การทำนายปริมาณสารส้มที่ใช้ในกระบวนการโคเอคกูเลชั่นด้วยโปรแกรม Weka: กรณีศึกษาในนครหลวงเวียงจันทน์ สาธารณรัฐประชาธิปไตยประชาชนลาว

นายขุมคำ รัดศาวงศ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมสิ่งแวดล้อม ภาควิชาวิศวกรรมสิ่งแวดล้อม
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2559

Prediction of Alum Dosage in Coagulation Process by Weka Program:
A Case Study in Vientiane Capital, Lao PDR

Mr. Khoumkham Ladsavong

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Environmental Engineering
Department of Environmental Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2016

Thesis Title	Prediction of Alum Dosage in Coagulation Process by Weka Program: A Case Study in Vientiane Capital, Lao PDR
By	Mr. Khoumkham Ladsavong
Field of Study	Environmental Engineering
Thesis Advisor	Associate Professor Petchporn Chawakitchareon, Ph.D.
Thesis Co-Advisor	Professor Yasushi Kiyoki, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

....Dean of the Faculty of Engineering
(Associate Professor Supot Teachavorasinskun, Ph.D.)

THESIS COMMITTEE

....Chairman
(Associate Professor Chavalit Ratanatamskul, Ph.D.)

....Thesis Advisor
(Associate Professor Petchporn Chawakitchareon, Ph.D.)

....Thesis Co-Advisor
(Professor Yasushi Kiyoki, Ph.D.)

....Examiner
(Associate Professor Sirima Panyametheekul, Ph.D.)

....Examiner
(Assistant Professor Sarun Tejasen, Ph.D.)

....External Examiner
(Assistant Professor Suwisa Mahasandana, D.Eng.)

ขุมคำ รัดศาวงศ์ : การทำนายปริมาณสารส้มที่ใช้ในกระบวนการโคเอคกูเลชั่นด้วยโปรแกรม Weka:กรณีศึกษาในนครหลวงเวียงจันทน์ สาธารณรัฐประชาธิปไตยประชาชนลาว (Prediction of Alum Dosage in Coagulation Process by Weka Program: A Case Study in Vientiane Capital, Lao PDR) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร.เพ็ชรพร เชาวกิจเจริญ, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ศ. ดร.ยาซูชิ คิโยกิ, 176 หน้า.

หัวข้อการวิจัยเรื่อง การทำนายปริมาณสารส้มในขั้นตอนการตกตะกอน โดยใช้โปรแกรม WEKA ข้อมูลที่นำมาใช้นั้น เก็บมาจากโรงประปา 2 แห่ง คือ จิ่นายโม้ และดงหมากคายข้อมูลดังกล่าวนั้นถูกเก็บหลังจากการทำการทดลอง จาร์เทสต์(Jar test) สำหรับ CWTP นั้น เก็บข้อมูลจากปี พุทธศักราช 2552-2559 โดยมีการบันทึกไว้ 2038 ครั้ง ส่วนข้อมูลอีกชุดจาก จาก DWTP เก็บข้อมูลจากปีพุทธศักราช 2551-2559 โดยมีการบันทึกไว้ 2802 ครั้ง โมเดลที่ถูกสร้างขึ้นมานั้นสำหรับปริมาณการเติมสารส้มในกระบวนการตกตะกอน โดยจะใช้ 4 วิธี คือ MLP, M5rules, M5P และ REPTree ส่วนข้อมูลที่จะนำมาทำโมเดลนั้น เราได้แบ่งเป็น 2 กลุ่ม กลุ่มแรกนั้น ได้ทำการแทนที่ข้อมูลที่หายไป โดยใช้ค่าเฉลี่ยของตัวแปรนั้นของแต่ละเดือน ส่วนข้อมูลกลุ่มที่สองนั้นในการบันทึกแต่ละวัน ถ้ามีข้อมูลตัวแปรใดหายไป ก็จะลบข้อมูลของการบันทึกค่าตัวแปรของวันนั้นออกทั้งหมด เพื่อลดการผิดเพี้ยนของโมเดล ผลลัพท์นั้นแสดงให้เห็นว่า โมเดลสำหรับการทำนายปริมาณสารส้ม โดยวิธี M5Rules จากข้อมูลกลุ่มที่ 1 ของ CWTP นั้น ให้ค่า RMSE เท่ากับ 4.043 ซึ่งน้อยกว่าวิธีอื่นๆ เมื่อเราใช้วิธีนี้ ในการทำนายค่าจริงของปริมาณสารส้ม ดังนั้น วิธี M5Rules จึงมีความแม่นยำ และ ความน่าเชื่อถือมากกว่า วิธีอื่นๆ แต่ในทางตรงกันข้าม ที่ DWTP พบว่า การสร้างโมเดลสำหรับการทำนายค่าปริมาณสารส้ม โดยใช้วิธี MLP จากข้อมูลกลุ่มที่ 1 มีค่า RMSE เท่ากับ 1.849 ซึ่งน้อยมาก เมื่อเราใช้วิธีนี้ ในการทำนายค่าจริงของปริมาณสารส้ม เพราะฉนั้น วิธี MLP โดยใช้ข้อมูลจาก DWTP ให้ผลลัพท์ ที่แม่นยำ และ น่าเชื่อถือ สูงกว่าวิธีอื่นๆ สุดท้ายนี้ ผลการวิจัยแสดงให้เห็นว่า โมเดลมีความแม่นยำสูงในหน้าแล้ง มากกว่าในหน้าฝน.

ภาควิชา วิศวกรรมสิ่งแวดล้อม ลายมือชื่อนิสิต ....

สาขาวิชา วิศวกรรมสิ่งแวดล้อม ลายมือชื่อ อ.ที่ปรึกษาหลัก ....

ปีการศึกษา 2559 ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

5870326221 : MAJOR ENVIRONMENTAL ENGINEERING
KEYWORDS: COAGULATION PROCESS / ALUM / DATA MINING / WEKA DATA MINING SOFTWARE

KHOUMKHAM LADSAVONG: Prediction of Alum Dosage in Coagulation Process by Weka Program: A Case Study in Vientiane Capital, Lao PDR. ADVISOR: ASSOC. PROF. PETCHPORN CHAWAKITCHAREON, Ph.D., CO-ADVISOR: PROF. YASUSHI KIYOKI, Ph.D., 176 pp.

This research topic entitled of "Prediction of Alum Dosage in the Coagulation Process using Weka Program". The data resources collected from 2 water treatment plants i.e. Chinaimo Water Treatment Plant (CWTP) and Dongmarkkaiy Water Treatment Plant in Vientiane Capital, Lao PDR. Those data resources were collected from the previous Jar-Test experimental. For the CWTP, the data resources collected from 2009 to 2016, we selected 2,038 records. For the DWTP, the data resources collected from 2008 to 2016, we selected 2,802 records. The model building for alum dosage prediction, we used 4 methods i.e. Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree with 2 data groups i.e. the first data group, we substituted all missing values of each parameter by the average values of that parameter. For the second data group, we have cut off the missing values to reduce bias. The results indicated that the model building for alum dosage prediction by M5Rules method from the model group 1 of the CWTP realizes the less RMSE of 4.043 than another method when we have used this method to predict the alum dosage in the real applications. Thus, the M5Rules method realizes higher precision and credibility than other methods. On the other hand, in the DWTP, we found that the model building for alum dosage prediction by using MLP method from model group 1 realizes the less RMSE of 1.849 when we used it to predict the alum dosage in the real applications. Therefore, the MLP method yields the higher accuracy and dependability than other methods of the DWTP. Finally, the model had the highest precision in the drying season than raining season.

Department: Environmental Engineering
Field of Study: Environmental Engineering
Academic Year: 2016

Student's Signature ....
Advisor's Signature ....
Co-Advisor's Signature ....

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to many people who encouraged and provided help and support to conduct my study. First of all, I would like to thanks my thesis advisor, Assoc. Prof. Petchporn Chawakitchareon, Ph.D., who has never given up on me and been very supportive and encouraged me to overcome all the obstacles through these years of study at Chulalongkorn University.

I would like to thanks my co-advisor, Prof. Yasushi Kiyoki, Ph.D., who has always given me the appreciated commendation. I would like to sincerely thank my thesis committee, Assoc. Prof. Chawalit Ratanathamsakul, Ph.D., Assoc. Prof. Sirima Panyametheekul, Ph.D., Asst. Prof. Sarun Tejasen, Ph.D., and Asst. Prof. Suwisa Mahasandana, D. Eng. for their valuable time in reading and providing suggestions regarding my thesis.

I would like to extend my gratitude to all instructors, staff, and friend at the Environmental Engineering Department, Faculty of Engineering. I also thankful to the employees at the Vientiane Capital Water Supply State Enterprise of Lao PDR, especially, Chinaimo and Dongmarkkaiy Water Treatment Plant who fully collaborated in and gave any conveniences to me. Without their kind cooperation, I might not be able to complete my study.

I would like to sincerely thankful to AUN/SEED-Net Scholarship, who encourage the convenience things for my education and give advice me.

Finally, I would like to greatly thankful to my parents who always support and achieve my ultimate goal of life.

CONTENTS

Page

Page

Page

Page

Page

LIST OF TABLES

Page

Page

Page

Page

Page

LIST OF FIGURES

Page

Page

Page

Page

CHAPTER I
INTRODUCTION

1.1 Background and importance of the study

The basic Water Treatment Plant (WTP) has many processes i.e. coagulation, flocculation, sedimentation, filtration, and disinfection. The coagulation process is one of WTP also. In this case, the term coagulation process is referred to the process of inducing contacts between some chemical and colloidal particles to affect a reaction. The reaction product is called here, a *micro-floc* (Hendricks, 2006). Flash mixing is an integral part of coagulation, the purpose of flash mixing is to disperse water-treating chemicals quickly and uniformly throughout the process water. Effective flash mixing is especially important when using metal coagulants such as alum and ferric chloride since their hydrolysis occurs within a second, and subsequent adsorption to colloidal particles is almost immediate (Kawamura, 2000).

Common coagulants used in conventional water treatment include alum, ferric chloride, ferric sulfate, sodium aluminate, and various cationic polymers. Coagulant aids are used to improve the coagulation process and build stronger, more easily settled floc (Pizzi, 2011).

Many years ago, the researchers tried to use the computer software to determine the optimum coagulants dosage such as alum. For example, (Maier et al., 2004) studied about *"Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters"*. The performance of the models is found to be very good, with correlation (R^2) values ranging from 0.90 to 0.98 for the process models predicting treated water turbidity, color and ultraviolet absorbance at a wavelength of 254 nm (UVA-254).

Nowadays, Vientiane capital, Lao PDR has 4 water treatment plants i.e. *Dongmarkkiay Water Treatment Plant (case study), Dongbung Water Treatment Plant, Chinaimo Water Treatment Plant (case study),* and *Kaoliaow Water Treatment Plant* are supported the water supply to the population. All those water treatment plants determine the coagulant dosage by using Jar-Test experiment. Jar-Test is perhaps least understood, but the most useful of process tools available to operators. A few simple ideas and techniques need to be mastered. When the plant staff becomes efficient in this process control strategy, a more rapid response to treatment upsets is seen (Pizzi, 2011).

For the Jar-Test experiment has some disadvantages. As this method use 2-3 hours for chemical analysis. Therefore, in this research, we would like to apply the Weka data mining software to predict the alum dosage in the coagulation process and evaluated the effective of water treatment plant by analysis and prediction of alum dosage in coagulation process. Thus, in this research, we are compared various alum dosage prediction methods, i.e. *Multilayer Perceptron (MLP), M5Rules, M5P,* and *REPTree* to find the method which gives the highest accuracy.

1.2 Objectives

1.2.1 Apply Weka data mining software to predict alum dosage in coagulation process.

1.2.2 To evaluate the effective of water supply production processes by analysis and prediction of chemical quantity for coagulation.

1.3 Scope of the study

In this research, we studied in Vientiane Capital, Lao PDR. Vientiane Capital is a capital city of Lao PDR that it located in the central of Lao PDR. We studied in two plants i.e. *Dongmarkkaiy Water Treatment Plant (DWTP)* and *Chinaimo Water Treatment Plant (CWTP)* which the DWTP is used Nam Ngum River and the CWTP is used Mekong River to be the raw water. For the location of both plants is indicated in **Figure 1.1**.

For building the model, three main parameters, i.e. *Turbidity, pH,* and *Alkalinity,* and an *Alum* dosage are put in the Weka data mining software. In this case, turbidity, pH, and alkalinity are independent variables and an alum is a dependent variable.

For the data that we put in the model, we collected the previous data of Jar-Test experiment. For the DWTP, we collected from 2008-2016 and 2009-2016 for the CWTP respectively.

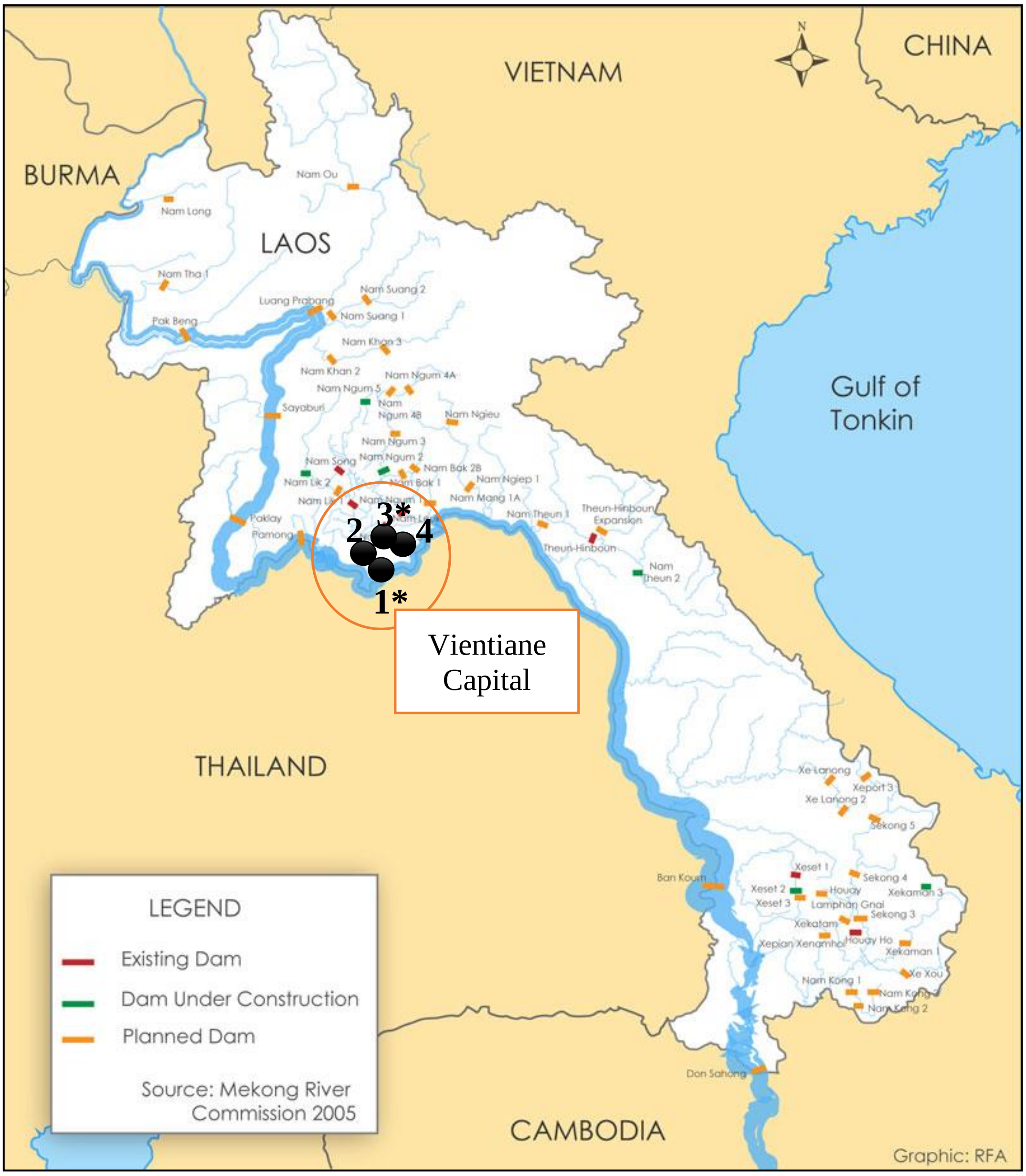


***Figure 1.1** The case study location in Vientiane Capital, Lao PDR.*

1*: Chinaimo Water Treatment Plant (CWTP) is used Mekong River to be the raw water.

2: Kaoliaow Water Treatment Plant is used Mekong River to be the raw water.

3*: Dongmarkkaiy Water Treatment Plant (DWTP) is used Nam Ngum River to be the raw water.

4: Dongbung Water Treatment Plant is used Nam Ngum River to be the raw water.

Note: *Number 1* and 3* water treatment plants are our case study.*

1.4 Expected outcome

1.4.1 Collecting data of raw water.

1.4.2 Adjusted model can predict the amount of chemical for adding to the coagulation and can use in the real plant.

1.5 Key words

Coagulation process

Alum

Data mining

Weka data mining software

CHAPTER II
LITERATURE REVIEW

2.1 Coagulation Process

As we knew the water treatment process is had many processes i.e. coagulation, flocculation, sedimentation, filtration, and disinfection. Coagulation is necessary to reduce or eliminate the energy requirement to destabilize a hydrophobic colloidal suspension to permit coagulation to occur (Scholz, 2015).

Coagulation involves the addition of a chemical coagulant or coagulants for conditioning the suspended, colloidal, and dissolved matter for subsequent processing by flocculation or to create conditions that will allow for the subsequent removal of particulate and dissolved matter (Crittenden et al., 2012) .Coagulation and flocculation can also be differentiated based on the time required for each of the processes. Coagulation typically occurs in less than 10 s, whereas flocculation occurs over a period of 20 to 45 min (Crittenden et al., 2012). The overview of the coagulation process is provided in the **Figure 2.1** below.

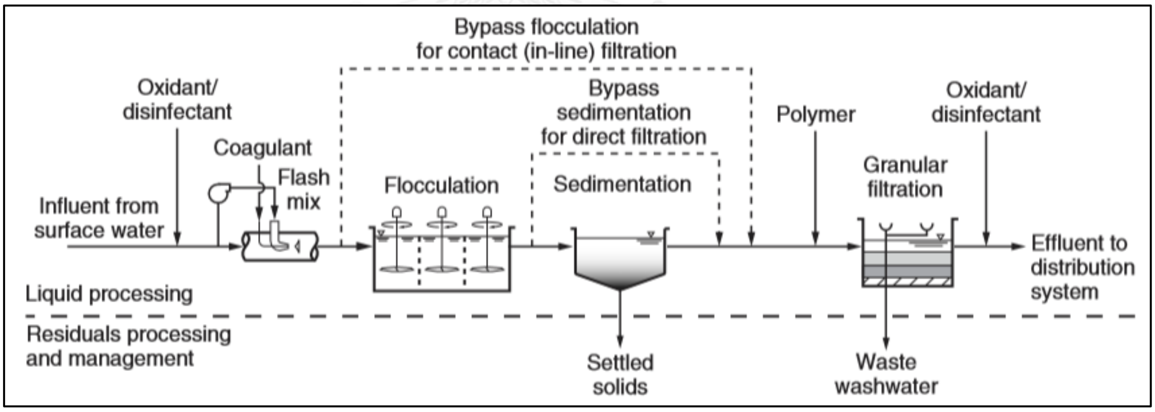


Figure 2.1 *Typical water treatment process (Crittenden et al., 2012)*

The principle of the coagulation process is depended on the source of the water, the principle method are as follows:

1) Reduction or neutralization of the charges on the colloid (Scholz, 2015);

2) Adsorption and/or reaction of portions of the colloidal and dissolved NOM to particles (Crittenden et al., 2012); and

3) Creation of flocculant particles that will sweep through the water to be treated, enmeshing small suspended, colloidal, and dissolved material as they settle (Crittenden et al., 2012)

2.1.1 Destabilization

Destabilization means destroying the stability of particles that has mechanism as below:

1) Diffuse layer: Addition of ions having a charge opposite to that of colloid produce correspondingly high concentrations of counter ions in diffuse layer as shown in **Figure 2.2.** The volume of diffuse layer necessary to maintain electronuetraility is lowered. The amount of electrolyte required to achieve coagulation by double layer compression is independent of the concentration of colloids in liquid. It is not possible to cause charge reversal on a colloid regardless of how much electrolyte is added.

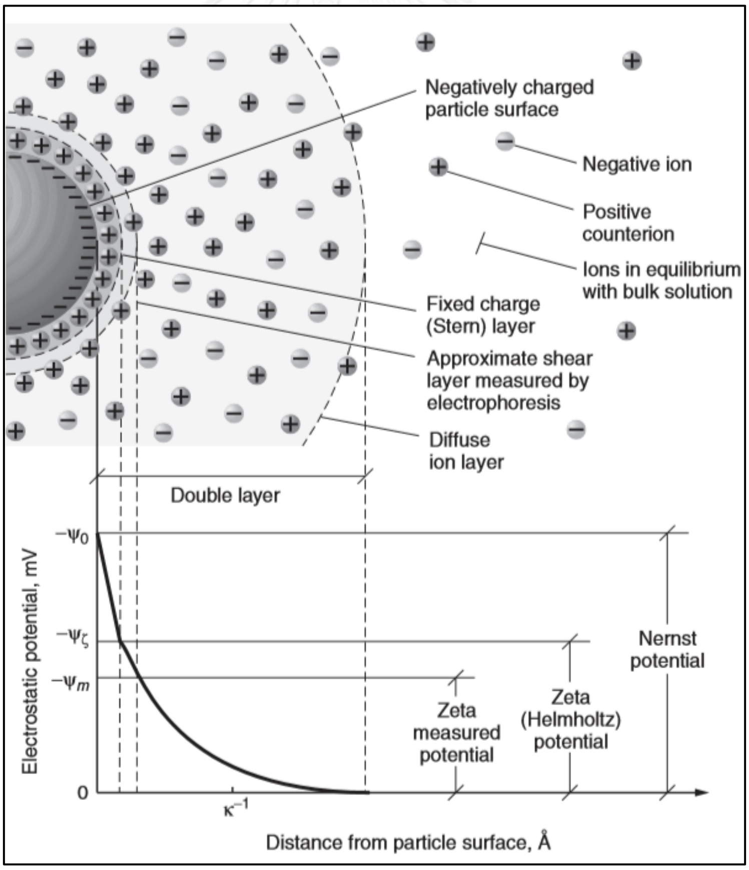

Negatively charged particle surface
Negative ion
Positive counterion
Ions in equilibrium with bulk solution
Fixed charge (Stern) layer
Approximate shear layer measured by electrophoresis
Diffuse ion layer
Double layer
$-\psi_0$
$-\psi_\zeta$
$-\psi_m$
0
Electrostatic potential, mV
κ^{-1}
Distance from particle surface, Å
Zeta measured potential
Zeta (Helmholtz) potential
Nernst potential

Figure 2.2 *Double Layers (Crittenden et al., 2012)*

2) Adsorption and Charge Neutralization: Some chemical species capable is adsorbed at the surface colloidal particles, if the adsorbed species carry a charge opposite to that of colloid will reduction of surface potential destabilization of colloidal particle.

3) Enmeshment in a Precipitate (Sweep-Floc Coagulation): Inverse relationship between the coagulant dosage and the concentrations of colloids to be removed. At large excess of coagulant is required to produce a large amount of precipitate that will enmesh the relatively few colloidal particles as it settles if low concentrations. On the other hand, coagulation will occur at a lower chemical dosage because the colloids serve as nuclei to enhance precipitate formation. It is sometimes advantageous to add turbidity to the dilute colloidal suspensions.

4) Interparticle Bridging: Natural organic polymers (e.g., starch, cellulose), synthesis organic, and polymeric compounds are effective coagulant agents; they have large molecular size and maybe anionic, cationic or non-ionic. Polymer molecule becomes attached to a colloidal particle at one or more site due to: coulombic attraction (if the polymer and particle are opposite charge). The “tail” of the adsorbed polymer will extend out into the bulk solution and can become attached to vacant sites on the surface of another particle to form a chemical bridge. This bridging action results in the formation of a floc particle having favorable settling characteristics. Overdosing of polymer saturate the surface of colloidal no site is available for the formation of polymer bridge. Intense and prolonged may destroy previously formed bridges agitation.

2.1.2 Coagulant Aids

A coagulant aid is a chemical added during coagulation to achieve one or more of the following results:

- Improve coagulation;
- Build stronger, more settleable floc;
- Overcome of the effect of temperature drops that slow coagulation;
- Reduce the amount of coagulant needed; and
- Reduce the amount of sludge produced.

Three common coagulant aids are Alum, PACl, and Ferric Chloride.

