Thai Environmental Engineering Journal Vol. 34 No. 1 (2020) : 1-8
www.eeat.or.th

# Alum Dosage Reduction and Sensitivity Analysis in Water Treatment System using Data Mining Software: Case study of Provincial Waterworks Authority, Udonthani, Thailand

**Jintawat Lachinla and Petchporn Chawakitchareon***
Environmental Engineering Department, Faculty of Engineering,
Chulalongkorn University, Bangkok, 10330, Thailand
*E-mail : petchporn.c@chula.ac.th

## Abstract

This research applied Rapid Miner V.9.2 for alum dosage reduction and sensitivity analysis in water supply system of PWA, Udonthani, Thailand. The input parameters were pH and turbidity of raw water, pH and turbidity of pre-filtered water. The output parameter was alum dosage. The data were used from October, 2004 to April, 2019 that collected 5,118 records. The theory used W-Linear Regression W-MLP W-REPTree W-M5P W-M5Rules and GBT for modeling, alum dosage prediction, apply to alum dosage reduction and sensitivity analysis. From all 24 scenarios experiment, in conclusion, 10 models could be the alum dosage prediction. When applied to reduce the alum using dosage and sensitivity analysis, it was found model can the most alum reduction was model in summer by W-M5P theory and model in winter by GBT. All two models were used in Banthon WTP to reduce the alum dosage up to 21.69 percentage per year or 243,230 baht per year. The input parameters affected the most sensitive model that were pH and turbidity of raw water, pH of pre-filtered water. Therefore, this model could be applied to reduce cost of alum for PWA, Udonthani.

**Keywords :** Prediction; Raw Water Quality; Alum; Chemical; Water Treatment System; Sensitivity

## Introduction

Tap water has a need for our daily life such as consumption, using tap water in product process, cleaning of raw material, so tap water has increase demand every day. Consequently, tap water has important in every section. Water supply system includes raw water source, water treatment system, water transportation system and water distribute system [1]. In Thailand, the most water supply system includes coagulation, flocculation, sedimentation, filtration and disinfection. For Provincial Waterworks Authority, Udonthani, it is the surface water supply system (conventional) that it is used commonly in large community [2, 3]. Raw water source uses from river, canal, reservoir or dam because there is more water volume for the water demand. Most of the raw water is high turbidity; then, the water treatment system uses alum for turbidity reduction of raw water [3-5]. The data of Provincial Waterworks Authority, Udonthani is collected from October, 2013 to April, 2019. It finds that much volume of alum using in water treatment system because the pre-filtered water was 0.5 to 6.5 NTU. Most of pre-filtered water was 0.5 to 2 NTU (65% of all data) which the alum using was over necessary. The pre-filtered water criteria of Provincial Waterworks Authority must be less than 10 NTU [6]. If Provincial Waterworks Authority, Udonthani must reduce the volume of alum using, they will must control the pre-filtered water approximately 10 NTU. As a result, this research aimed to alum dosage reduction and sensitivity analysis in water treatment system using data mining software: Case study of Provincial Waterworks Authority, Udonthani by using RapidMiner V.9.2.

Previous research tried to use the WEKA program to predict alum concentration demand. The study used four prediction methods namely Multilayer Perception, M5P, M5Rules, and REPTree with six parameters as an input including turbidity, water hardness, pH, electroconductivity, color, and, total suspended solids. The best result from the research was the M5Rules method, which was then used to create an alum concentration predicting model that helped reduced time used and limitations in jar test [7]. Another research study the relationship between pH and the amount of lime used using Regression theory, which was done using pH as an input into the Matlab program. The result was an 89 percent effective correlation equation between pH and the amount of lime used [8]. Moreover, researcher studied the weight-based growth model for Nile tilapia in waste water treatment pond without feeding. It was growth model for forecasting of tilapia weight and sensitivity analysis of model for change the recording of important information. The result was that most sensitivity variables was pH value [9]. Consequently, from literature review, it was found that parameters that affect alum dosage, Modeling methods, and the sensitivity analysis of the model.

