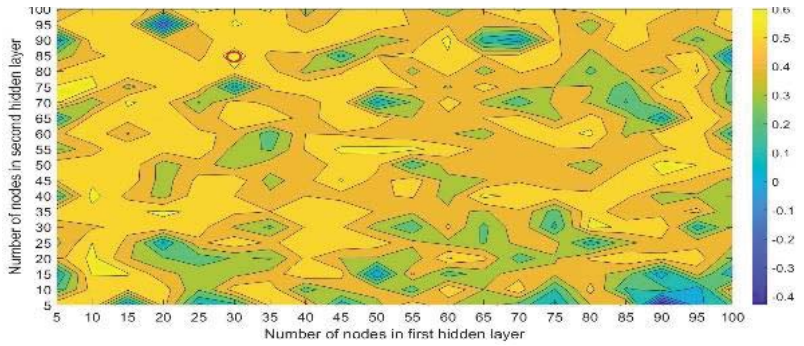


(b) R^2

Fig. 10 The PACL prediction model with one hidden layer

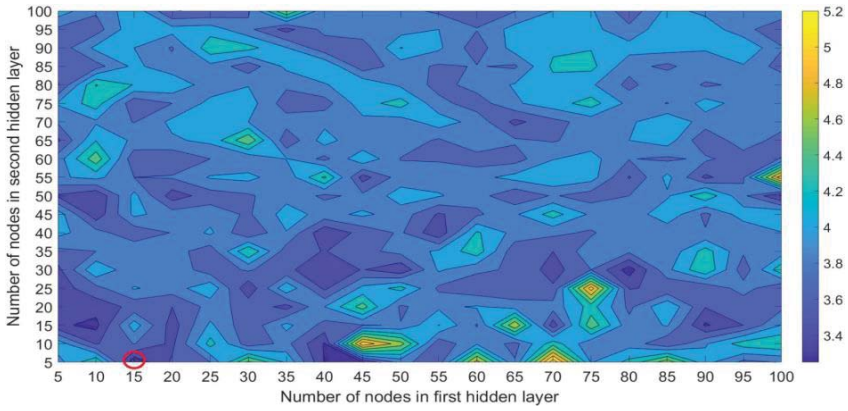


(a) MAE

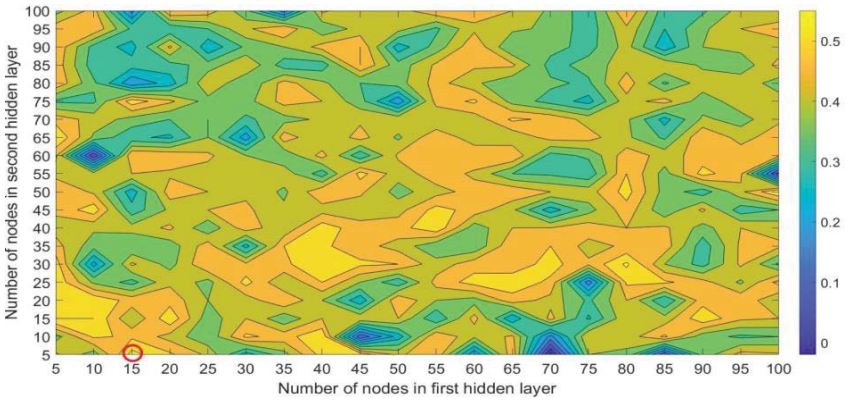


(b) R^2

Fig. 11 The model predicts alum with two hidden layers



(a) MAE



(b) R^2

Fig. 12 The model predicts PACL with two hidden layers

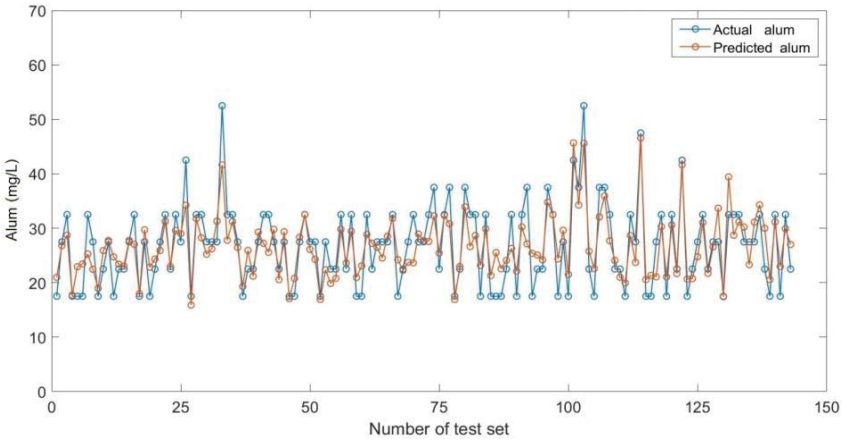


Fig. 13 Prediction of alum in the test set

V. Results and Discussion

The results of (a) MAE and (b) R^2 of the models are shown in Figs. 9 and 10. It can be seen that the model predicted the amount of alum with the lowest (a) MAE of 3.18 mg/L and the largest (b) R^2 value of 0.73 for the model using 25 nodes with one hidden layer shown in Fig. 9. On the other hand, Fig. 10 shows that the model predicted the amount of PACL with the lowest (a) MAE of 3.21 mg/L and the highest (b) R^2 value of 0.59 for the model using five nodes with one hidden layer.