1) **Alum (Aluminum Sulfate)**

Alum has molecular formula as $Al_2(SO_4)_3 * H_2O$ that * has value from 14.3 to 18. Because alum is the most common coagulant used for water treatment, it is important to understand how it promotes floc formation and settling. The process that takes place is as follows:

- Alum added to raw water reacts with the alkalinity naturally present to form jellylike floc particles of aluminum hydroxide, $Al(OH)_3$. A certain level of alkalinity is necessary for the reaction to occur. If not enough is naturally present, the alkalinity of the water must be increased. Alum will consume alkalinity in the ratio of 1 to 0.5: i.e., 1 mg/L dry-basis alum will consume 0.5 mg/L alkalinity as $CaCO_3$. One mg/L dry basis dosage will produce 0.26 mg/L sludge.
- The positively charge trivalent aluminum ion neutralizes the negatively charged particles of color or turbidity. This occurs within 1or 2 second after the chemical is added to the water, which is why rapid, thorough mixing is critical to good coagulation.
- Within a few second, the particles begin to attach to each other to form larger particles.
- The floc that is first formed is made up of microfloc that still has a positive charge from the coagulant; the floc particles continue to neutralize negatively charged particles until they become neutral particles themselves. This is referred to as charge neutralization.

Finally, the microfloc particles begin to collide and stick together (agglomerate) to form larger, settleable floc particles.

Many physical and chemical factors can affect the success of a coagulant, including mixing conditions; pH, alkalinity, and turbidity levels; and water temperature. Alum works best in a pH range of 5.8 to 8.5 as indicated in **Figure 2.3**. If it is used outside this range, the floc either does not form completely or it may form and then dissolve back into the water.

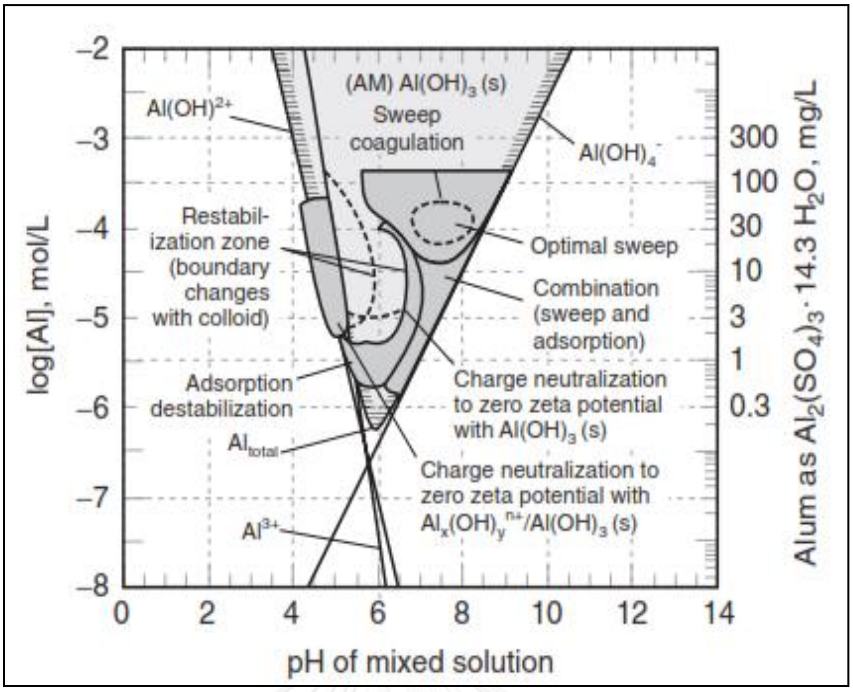


Figure 2.3 *Diagram used to design and control the coagulation by alum (Amirtharajah and Mills, 1982)*

2) PACl (Polyaluminum Chloride)

PACl has molecular formula as $Al_n(OH)_mCl_{3n-}$m. PACl is famous coagulant that used since 1970 to currently; it is famous coagulant in Japan and some countries in Europe. PACl is made from Al_2O_3 reacted with HCl to become $AlCl_3$ at higher temperature, after that it will react with the base at higher temperature and pressure for gather and become aluminum polymer. If we added alum to this solution, the alum will reduce intensity of Al_2O_3 to 10-11% and added sulfate to them for be connect 2 atoms of aluminum.

3) Ferric Chloride

Ferric chloride has molecular formula $FeCl_3$*$6H_2O$ or $FeCl_3$. Ferric chloride has crystals are yellow or brown, green or black pigment, and has yellowish brown solution. When it soluble in the water, it will destroy particles stability that have negative charge and made the floc of Ferric Hydroxide and reacted with alkalinity as formula below:

$$2FeCl_3 + 3Ca(HCO_3)_2 + 3H_2O \rightarrow 2Fe(OH)_3 + 3CaCl_2 + 3HCO_3^- + 3H^+$$

$$2FeCl_3 + 3Ca(OH)_2 \rightarrow 2Fe(OH)_3 + 3CaCl_2$$

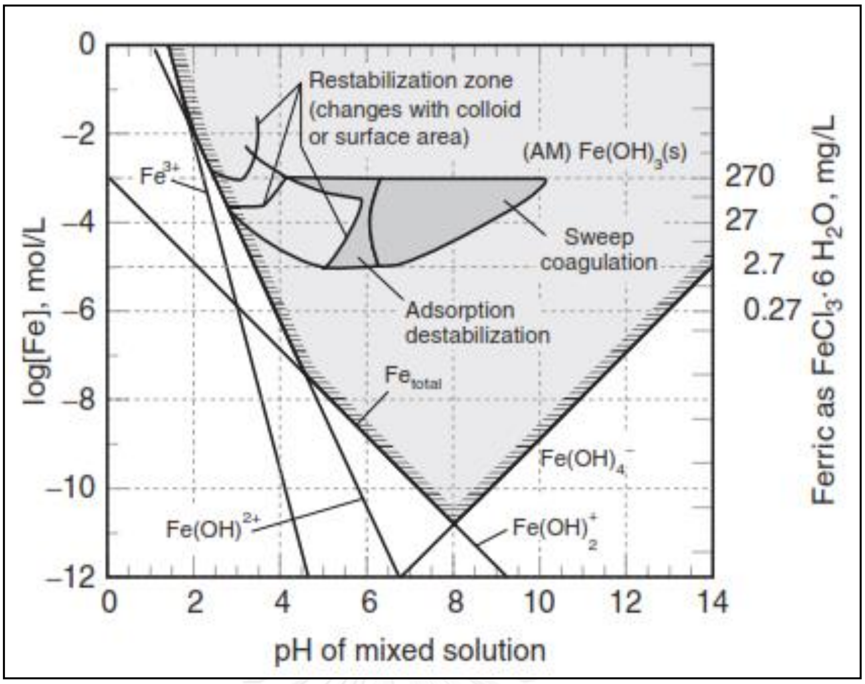


Figure 2.4 *Diagram used to design and control the coagulation by ferric chloride (Amirtharajah and Mills, 1982)*

2.2 Jar-Test

The jar test is as much art as it is science. A different coagulant dose is added to each of the 4 or 6 jars. A short period of rapid mixing (for coagulation) and then a longer period of slow mixing (flocculation) occur. Last, a no-stirring quiescent period permits settling. Chemicals for pH adjustment, coagulant aids; ballasting substances (carbon, clay, etc.) also may be added to the jars (Engelhardt, 2010).

During stirring and the quiescent periods the operator or lab tech will observe the jar for floc formation and settling rate and use this information to then make chemical dose changes to the process (Engelhardt, 2010). Three main parameters must to measure are pH, Turbidity, and Alkalinity.

It is as much an art as a science because operators must learn to interpret “when my little jar looks this way, my big basin will look this way.” The more measurements are made; the better the operator or lab person can interpret the jar test results – based more on measurement (science) and less on art. This is important because filter performance is directly affected by how well the floc forms, settles and withstands shearing effects during mixing and filtration (Engelhardt, 2010).

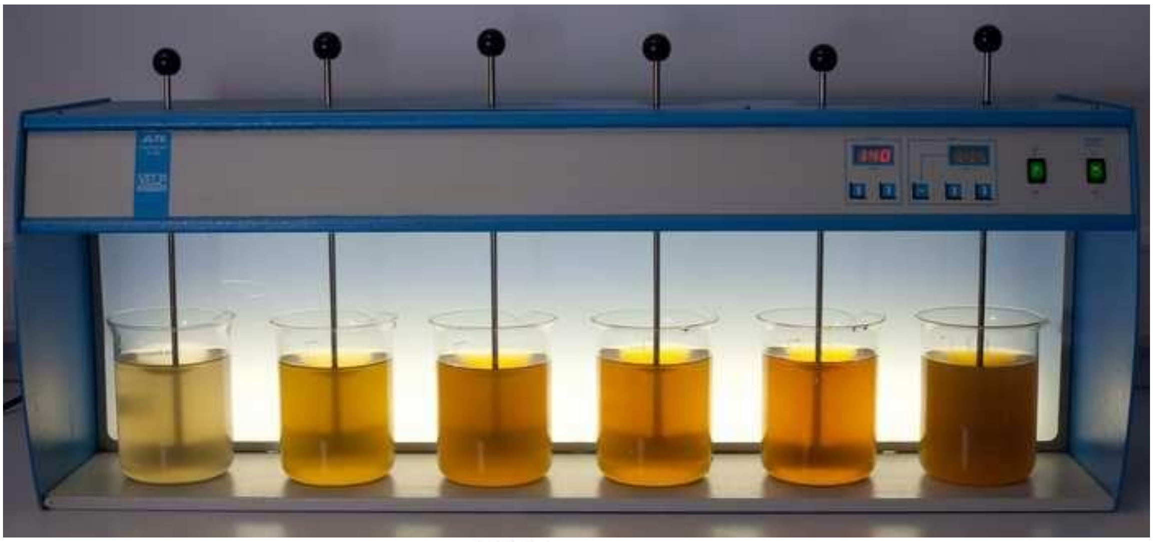


Figure 2.5 *Jar test experiment*

2.3 The relationship of turbidity, pH, and alkalinity with alum

As we have known, we would like to separate the particles out of the water by sedimentation. Turbidity describes the amount of light scattered or blocked by suspended particles in a water sample. For the pH, the alum works best in a pH rang of 5.8 to 8.5. Therefore, it is regarded as an acid salt, and the water must contain enough alkalinity (natural or added) to react with the acid as it forms to maintain the pH within the desired range for good coagulation and flocculation (Crittenden et al., 2012). Thus, turbidity, pH, and alkalinity has relationship with alum.

2.4 Introduction of the case study

2.4.1 Chinaimo Water Treatment Plant (CWTP)

CWTP was located at Boroh Village, Hardxaiyfong District, and Vientiane Capital. CWTP was divided into 2 phase as *Phase 1*, CWTP was founded in 1978 and finished construction in 1980, the capacity is 40,000 m^3/day; *Phase 2*, CWTP was founded in 1993 and finished construction in 1996, the capacity is 40,000 m^3/day. Thus, the capacity of this plant is 80,000 m^3/day. Normally, the produced capacity of CWTP is about 70,036,673 m^3/day in every year. The Mekong River is used to be the raw water of this plant, the Mekong River is the big river in Lao PDR and we can use the raw water from this river all seasons or all year. The CWTP distribute the water supply to 4 districts as Chanthabuli, Sysuttanark, Xaysettha, and Hardxaiyfong District (Vientiane Capital).

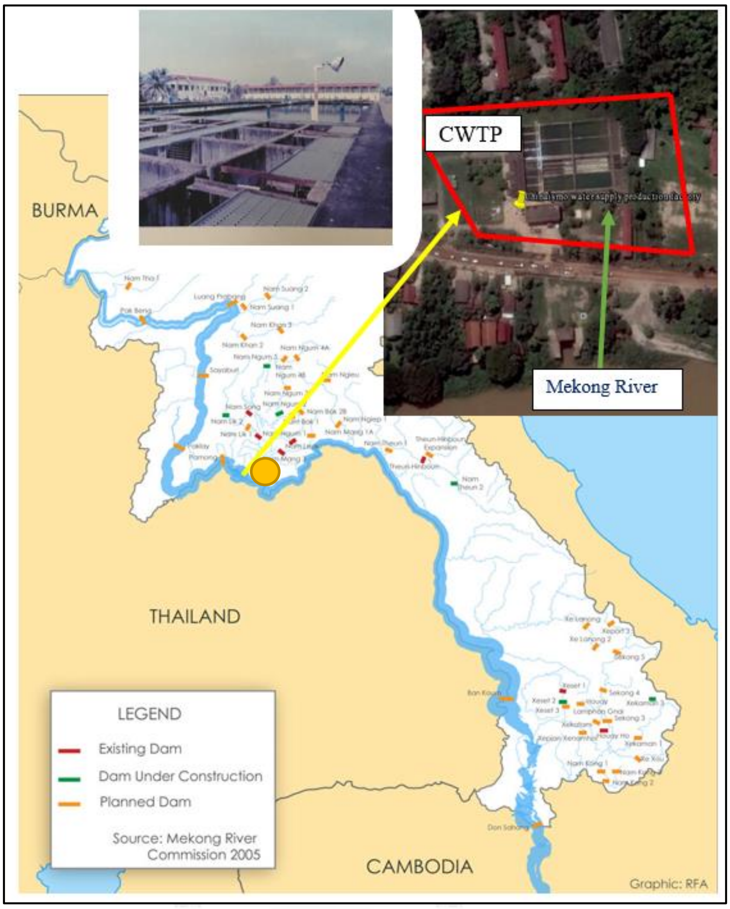


Figure 2.6 *The Chinaimo water treatment plant (CWTP) location*

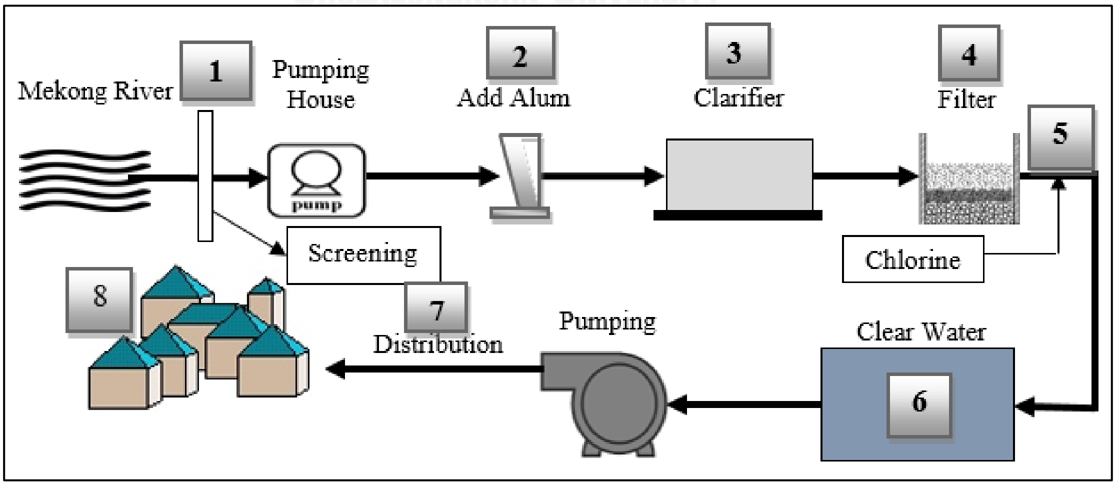


Figure 2.7 *Water treatment processes of Chinaimo water treatment plant (CWTP)*

1. Improving raw water quality

Before the raw water past to the pumping house have had screening for the screen the plastics, woods or other materials that suspended with the raw water.

2. Adding the chemical

Before the raw water will pump to the clarifier have to add the chemical as we have to know the Alum for help the particles arrest to the floc and can settling.

3. Clarifier

When we have added the chemical, the raw water will drain to *Clarifier*, in this process, the chemical will react with the particles or turbidity, the particles will arrest to floc and settling. The bright water was been in the top of clarifier will drain to filtration. For the duration time of settling is about 2 hours and the turbidity has value not over 5 NTU. If the rainy season will have high turbidity, they will add PACl for help the settling.

4. Filtration

The water was clarified will drain to the sand filter, the water was filtered will have turbidity is not over 2 NTU.

5. Disinfection

The water that we have gotten from the filtration will have the disinfection for the safety and clean for the consumption by adding the chlorine that it can disinfect almost diseases and good destroy the organic compounds, odor, color, and iron.

6. Clear water

The water has gotten from the disinfection will storage in the clear water for distribution.

7. Distribution

Pump the water supply to the pipe system for distribution to the population.

8. The management of water supply quality

Raw water, water in the production processes, and water in the distribution system will usually have exhaustive analysis the quality for control the quality have to be in the standard before distribution to the population.

2.4.2. Dongmarkkaiy water treatment plant (DWTP)

DWTP was located at Donetiw Village, Xaythany District, and Vientiane Capital. The DWTP was founded in 2003 and has been finished in 2006. The DWTP has had capacity is 23,000 m^3/day. Raw water is from Nam Ngum River.

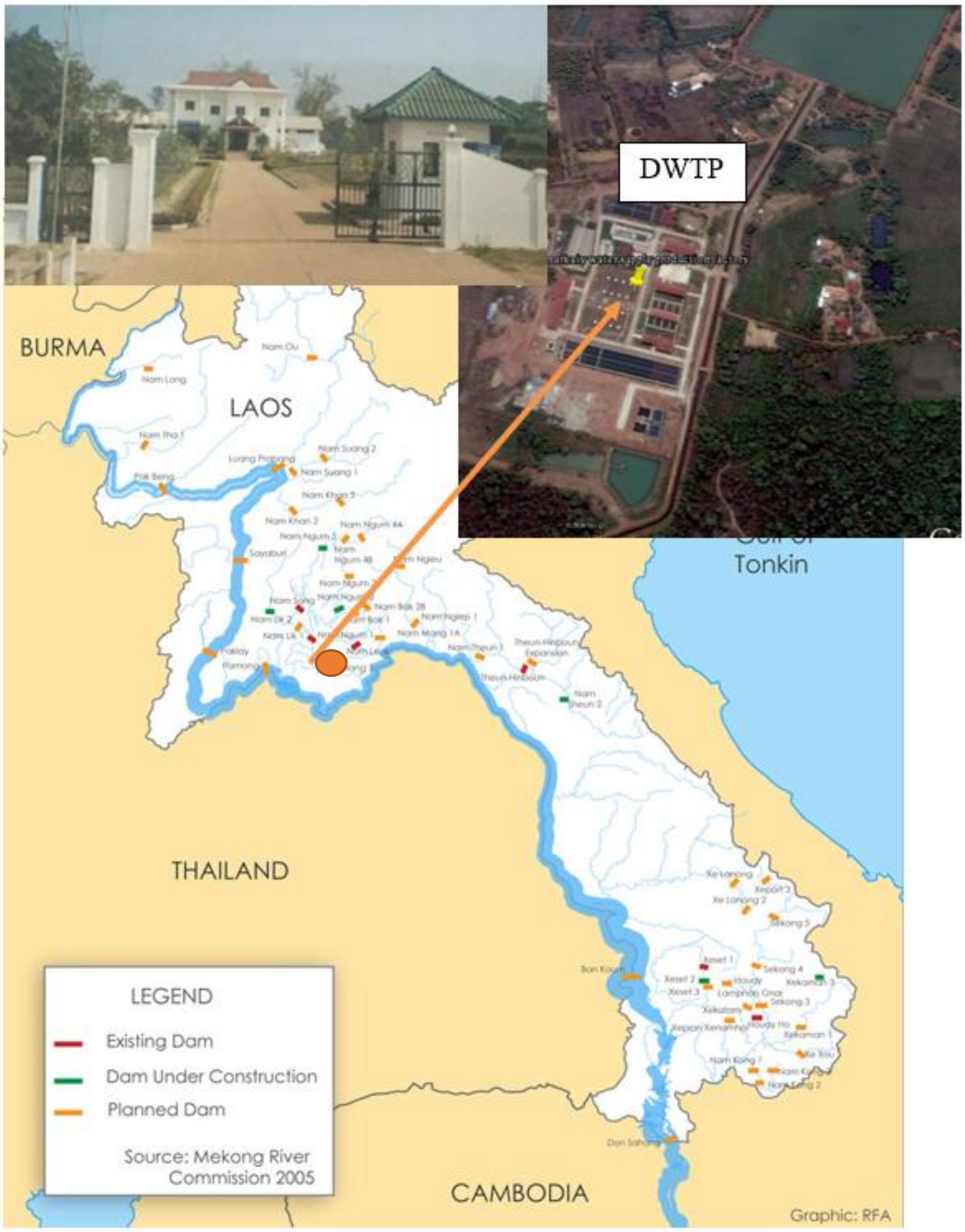


Figure 2.8 *The Dongmarkkaiy water treatment plant (DWTP) location*

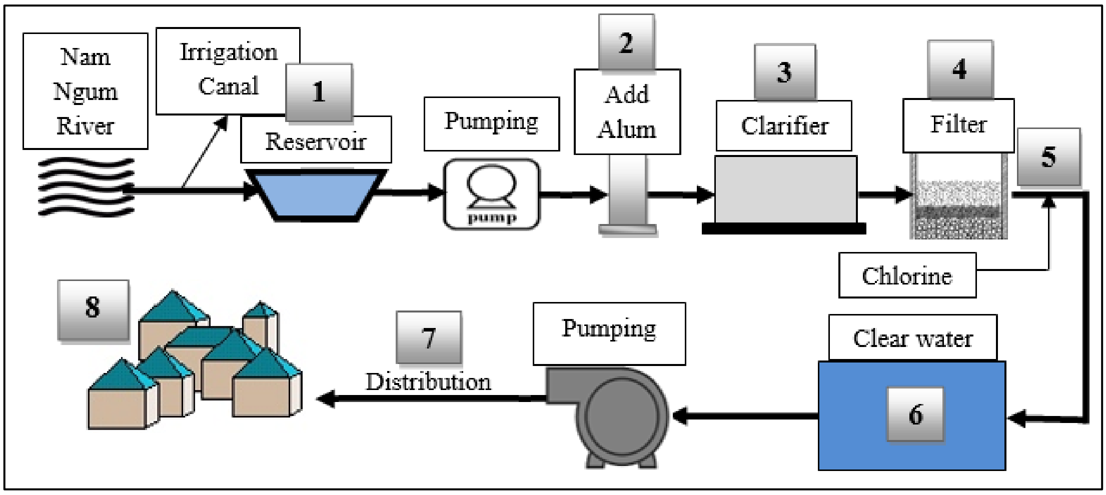


Figure 2.9 *Water treatment processes of Dongmarkkaiy water treatment plant (DWTP)*

1. **Improving raw water quality**

This plant pumps the raw water form Nam Ngum River, has a little bit settling while flow in the irrigation canal. Before the raw water past to the pumping house have had screening for the screen the plastics, woods or other materials that suspended with the raw water.

2. **Adding the chemical**

Before the raw water will pump to the clarifier have to add the chemical as we have to know the Alum for help the particles arrest to the floc and can settling.

3. **Clarifier**

When we have added the chemical, the raw water will drain to *Clarifier*, in this process, the chemical will react with the particles or turbidity, the particles will arrest to floc and settling. The bright water was been in the top of clarifier will drain to filtration. For the duration time of settling is about 2 hours and the turbidity has value not over 5 NTU. If the rainy season will have high turbidity, they will add PACl for help the settling.

4. **Filtration**

The water was clarified will drain to the sand filter, the water was filtered will have turbidity is not over 2 NTU.

5. Disinfection

The water that we have gotten from the filtration will have the disinfection for the safety and clean for the consumption by adding the chlorine that it can disinfect almost diseases and good destroy the organic compounds, odor, color, and iron.

6. Clear water

The water has gotten from the disinfection will storage in the clear water for distribution.

7. Distribution

Pump the water supply to the pipe system for distribution to the population.

8. The management of water supply quality

Raw water, water in the production processes, and water in the distribution system will usually have exhaustive analysis the quality for control the quality have to be in the standard before distribution to the population.

2.5 Water quality characteristics

Lao People's Democratic Republic (Lao PDR) located in the heart of the Indochinese peninsular, in Southeast Asia, at a latitude of 14 to 23 degrees north and longitude 100 to 108 degrees east. Lao PDR covers a land of area around 236,800 square kilometers, three-quarters of which is mountains and plateau. The country has three distinct regions that consist of the north, central, and south. Lao PDR shares a 505 Km border with China to the north, 535 Km of border with Cambodia to the south, 2,069 Km of border with Vietnam to the east, 1,835 Km of border with Thailand to the west, and a 236 Km border with Myanmar to the northwest. The country stretches 1,700 Km from north to South, with an east-west width of over-500 Km at its widest and only 140 Km at the narrowest point. Geographical characteristics in Lao PDR can be divided into 12 main river basins, shown as **Table 2.1**.