## Methodology

**Searching, collecting and requesting information**

This research's data was data of Provincial Waterworks Authority, Udonthani from October, 2013 to April, 2019 included 1) daily report of raw water quality 2) daily report of process water quality and 3) daily report of jar-test.

**The input parameters, output parameters and data preparation**

From research review about the chemical dosage prediction and the coagulation process, it was found that the coagulation process was to remove turbidity from raw water. The coagulation process occurred that the pH after the coagulation process was neutral [1]. In this research, the researchers considered pH and turbidity of pre-filtered water. So, the researcher had determined the input parameters which were pH of raw water, pH of pre-filtered water, turbidity of raw water (NTU) and turbidity of pre-filtered water (NTU). The output parameter was alum dosage (mg/l). Data preparation was two excel files, as follows.

The set data $1^{st}$ was used for parameter adjustment each theory, modeling for alum dosage prediction in 24 cases that consideration was overall of three WTP, each WTP, each raw water source and each season. The set data $1^{st}$ was 4,029 data since October, 2013 to April, 2018.

The set data 2nd was used for the alum dosage prediction verification in water supply system in 24 cases. The set data 2nd was 1,089 data since May, 2018 to April, 2019. In both excel data file, the researchers cut the set data that some input parameters were distorted or missing. In addition, the data set couldn't have all data for reduction the bias that would occur [10-11, 14].

## Theory of experiment

There are six theories in total. W-Linear Regression is a linear relationship between input and the modeled parameters where the output is continuous [12]. W-Multilayer Perceptron is a multilayer perceptron where the first layer received and calculated the input and weight of the incoming data, then transfer it to the next layer [13]. W-REPTree create a tree-like structure from information gained, variance, and cutting, which is similar to the C4.5 theory, but REPTree is less prone to errors and faster [14]. W-M5P is also a tree-like structure, but with linear regression function replacing the last leaf of the model. W-M5P is a way to predict numerical data. The value of the node weighs less than the variable resulting in a reduced error rate, which makes this theory the most popular. Lastly, W-M5Rules, based on the M5 theory, is a theory where a regression term in the form of If-Then Rules is created. [15] Gradient boosted trees (GBT) is a boosting algorithm using decision tree as weak learners. So, the best of GBT is the ability to learn from both discrete and continuous data due to its tree-based structure. [16]

## Experiment

First, this research used the Rapid Miner V.9.2 program to adjust parameter each theory such as W-Linear Regression, W-MLP, W-REPTree, W-M5P, W-M5Rules and gradient boosted tree (GBT), so considered each of all 24 cases. Second, the optimal parameter each theory was applied to model creation for alum dosage prediction. Then, the model was adjusted parameter each theory to the operator addition, such as Retrieve, Filter example, Select attributes and Apply model. As show in Figure 1, for alum dosage prediction, alum dosage reduction and sensitivity analysis in water treatment system of Provincial Waterworks Authority, Udonthani.

After that, we applied all model for alum dosage reduction in Provincial Waterworks Authority, Udonthani, which turbidity of pre-filtered water was selected that is 2 to 6.5 NTU. Finally, sensitivity analysis of all four parameters would have an effect that how much alum was change (mg/l). This consideration would change each parameter to ±5 ±10 ±20 ±50 and ±90 percent respectively, so the result was measured from MAE of decreased variable, increased variable and total. If the average of MAE is high, the model is very sensitive. Nevertheless, if the average of MAE is low, the model is less sensitive.

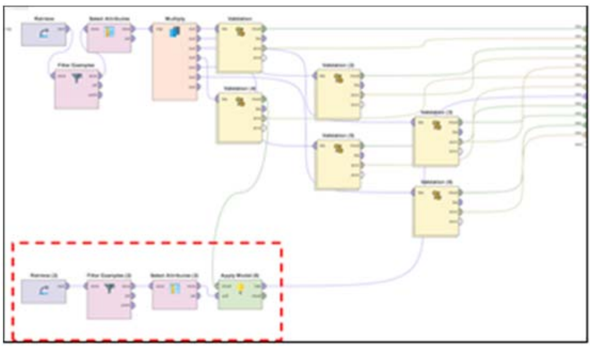

**Figure 1** The model for alum dosage prediction, alum dosage reduction and sensitivity analysis

**Matrix in determining the effectiveness of the model**

1. Root Mean Square Error is a value of error or difference between the predicted data and the actual data. The closer the RMSE is to 0 the more accurate the model is. The equation for the RMSE is shown in equation (1) [15].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x)^2}{n}} \qquad (1)$$