For the two hidden layer type model, Fig. 11 illustrates a model performance evaluation for predicting the alum dosage, showing that the lowest (a) MAE was 2.94 mg/L and the largest (b) R^2 was 0.66 for the model with 30 nodes in the first layer and 85 nodes in the second layer or 30-85 for short. Fig. 12

shows the evaluation of the PACL prediction model with the maximum (b) R^2 of 0.57 and the lowest (a) MAE of 3.25 mg/L for the model (15-5). Figs. 13 and 14 illustrate the plots of the predicted and actual alum and PACL dosages on the test set, respectively.

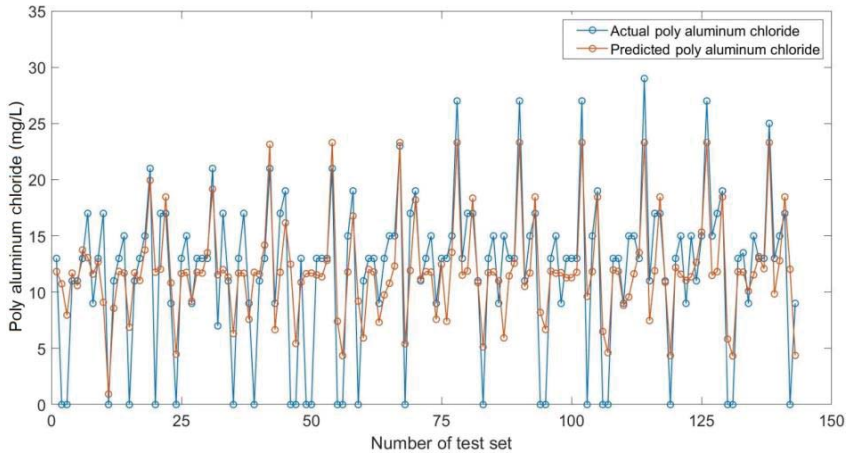


Fig. 14 Prediction of PACL in the test

The results show that the best candidate model for alum prediction is from two hidden layer (30-85) with R^2 = 0.66 and MAE = 2.94 mg/L. On the other hand, the best model for PACL was from five nodes with one hidden layer with R^2 = 0.59 and MAE = 3.21 mg/L.

From the performance evaluation, it was found that both ANN models were able to simulate the jar test; however, the alum model performed better than the PACL one. These models still need to be improved because of the relatively small R^2 and large MAE. It can be seen that the MAE of alum is about 12 percent compared to the actual mean of 26.45 mg/L as shown in Table IV and for PACL it is about 28% compared to the actual mean of 11.45 mg/L, as shown in Table IV.

If a 10% error is set as a threshold [1] (i.e., the calibration error is about 10%), the alum model has failed slightly and the PACL model has failed entirely. This is because the data feed to the model was randomly divided into three groups: training set, test set, and validation set with a ratio of 2/3, 1/6, and 1/6, respectively. Even though the training group is the largest one with random selection, there is no guarantee that all test and validation sets are subsets of the training set (i.e., the range of training set covers all that of test and validation sets). In general, the ANN model promises better performance in interpolation which is more preferable to the extrapolation range [1]. These random selection results in small R^2 and large MAE of alum models and even worse in the narrow range of the inputs PACL models. Therefore, using the ANN model to replace the traditional Jar test requires a larger range of training set in order to fully educate the model. The larger the range of the training set can ensure that the prediction mode works in a corrected fashion.

VI. Conclusion

Using the ANN model instead of the traditional Jar test still has several aspects to explore and improve. It is found that the predictive errors are about 12% and marginally fail to meet the threshold of 10%. However, if the errors of the predicted models are tolerated, the SJT model can be put into action.

Additionally, this approach is worthwhile for further development. The model can be optimized using clustering techniques to ensure that the ANN model operates in the range called interpolation or using multiple models in predictions.

Acknowledgments

We appreciate the Bangkhen Water Treatment Plant (BKWTP), Metropolitan Waterworks Authority for the data used in this research. This publication of this research was funded by the Graduate School at Srinakharinwirot University [budget-year 2564].

References

[1] S. Sasananan, "Water Treatment Plant Clarifier Control: An Artificial Intelligence Approach," Doctoral dissertation, University of Tasmania, Australia, 2009.
[2] E. E. Arasmith, Jar Test. Operational Control Tests for Wastewater Treatment Facilities. Instructor's Manual (and) Student Workbook. Linn-Benton Community College, 1981.
[3] M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed. New York: Addison Wesley, 2011.pp. 205-215
[4] O. Degremont, Water Treatment Handbook, 6th ed. France, 1991.
[5] H. Maier, N. Morgan, and C. Chow, "Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters," Environmental Modelling and Software, vol. 19, pp. 485-494, 05/01 2004, doi: 10.1016/S1364-8152(03)00163-4.
[6] Mark Hudson Beale, Martin T. Hagan, and H. B. Demuth, Deep Learning Toolbox™ User's Guide. 2020.
[7] W. Kuanthong, W. Liamlaem, and S. Sasananan, "Clarifier Models using Artificial Neural Networks - Case Study: Bangkhen Water Treatment Plant," SWU Engineering Journal, vol. 10, no. 1, pp. 32-44, 2015.