Table 2.1 *Major River Basins of Lao PDR with catchment area exceeding 4,500 Km^2*

No	Basin	Total area (km^2)	Location	
			Latitude	Longitude
1	Nam Ou	24,637	19.983333-22.500000	101.56667-103.08333
2	Sekong[1]	22,179	14.000000-16.333333	105.93333-107.71667
3	Nam Kading	14,820	17.600000-19.133333	103.76667-105.78333
4	Sebanghieng	19,223	15.883333-17.166667	104.73333-106.46667
5	Nam Ngum[2]	16,841	18.016667-19.800000	101.85000-103.41667
6	Sebangfay	10,345	17.925833-16.809722	104.72750-106.48972
7	Nam Tha	8,917	19.983333-21.266667	100.58333-101.96667
8	Nam Khan	7,490	19.350000-20.350000	101.93333-103.71667
9	Sedone	7,229	15.116667-16.183333	105.63333-106.73333
10	Nam Suang	6,578	19.800000-20.966667	102.23333-103.35000
11	Nam Ma	5,947	20.233333-21.016667	103.13333-104.66667
12	Nam Ngiep	4,577	18.366667-19.233333	103.50000-104.10000

Notes:

1: Upstream of Lao PDR-Cambodia boarder only
2: Includes Nam Lik River and Nam Song River
Source: Draft National Water Resources Profile (2006)

The country is generally rich in water resources. Total available surface water resources (including the flow of the Mekong River and its tributaries) are 55,000 m^3 on an annual per capita basis, the highest in Asia. However, little of the national water supply has been developed. Total storage capacity of large reservoirs is less than 3% of annual surface flow. With the pressure of rapid demographic growth, economic development and urbanization, water quality is increasingly likely to deteriorate. The major issues which may arise include:

- The installation of hydropower schemes poses some important water quality problems or risks. In most deep-water reservoirs in the tropics, in the first few years following impoundment, oxygen depletion will take place in the lower part of the reservoir. This situation is mainly due to thermal stratification and the decomposition of submerged biomass or organic matter. If the turbine water comes from a single, low-level discharge from the reservoir, it will be low in dissolved oxygen (anoxic) and maybe high in noxious compounds (methane, mercury, etc.).

- Lao PDR is rich in mineral resources and is increasing its exploitation of these resources. Mining uses water in both the mining and ore processing stages, although little information is currently available on the amount of water which is consumed. Water use is not included in mining licenses. In some cases, mines, processing areas and mine tailings (waste) storage areas are close to rivers and reservoirs.
- Population growth in cities, towns and villages lead to extensive municipal waste and organic matter release to waterways. No urban centers have access to comprehensive piped sewerage systems. Urban drains act as secondary sewers, carrying industrial discharges and septic tank seepage and overflow in the rainy season. As the result, water in the drainage system is invariably contaminated with fecal matter from latrines and coliform from septic tank effluent.
- The growing number of industries has increase the incidence and risk of pollution. The larger mills and industries of concern in Lao PDR are pulp and paper, timber, food processing, garment manufacturing and cement factories and gravel pits. Most of these have only limited wastewater treatment systems for reducing waste concentrations and loads in the final effluent discharge to waterways.
- Organic and nutrient pollution and sediment can be discharged from agricultural areas. The use of agricultural chemicals in Lao PDR is still relatively low and is expected to remain so during coming years, apart from areas of more intensive, commercial production, including animal production. Increased irrigation can lead to increased nutrients, pesticides and sediment entering waterways through agricultural drainage. An increase in the extent of irrigation can also open new areas for waterborne disease vectors (mosquitoes, snails).
- In the mountainous areas, forest cover has been reduced by slash-and-burn agriculture, conversion of the land to agriculture, road construction and logging. The main slash-and-burn systems has been a rapid decline in the length of fallow periods due to an increased demand for land and resource. Rotational cycles have declined to as low as 3-5 years. Such short rotation ultimately degrades the soil and increase the time that steep slopes are exposed and susceptible to serious erosion, leading to sedimentation, changes in the downstream flow pattern and other impacts on the downstream water ecosystem.

The characteristics of physical, chemical and microbiological are used to determine water quality. In the river, water quality is characterized with wide various parameters. The physical characteristics are caused by some particles that we can be aware of 5 senses of human, such as Conductivity, Taste-Odor, Temperature, Turbidity, Total Dissolved Solids (TDS) and Total Solid (TS). The chemical characteristics caused by mineral or chemical compound that are dissolved in the water, rock or/and soils, are reflection in the water, such as Alkalinity, Biochemical Oxygen Demand (BOD), Dissolved Oxygen (DO), Hardness, Nitrate-nitrogen (NO_3-N) and Potential of Hydrogen ion (pH). The microbiological characteristics are mainly important in water because it causes disease in human, like “Water Born Disease” such cholera disease and typhoid disease. The water quality filed usually analyzes biological characteristics in term of Total Coliform Bacterial and Fecal Coliform Bacterial. The different purpose of water-use should be estimated in term of the specific water-quality parameters for the effect of water-use. The characteristics of physical, chemical, and microbiological are shown in **Table 2.2.**

The water contamination with heavy metals, chemical and microbial has been the factor in determining human care. Usually, the water system containing different anions form and heavy metals (include Cadmium, Chromium, Manganese, Iron and Lead) is useful and gives effects on human care. Some of heavy metals are essentially required for body growth, such as cobalt, copper, and zinc, while the high concentration of other metals is toxic, such as cadmium, chromium, manganese, iron, and lead.

Table 2.2 *The definitions of characteristic of physical, chemical, and biological (Veesommai et al., 2016)*

Parameter	Symbol	Definition
Alkalinity	-	Alkalinity is a chemical measurement of a water's ability to neutralize acids. Alkalinity is also a measure of a water's buffering capacity or its ability to resist changes in pH upon the addition of acids or bases. This parameter is reported in as (mg/L $CaCO_3$)
Biochemical Oxygen Demand	BOD	Biochemical Oxygen Demand (BOD) refers to the amount of oxygen that would be consumed if all the organics in one liter of water were oxidized by bacteria and protozoa. This parameter is reported in (mg/L)
Conductivity	-	The conductivity is related between Total Dissolved Solids (TDS) and Electrical Conductivity. This parameter is reported in (μS)
Coliform Bacterial	-	The kind of bacterial that live in intestines of warm-blooded animal. This parameter used to present the pathogenic organisms of human, and this parameter is reported in (MPN/100 mL or CFU/100 mL)
Dissolved Oxygen	DO	The concentration of oxygen that require by microorganisms, fishes and another aqueous life in aquatic system. This parameter is reported in (mg/L)
Fecal Coliform Bacterial	-	The king of bacterial that growth and live with animal or/and human waste. This parameter used to present the pathogenic organisms of human and this parameter is reported in (MPN/100 Ml or CFU/100 mL)
Hardness	-	Hardness is defined as the sum of the concentrations of calcium and magnesium ions dissolved in water. This parameter is reported in as (mg/L $CaCO_3$)
Nitrate Nitrogen/Nitrite Nitrogen	NO_3-N/ NO_2-N	The nitrate anions are resulted of the bacteriological oxidation nitrogenous in soil. The nitrate anions are one of the indicators for the degree of the pollution in water with nitrate-content substances (the highly values of nitrate anion can be caused "Algae Bloom Crisis" and "Acid Precipitation"). This parameter is reported in (mg/L)
Oxidation Reduction Potential	ORP	ORP measures an aqueous system's capacity to either release or accept electrons from chemical reactions. When a system tends to accept electrons, it is an oxidizing system and when it trends to release electrons, it is a reducing system. This parameter is reported in (mV)
Potential of Hydrogen ion	pH	The measurement of acidity and basicity in aqueous solution. From the theory, pH in water should be between 0-14 and pure water should be in pH=7
Total Dissolved Solids	TDS	TDS which refers to solid compound or article in the solid phase is dissolved in aqueous, such inorganic acid and organic compound. This parameter is reported in (mg/L)
Total Solids	TS	TS can refer to total solid compounds or article in the solid phase in aqueous, after evaporation of the water and dry the solid compound or article in the solid phase at 103 °C – 105 °C. This parameter is reported in (mg/L)
Suspended Solids	SS	The SS which refer to solid compounds or article in the solid phase isn't dissolved in aqueous and suspended in aqueous. This parameter is reported in (mg/L)
Turbidity	-	The turbidity can by caused by infection of soil, sand, algae, plankton, diatom and colloidal, and is an efficiency indicator for water analysis in Environmental field, which measured by the light-transmitting properties in the water. This parameter is reported in (NTU)

2.6 5D World Map (5DWM) System

2.6.1 Spatio-Temporal and Semantic Computing

We have introduced the architecture of a multi-visualized and dynamic knowledge representation system “5D World Map System (Kiyoki et al., 2016; Kiyoki et al., 2012; Sasaki et al., 2010), applied to environmental analysis and semantic computing. The basic space of this system consists of a temporal (1st dimension), spatial (2nd, 3rd and 4thdimensions) and semantic dimensions (5th dimension, representing a large-scale and multiple-dimensional semantic space that is based on our semantic associative computing system (MMM). This space memorizes and recalls various multimedia information resources with temporal, spatial and semantic correlation computing functions, and realizes a 5D World Map for dynamically creating temporal-spatial and semantic multiple views applied for various “environmental multimedia information resources.”

2.6.2 Semantic Computing in 5D World Map System

We apply the dynamic evaluation and mapping functions of multiple views of temporal-spatial metrics, and integrate the results of semantic evaluation to analyze environmental multimedia information resources. MMM is applied as a semantic associative search method (Barker et al., 2003; Kiyoki et al., 1994; Kiyoki et al., 2012; Valentin et al., 1999) for realizing the concept that "semantics" and “impressions” of environmental multimedia information resources, according to the "context". The main feature of this system is to create world-wide global maps and views of environmental situations expressed in multimedia information resources (image, sound, text and video) dynamically, according to user's viewpoints. Spatially, temporally, semantically and impressionably evaluated and analyzed environmental multimedia information resources are mapped onto a 5D time-series multi-geographical space. The basic concept of the 5D World Map System is shown in Figures 3 and 4. The 5D World Map system applied to environmental multimedia computing visualizes world-wide and global relations among different areas and times in environmental aspects, by using dynamic mapping functions with temporal, spatial, semantic and impression-based computations (Kiyoki et al., 2016; Sasaki et al., 2010).

2.6.3 SPA: Sensing, Processing and Analytical Actuation Functions in 5D World Map

"SPA" is a fundamental concept for realizing environmental system with three basic functions of "Sensing, Processing and Analytical Actuation" to design a global environmental system with Physical-Cyber integration. "SPA" is effective and advantageous to detect environmental phenomena as real data resources in a physical-space (real space), map them to cyber-space to make analytical and semantic computing, and actuate the analytically computed results to the real space by visualization for expressing environmental phenomena with causalities and influence. This concept is applied to our semantic computing in 5D World Map System, as shown in **Figures 2.10**.

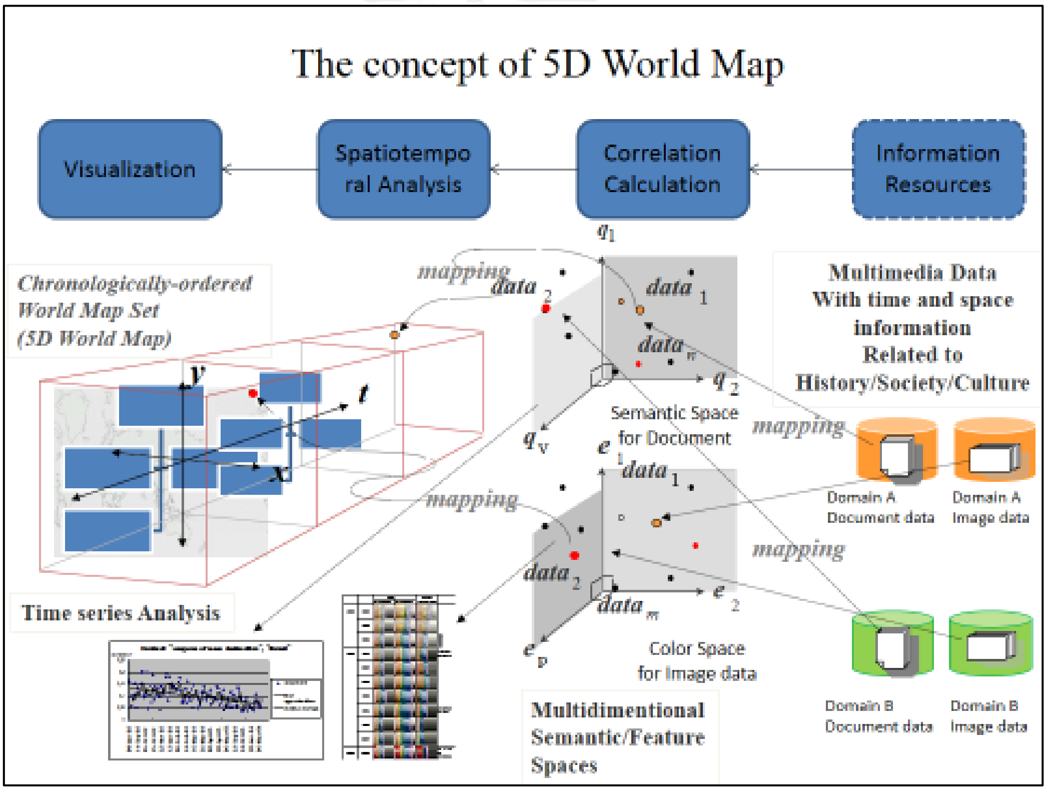


Figure 2.10 *5D World Map System for world-wide viewing for Global Environmental Analysis (Kiyoki et al., 2016)*

2.7 Data mining

Data mining is a field of study that emerges from statistics, machine learning, and database systems. As a discipline for data analysis, statistics contributes significantly towards data mining in terms of fundamental theories and methods for data analysis, measures for evaluating significance and relevance of patterns, and so on.

Machine learning, particularly inductive learning from data as a branch of artificial intelligence, has a long history dating from the 1950s. Over the decades, many machine-learning methods and algorithms have been developed. The application of these methods to data mining problems is a major issue of interest. Most of the time, these methods and algorithms need to be modified in terms of performance efficiency to scale up and solve problems with real-life databases.

Data mining has become a popular and interesting subject of computing in recent years. Since its conception in the early 1990s, the subject has received a huge amount of attention from the research community, the IT industry and beyond. Knowledge matured over the past two decades has started to flow from research and practice into postgraduate and undergraduate degree programmers.

2.7.1 Objectives of data mining

The main objectives of data mining can be broadly categorized into *classification, estimation, prediction,* and *data description.* Objects are classified into one of a set of pre-defined class.

To do this, a classification model is built from a set of data examples. The accuracy of the classification by the model is then evaluated to give some degree of confidence to the result. Once a reliable classification model has been developed, it is then used to classify data records whose class outcomes are unknown. For instance, a classification model for determining whether a credit card application should be granted can be built by using historical credit card application records. The model can then be used to determine whether to accept or reject an application.

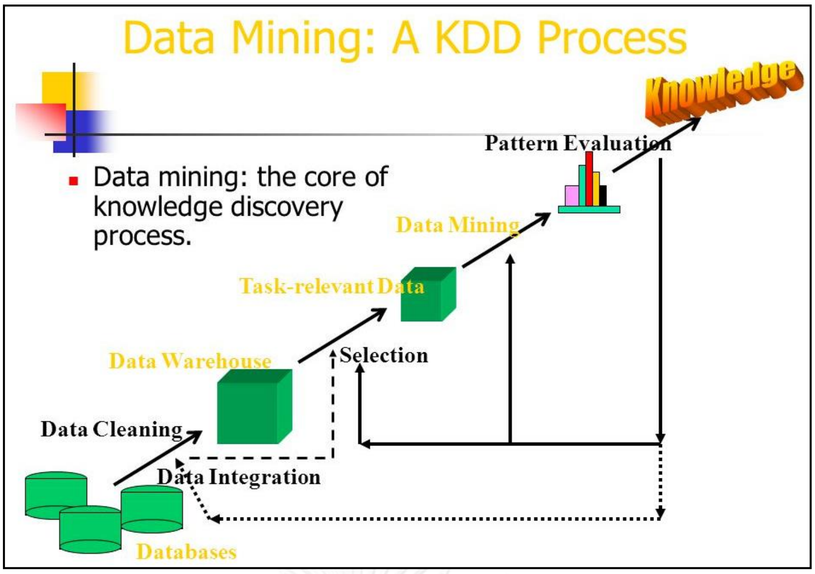
Estimation is like classification. Instead of classifying an object into a discrete class, this task involves building a model, again based on a set of data examples, to estimate the value of a continuous outcome variable. For example, an estimation model can be built on records about house sales. The model produces an estimated value of a house per features such as the number of bedrooms, the facilities (e.g. en-suite, and garage) and the total area of floor space.

Prediction overlaps significantly with the classification and estimation. Prediction is more concerned with a future outcome of the output variable. For instance, historical data recordings on weather conditions are used to predict tomorrow's weather. Solutions for classification and estimation are widely used for prediction.

Data description is about describing general or specific features of the selected data set. It includes summary statistics, clustering, and characteristic rule mining. One powerful data description method is data visualization – using visual vocabulary to describe features and trends in a data set.

2.7.2 Data Mining and Knowledge Discovery in Databases

Knowledge discovery in database (KDD) refers to the efficient process of searching through large volumes of raw data in databases to find useful information patterns that are implicitly embedded in the raw data. Strictly speaking, the term KDD tends to refer to the complete cycle of discovery from unprocessed raw data to knowledge. The term *data mining* normally refers to the integral step of the KDD process that discovers and outputs hidden information patterns from prepared raw data. In practice, however, the two terms are often used interchangeably, causing some degree of confusion. Some key phrases in that description need further explanation.



Data Mining: A KDD Process

- Data mining: the core of knowledge discovery process.

Figure 2.11 KDD process (Du, 2010)

2.7.3 Data mining process

Generally, the data mining process consists of three key steps: preparation of input data, mining of data, and post-processing of output patterns.

2.7.3.1 Data preparation

The data preparation step is a complex process that may involve data collection and selection, pre-processing and formatting. Data collection involves the identification of data sources and the gathering of relevant data details from the sources. Data may be collected from different parts of the same database, different databases within the same organization, or even from external data sources.

2.7.3.2 Data mining

The second step of data mining is the actual mining from the input data to patterns. At this stage, a sensible data mining task must be properly designed to comply with the objectives of the investigation.

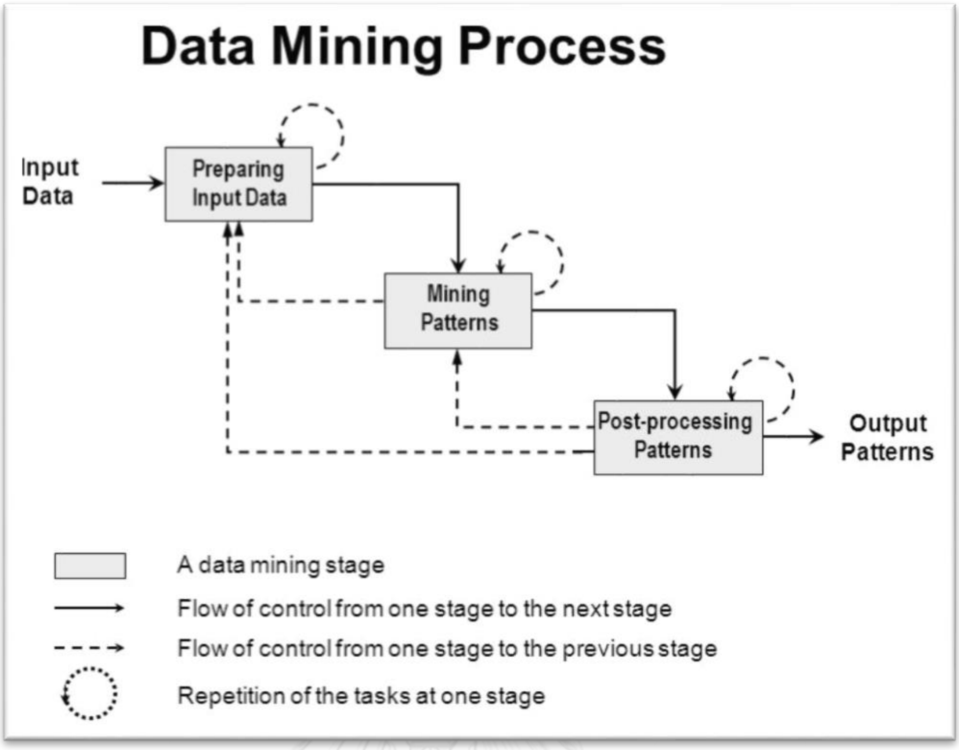

Data Mining Process
Input Data
Preparing Input Data
Mining Patterns
Post-processing Patterns
Output Patterns
A data mining stage
Flow of control from one stage to the next stage
Flow of control from one stage to the previous stage
Repetition of the tasks at one stage

Figure 2.12 *Data mining process (Du, 2010)*

2.7.3.3 Post-Processing of Patterns

The post-processing stage refers to any further processing of the discovered patterns after mining. The post-processing includes pattern evaluation, pattern selection, and pattern interpretation. First, the credibility and significance of the patterns are vital for data mining, and hence must be evaluated objectively using appropriate methods. Often, not all patterns are of interest, and therefore a further selection may be needed. A ranking criterion for *interestingness* may be deployed in the selection process. It is then important that the credible, significant and interesting patterns are understood and interpreted correctly. Appropriate visualization of the patterns can assist the interpretation. This is because human eyes are a powerful tool to identify visual patterns and trends.

2.7.4 Promises and challenges

Data mining technology has a wide range of applications. The following list outlines some major areas:

- Finance and insurance, including discovery of financial and insurance frauds, investment risk analysis, and credit history analysis;
- Marketing and sales, including customer profiling, computer-aided marketing and promotion, and sales analysis;
- Medicine, including diagnosis of disease, analysis of functions of genes, and analysis of the effects of new drugs;
- Agriculture, including diagnosis of plant diseases and the planning of agricultural produce within a region;
- Social development and economics, including city resource planning, national and local government policy making, and local and global economy monitoring;
- Engineering and manufacturing, including evaluation of computer-aided design and manufacturing and fault detection in production lines;
- Natural sciences, including study of observed of experimental data, generation of new hypotheses, induction of new theories, and relationships between important variables;
- Military and intelligence, an area that is highly classified;
- Law enforcement, including criminal profiling, identification of criminals and terrorists, and detection of money-laundering activities.

2.7.5 Data Mining Tools and Technologies

2.7.5.1 Artificial Neural Network

An artificial neural network (ANN) is a connected network of artificial neuron nodes, emulating the network of biological neurons of the human brain. Artificial neural networks have been used successfully in various application areas. One important area is classification: it remains as one of the most important approaches to classification and it would, therefore, be inappropriate not to mention it.

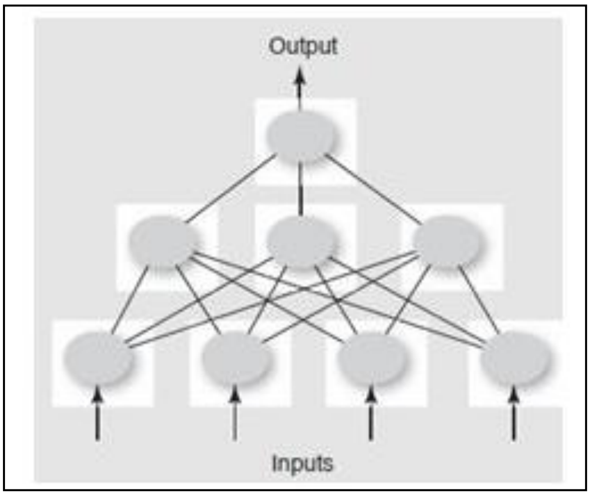


***Figure 2.13** Artificial Neural Networks (Du, 2010)*

2.7.5.2 Decision Trees

A typical decision tree consists of leaf nodes, internal nodes, and links. A leaf node represents a class label. An internal node represents the name of an attribute. The link from a parent node to a child node represents a value of the attribute of the parent node. The decision tree approach for classification *induces*, from a set of training examples, a decision tree as the classification model.

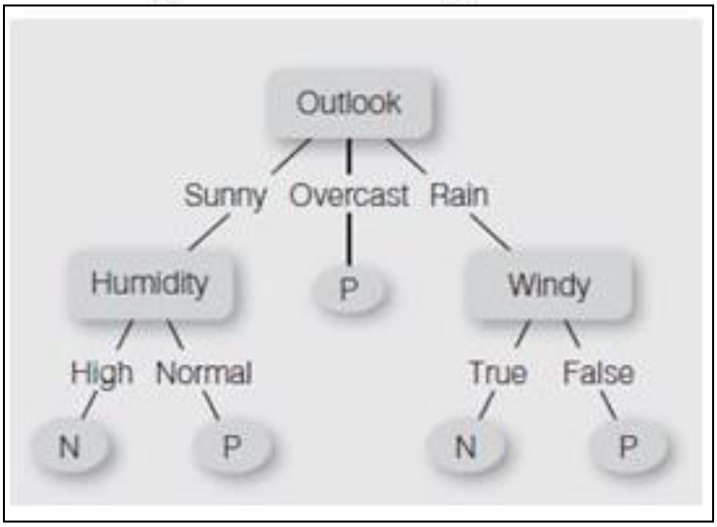


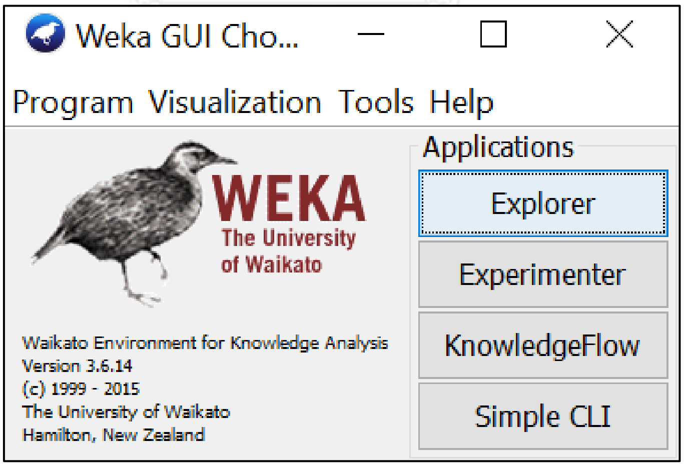
***Figure 2.14** Decision Tree (Du, 2010)*

2.7.5.3 k-Nearest Neighbors (kNN)

The purpose of the k-Nearest Neighbors (kNN) algorithm is to use a database in which the data points are separated into several separate classes to predict the classification of a new sample point. This sort of situation is best motivated through examples.