2. Mean Absolute Error is a mean value of the absolute error value of the predicted data. The closer it is to 0, the more accurate the model is. The equation to calculate MAE is shown in equation (2) [14].

$$MAE = \sum_{i=1}^{n}\frac{|x_i - x|}{n} \qquad (2)$$

$x_i$ = Actual alum (mg/l)
$x$ = Predicted alum (mg/l)
$n$ = Number of data

3. Consider the amount of predicted values set within the error ±5 mg/l more than 80% in both parts. Consideration of the best model is the least RMSE and MAE from all six theories of each case. Unless it has not the best model, we must consider the amount of predicted values set within the error ±5 mg/l more than 80% in both parts.

## Results and Discussion

From modeling and the optimum alum dosage prediction, all 24 formats (1,440 cases), the best efficiency model for each format was showed in table 1. The result was found that 10 models had passed the criteria and the best qualities each case.

From Figure 2, these was turbidity of pre-filtered water in Provincial Waterworks Authority, Udonthani from May, 2018 to April, 2019 that turbidity of pre-filtered water was 0.5 to 6.5 NTU. Most of turbidity of pre-filtered water was 0.5 to 2 NTU that was 65 percent of all data, so these was the overdose alum using in coagulation, flocculation and sedimentation. Since Provincial Waterworks Authority defined criteria for turbidity of pre-filtered water less than 10 NTU [6]; therefore, 10 models were applied for alum dosage reduction in Provincial Waterworks Authority, Udonthani. Turbidity of pre-filtered water was selected that was 2 to 6.5 NTU, which was approximately 35 percent. So, it was not to affect the model efficiency.

**Table 1** The model was passed the criteria each case

| Model | Condition | Theory | RMSE | MAE |
|---|---|---|---|---|
| 1 | Banthon WTP (Format 1-3) | GBT | 4.784 | 3.259 |
| 2 | Bannikom WTP (Format 1-4) | GBT | 3.949 | 2.751 |
| 3 | Raw water from Huailuang dam (Format 1-5) | GBT | 4.886 | 3.338 |
| 4 | 3 WTP (summer) (Format 1-7) | W-M5P | 5.407 | 4.015 |
| 5 | Banthon WTP (summer) (Format 1-13) | W-M5P | 4.140 | 3.208 |
| 6 | Banthon WTP (rains) (Format 1-14) | W-M5P | 3.651 | 1.642 |
| 7 | Banthon WTP (winter) (Format 1-15) | GBT | 3.306 | 2.797 |
| **8** | **Bannikom WTP (summer) (Format 1-16)** | **GBT** | **2.049** | **1.264** |
| 9 | Bannikom WTP (winter) (Format 1-18) | GBT | 3.351 | 2.286 |
| 10 | Raw water from Huailuang dam (winter) | W-M5P | 4.615 | 1.968 |