2.8 Weka data mining software

The Waikato Environment for Knowledge Analysis (WEKA) is a machine learning toolkit developed at the University of Waikato in Hamilton, New Zealand. The software provides many machine-learning, statistics and other data mining solutions for various types of data mining task, such as classification, cluster detection, association rule discovery and attribute selection. The software is also equipped with data pre-processing and post-processing tools and visualization tools so that complete data mining projects can be conducted via a number of different styles of user interface. The toolkit is written in Java and can, therefore, run on various platforms, such as Linux, Windows, and Macintosh. It is distributed under the terms and conditions of the GNU General Public License.



Waikato Environment for Knowledge Analysis
Version 3.6.14
(c) 1999 - 2015
The University of Waikato
Hamilton, New Zealand

Figure 2.15 *Weka machine-learning toolkits for data mining (Bouckaert et al., 2016)*

The ***Explorer*** route provides an interactive way of performing a data mining investigation. Through a simple yet effective graphical user interface, the end user can open an input data set and observe and understand its features via controls on the *Pre-processing* and *visualize* tab pages. The user can also select a data pre-processing operation to prepare the data before mining. Through the Explorer, the user performs a mining task by selecting a mining solution and setting the relevant parameters. The discovered patterns and the evaluation results are displayed, and some patterns can be visualized. The Explorer has limits. By default, it can only deal with an input data set of several thousands of data records, because the entire data set is loaded into the main memory. Although it is possible to change the default setting of memory sizes, the amount of available memory may not be able to accommodate the entire data set.

The ***Experimenter*** interface is particularly designed for evaluation and selection of classification techniques. It allows the user to automate the process by setting different learning algorithms with parameters upon several chosen data set, collecting performance statistics, and testing the significance of the accuracy of the classification models produced by different classification solutions.

The ***KnowledgeFlow*** interface allows the user to set up a more serious batch-processing mining task for larger data sets. Via the controls on the graphical user interface, the user can specify a sequence of pre-processing and mining tasks in the form of a task flow chart. Functions in the Knowledge Flow mode of the system have incremental algorithms behind them to overcome the memory space limits.

The *Simple CLI* route offers an interface that allows the end user to call a Java function by issuing commands with command-line parameters to start mining or data pre-processing functions.

2.9 Weka Algorithms

On the *Classify* panel, when you select a learning algorithm using the *Choose* button the command-line version of the classifier appears in the line beside the button, including the parameters specified with minus signs. To change the parameters, click that line to get an appropriate object editor. They are divided into Bayesian classifiers, trees, rules, functions, lazy classifiers, multi-instance classifiers, and a final miscellaneous category.

2.9.1 Multilayer Perceptron (MLP)

MultilayerPerceptorn need not be run through the graphical interface. Several parameters can be set from the object editor to control its operation. If you are using the graphical interface, they govern the initial network structure, which you can override interactively. With *the autoBuild* set, hidden layers are added and connected. The default is to have the one hidden layer shown in Figure 3.17; however, without *autoBuild*, this would not appear and there would be no connections. The *hidden layers* parameter defines what hidden layers are present and how many nodes each one contains.

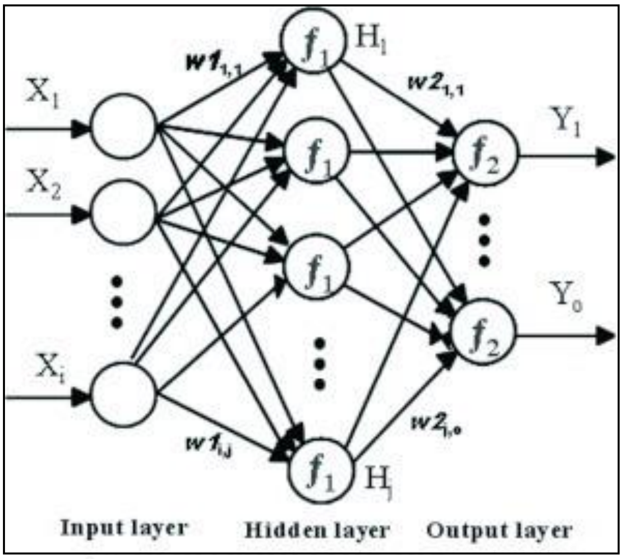


Figure 2.16 *Multilayer perceptron (MLP) (Stehlé et al., 2010)*

2.9.2 M5Rules

M5Rules obtains regression rules from model trees built using M5'. *Ridor* learns rules with exceptions by generating the default rule, using incremental reduced-error pruning to find exceptions with the smallest error rate, finding the best exception, and iterating.

2.9.3 M5P

M5P is model tree learner; trees that are used for numeric prediction are just like ordinary decision trees, except that at each leaf they store either a class value that represents the average value of instances that reach the leaf, in which case the tree is called a *regression tree,* or a linear regression model that predicts the class value of instances that reach the leaf, in which case it is called a *model tree*. In what follows we will talk about model trees because regression trees are really a special case.

Regression and model trees are constructed by first using a decision tree induction algorithm to build an initial tree.

When the model tree is used to predict the value for a test instance, the tree is followed down to a leaf in the normal way, using the instance's attribute value to make routing decisions at each node. The leaf will contain a linear model based on some of the attribute values, and this is evaluated for the test instance to yield a raw predicted value.

2.9.4 REPTree

REPTree builds a decision or regression tree using information gain/variance reduction and prunes it using reduced-error pruning. Optimized for speed, it only sorts values for numeric attributes once. It deals with missing values by splitting instance into pieces, as C4.5 does. You can set the minimum number of instances per leaf, maximum tree depth (useful when boosting trees), the minimum proportion of training set variance for a split (numeric classes only), and the number of folds for pruning.

2.10 Related researches

(Nahm et al., 1996) studied the using chemical for settling of particles in the water treatment processes by using Jar Test and the chemical for settling of particles is Polyaluminum Chloride (PACl) in this research. This research studied the relationship between the 6 independent variables with PACl experimentation. The experiment of this research used the neural network to determine the relationship of the data and compared the data with Jar Test. The results from a neural network have very less discrepant value when compared with the values of Jar Test.

(Gagnon et al., 1997) studied the models of chemical for settling in the water treatment processes by using the neural network to analyze the amount of the chemical for settling. Assigned 4 input variables; they are pH, turbidity, temperature, and conductivity parameters. Alum is dependent variable; used cross-correlation coefficient to customize input data. For the analysis was divided into three groups as Learning, Validation, and Testing data. Used to 2 statistical evaluation types as SD and MAE, the result of Validation has very fewer values of SD and MAE.

(Chun et al., 1999) studied the ANFIS (Adaptive Neuro-Fuzzy Inference Systems) system of chemical for settling of particles in the water treatment processes. Used to algorithm fuzzy logic to analyze the amount of chemical for settling of particles by assigned 4 independent variables, a dependent variable is PACl and compared with other forms of analysis showed that ANFIS system has been accurate analysis more than other forms of analysis to be compared.

(Bae et al., 2004) studied about data mining and artificial modeling to predict the amount of chemical for settling in wastewater treatment processes. They assigned 5 independent and 3dependent variables, the dependent variables were *PACl, PASS,* and *PSO-M*. Doing data mining will easier extract the data and can be applied to describe the relationship between the data. From doing the data mining can be divided the data into 6 groups as 1. training 70%: testing 30% 2. training 50%: testing 50% for all three variables. The result is more efficiently as used of decision trees for choosing the group of chemicals for settling and used neural network analysis. The discrepant value will have very fewer values.

(Charutragulchai, 2006) studied about the addition of alum in the Bangkhen water supply production processes by trees. In the adjusting, the data will be using *10-fold cross-validation method* and 7 independent variables. In the analysis amount of alum by using decision tree analysis was compared with 3 other analysis as *SVM Single, decision tree,* and *neural network*. The results of the RMSE analysis of trees will have minimum values.

(Hannouche et al., 2011)studied about the relationship between turbidity and suspended solids in the drainage system. They collected the samples in the LCPC laboratory that has monitoring parameters in the two drainage systems of *Saint-Mihiel* and *Cordon Bleu*. They divided the sample into two seasons are the dry season and the rainy season. For the analysis, the relationship between turbidity and suspended solids used Mie's theory equations to calculate the proportional relationship, the results, the season was not effected on the relationship between turbidity and suspended solids.

(Daphne et al., 2011) studied the relationship between turbidity and suspended solids in Singapore Rivers. They have collected the sampling in the range of 50 to 100 meters and collected both before and after the rainy day. They have been separated the analysis into two characteristics as 1). General sample: they collected 48 samples, and 2). Choose Samples with suspended solids. They were used *R-square* to assess the precise; for the result, the data after the heavy rain had the R-square changed values more than the data before the rain of up to 1.8 times, but the relationship in both cases, the two that are relatively reliable.

(García-Laencina et al., 2013) studied about the classification of *missing value types* by *multi-task learning perceptron*. This research, they provided data into2 series as *no missing value* to monitoring missing value group to tell us that will be affected much or not and compared with other 4 methods as *K nearing neighbor, Self-organizing map, MLP and Gaussian mixture model*. For corrected methodology is in% mean + SD. They have gotten the result as the *task learning perceptron* has the highest precision; the second is the Gaussian mixture model.

(Bagheri et al., 2015)studied the SBR modeling for wastewater treatment by using *multilayer perceptron (MLP)* and *RBFANN methods*. In this research, has been had 6 independent variables, they are COD, NH4, TP, TSS, FT, RT, Al, and MLVSS.MLP analysis was used the hidden layer 10 layers, but RBFANN analysis was used less than 6 layers; the results of both analytical can make SBR modeling which the model of TSS, TP, COD, and NH4 were best displayed. When looking at the results of the R-square of both models' values were at 0.9 to 0.99 and RMSE values were very nearly zero when compared with themselves of both models as MLP method has precision but RBFANN has gotten from the result of the R-square and RMSE values.

(Kalmegh, 2015) studied about the classification of news in India Country by conducting data mining by using the *REPTree, CART*, and *RandomTree Methods*. In this research has 649 data, separated all those data into 7 groups as business news, crime news, education news, medical news, politics news, sports news, and technology news. The goal was to compare those3 methods as which methods can best classification the information. The results of RandomTree were had precision in the classification the information. Besides that, two methods can classification the politics news is the best, but for other news were had more distortion in the classification.

(Chawakitchareon et al., 2017) studied about prediction of alum dosage in water supply by Weka data mining software. This research presented a comparison of M5P, M5Rules and REPTree to the results from multilayer perceptron (MLP). They input 6 parameters i.e. turbidity, alkalinity, pH, conductivity, color, and suspended solids. The data had been collected from 1st January 2002 to 31st July 2015 at Bangkhen water treatment plant, Thailand. The results indicated that the M5Rules method yielded the highest precision to predict the alum dosage.

CHAPTER III
RESEARCH METHODOLOGY

3.1 Water sampling collection

We measured the water quality characteristics in term of physical and chemical by using the equipment.

1) First equipment called Horiba Sensor U50, this equipment was measured the water quality in term of physical and it can measure 8 parameters i.e. *Temperature, pH, Oxidation Reduction Potential (ORP), Turbidity, Conductivity, Dissolved Oxygen (DO), Total Dissolved Solids (TDS),* and *Salinity*. The collected method, we measured in 4 different depths at 5m, 3m, 1m, and 0.5m.

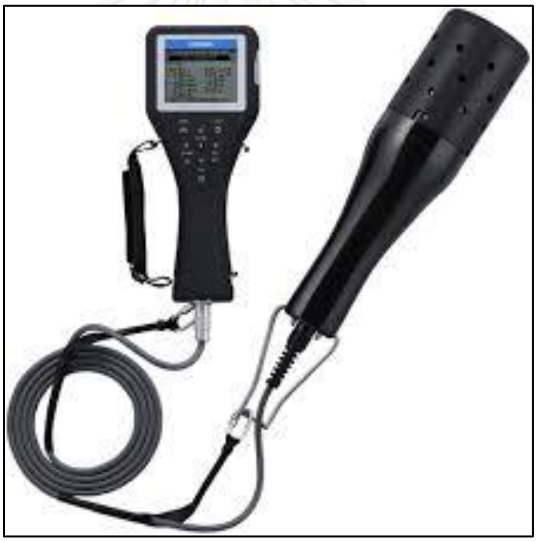


Figure 3.1 *Horiba Sensor U50*

2) We measured the water quality in term of chemical by using the *Horiba's compact ion meter (LAQUAtwin)*, this equipment measured the ions i.e. Ca^+, NO_3^-, Na^+, and K^+. On the hand, we used the *Fluorimeter AND1100* to measure the heavy metal i.e. Zinc (Zn), Cadmium (Cd), Lead (Pb), Uranium (U), Mercury (Hg), and Copper (Cu).

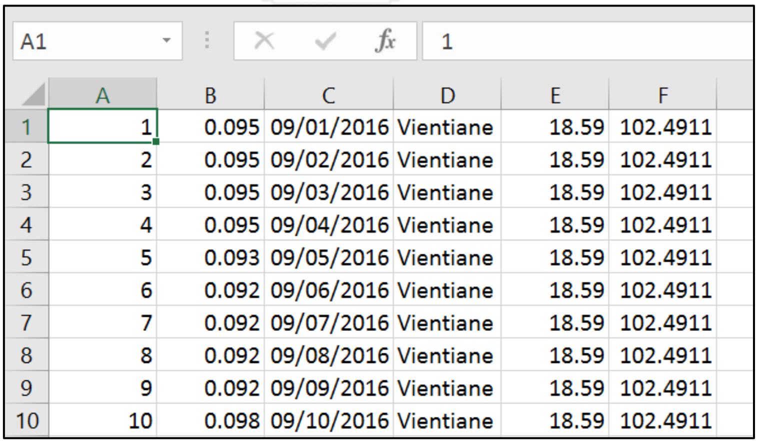
Figure 3.3 *Laquatwin*

Figure 3.2 *Fluorimeter*

3.2 5D World Map (5DWM) System

3.2.1 Preparing the data for uploading and visualization by 5DWM system

We collected 48 files in CSV from (8 parameter from 3 rivers (Mekong River, Nam Ngum River, and Nam Lik River)) and added semantic and spatiotemporal metadata, such Category, Location, Date, and Description for each data in which the data structure as shown in **Figure 3.4**, are based on 5D Word Map System.



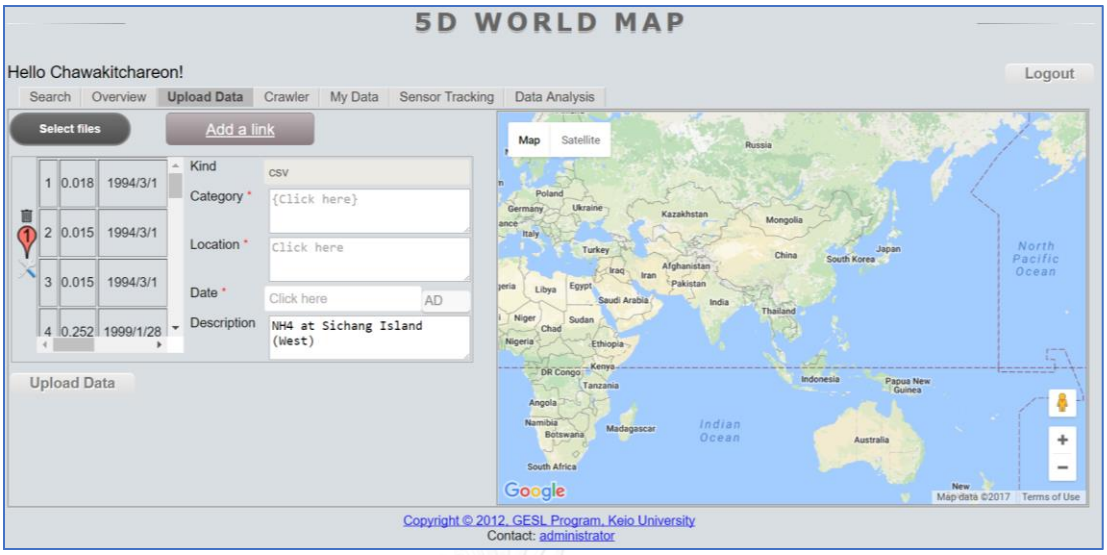
A1 | fx | 1

	A	B	C	D	E	F
1	1	0.095	09/01/2016	Vientiane	18.59	102.4911
2	2	0.095	09/02/2016	Vientiane	18.59	102.4911
3	3	0.095	09/03/2016	Vientiane	18.59	102.4911
4	4	0.095	09/04/2016	Vientiane	18.59	102.4911
5	5	0.093	09/05/2016	Vientiane	18.59	102.4911
6	6	0.092	09/06/2016	Vientiane	18.59	102.4911
7	7	0.092	09/07/2016	Vientiane	18.59	102.4911
8	8	0.092	09/08/2016	Vientiane	18.59	102.4911
9	9	0.092	09/09/2016	Vientiane	18.59	102.4911
10	10	0.098	09/10/2016	Vientiane	18.59	102.4911

Figure 3.4 *The preparing file for uploading to 5DWM*

3.2.2 The data uploading to 5DWM system

Open the 5DWM system and go to the "Upload Data". After that, select the file that already prepared. The display will show in **Figure 3.5** and we fill the information follow the Category, Location, and Date. For the category box, we choose the "water pollution". Finally, click "Upload Data" and the uploaded data will show in "My Data" as **Figure 3.6**.



Hello Chawakitchareon!

1	0.018	1994/3/1
2	0.015	1994/3/1
3	0.015	1994/3/1
4	0.252	1999/1/28

Kind: csv

Category *: {Click here}

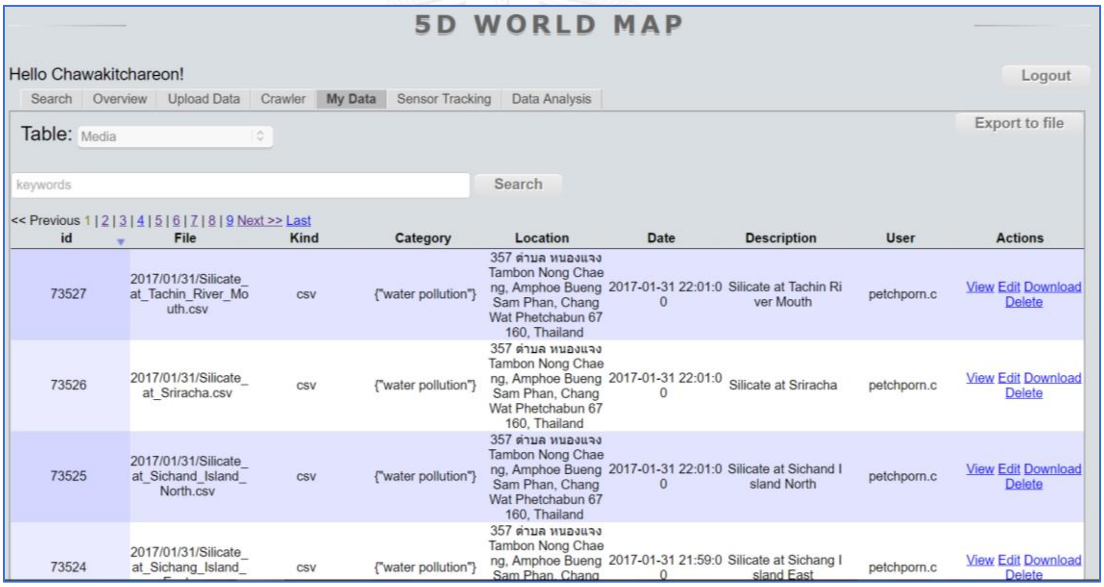
Location *: Click here

Date *: Click here AD

Description: NH4 at Sichang Island (West)

Copyright © 2012, GESL Program, Keio University
Contact: administrator

***Figure 3.5** The display of the data uploading in 5DWM system*



Hello Chawakitchareon!

Table: Media

keywords

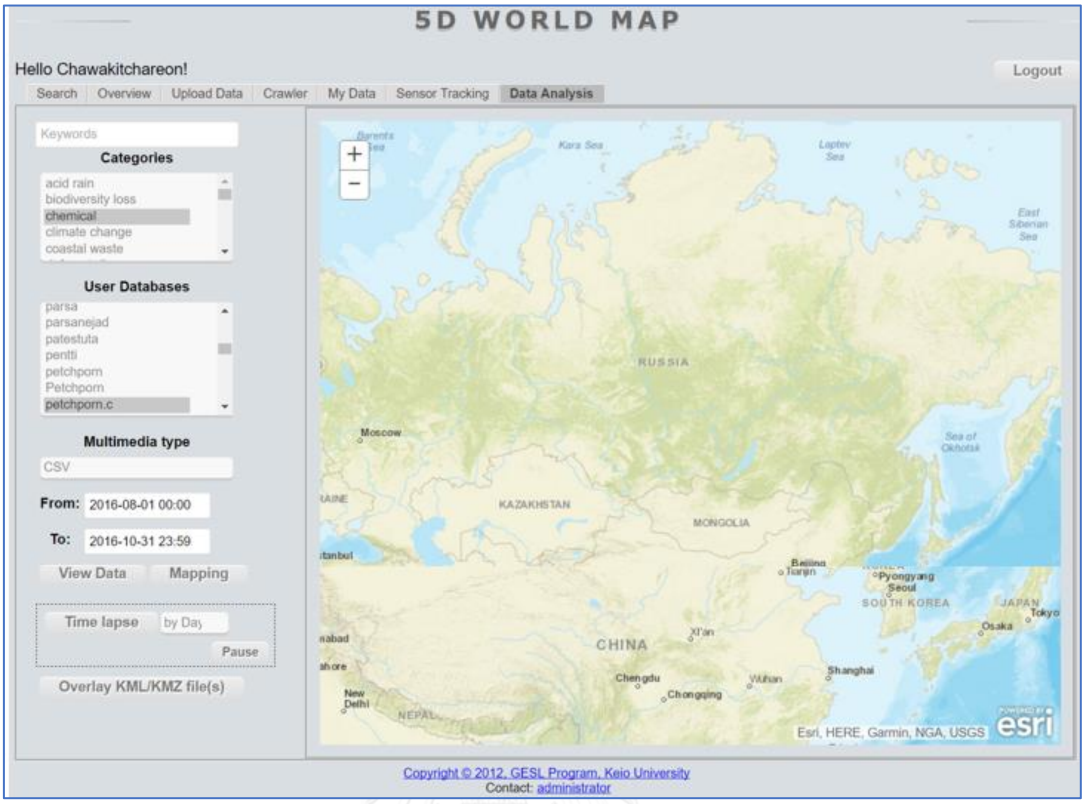
<< Previous 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 Next >> Last

id	File	Kind	Category	Location	Date	Description	User	Actions
73527	2017/01/31/Silicate_at_Tachin_River_Mouth.csv	csv	{"water pollution"}	357 ตำบล หนองแจง Tambon Nong Chaeng, Amphoe Bueng Sam Phan, Chang Wat Phetchabun 67160, Thailand	2017-01-31 22:01:00	Silicate at Tachin River Mouth	petchporn.c	View Edit Download Delete
73526	2017/01/31/Silicate_at_Sriracha.csv	csv	{"water pollution"}	357 ตำบล หนองแจง Tambon Nong Chaeng, Amphoe Bueng Sam Phan, Chang Wat Phetchabun 67160, Thailand	2017-01-31 22:01:00	Silicate at Sriracha	petchporn.c	View Edit Download Delete
73525	2017/01/31/Silicate_at_Sichand_Island_North.csv	csv	{"water pollution"}	357 ตำบล หนองแจง Tambon Nong Chaeng, Amphoe Bueng Sam Phan, Chang Wat Phetchabun 67160, Thailand	2017-01-31 22:01:00	Silicate at Sichand Island North	petchporn.c	View Edit Download Delete
73524	2017/01/31/Silicate_at_Sichang_Island_	csv	{"water pollution"}	357 ตำบล หนองแจง Tambon Nong Chaeng, Amphoe Bueng Sam Phan, Chang	2017-01-31 21:59:00	Silicate at Sichang Island East	petchporn.c	View Edit Download Delete

***Figure 3.6** The data already uploaded in 5DWM system*

3.2.3 Data analysis and visualization by 5DWM

For the data analysis, we go to the "Data Analysis" function. The display will show in **Figure 3.7.** In this case, in the category box choose "water pollution", the user databases choose your user name such as our user name is "Petchporn.c", the multimedia type choose "CSV", fill the date of the data collected in the "From" box and "To" box. After that, click view data and the data view show like **Figure 3.8.** Finally, the system will visualize the sampling points in term of color that choose by itself like **Figure 3.9**.

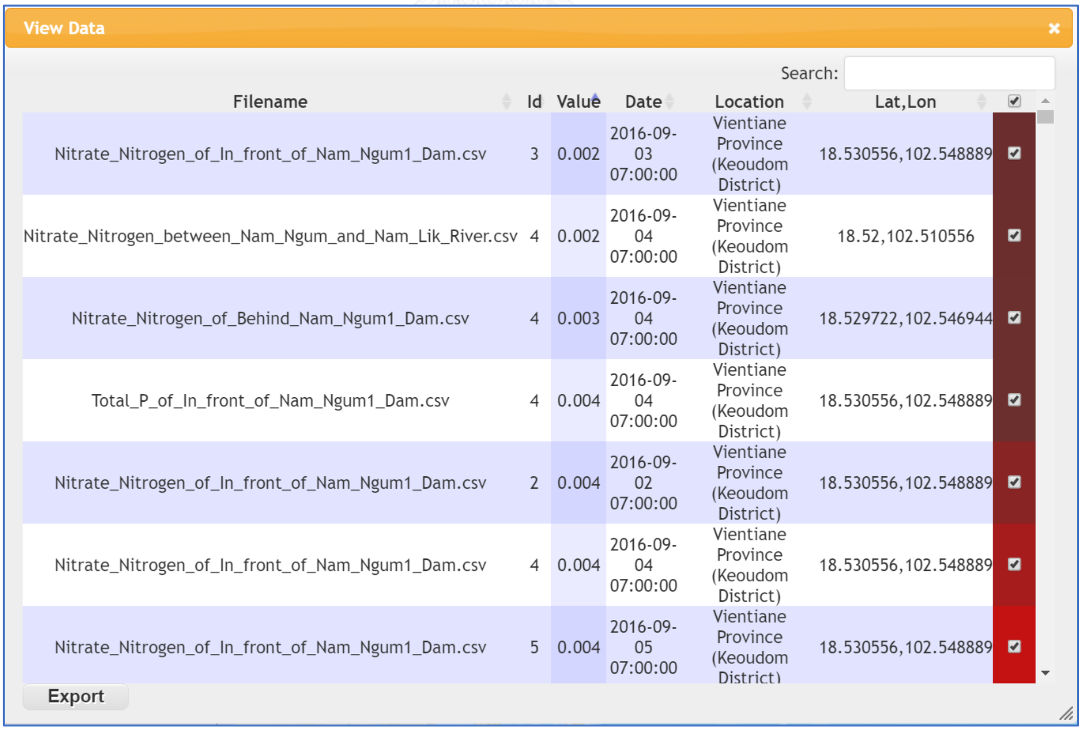


Multimedia type: CSV

From: 2016-08-01 00:00

To: 2016-10-31 23:59

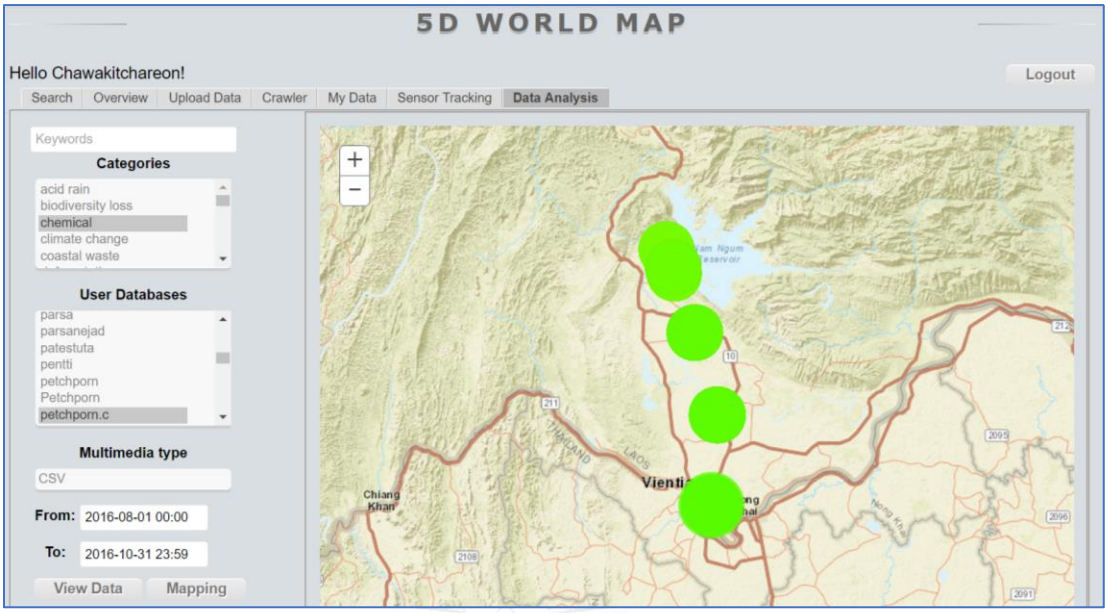
Figure 3.7 *The display of data analysis in 5DWM*



Search:

Filename	Id	Value	Date	Location	Lat,Lon	☑
Nitrate_Nitrogen_of_In_front_of_Nam_Ngum1_Dam.csv	3	0.002	2016-09-03 07:00:00	Vientiane Province (Keoudom District)	18.530556,102.548889	☑
Nitrate_Nitrogen_between_Nam_Ngum_and_Nam_Lik_River.csv	4	0.002	2016-09-04 07:00:00	Vientiane Province (Keoudom District)	18.52,102.510556	☑
Nitrate_Nitrogen_of_Behind_Nam_Ngum1_Dam.csv	4	0.003	2016-09-04 07:00:00	Vientiane Province (Keoudom District)	18.529722,102.546944	☑
Total_P_of_In_front_of_Nam_Ngum1_Dam.csv	4	0.004	2016-09-04 07:00:00	Vientiane Province (Keoudom District)	18.530556,102.548889	☑
Nitrate_Nitrogen_of_In_front_of_Nam_Ngum1_Dam.csv	2	0.004	2016-09-02 07:00:00	Vientiane Province (Keoudom District)	18.530556,102.548889	☑
Nitrate_Nitrogen_of_In_front_of_Nam_Ngum1_Dam.csv	4	0.004	2016-09-04 07:00:00	Vientiane Province (Keoudom District)	18.530556,102.548889	☑
Nitrate_Nitrogen_of_In_front_of_Nam_Ngum1_Dam.csv	5	0.004	2016-09-05 07:00:00	Vientiane Province (Keoudom District)	18.530556,102.548889	☑

Figure 3.8 *The display of 5DWM when we viewed the data*



Hello Chawakitchareon!

Multimedia type: CSV

From: 2016-08-01 00:00

To: 2016-10-31 23:59

Figure 3.9 The display of 5DWM when the system already visualized

3.3 Weka Data Mining Software

Weka developed from Java, it has many algorithms. In this research, we used Weka version of 3.6.14. In the Weka Explorer has 6 functions to use i.e. preprocess, classify, cluster, associate, select attributes, and visualize.

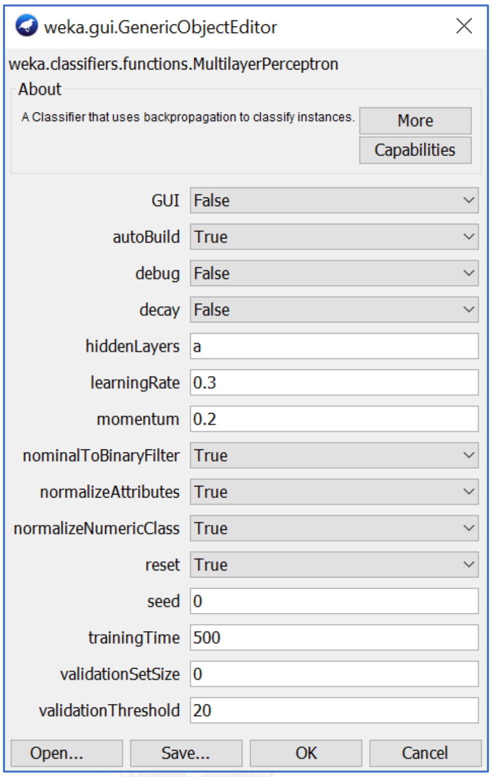
3.3.1 Advantages of Weka

- It has many algorithms to the data mining creating.
- Can directly download because it is a public software.
- It is not a heavy software.
- Include the preprocessing and any techniques for the model building.
- Easy to use and has the graphical type in the software for the user.

3.3.2 Functions for the model building and the functions working

3.3.2.1 Multilayer Perceptron (MLP)

A Classifier that uses backpropagation to classify instances. This network can be built by hand, created by an algorithm or both. The network can also be monitored and modified during training time. The nodes in this network are all sigmoid (except for when the class is numeric in which case the output nodes become unthresholded linear units). In the MLP method has many functions to adjust by ourselves, these functions indicated in **Figure 3.10**.



weka.gui.GenericObjectEditor

weka.classifiers.functions.MultilayerPerceptron

About

A Classifier that uses backpropagation to classify instances.

GUI	False
autoBuild	True
debug	False
decay	False
hiddenLayers	a
learningRate	0.3
momentum	0.2
nominalToBinaryFilter	True
normalizeAttributes	True
normalizeNumericClass	True
reset	True
seed	0
trainingTime	500
validationSetSize	0
validationThreshold	20

***Figure 3.10** The MLP function in Weka*

In this research, we especially adjust the hidden layer, learning rate, momentum, seed, and training time. For the definition of these function shown in **Table 3.1** as below.

***Table 3.1** The definition of some MLP functions*

Options	Definition
HiddenLayer	This defines the hidden layers of the neural network. This is a list of positive whole numbers. his will only be used if auto build is set. There are also wildcard values 'a' = (attribs + classes) / 2, 'i' = attribs, 'o' = classes, 't' = attribs + classes.
LearningRate	The amount the weights are updated.
momentum	Momentum applied to the weights during updating.
seed	Seed used to initialise the random number generator. Random numbers are used for setting the initial weights of the connections between nodes, and also for shuffling the training data.
trainingTime	The number of epochs to train through. If the validation set is non-zero, then it can terminate the network early.

The working of MLP method is the same the Neural Network that shown in the **Figure 3.11**.

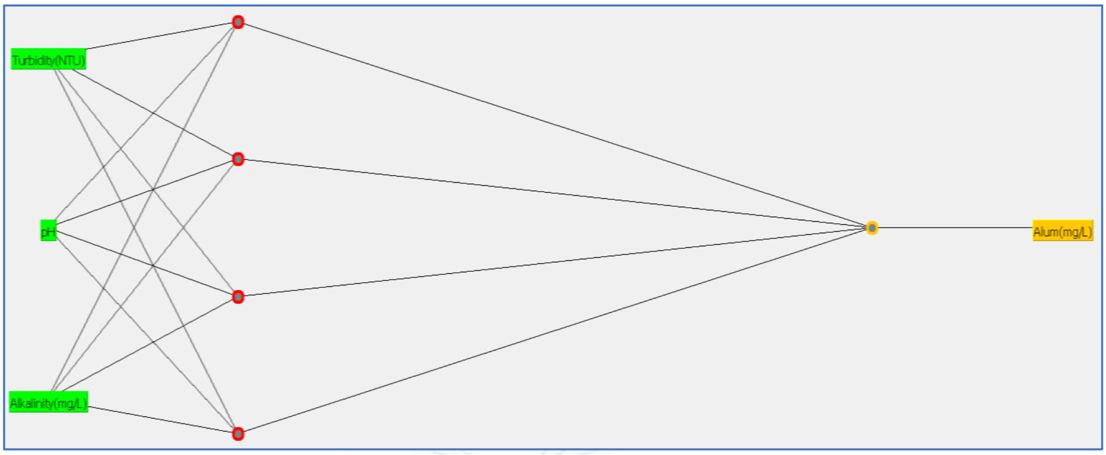
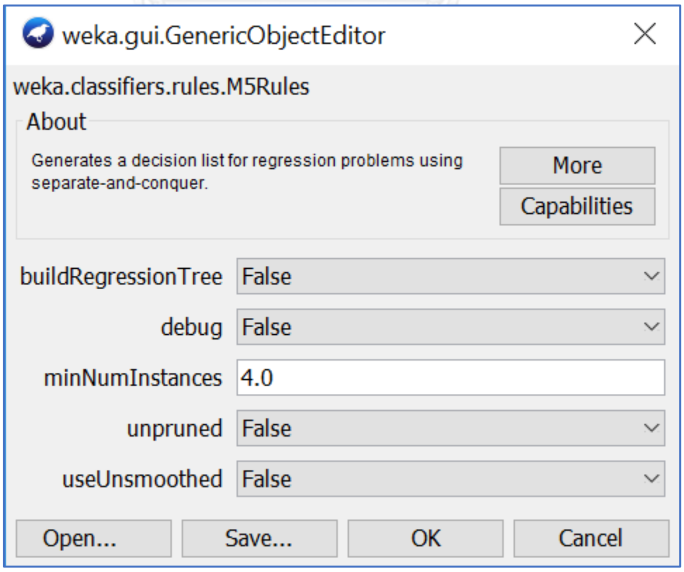


Figure 3.11 *Shown the MLP working*

3.3.2.2 M5Rules

Generates a decision list for regression problems using separate-and-conquer. In each iteration, it builds a model tree using M5 and makes the "best" leaf into a rule.

The M5Rules options are shown in **Figure 3.12** as below.


weka.gui.GenericObjectEditor

weka.classifiers.rules.M5Rules

About

Generates a decision list for regression problems using separate-and-conquer.

buildRegressionTree False

debug False

minNumInstances 4.0

unpruned False

useUnsmoothed False

Figure 3.12 *The M5Rules functions in Weka*

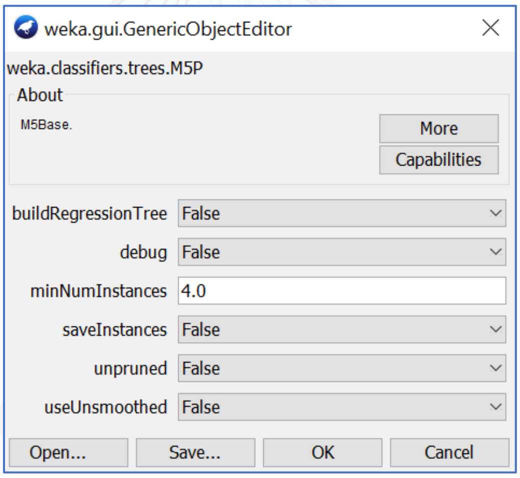
The definition of M5Rules options i.e. buildRegressionTree, debug, minNuminstances, unpruned, and useUnsmoothed are indicated in **Table 3.2**

Table 3.2 *The definition of M5Rules options*

Options	Definition
buildRegressionTree	Whether to generate a regression tree/rule instead of a model tree/rule.
debug	If set to true, classifier may output additional info to the console.
minNumInstance	The minimum number of instances to allow at a leaf node.
unpruned	Whether unpruned tree/rules are to be generated.
useUnsmoothed	Whether to use unsmoothed predictions.

3.3.2.3 M5P

Implements base routines for generating M5 Model trees and rules. The original algorithm M5 was invented by R. Quinlan (Quinlan, 1992) and Yong Wang made improvements (Wang and Witten, 1996). The M5P options are shown in **Figure 3.13**.



weka.gui.GenericObjectEditor

weka.classifiers.trees.M5P

About

M5Base.

buildRegressionTree False

debug False

minNumInstances 4.0

saveInstances False

unpruned False

useUnsmoothed False

Figure 3.13 *The M5P options in Weka*

The definition of M5P options indicated in **Table 3.3** as below:

Table 3.3 *The definition of M5P options*

Options	Definition
buildRegressionTree	Whether to generate a regression tree/rule instead of a model tree/rule.
debug	If set to true, classifier may output additional info to the console.
minNumInstances	The minimum number of instances to allow at a leaf node.
saveInstances	Whether to save instance data at each node in the tree for visualization purposes.
unpruned	Whether unpruned tree/rules are to be generated.
useUnsmoothed	Whether to use unsmoothed predictions.

The visualization tree of M5P method is shown in **Figure 3.14** as below:

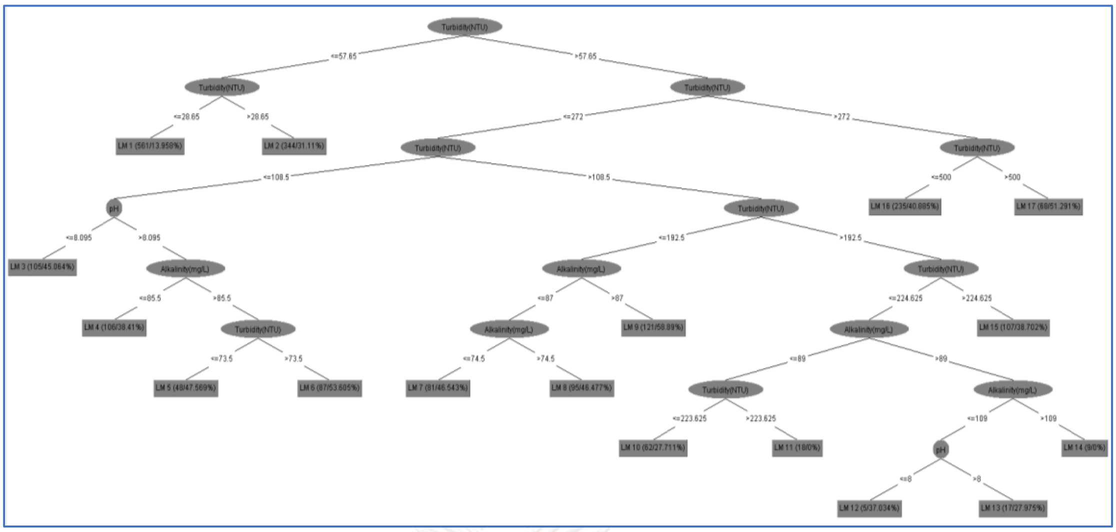
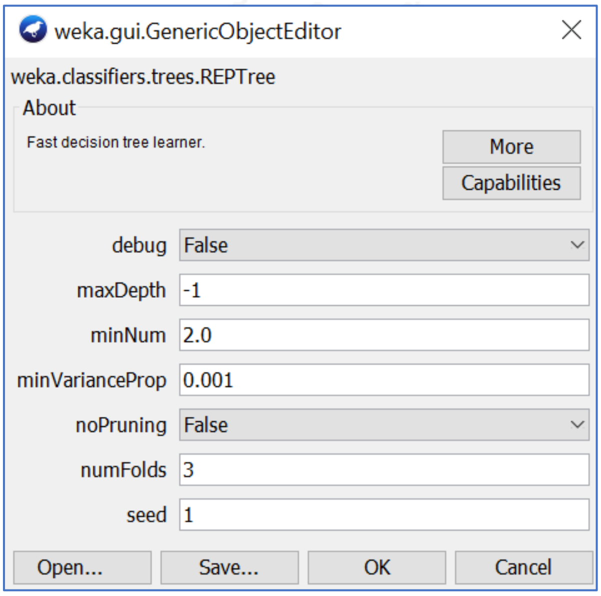


Figure 3.14 *The visualization tree of M5P method*

3.3.2.4 REPTree

Fast decision tree learner. Builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with back fitting). Only sorts values for numeric attributes once. Missing values are dealt with by splitting the corresponding instances into pieces (i.e. as in C4.5). The REPTree options are indicated in **Figure 3.15** as below.


weka.gui.GenericObjectEditor

weka.classifiers.trees.REPTree

About

Fast decision tree learner.

debug False

maxDepth -1

minNum 2.0

minVarianceProp 0.001

noPruning False

numFolds 3

seed 1

Figure 3.15 *The REPTree options in Weka*

Table 3.4 The definition of REPTree options

Options	Definition
debug	If set to true, classifier may output additional info to the console.
maxDepth	The maximum tree depth (-1 for no restriction).
minNum	The minimum total weight of the instances in a leaf.
minVarianceProp	The minimum proportion of the variance on all the data that needs to be present at a node in order for splitting to be performed in regression trees.
noPruning	Whether pruning is performed.
numFolds	Determines the amount of data used for pruning. One-fold is used for pruning, the rest for growing the rules.
seed	The seed used for randomizing the data.

The visualized tree of REPTree indicated in **Figure 3.16** as below.

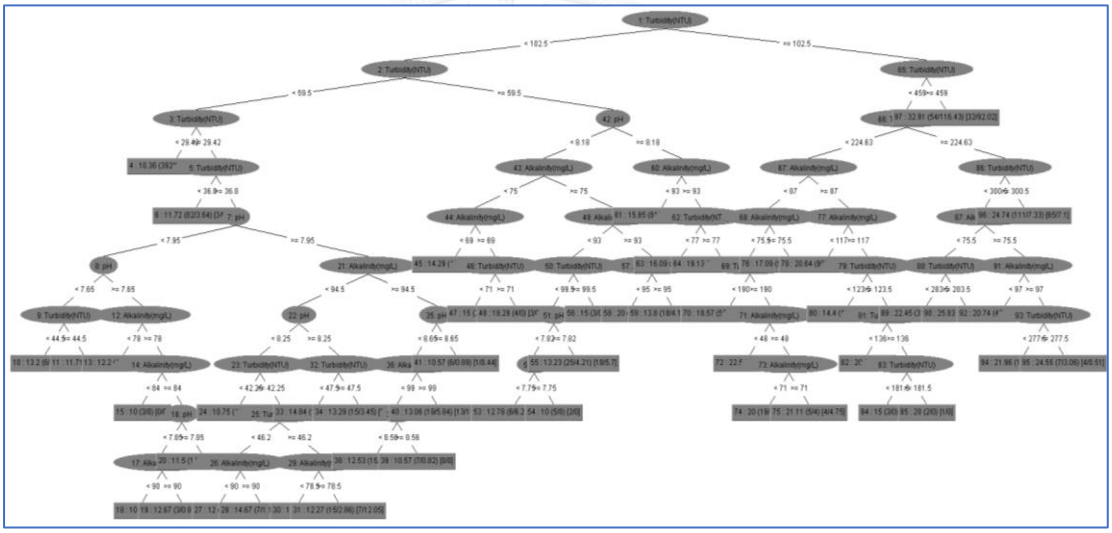
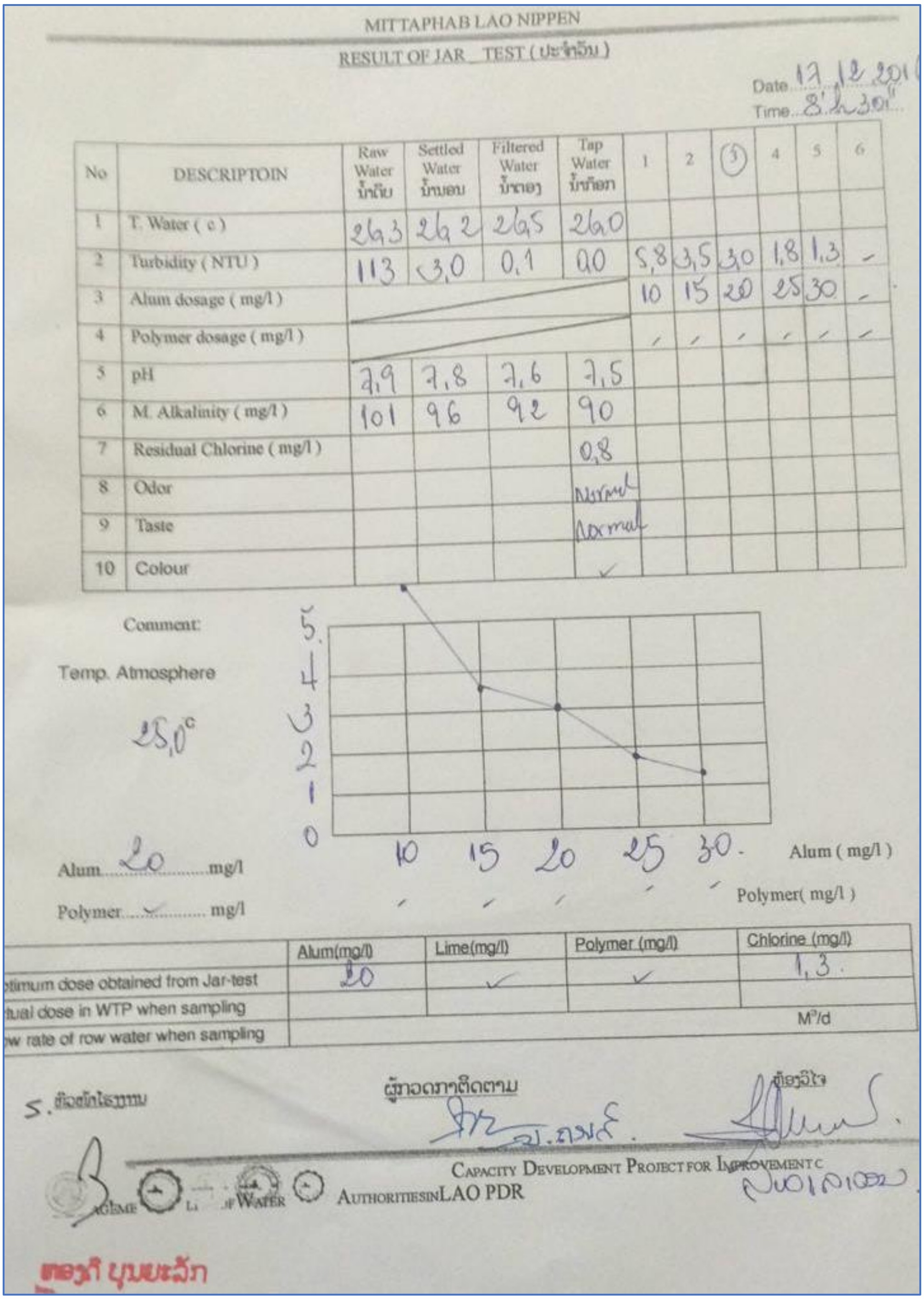