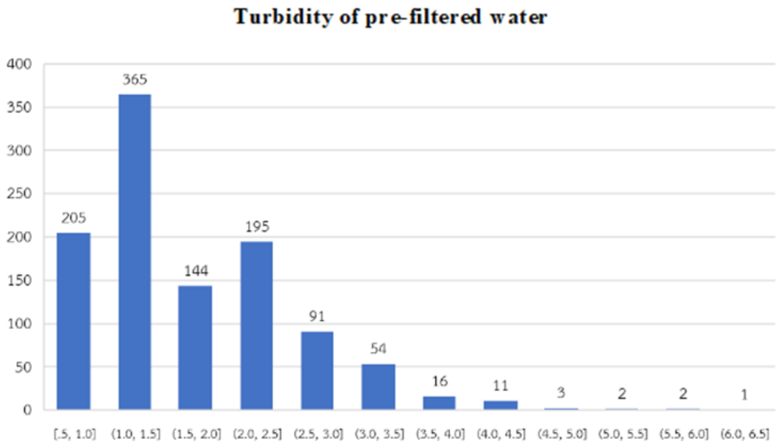


**Figure 2** The turbidity graph of pre-filtered water (May, 2018 to April, 2019)

**Table 2** Comparison between the actual and predicted alum dosage

| Model | Actual Alum | | | Alum Prediction | | | ±Price (Baht) | ± percent |
|---|---|---|---|---|---|---|---|---|
| | Dosage (kg.) | Avg. per day (kg./day) | Price (Baht) | Dosage (kg.) | Avg. per day (kg./day) | Price (Baht) | | |
| 1 | 305,339 | 843 | 1,885,255 | 269,703 | 745 | 1,665,227 | -220,028 | -11.67 |
| 2 | 290,094 | 795 | 1,791,126 | 263,613 | 722 | 1,627,628 | -163,498 | -9.13 |
| 3 | 621,408 | 667 | 3,836,758 | 591,937 | 635 | 3,654,794 | -181,964 | -4.74 |
| 4 | 152,505 | 571 | 941,613 | 145,568 | 545 | 898,783 | -42,830 | -4.55 |
| 5 | 74,782 | 840 | 461,724 | 62,121 | 698 | 383,554 | -78,170 | -16.93 |
| 6 | 123,695 | 825 | 763,730 | 163,335 | 1,089 | 1,008,477 | 244,747 | 32.05 |
| 7 | 106,862 | 869 | 659,801 | 80,129 | 651 | 494,743 | -165,058 | -25.02 |
| 8 | 65,674 | 738 | 405,489 | 73,042 | 821 | 450,981 | 45,492 | 11.22 |
| 9 | 91,394 | 743 | 564,293 | 87,035 | 708 | 537,380 | -26,913 | -4.77 |
| 10 | 203,356 | 711 | 1,255,582 | 199,548 | 698 | 1,232,070 | -23,513 | -1.87 |

Annotation: Price of Aluminium Sulphate was 6,174.30 baht per metric ton (excluding vat)

From table 2, comparison between the actual alum dosage and the predict alum dosage showed that.

- Model 1st used Banthon WTP data that alum dosage reduced 11.67 percent per year or 220,028 baht per year.
- Model 2nd used Bannikom WTP data that alum dosage reduced 9.13 percent per year or 163,498 baht per year.
- Model 3rd used raw water data from Huailuang dam that alum dosage reduced 4.74 percent per year or 163,498 baht per year.
- Model 4th used all three WTP (summer) data that alum dosage reduced 4.55 percent per year or 181,964 baht per year.
- Model 5th used Banthon WTP (summer) data that alum dosage reduced 16.93 percent per year or 78,170 baht per year.

- Model 6th used Banthon WTP (rains) data that alum dosage increased 32.05 percent per year or 244,747 baht per year.
- Model 7th used Banthon WTP (winter) data that alum dosage reduced 25.02 percent per year or 165,058 baht per year.
- Model 8th used Bannikom WTP (summer) data that alum dosage increased 11.22 percent per year or 45,492 baht per year.
- Model 9th used Bannikom WTP (winter) data that alum dosage reduced 4.77 percent per year or 26,913 baht per year.
- Model 10th used raw water data from Huailuang dam (winter) that alum dosage reduced 1.87 percent per year or 23,513 baht per year.

In summary, Model 1st, 2nd, 3rd, 4th, 5th, 7th, 9th and 10th can reduce the actual alum dosage. Model 5th uses Banthon WTP (summer) data that uses W-M5P theory to modeling and Model 7th uses Banthon WTP (winter) data that uses GBT theory to modeling. Consequently, both the model is used in Banthon WTP that can reduce the most actual alum dosage 21.69 percent per year or 243,230 baht per year. However, Model 10th uses raw water data from Huailuang dam (winter) that uses W-M5P theory to modeling, so Model 10th can reduce the least actual alum dosage 1.87 percent per year or 23,513 baht per year. On the other hand, Model 6th and 8th cannot reduce the actual alum dosage.

**Table 3** Sensitivity Analysis each model

| Model | Input parameters | MAE | | |
|---|---|---|---|---|
| | | Decreased parameter | Increased parameter | Total |
| 1 | pH of raw water | 11.14 | 22.44 | 16.79 |
| | pH of pre-filtered water | 16.64 | 6.06 | 11.35 |
| | Turbidity of raw water | 5.25 | 6.98 | 6.11 |
| | Turbidity of pre-filtered water | 2.71 | 2.42 | 2.57 |
| 2 | pH of raw water | 1.59 | 6.52 | 4.06 |
| | pH of pre-filtered water | 5.35 | 28.41 | 16.88 |
| | Turbidity of raw water | 6.45 | 7.20 | 6.83 |
| | Turbidity of pre-filtered water | 3.69 | 8.60 | 6.14 |
| 3 | pH of raw water | 5.64 | 8.75 | 7.20 |
| | pH of pre-filtered water | 11.59 | 4.66 | 8.12 |
| | Turbidity of raw water | 3.58 | 4.66 | 4.12 |
| | Turbidity of pre-filtered water | 3.89 | 4.24 | 4.07 |
| 4 | pH of raw water | 4.93 | 21.03 | 12.98 |
| | pH of pre-filtered water | 7.43 | 38.85 | 23.14 |
| | Turbidity of raw water | 2.18 | 5.84 | 4.01 |
| | Turbidity of pre-filtered water | 0.89 | 0.70 | 0.80 |
| 5 | pH of raw water | 16.09 | 16.66 | 16.38 |
| | pH of pre-filtered water | 16.51 | 16.09 | 16.30 |
| | Turbidity of raw water | 17.11 | 11.62 | 14.36 |
| | Turbidity of pre-filtered water | 16.94 | 14.69 | 15.82 |

| Model | Input parameters | MAE | | |
|---|---|---|---|---|
| | | Decreased parameter | Increased parameter | Total |
| 6 | pH of raw water | 36.13 | 99.00 | 67.57 |
| | pH of pre-filtered water | 18.95 | 37.02 | 27.99 |
| | Turbidity of raw water | 34.51 | 47.25 | 40.88 |
| | Turbidity of pre-filtered water | 35.59 | 36.33 | 35.96 |
| 7 | pH of raw water | 14.69 | 40.54 | 27.61 |
| | pH of pre-filtered water | 9.30 | 6.36 | 7.83 |
| | Turbidity of raw water | 13.59 | 10.85 | 12.22 |
| | Turbidity of pre-filtered water | 11.91 | 9.56 | 10.73 |
| 8 | pH of raw water | 9.35 | 8.97 | 9.16 |
| | pH of pre-filtered water | 7.68 | 9.82 | 8.75 |
| | Turbidity of raw water | 10.34 | 10.50 | 10.42 |
| | Turbidity of pre-filtered water | 9.75 | 9.39 | 9.57 |
| 9 | pH of raw water | 3.15 | 2.58 | 2.87 |
| | pH of pre-filtered water | 16.62 | 13.30 | 14.96 |
| | Turbidity of raw water | 0.66 | 0.46 | 0.56 |
| | Turbidity of pre-filtered water | 2.26 | 7.23 | 4.75 |
| 10 | pH of raw water | 4.14 | 20.77 | 12.45 |
| | pH of pre-filtered water | 12.24 | 13.67 | 12.96 |
| | Turbidity of raw water | 21.26 | 13.20 | 17.23 |
| | Turbidity of pre-filtered water | 17.64 | 8.91 | 13.27 |

From table 3, sensitivity analysis uses 10 models that is the best performance in each case. To sum up, the pH of raw water is the most sensitive to model 1st, 5th, 6th and 7th. The pH of pre-filtered water is the most sensitive to model 2nd, 3rd, 4th and 9th. Moreover, the turbidity of raw water is the most sensitive to model 8th and 10th.