***Figure 3.16** The visualized tree of REPTree*

3.4 Data type

The data information for the processing model is the raw water data in the intake of the Chinaimo and Dongmarkkaiy Water Treatment Plant. For the Chinaimo Water Treatment Plant (CWTP), we collected from 2009-2016 and we collected the 2,038 records. For the Dongmarkkaiy Water Treatment Plant (DWTP), we collected from 2008-2016 and we collected the 2,802 records. For the collected data is the hard copy of the Jar-Test result as shown in **Figure 3.17**. The parameters of raw water are temperature, turbidity, pH, and alkalinity. For the coagulant aid is alum. Therefore, we must key those parameters by hands that the alum must be in the last column because we would like to predict the alum dosage as indicated in **Figure 3.18.**



MITTAPHAB LAO NIPPEN

RESULT OF JAR _ TEST (ປະຈຳວັນ)

Date 17.12.2016

Time 8h30

No	DESCRIPTOIN	Raw Water ນ້ຳດິບ	Settled Water ນ້ຳນອນ	Filtered Water ນ້ຳກອງ	Tap Water ນ້ຳກ໊ອກ	1	2	③	4	5	6
1	T. Water (c)	26,3	26,2	26,5	26,0						
2	Turbidity (NTU)	113	<3,0	0,1	0,0	5,8	3,5	3,0	1,8	1,3	-
3	Alum dosage (mg/l)					10	15	20	25	30	-
4	Polymer dosage (mg/l)					/	/	/	/	/	/
5	pH	7,9	7,8	7,6	7,5						
6	M. Alkalinity (mg/l)	101	96	92	90						
7	Residual Chlorine (mg/l)				0,8						
8	Odor				Normal						
9	Taste				Normal						
10	Colour				✓						

Comment:

Temp. Atmosphere

25,0°c

Alum 20 mg/l

Polymer ✓ mg/l

	Alum(mg/l)	Lime(mg/l)	Polymer (mg/l)	Chlorine (mg/l)
ptimum dose obtained from Jar-test	20	✓	✓	1,3.
tual dose in WTP when sampling				
w rate of row water when sampling				M³/d

ຫົວໜ້າໂຮງງານ

ຜູ້ກວດກາຕິດຕາມ

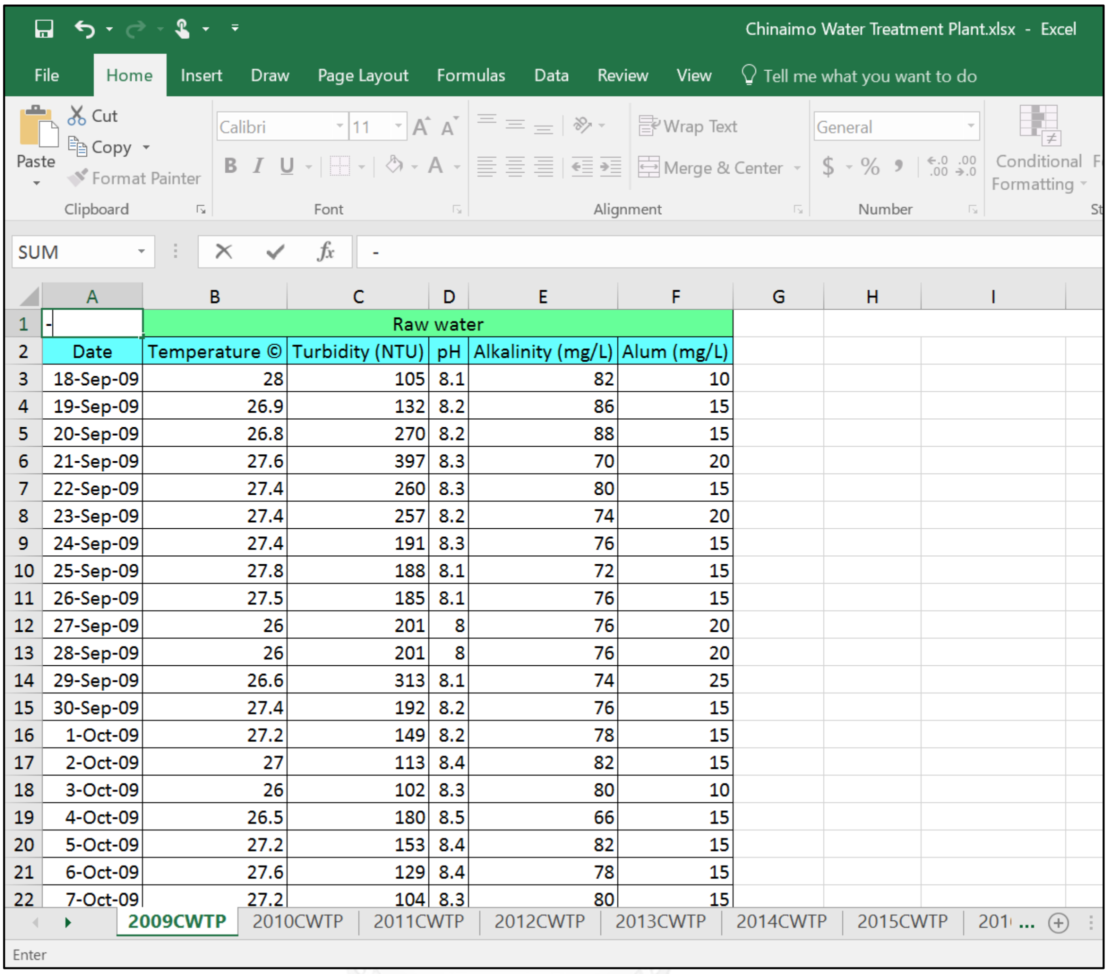
ຜູ້ວິໄຈ

CAPACITY DEVELOPMENT PROJECT FOR IMPROVEMENT C

AUTHORITIESINLAO PDR

ທອງວີ ບຸນຍະລັກ

Figure 3.17 The data is the hard copy

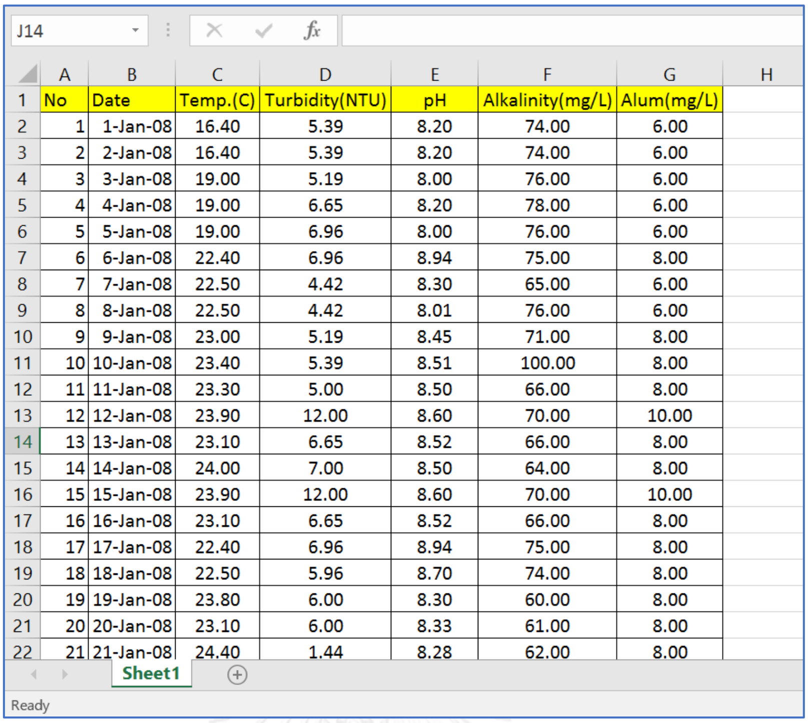


	A	B	C	D	E	F
1	-	Raw water				
2	Date	Temperature ©	Turbidity (NTU)	pH	Alkalinity (mg/L)	Alum (mg/L)
3	18-Sep-09	28	105	8.1	82	10
4	19-Sep-09	26.9	132	8.2	86	15
5	20-Sep-09	26.8	270	8.2	88	15
6	21-Sep-09	27.6	397	8.3	70	20
7	22-Sep-09	27.4	260	8.3	80	15
8	23-Sep-09	27.4	257	8.2	74	20
9	24-Sep-09	27.4	191	8.3	76	15
10	25-Sep-09	27.8	188	8.1	72	15
11	26-Sep-09	27.5	185	8.1	76	15
12	27-Sep-09	26	201	8	76	20
13	28-Sep-09	26	201	8	76	20
14	29-Sep-09	26.6	313	8.1	74	25
15	30-Sep-09	27.4	192	8.2	76	15
16	1-Oct-09	27.2	149	8.2	78	15
17	2-Oct-09	27	113	8.4	82	15
18	3-Oct-09	26	102	8.3	80	10
19	4-Oct-09	26.5	180	8.5	66	15
20	5-Oct-09	27.2	153	8.4	82	15
21	6-Oct-09	27.6	129	8.4	78	15
22	7-Oct-09	27.2	104	8.3	80	15

Figure 3.18 *The data information was typed by hands in excel*

3.5 Data-Preparation

We gathered all data of each year in the same file by using Microsoft Excel. For the data ranking is ranked from the oldest data to the newest data that indicated in **Figure 3.19**. Do not switch the data such as the turbidity value can’t put in pH value or the parameter value of 1st October 2009 can’t put in 3rd October 2009.



No	Date	Temp.(C)	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
1	1-Jan-08	16.40	5.39	8.20	74.00	6.00
2	2-Jan-08	16.40	5.39	8.20	74.00	6.00
3	3-Jan-08	19.00	5.19	8.00	76.00	6.00
4	4-Jan-08	19.00	6.65	8.20	78.00	6.00
5	5-Jan-08	19.00	6.96	8.00	76.00	6.00
6	6-Jan-08	22.40	6.96	8.94	75.00	8.00
7	7-Jan-08	22.50	4.42	8.30	65.00	6.00
8	8-Jan-08	22.50	4.42	8.01	76.00	6.00
9	9-Jan-08	23.00	5.19	8.45	71.00	8.00
10	10-Jan-08	23.40	5.39	8.51	100.00	8.00
11	11-Jan-08	23.30	5.00	8.50	66.00	8.00
12	12-Jan-08	23.90	12.00	8.60	70.00	10.00
13	13-Jan-08	23.10	6.65	8.52	66.00	8.00
14	14-Jan-08	24.00	7.00	8.50	64.00	8.00
15	15-Jan-08	23.90	12.00	8.60	70.00	10.00
16	16-Jan-08	23.10	6.65	8.52	66.00	8.00
17	17-Jan-08	22.40	6.96	8.94	75.00	8.00
18	18-Jan-08	22.50	5.96	8.70	74.00	8.00
19	19-Jan-08	23.80	6.00	8.30	60.00	8.00
20	20-Jan-08	23.10	6.00	8.33	61.00	8.00
21	21-Jan-08	24.40	1.44	8.28	62.00	8.00

Figure 3.19 *Data-preparation*

3.6 Data-Preprocessing Technique

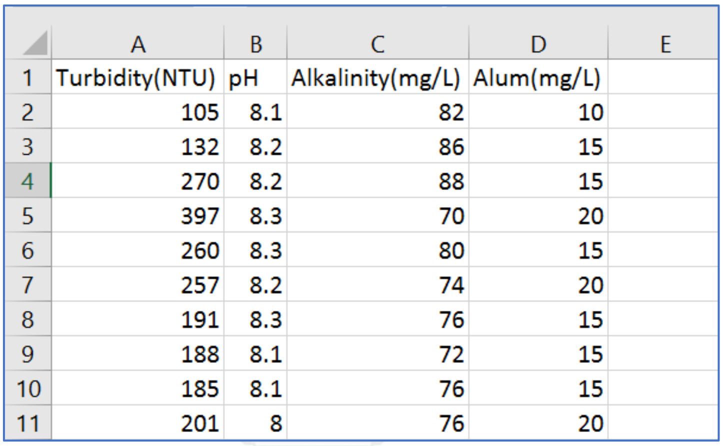
For the data-preprocessing technique is upon the discretion of the researchers as how they adjust the data information, which data should cut out, and which data should substitute by which data. In the data preprocessing do not put these symbols i.e. For this research, we divided the data preprocessing out into 2 groups as below:

- ***1st data preprocessing***: We substituted all missing values of each parameter by the average value of that parameter, computed by each month. After the data preprocessing, we get the 1st data group of 2,069 records from 2,038 records for the CWTP and for the DWTP, we got the 1st data group of 2,861 records from 2,802 records.
- ***2nd data preprocessing***: We cut off the missing value to reduce bias. After the data preprocessing, we got the 2nd data group of 2,022 records from 2,038 records for the CWTP and for the DWTP, we got the 2nd data group of 2,284 records from 2,802 records.

3.7 Preparation the database file for classifier, Processing model, and Model adjustment

3.7.1 Prepare database file

When the file already passed the preprocessing, we deleted the No, Date, and Temperature column out because we need only turbidity, pH, alkalinity, and alum for the model building as indicated in **Figure 3.20**. Finally, save file in the CSV file as “Name-model building.csv”

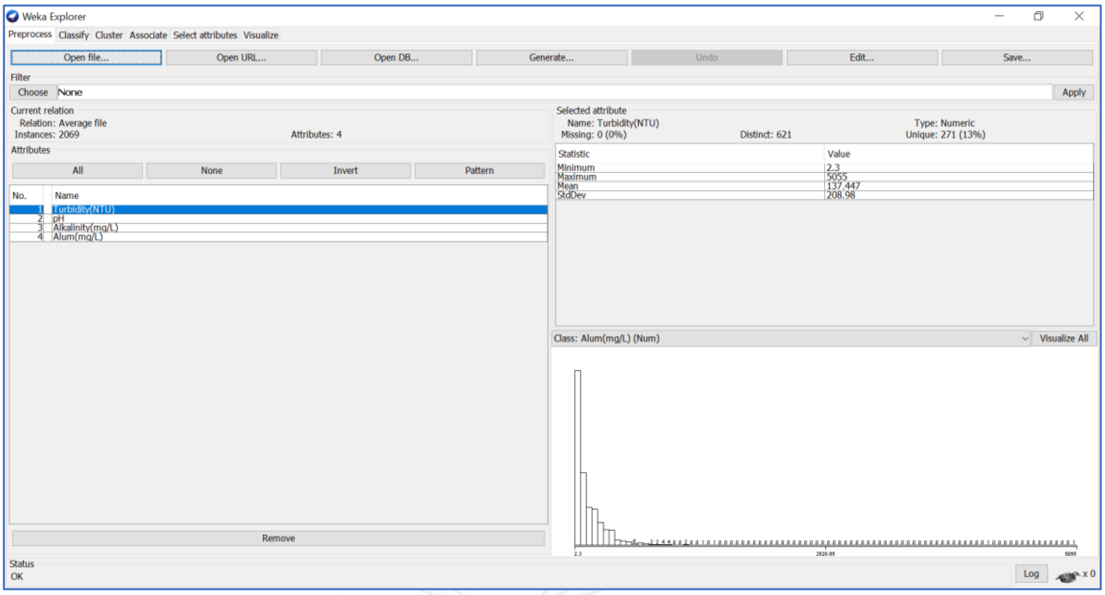


	A	B	C	D	E
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)	
2	105	8.1	82	10	
3	132	8.2	86	15	
4	270	8.2	88	15	
5	397	8.3	70	20	
6	260	8.3	80	15	
7	257	8.2	74	20	
8	191	8.3	76	15	
9	188	8.1	72	15	
10	185	8.1	76	15	
11	201	8	76	20	

***Figure 3.20** The file preparation for the model building*

3.7.2 Processing model

Open the Weka software, click on “Preprocess”, choose “Open file” that we already prepared. If the file is good file or well prepared in the preprocessing, we can open the file as **Figure 3.21.** If do not open the file, please go back to data preprocessing again because in the file maybe have these symbols i.e. = / - \ []. _ , $ # ! @ ^ * ‘’ “” or space in the database file.



Filter: None

Current relation
Relation: Average file
Instances: 2069
Attributes: 4

Attributes

No.	Name
1	Turbidity(NTU)
2	pH
3	Alkalinity(mg/L)
4	Alum(mg/L)

Selected attribute
Name: Turbidity(NTU)
Missing: 0 (0%)
Distinct: 621
Type: Numeric
Unique: 271 (13%)

Statistic	Value
Minimum	2.3
Maximum	5055
Mean	137.447
StdDev	208.98

Class: Alum(mg/L) (Num)

Status
OK

Figure 3.21 *The display of Weka that can open the file*

In the display can open the file, we can visualize the parameters that they input to the software as **Figure 3.22**.

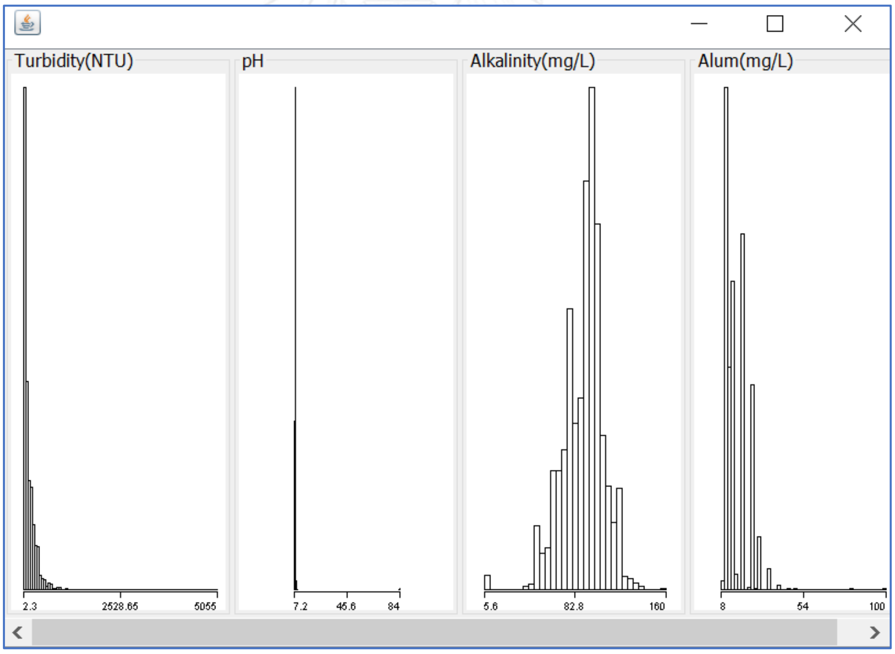
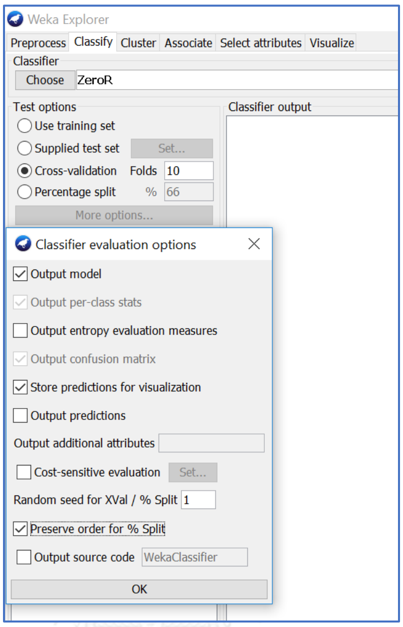


Figure 3.22 *The visualization of parameters*

When we can open the file, click on "Classify" and choose "Test options" to be "Cross-validation Fold = 10". After that click on "More options…" and mark on *"Output model"* and *"Preserve order for % Split"* as **Figure 3.23**. In the "Classifier", change to be Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree respectively.

After that, click on “Start” and wait until the chicken stop that it is in the right across of the Weka display. The created model will show in the “Classifier output” as **Figure 3.24**. Finally, right-click on the created model in the “Result list” and click “Save model”.

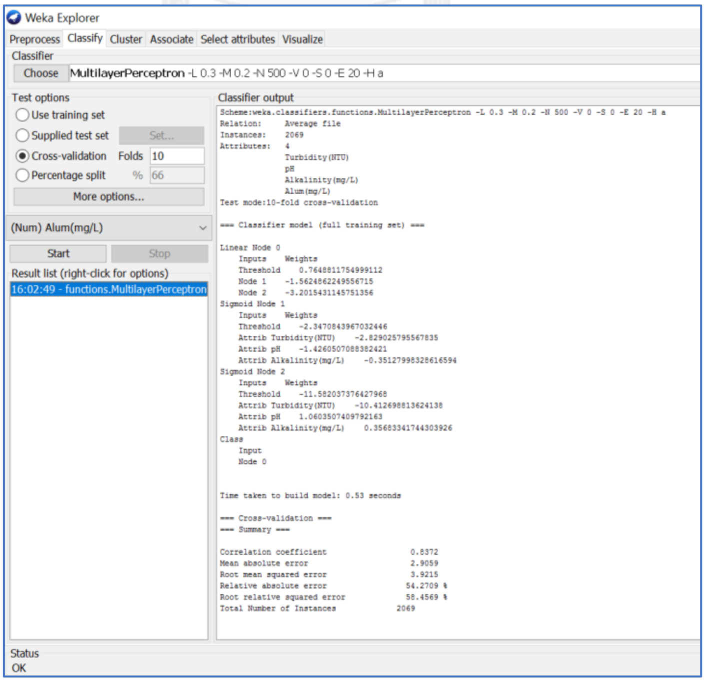


Classifier: ZeroR

Test options
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

☑ Output model
☑ Output per-class stats
☐ Output entropy evaluation measures
☑ Output confusion matrix
☑ Store predictions for visualization
☐ Output predictions
Output additional attributes
☐ Cost-sensitive evaluation
Random seed for XVal / % Split 1
☑ Preserve order for % Split
☐ Output source code WekaClassifier

__Figure 3.23__ The classifier evaluation options



Classifier: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Num) Alum(mg/L)

16:02:49 - functions.MultilayerPerceptron

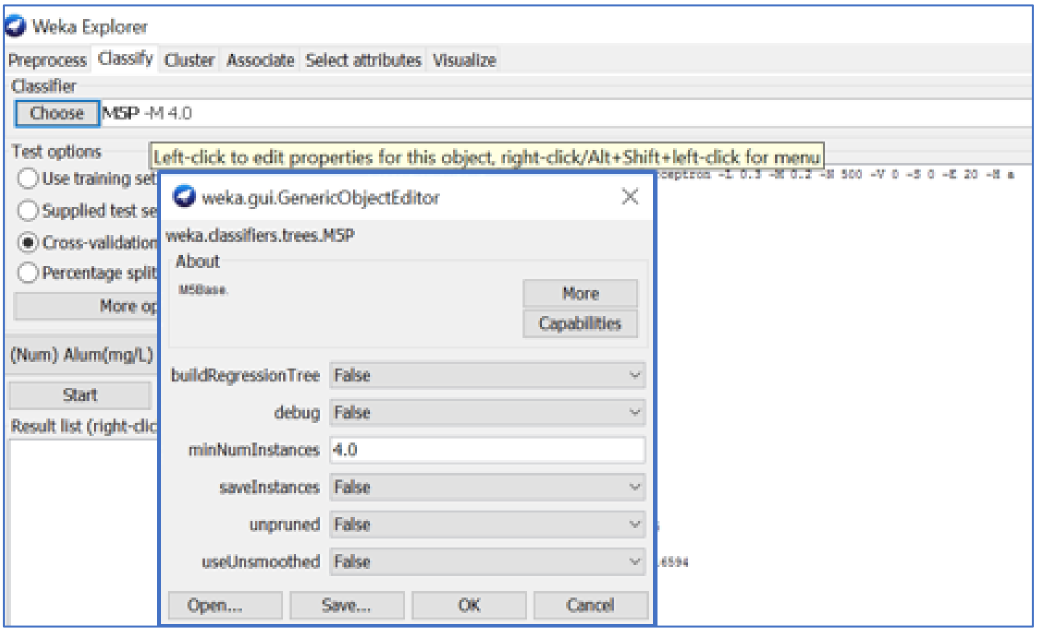
Classifier output

```
Scheme:weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Relation:     Average file
Instances:    2069
Attributes:   4
              Turbidity(NTU)
              pH
              Alkalinity(mg/L)
              Alum(mg/L)
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

Linear Node 0
    Inputs    Weights
    Threshold    0.7648811754999112
    Node 1    -1.5624862249556715
    Node 2    -3.2015431145751356
Sigmoid Node 1
    Inputs    Weights
    Threshold    -2.3470843967032446
    Attrib Turbidity(NTU)    -2.829025795567835
    Attrib pH    -1.4260507088382421
    Attrib Alkalinity(mg/L)    -0.35127998328616594
Sigmoid Node 2
    Inputs    Weights
    Threshold    -11.582037376427968
    Attrib Turbidity(NTU)    -10.412698813624138
    Attrib pH    1.0603507409792163
    Attrib Alkalinity(mg/L)    0.35683341744303926
Class
    Input
    Node 0


Time taken to build model: 0.53 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.8372
Mean absolute error                      2.9059
Root mean squared error                  3.9215
Relative absolute error                 54.2709 %
Root relative squared error             58.4569 %
Total Number of Instances             2069
```

__Figure 3.24__ The completely of the model building

3.7.3 Model adjustment

Each algorithm in the classifier of Weka can synthesize the model by statistical and give any configurations that each configuration give different RMSE value. The very less RMSE indicate that the model give the predictive value is nearly the actual value. In this case, we can adjust in any configurations but each method that used to build the model use different synthesis time. We can adjust the model by left-click on the algorithms that use in the classifier and adjust all the options in that algorithm which configuration will give the less RMSE and save the adjusted value for the comparing the result. The model adjustment indicated in **Figure 3.25** as below.



M5P -M 4.0

Left-click to edit properties for this object, right-click/Alt+Shift+left-click for menu

(Num) Alum(mg/L)

weka.classifiers.trees.M5P

M5Base.

buildRegressionTree: False

debug: False

minNumInstances: 4.0

saveInstances: False

unpruned: False

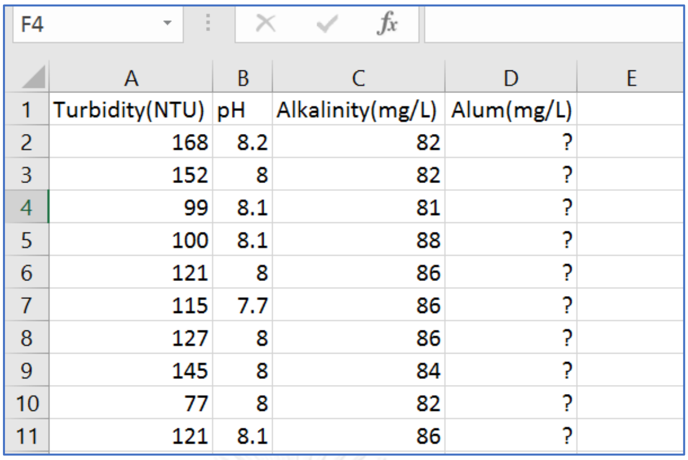
useUnsmoothed: False

__Figure 3.25__ The options of algorithms for the model adjustment

3.8 Supply Test Set and Prediction

3.8.1 Preparing file for the prediction

The predictive file is the same standard of the file for the model building. The parameter name of the predictive file is the same parameter name of the file for the model creating. The file must be CSV file. For the predictive file rank the parameter from Turbidity(NTU) (1st column), pH (2nd column), Alkalinity(mg/L) (3rd column), and Alum(mg/L) (4th column). For the alum column must substitute the value by question mark “?” as **Figure 3.26** and save file in “Name-prediction.csv”. In this research, we divided the data into 2 seasons, thus, we must prepare the predictive file in term of drying and rainy season.

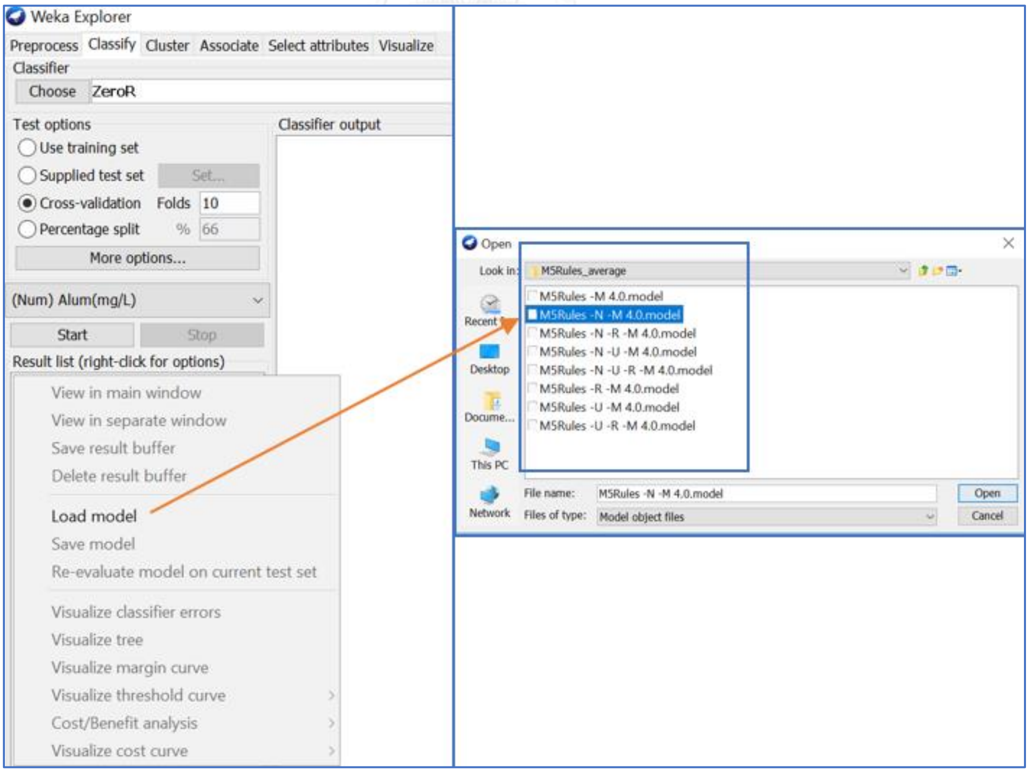


	A	B	C	D	E
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)	
2	168	8.2	82	?	
3	152	8	82	?	
4	99	8.1	81	?	
5	100	8.1	88	?	
6	121	8	86	?	
7	115	7.7	86	?	
8	127	8	86	?	
9	145	8	84	?	
10	77	8	82	?	
11	121	8.1	86	?	

Figure 3.26 *The file preparation for the prediction*

3.8.2 The model opening for the prediction

1. In the "Result list", right-click and choose "Load model", choose the created model as shown in **Figure 3.27.**



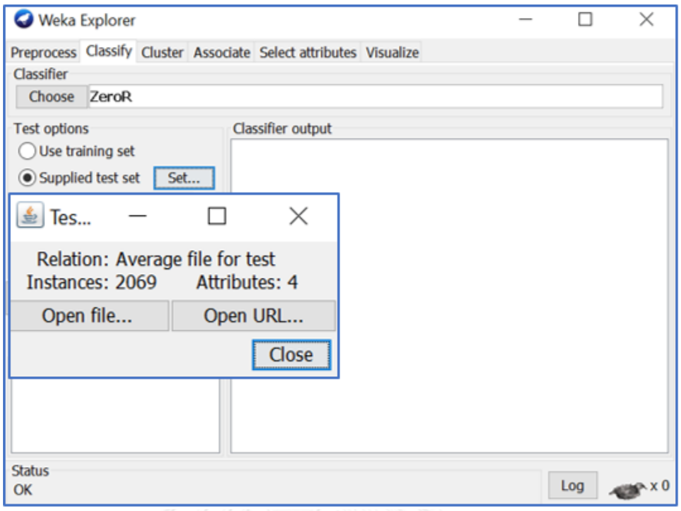
Look in: M5Rules_average

M5Rules -M 4.0.model
M5Rules -N -M 4.0.model
M5Rules -N -R -M 4.0.model
M5Rules -N -U -M 4.0.model
M5Rules -N -U -R -M 4.0.model
M5Rules -R -M 4.0.model
M5Rules -U -M 4.0.model
M5Rules -U -R -M 4.0.model

File name: M5Rules -N -M 4.0.model
Files of type: Model object files

Figure 3.27 *Opening the built model*

- At the “Test options”, mark on “Supplied test set” and click “Set…”. After that, the “Test Instances” will show, click “Open file…”, open the predictive file, if the file correct prepared, the software will count the instances and attributes as shown in **Figure 3.28**. Finally, click “Close”.



ZeroR

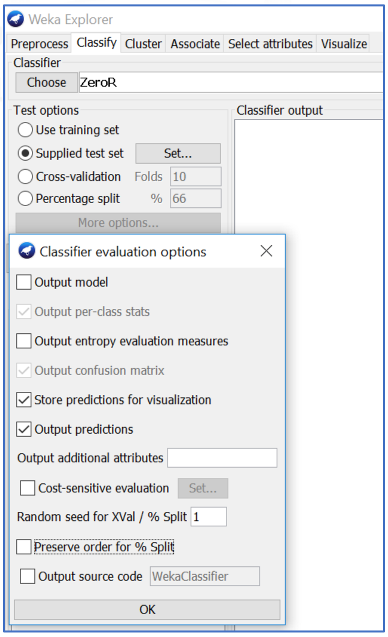
○ Use training set
● Supplied test set

Relation: Average file for test
Instances: 2069 Attributes: 4

OK

Figure 3.28 *Opening the predictive file*

- At the “Test options”, click on “More options…”, the “Classifier evaluation options” will show, mark on “Output predictions” and click “OK” as shown in **Figure 3.29**.



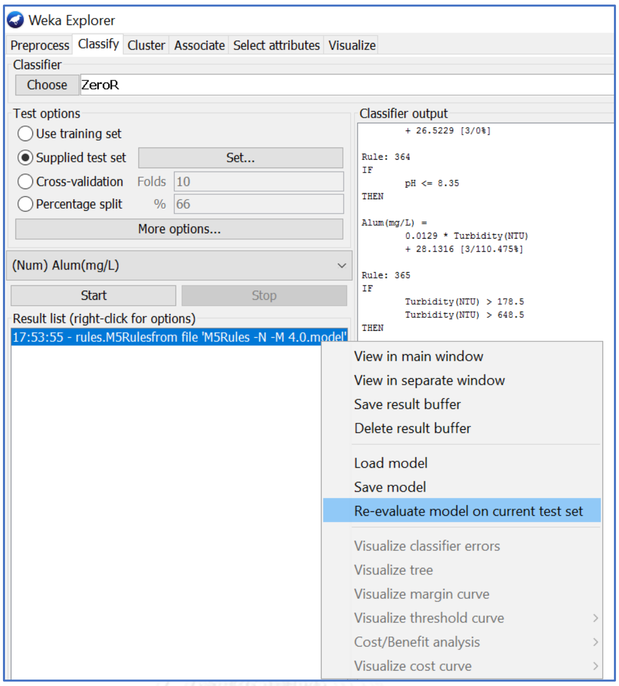
ZeroR

○ Use training set
● Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

☐ Output model
☑ Output per-class stats
☐ Output entropy evaluation measures
☑ Output confusion matrix
☑ Store predictions for visualization
☑ Output predictions
Output additional attributes
☐ Cost-sensitive evaluation
Random seed for XVal / % Split 1
☐ Preserve order for % Split
☐ Output source code WekaClassifier

Figure 3.29 *The prediction procedure*

- Right-click on the model that already loaded, the options will show and choose "Re-evaluate model and current test set as indicated in **Figure 3.30**.



Classifier: ZeroR

○ Use training set
● Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

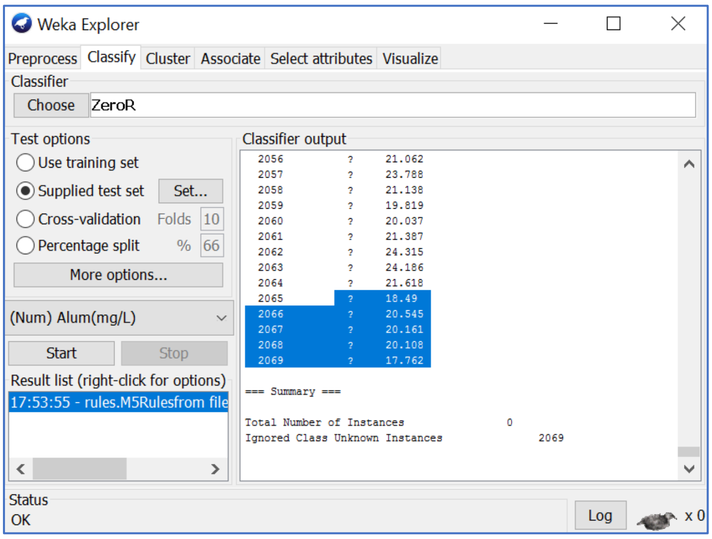
(Num) Alum(mg/L)

17:53:55 - rules.M5Rulesfrom file 'M5Rules -N -M 4.0.model'

```
        + 26.5229 [3/0%]

Rule: 364
IF
        pH <= 8.35
THEN

Alum(mg/L) =
        0.0129 * Turbidity(NTU)
        + 28.1316 [3/110.475%]

Rule: 365
IF
        Turbidity(NTU) > 178.5
        Turbidity(NTU) > 648.5
THEN
```

***Figure 3.30** The Re-evaluate model by supplied test set*

- When the model already predicted, the results will show in "Classifier output" as shown in **Figure 3.31**. After that, we save the results in the Notepad for the calculation the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).



Classifier: ZeroR

○ Use training set
● Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Num) Alum(mg/L)

17:53:55 - rules.M5Rulesfrom file

```
  2056          ?     21.062
  2057          ?     23.788
  2058          ?     21.138
  2059          ?     19.819
  2060          ?     20.037
  2061          ?     21.387
  2062          ?     24.315
  2063          ?     24.186
  2064          ?     21.618
  2065          ?     18.49
  2066          ?     20.545
  2067          ?     20.161
  2068          ?     20.108
  2069          ?     17.762

=== Summary ===

Total Number of Instances               0
Ignored Class Unknown Instances              2069
```

Status
OK

***Figure 3.31** The predictive alum dosage results*

3.9 Testing the model precision

The model precision is evaluated by using the statistical methods as the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

3.9.1 Root Mean Square Error (RMSE)

RMSE is a measurement of the difference between the actual values and the predictive values from the model, if RMSE has the less value, the result indicated that the predictive values are nearly the actual values. Therefore, the best value of the RMSE is zero.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a-b)^2} \qquad (3.1)$$

Where: $a =$ Amount of actual chemical value (mg/L) in the process.

$b =$ Amount of predictive chemical value (mg/L) from the model.

$n =$ Number of record.

3.9.2 Mean Absolute Error (MAE)

MAE has the characteristic look like RMSE, if the MAE has less value, the model has the highest accuracy. The best MAE value is zero too.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|a-b| \qquad (3.2)$$

Where: $a =$ Amount of actual chemical value (mg/L) in the process.

$b =$ Amount of predictive chemical value (mg/L) from the model.

$n =$ Number of record.

CHAPTER IV
RESULTS AND DISCUSSION

4.1 Water quality analysis and display by 5DWM

4.1.1 The experiments

4.1.1.1 The Horiba sensor U50

The conductivity, dissolved oxygen (DO), oxidation reduction potential (ORP), turbidity, pH, temperature, total dissolved solids (TDS), and salinity are measured by Horiba sensor U50, the results are shown in **Table 4.1-4.4**. The results indicated that all parameters are good quality. However, the turbidity value of Mekong (1st sampling point) is higher than Nam Ngum river (2nd, 3rd, and 4th sampling points) and Nam Lik river (5th and 6th sampling points).

Table 4.1 *The Horiba sensor U50 results in depth of 5 m*

	Parameters							
Spot	**Temperature (C)**	**pH**	**ORP (mV)**	**Conductivity (µS/cm)**	**Turbidity (NTU)**	**DO (mg/L)**	**TDS (g/L)**	**Salinity**
1st	28	7.3	307	0.21	483	16.8	0.14	0.1
2nd	28	6.5	335	0.12	38	6.6	0.08	0.1
3rd	27	6.6	317	0.14	21.3	7.7	0.09	0.1
4th	26	6.5	281	0.14	6.34	7.8	0.09	0.1
5th	27	6.7	318	0.14	34.1	7.9	0.09	0.1
6th	27	7.1	298	0.14	24.1	7.12	0.09	0.1

Table 4.2 *The Horiba sensor U50 results in depth of 3 m*

	Parameters							
Spot	**Temperature (C)**	**pH**	**ORP (mV)**	**Conductivity (µS/cm)**	**Turbidity (NTU)**	**DO (mg/L)**	**TDS (g/L)**	**Salinity (ppt)**
1st	28	7.3	313	0.21	487	7	0.14	0.1
2nd	28	6.6	336	0.12	37	5.4	0.08	0.1
3rd	27	6.7	326	0.14	20	5.7	0.09	0.1
4th	26	6.6	240	0.14	6.2	3.6	0.09	0.1
5th	27	6.7	322	0.14	31.8	6.8	0.09	0.1
6th	27	7.2	301	0.15	43.1	7	0.10	0.1

Table 4.3 *The Horiba sensor U50 results in depth of 1 m*

	Parameters							
Spot	Temperature (C)	pH	ORP (mV)	Conductivity (µS/cm)	Turbidity (NTU)	DO (mg/L)	TDS (g/L)	Salinity (ppt)
1st	28	7.4	314	0.22	404	6.5	0.14	0.1
2nd	28	6.7	332	0.12	33.9	4.9	0.08	0.1
3rd	27	6.7	327	0.14	19.2	5	0.09	0.1
4th	26	6.6	233	0.14	11	2.8	0.09	0.1
5th	27	7	321	0.15	35	6.3	0.1	0.1
6th	27	7.8	204	0.15	45.3	6.7	0.1	0.1

***Table 4.4** The Horiba sensor U50 results in depth of 0.5 m*

	Parameters							
Spot	Temperature (C)	pH	ORP (mV)	Conductivity (µS/cm)	Turbidity (NTU)	DO (mg/L)	TDS (g/L)	Salinity (ppt)
1st	28	7.5	313	0.21	393	6.3	0.14	0.1
2nd	28	6.8	326	0.12	34.8	4.7	0.08	0.1
3rd	27	6.8	321	0.14	18.8	4.6	0.09	0.1
4th	26	6.7	227	0.14	6.6	2.4	0.09	0.1
5th	27	7.6	287	0.15	36.4	6	0.1	0.1
6th	27	7.8	202	0.15	47.4	6.3	0.1	0.1

Because the water sampling collection was in the rainy season, the turbidity was also high. When we compared the results of 4 depths i.e. 5 m, 3 m, 1 m, and 0.5 m, the parameters of each depth are nearly the same value. Therefore, the average results of each parameter are indicated in **Table 4.5.** As we knew, the Mekong River is the big river and it has many sub basin, thus, the turbidity is quite high.

***Table 4.5** The average results of Horiba sensor U50*

	Parameters							
Spot	Temperature (C)	pH	ORP (mV)	Conductivity (µS/cm)	Turbidity (NTU)	DO (mg/L)	TDS (g/L)	Salinity (ppt)
1st	28	7.4	312	0.21	442	9.2	0.14	0.1
2nd	28	6.7	332	0.12	35.9	5.4	0.08	0.1
3rd	27	6.7	323	0.14	19.8	5.8	0.09	0.1
4th	26	6.6	245	0.14	7.5	4.2	0.09	0.1
5th	27	7	312	0.15	34.3	6.8	0.1	0.1
6th	27	7.5	251	0.15	40	6.8	0.1	0.1

4.1.1.2 The Fluorimeter AND1100 and Horiba's compact ion meter (LAQUAtwin)

This LAQUAtwin equipment measured the ions as Ca^{2+}, NO_3^-, Na^+, and K^+. The Fluorimeter AND1100 measured the heavy metals as Zinc (Zn), Cadmium (Cd), Lead (Pb), Uranium (U), Mercury (Hg), and Copper (Cu). The results are shown in **Table 4.6, Table 4.7**. The Lao's surface water quality standards are shown in **Table 4.8**.

Table 4.6 *The LAQUAtwin results*

	Ions			
Spot	**Calcium (Ca^{2+})**	**Nitrate (NO_3^-)**	**Sodium (Na^+)**	**Potassium (K^+)**
1st	50	**120**	26	**150**
2nd	63	**110**	190	9
3rd	64	**10**	19	1
4th	12	**92**	2	3
5th	370	**240**	8	6
6th	34	**13**	85	**46**

Table 4.7 *The Fluorimeter AND1100 results*

	Heavy Metals					
Spot	**Zinc (Zn)***	**Cadmium (Cd)**	**Lead (Pb)**	**Uranium (U)**	**Mercury (Hg)***	**Copper (Cu)**
1st	**0.006**	0.1	2	2	**<0.2**	40
2nd	**<0.005**	0.1	2	2	**<0.2**	40
3rd	**0.006**	0.1	2	2	**<0.2**	65
4th	**0.006**	0.1	2	2	**<0.2**	42
5th	**0.006**	0.1	2	2	**<0.2**	44
6th	**0.006**	0.1	2	2	**<0.2**	40

***:** Zn and Hg were rechecked by Atomic Absorption Spectrophotometer

Table 4.8 *Laos's surface water quality standard*

	Nitrate (NO_3^-)	**Sodium (Na^+)**	**Potassium (K^+)**	**Zinc (Zn)**	**Cadmium (Cd)**	**Lead (Pb)**	**Mercury (Hg)**	**Copper (Cu)**
Standard	<5	200	10	1	5	50	2	100

Source: Water Resources and Environment Administration, "Agreement on the National Environmental Standards", Vientiane Province, 7 December 2009

The analysis results in **Table 4.6** indicated that all of sampling points have nitrate (NO_3^-) values over the standard limit. Nitrate is the most common contaminant of surface water used for human consumption. Once in the stomach, nitrate (NO_3^-) is converted to nitrite (NO_2^-), and this form reduces the capacity of blood to carry oxygen to cells. This can result in "blue baby" disease in infants (Schröder, J. J. et al., 2004). The Potassium (K^+) values at the 1st and 6th sampling points are over the standard too. For Ca^{2+} and Na^{+} values are within the standard limit. Besides that, the all heavy metals are within the standard limit.

Because we met the nitrate and potassium are over the surface water quality standard that in the upstream area of those spots is had a lot of the fertilizer using for the agricultural.

4.1.2 5DWM system

The physical and chemical parameters of water quality on September 2016 was analyzed and visualized on 5DWM, and the results are shown in **Figure 4.1-4.3.** The overview of 3 rivers in 5DWM as shown in **Figure 4.4.** The parameters are conductivity, dissolved oxygen (DO), oxidation reduction potential (ORP), Ph, temperature, turbidity, total dissolved solids (TDS), and salinity. They are displayed by 5DWM in green color.

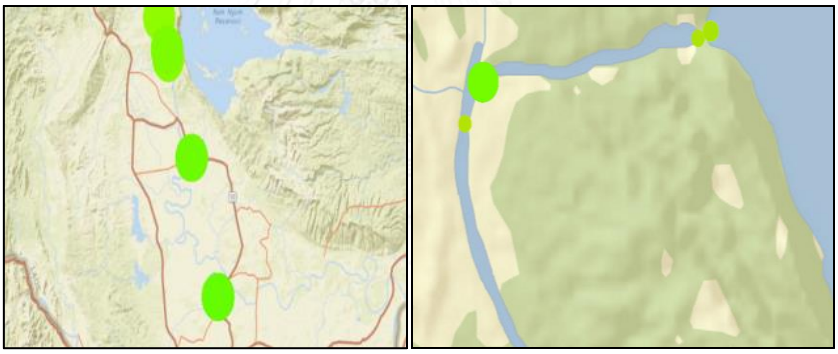


Figure 4.1 *Nam Ngum River*

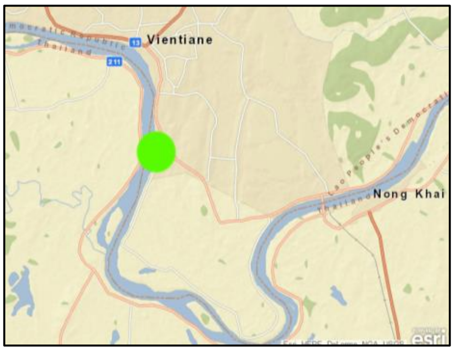


Figure 4.2 *Mekong River*

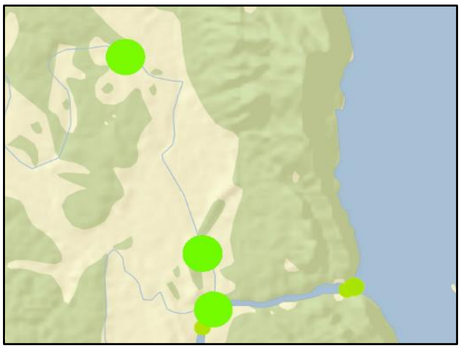


Figure 4.3 *Nam Lik River*

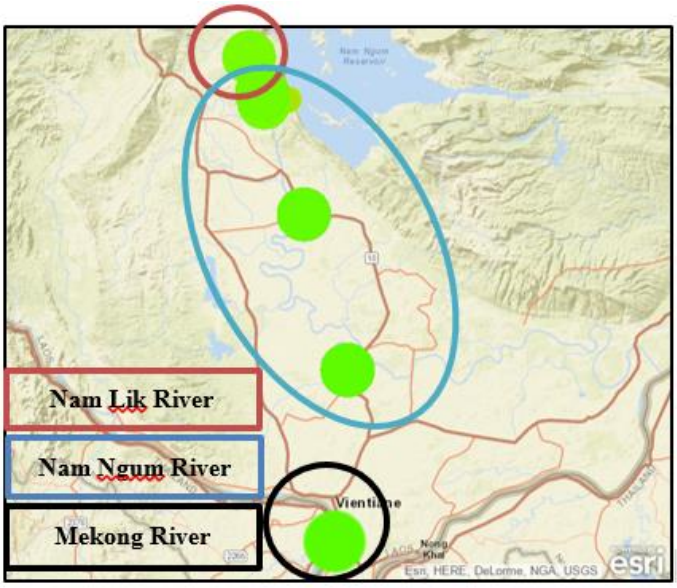


Figure 4.4 *Overview of 3 rivers in 5DWM*

For the color of spot is automatically chose by 5DWM. The color doesn't have meaning but we can give its definition by the comparing with the surface water quality standard.