## Conclusion

Model can use to alum dosage prediction in water supply system of Provincial Waterworks Authority, Udonthani. So, from all 24 formats (1,440 cases), the best efficiency model for each format. The result finds that 10 models have passed the criteria and the best qualities each case. For the alum dosage reduction in water treatment system of Provincial Waterworks Authority, Udonthani using data mining software, the result finds Model 5th uses Banthon WTP (summer) data that uses W-M5P theory to modeling and Model 7th uses Banthon WTP (winter) data that uses GBT theory to modeling. Consequently, both the model is used in Banthon WTP that can reduce the most actual alum dosage 21.69 percent per year or 243,230 baht per year. For the sensitivity analysis, the result finds three parameters that are the most sensitive to model, such as pH of raw water, pH of pre-filtered water and turbidity of raw water. As a result, the operator must accurately record data and verify the accuracy of water quality instruments every six months to data isn't error.

## Acknowledgment

The author would like to thank the Provincial Waterworks Authority for their financial support and providing raw data information's. This research is also in part supported by Graduate School of Chulalongkorn University.

## References

[1] Tanthunwet, M. 1999. Water Supply Engineering. 1000, 3. Bangkok: Chulalongkorn University.

[2] Provincial Waterworks Authority. 2018. Guide: Process of Water Supply and Water Quality Control.

[3] Wangpyasan, T. 2014. Water Supply Engineering. 2000, 2. Bangkok: Chulalongkorn University.

[4] Somkraisornkij, P. 2007. Coagulation and Flocculation Process [Online]. Available from: https://www.mwa.co.th/ewt_dl_link.php?nid=441 [2019, April 28].

[5] Crittenden, J. C., Trussell, R. R., Hand, D. W., Howe, K. J. and Tchobanoglous, G. 2012. MWH's water treatment: principles and design. 3. Canada: John Wiley & Sons, Inc., Hoboken, New Jerscy.

[6] Provincial Waterworks Authority. 2017. Method: Jar-Test [Online]. Available from: http://www.oic.go.th/FILEWEB/CABINFOCENTER38/DRAWER032/GENERAL/DATA0000/00000054.PDF [2019, June 1].

[7] Chawakitchareon, P., Boonnao, N. and Charutragulchai, P. 2017. Prediction of Alum Dosage in Water Supply by WEKA Data Mining Software. Information Modelling and Knowledge Bases XXVIII, 2: 83-92.

[8] Abdullahi, M. E., Folorunsho, A. D., Agaie, B. G. and Jibril, M. 2012. Predictive Model for Lime Dosage in Water Treatment Plant. International Journal of Scientific and Research Publications (IJSRP), 2: 230-233.

[9] Natitaweetrab, C. 2012. Weight-Based Growth Model for Nile Tilapia in Waste Water Treatment Pond without Feeding. Master's thesis. Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University.

[10] Auwatanamongkol, S. 2018. Data Mining. 562, 2. Bangkok: National Institute of Development Administration (NIDA).

[11] Yimngam, A. 2019. Data Mining. [Online]. Available from: http://compcenter.bu.ac.th/news-information/data-mining [2019, April 28]

[12] Katanyukul, T. 2018. Basic Machine Learning. 100, 1. Khonkaen: Faculty of Engineering, Khon kaen University.

[13] Wareeprasert, N. and Ladi, N. 2009. Artificial intelligence. 1. Bangkok : KTP Comp & Consulting.

14] Sinsomboonthong, S. 2018. Data Mining 1. 2000, 2. Bangkok: Chamchuri Products.

[15] Thipnang, A. 2018. Enhancing Alum Dosage Prediction Accuracy in Coagulation Process by Data Mining Software: A Case Study of Bang khen Water Supply Plant, Bangkok, Thailand. Master's thesis. Department of Environmental and Energy Engineering for Sustainability, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang.

[16] Nuengkham, C. 2019. Gradient Boosting Machine[Online]. Available from: https://medium.com/cw-quantlab/optimize-trading-system-with-gradient-boosting-machine-8db7d392f92d [2019, September 16].