4.2 The model for alum dosage prediction using Weka data mining software

4.2.1 Chinaimo Water Treatment Plant (CWTP)

The models for alum dosage prediction by using Weka data mining software are built from 2 data groups with 4 methods i.e. Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree. The *first data group*, we substituted all missing values of each parameter by the average value of that parameter, computed by each month and this group has 2,069 records. The *second data group*, we cut off the missing value to reduce bias and this group has 2,022 records.

4.2.1.1 The model building and adjustment from the first data group of the CWTP

1) Multilayer Perceptron (MLP) method

The model building for alum dosage prediction by using MLP method from 1st data group of CWTP, we found 3 methods that they gave the less RMSE value as shown in the gray rows in **Table 4.9**, they are:

- The 1st method gave the *RMSE of 2.9781* (No.45 in Table 4.9) that it is completely built from the *12-hidden layer of, 0.3-learning rate, 9-seed, 0.2-momentum, 5000-training time, and 20-validation threshold*. This method symbol is named of the "*MLP-L0.3-M0.2-N5000-V0-S9-E20-H12*";
- The 2nd method gave the *RMSE of 2.9790* (No.35 in Table 4.9) that it is successfully set from the *8-hidden layer, 0.3-learning rate, 6-seed, 3000-training time, 0.2-momentum, and 20-validation threshold*. This method symbol is entitled of the "*MLP-L0.3-M0.2-N3000-V0-S6-E20-H8*"; and
- the 3rd method gave the *RMSE of 3.0280* (No.31 in Table 4.9) and this method already built from the *4-hidden layer, 0.3-learning rate, 5000-training time, 3-seed, 0.2-momentum, and 20-validation threshold*. This method symbol is named of the "*MLP -L0.3 -M0.2 -N5000 -V0 -S3 -E20 -H4*".

For the details of the models building and adjustment are indicated in **Table 4.9**.

Table 4.9 *The adjustment of each algorithms of Multilayer Perceptron (MLP) method in the model building from the 1st data group of the CWTP*

Multilayer Perceptron (MLP) algorithms									
No	**Hidden layers**	**Learning rate**	**Momentum**	**Seed**	**Training time**	**Validation Threshold**	**Correlation Coefficient**	**MAE**	**RMSE**
1	a	0.3	0.2	0	500	20	0.8372	2.9059	3.9215
2	a	0.3	0.2	0	1000	20	0.8515	2.8758	3.7833
3	a	0.3	0.2	0	2000	20	0.8524	2.8749	3.7705
4	a	0.3	0.2	0	3000	20	0.8524	2.847	3.7666
5	a	0.3	0.2	0	4000	20	0.8527	2.8724	3.7619
6	a	0.3	0.2	0	5000	20	0.8530	2.8704	3.7578
7	a	0.3	0.2	0	10000	20	0.8535	2.8648	3.7477
8	i	0.3	0.2	0	500	20	0.8620	2.7324	3.5394
9	i	0.3	0.2	0	1000	20	0.8663	2.7027	3.4909
10	i	0.3	0.2	0	2000	20	0.8711	2.6736	3.4493
11	i	0.3	0.2	0	3000	20	0.8718	2.6704	3.4422
12	i	0.3	0.2	0	4000	20	0.8721	2.6689	3.4363
13	i	0.3	0.2	0	5000	20	0.8742	2.6512	3.3987
14	o	0.3	0.2	0	500	20	0.8127	2.9340	4.0271
15	o	0.3	0.2	0	1000	20	0.8127	2.9325	4.0266
16	o	0.3	0.2	0	2000	20	0.8127	2.9325	4.0266
17	o	0.3	0.2	0	3000	20	0.8127	2.9325	4.0266
18	o	0.3	0.2	0	4000	20	0.8127	2.9325	4.0266
19	o	0.3	0.2	0	5000	20	0.8127	2.9325	4.0266
20	t	0.3	0.2	0	500	20	0.8643	2.7253	3.5195
21	t	0.3	0.2	0	1000	20	0.8693	2.6846	3.4679
22	t	0.3	0.2	0	2000	20	0.8722	2.6491	3.4103
23	t	0.3	0.2	0	3000	20	0.8739	2.6298	3.3823
24	t	0.3	0.2	0	4000	20	0.8745	2.6211	3.3694
25	t	0.3	0.2	0	5000	20	0.8745	2.6151	3.3626
26	4	0.3	0.2	3	500	20	0.8887	2.4235	3.1130
27	4	0.3	0.2	3	1000	20	0.8890	2.4003	3.1103
28	4	0.3	0.2	3	2000	20	0.8930	2.3596	3.0656
29	4	0.3	0.2	3	3000	20	0.8948	2.3399	3.0396
30	4	0.3	0.2	3	4000	20	0.8953	2.3334	3.0296
31	**4**	**0.3**	**0.2**	**3**	**5000**	**20**	**0.8953**	**2.3304**	**3.0280**
32	8	0.3	0.2	6	500	20	0.8929	2.3186	3.0581
33	8	0.3	0.2	6	1000	20	0.8935	2.3107	3.0499
34	8	0.3	0.2	6	2000	20	0.8970	2.2905	3.0116
35	**8**	**0.3**	**0.2**	**6**	**3000**	**20**	**0.8993**	**2.2707**	**2.9790**
36	8	0.3	0.2	6	4000	20	0.8993	2.2706	2.9801
37	8	0.3	0.2	6	5000	20	0.8983	2.2778	2.9953
38	12	0.3	0.2	9	500	20	0.8911	2.288	3.1164
39	12	0.3	0.2	9	1000	20	0.8927	2.2622	3.0848
40	12	0.3	0.2	9	2000	20	0.8953	2.2449	3.0447
41	12	0.3	0.2	9	3000	20	0.8984	2.2250	3.0000
42	12	0.3	0.2	9	4000	20	0.8997	2.2160	2.9837
43	**12**	**0.3**	**0.2**	**9**	**5000**	**20**	**0.9001**	**2.2139**	**2.9781**

Gray row: Good method

Because we would like to know about the precision of those method, we divided the 1st data group out in term of the drying (1,058 records) and raining (1,011 records) season. Thus, we calculated the RMSE of those method for finding the method precision and credibility. In this case, which season and which method will give the less RMSE value. As we knew, if which method give the less RMSE, that method will also give the precision. The results are shown in **Table 4.10.**

Table 4.10 *The RMSE and MAE value of 3 methods of MLP in term of drying and raining season for the model building from the 1st data group of the CWTP*

Multilayer Perceptron (MLP) method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
MLP-L0.3-M0.2-N5000-V0-S9-E20-H12	1.767	1.359	5.877	2.651
MLP-L0.3-M0.2-N3000-V0-S6-E20-H8	**1.619**	**1.294**	**5.521**	**2.588**
MLP-L0.3-M0.2-N5000-V-S3-E20-H4	2.186	1.687	6.575	2.817

From **Table 4.10**, we found that the *MLP-L0.2-M0.2-N3000-V0-S6-E20-H8* method gave the less RMSE in both of drying and raining season. In this case, it gave the RMSE of 1.619 in the drying season is less than the RMSE of 5.877 in the raining season. Therefore, the *MLP-L0.3-M0.2-N3000-V0-S6-E20-H8* method will give high accuracy when we use it to predict the alum dose in the drying season in the real application.

When we compared the RMSE and MAE value of the drying and raining season together, we met the RMSE and MAE value of the drying season are less than the RMSE and MAE value of the rainy season. Thus, in this case, the methods will give more the preciseness when we use them to predict the alum dosage in term of the drying season.

2) M5Rules method

For the model building from the 1st data group for the alum dosage prediction by using this M5Rules method, we used "*buildRegressionTree", "unpruned", "minNumInstances (given value=4.0)",* and *"useUnsmoothed"* function. We adjusted in eight method by using M5Rules options and we got three methods gave the less RMSE value that they are indicated in the gray rows of **Table 4.11**, they are:

- The 1st method gave us the RMSE of 2.9025 (No.2 in Table 4.11) and this method is completely set from the *buildRegressionTree is "FALSE", debug is "FALSE", 4-miniNumInstances, unpruned is "TRUE", and useUnsmoothed is "FALSE"*. This method symbol is named of the "*M5Rules-N-M4.0*";
- The 2nd method has provided us the RMSE of 3.1660 (No.6 in Table 4.11) and this method is successfully set from the *buildRegressionTree is "TRUE", debug is*

"FALSE", 4-minNumInstancesunpruned is "FALSE", and useUnsmoothed is "FALSE". This method symbol is entitled of the "*M5Rules-R-M4.0*"; and

- The 3rd method gave us the RMSE of 3.3270 (No.3 of Table 4.11) and it is already set from the *buildRegressionTree is "TRUE", 4-minNumInstances, debug is "FALSE", unpruned is "TRUE", and useUnsmoothed is "FALSE"*. Its method symbol is named of the "*M5Rules-N-R-M4.0*".

The results of the M5Rules algorithms adjustment are indicated in **Table 4.11** and each method symbol are shown in **Table 4.12** too.

Table 4.11 *The adjustment of each algorithms of M5Rules for the model building from 1st data group of the CWTP*

The adjustment of M5Rules algorithms								
No	**Build Regression Tree**	**Debug**	**Min Num Instances**	**unpruned**	**Use Unsmoothed**	**Correlation Coefficient**	**MAE**	**RMSE**
1	FALSE	FALSE	4	FALSE	FALSE	0.5848	2.0288	7.7570
2	**FALSE**	**FALSE**	**4**	**TRUE**	**FALSE**	**0.9023**	**1.9017**	**2.9025**
3	**TRUE**	**FALSE**	**4**	**TRUE**	**FALSE**	**0.8732**	**2.1320**	**3.3270**
4	FALSE	FALSE	4	TRUE	TRUE	0.8633	1.8169	3.5169
5	TRUE	FALSE	4	TRUE	TRUE	0.8633	1.8169	3.5169
6	**TRUE**	**FALSE**	**4**	**FALSE**	**FALSE**	**0.8820**	**2.0241**	**3.1660**
7	FALSE	FALSE	4	FALSE	TRUE	0.5733	2.0389	7.9954
8	TRUE	FALSE	4	FALSE	TRUE	0.8875	1.9226	3.0904

Gray row: Good method

Table 4.12 *Each symbol of M5Rules method*

M5Rules algorithms						
No	**Build Regression Tree**	**Debug**	**Min Num Instances**	**unpruned**	**Use Unsmoothed**	**Methods symbol**
1	FALSE	FALSE	4	FALSE	FALSE	M5Rules-M4.0
2	FALSE	FALSE	4	TRUE	FALSE	M5Rules-N-M4.0
3	TRUE	FALSE	4	TRUE	FALSE	M5Rules-N-R-M4.0
4	FALSE	FALSE	4	TRUE	TRUE	M5Rules-N-U-M4.0
5	TRUE	FALSE	4	TRUE	TRUE	M5Rules-N-U-R-M4.0
6	TRUE	FALSE	4	FALSE	FALSE	M5Rules-R-M4.0
7	FALSE	FALSE	4	FALSE	TRUE	M5Rules-U-M4.0
8	TRUE	FALSE	4	FALSE	TRUE	M5Rules-U-R-M4.0

The results indicated that the *M5Rules-N-M4.0* method gave the less RMSE of 2.9025 than another method. Thus, it will give the highest precision and credibility than another method when we would like to use it to predict the alum dose. Besides that, the method gave the higher RMSE of 7.9954 is *M5Rules-U-M4.0,* so it will give the low precision than another method when we use it to predict the alum dose.

Because we would like to know the preciseness of those 3 methods, we divided the 1st data group out in term of the drying (1,058 records) and raining (1,011 records) season. In this case, which season and which method will give the highest precision and credibility when we would like to use them to predict the alum dose that the RMSE value will decide them the precision. The results are indicated in **Table 4.13**.

Table 4.13 *The RMSE and MAE value of 3 methods of M5Rules in the drying and raining season for the model building from the 1st data group of CWTP*

M5Rules method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
M5Rules-N-M4.0	**0.957**	**0.925**	**3.197**	**1.881**
M5Rules-R-M4.0	1.638	1.290	5.397	2.414
M5Rules-N-R-M4.0	1.404	1.113	6.500	2.244

From **Table 4.13**, the M5Rules-N-M4.0 method gave the less RMSE of 0.957 and MAE of 0.925 than another method in the drying season. On the other hand, the M5Rules-N-M4.0 method also gave the less RMSE of 3.197 and MAE of 1.881 than another method too. For this reason, the M5Rules-N-M4.0 method will have the highest precision when we would like to use it to predict the alum dose in the drying season because the RMSE of 0.957 is nearly zero and it will give the highest accuracy and credibility than another method when we would like to use it to predict the alum dose in the rainy season. Therefore, we referenced the precision from RMSE value. Thus, the M5Rules-N-M4.0 method will give the highest precision in the drying season than rainy season.

3) M5P method

The M5P method is similarly the M5Rules. Thus, the functions used in M5P method is quite the same M5Rules functions but it has a *saveInstances* function is different from M5Rules functions. Normally, the M5P has six functions i.e. *buildRegressionTree, minNumInstances, unpruned, debug, saveInstances, and useUnsmoothed.* Because the *debug* and *saveInstances* function isn't have the effective to the model building or didn't have any changeable things when we used them with another function.

Therefore, the model building and adjustment, we especially used the functions like M5Rules method i.e. *buildRegressionTree, minNumInstances, useUnsmoothed,* and *unpruned.*

We adjusted in eight methods by using M5P algorithms and we got three methods gave the less RMSE value, they are:

- The 1st method gave the RMSE of 2.5955 (No.2 in Table 4.14) and this method is completely set from the *buildRegressionTree is "FALSE", debug is "FALSE", minNumInstances of 4, saveInstances is "FALSE", unpruned is "TRUE", and useUnsmoothed is "FALSE"*. This method symbol is named "*M5P-N-M4.0"*;
- The 2nd method gave the RMSE of 2.6300 (No.1 in Table 4.14) and this method is successfully built from the *buildRegressionTree is "FALSE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, unpruned is "FALSE", and useUnsmoothed is "FALSE".* This method symbol is "*M5P-M4.0*"; and
- The 3rd method gave the RMSE of 2.6600 (No.7 in Table 4.14) and this method is built from the *buildRegressionTree is "FALSE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, unpruned is "FALSE", and useUnsmoothed is "TRUE"*. This method symbol is named of the "*M5P-U-M4.0*".

The results of the model building and adjustment are completely indicated in **Table 4.14** and their models symbol are also shown in **Table 4.15**.

Table 4.14 *The adjustment of each algorithms by using M5P method in the model building from 1st data group of the CWTP.*

The adjustment of M5P algorithms									
No	Build Regression Tree	Debug	Min Num Instances	Save instances	Unpruned	Use unsmoothed	Correlation Coefficient	MAE	RMSE
1	**FALSE**	**FALSE**	**4**	**FALSE**	**FALSE**	**FLASE**	**0.9199**	**1.8384**	**2.6300**
2	**FALSE**	**FALSE**	**4**	**FALSE**	**TRUE**	**FALSE**	**0.9221**	**1.7590**	**2.5955**
3	TRUE	FALSE	4	FALSE	TRUE	FALSE	0.8893	1.8435	3.1009
4	FALSE	FALSE	4	FALSE	TRUE	TRUE	0.8852	1.7535	3.1692
5	TRUE	FALSE	4	FALSE	TRUE	TRUE	0.8852	1.7535	3.1692
6	TRUE	FALSE	4	FALSE	FALSE	FALSE	0.8829	1.8970	3.1794
7	**FALSE**	**FALSE**	**4**	**FALSE**	**FALSE**	**TRUE**	**0.9180**	**1.8407**	**2.6600**
8	TRUE	FALSE	4	FALSE	FALSE	TRUE	0.8823	1.8648	3.1611

Gray row: Good method

Table 4.15 *Each M5P method symbol*

M5P algorithms							
No	**Build Regression Tree**	**Debug**	**Min Num Instances**	**Save instances**	**Unpruned**	**Use unsmoothed**	**Configuration symbol**
1	FALSE	FALSE	4	FALSE	FALSE	FLASE	M5P-M4.0
2	FALSE	FALSE	4	FALSE	TRUE	FALSE	M5P-N-M4.0
3	TRUE	FALSE	4	FALSE	TRUE	FALSE	M5P-N-R-M4.0
4	FALSE	FALSE	4	FALSE	TRUE	TRUE	M5P-N-U-M4.0
5	TRUE	FALSE	4	FALSE	TRUE	TRUE	M5P-N-U-R-M4.0
6	TRUE	FALSE	4	FALSE	FALSE	FALSE	M5P-R-M4.0
7	FALSE	FALSE	4	FALSE	FALSE	TRUE	M5P-U-M4.0
8	TRUE	FALSE	4	FALSE	FALSE	TRUE	M5P-R-M4.0

From **Table 4.14**, the *M5P-N-M4.0* method gave the less RMSE of 1.7950 and MAE of 2.5955 than another method. Thus, this method will give the highest accuracy than another method when we would like to use it to predict the alum dose. On the other hand, the method gave the highest RMSE of 1.8970 and MAE of 3.1794 than another method is *M5P-R-M4.0* that it built from *buildRegressionTree* and *4-minNumInstances.* When we would like to use it to predict the alum dosage in the coagulation process, it will give the less precision due to the M5P-R-M4.0 method has the highest RMSE value.

Therefore, the M5P-N-M4.0 method successfully built from the *unpruned* and *minNumInstances of 4* will give us the predictive alum dosage is nearly the actual alum dosage. Because the RMSE and MAE value of each method (No.1, No.2, and No.7 in Table 4.14) is very nearly. Therefore, we must calculate the RMSE and MAE value of those methods in the drying and raining season for the looking for the precision. The RMSE and MAE value results are indicated in **Table 4.16**.

Table 4.16 *The RMSE and MAE value of 3 methods of M5P in the drying and raining season for the model building from the 1st data group in the CWTP*

M5P method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
M5P -M4.0	1.409	1.157	4.822	2.338
M5P -N -M4.0	**1.078**	**1.005**	**3.855**	**2.078**
M5P -U -M4.0	1.397	1.132	4.698	2.275

From Table 4.16, we found that the M5P-N-M4.0 method gave the less RMSE and MAE value than M5P-M4.0 and M5P-U-M4.0 method in both dying and raining season. In this case, the M5P-N-M4.0 method gave the RMSE of 1.078 and MAE of 1.005 in the drying season. On the other hand, it completely gave the RMSE of 3.855 and MAE of 2.078 in the rainy season. Thus, the M5P-N-M4.0 method will give the highest accuracy in the drying season than rainy season. Therefore, the M5P-N-M4.0 method will give us the highest precision when we would like to use it in the real application.

4) REPTree method

The model adjustment and building by using this method. Normally, the REPTree has seven functions i.e. *debug*, *maxDepth*, *minNum*, *minVarianceProp*, *noPruning*, *numFolds,* and *seed*.

We adjusted eight methods by using REPTree algorithms and we got three method that they gave the less RMSE value, they are:

- The 1st method gave the RMSE of 3.0137 (No.4 in Table 4.17) and this method is completely set from the *2-minNum, -1-maxDepth, 5-numFolds, 2-seed, debug is "FALSE", 0.001-minVarianceProp, and noPruning is "FALSE"*. This method symbol is named of the "REPTree-M2-V0.001-N5-S2-L-1";
- The 2nd method gave the RMSE of 3.1738 (No.6 in Table 4.17) that it is successfully built from the *2-minNum, -1-maxDepth, 7-numFolds, 5-seed, debug is "FALSE", 0.001-minVarianceProp, and noPruning is "FALSE"*. This method symbol is entitled of the "REPTree-M2-V0.001-N7-S5-L-1"; and
- The 3rd method gave the RMSE of 3.1935 (No.8 in Table 4.17) and this method is built from the *-1-maxDepth, 2-minNum, 9-numFolds, 7-seed, debug is "FALSE", 0.001-minVarianceProp, and noPruning is "FALSE"*. The method symbol is named of the "REPTree-M2-V0.001-N9-S7-L-1".

Each REPTree algorithms adjustment results are indicated in **Table 4.17** and each method symbol shown in **Table 4.18** too.

Table 4.17 *The adjustment of each REPTree algorithm for the model building from the 1st data group in the CWTP*

The adjustment of REPTree algorithms										
No	Debug	Max Depth	Min Num	Min Variance Prop	No Pruning	Num Folds	Seed	Correlation Coefficient	MAE	RMSE
1	FALSE	-1	2.0	0.001	TRUE	3	1	0.8667	1.8477	3.3540
2	FALSE	-1	2.0	0.001	FALSE	3	1	0.8766	2.0033	3.2296
3	FALSE	-1	2.0	0.001	TRUE	5	2	0.8667	1.8477	3.3540
4	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**5**	**2**	**0.8934**	**1.8182**	**3.0137**
5	FALSE	-1	2.0	0.001	TRUE	7	5	0.8667	1.8477	3.3540
6	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**7**	**5**	**0.8809**	**1.8152**	**3.1738**
7	FALSE	-1	2.0	0.001	TRUE	9	7	0.8667	1.8477	3.3540
8	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**9**	**7**	**0.8793**	**1.8508**	**3.1935**
9	FALSE	-1	2.0	0.001	TRUE	20	18	0.8667	1.8477	3.3540
10	FALSE	-1	2.0	0.001	FALSE	20	18	0.8760	1.8586	3.2350
11	FALSE	-1	2.0	0.0001	TRUE	20	18	0.8666	1.8471	3.3541
12	FALSE	-1	2.0	0.0001	FALSE	20	18	0.8760	1.8586	3.2350

Gray row: Good method

From **Table 4.17**, we found that the "*REPTree-M2-V0.001-N5-S2-L-1*" method (No.4 in Table 4.17) gave the less RMSE of 3.0137 than another method. Thus, this method will give the highest accuracy and credibility than another method.

Table 4.18 *Each REPTree method symbol*

REPTree algorithms								
No	Debug	Max Depth	Min Num	Min Variance Prop	No Pruning	Num Folds	Seed	Model symbol
1	FALSE	-1	2.0	0.001	TRUE	3	1	REPTree-M2-V0.001-N3-S1-L-1-P
2	FALSE	-1	2.0	0.001	FALSE	3	1	REPTree-M2-V0.001-N3-S1-L-1
3	FALSE	-1	2.0	0.001	TRUE	5	2	REPTree-M2-V0.001-N5-S2-L-1-P
4	FALSE	-1	2.0	0.001	FALSE	5	2	REPTree-M2-V0.001-N5-S2-L-1
5	FALSE	-1	2.0	0.001	TRUE	7	5	REPTree-M2-V0.001-N7-S5-L-1-P
6	FALSE	-1	2.0	0.001	FALSE	7	5	REPTree-M2-V0.001-N7-S5-L-1
7	FALSE	-1	2.0	0.001	TRUE	9	7	REPTree-M2-V0.001-N9-S7-L-1-P
8	FALSE	-1	2.0	0.001	FALSE	9	7	REPTree-M2-V0.001-N9-S7-L-1
9	FALSE	-1	2.0	0.001	TRUE	20	18	REPTree-M2-V0.001-N20-S18-L-1-P
10	FALSE	-1	2.0	0.001	FALSE	20	18	REPTree-M2-V0.001-N20-S18-L-1
11	FALSE	-1	2.0	0.0001	TRUE	20	18	REPTree-M2-V1.0E-4-N20-S18-L-1-P
12	FALSE	-1	2.0	0.0001	FALSE	20	18	REPTree-M2-V1.0E-4-N20-S18-L-1

Because we would like to know the precision of those three method in the drying and raining season, we divided the data in term of drying and raining season. The results indicated in **Table 4.19**.

Table 4.19 *The RMSE and MAE value of 3 REPTree methods in the drying and raining season for the model building from the 1st data group of the CWTP*

REPTree method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
REPTree-M2-V0.001-N5-S2-L-1	**1.093**	**0.915**	**4.098**	**2.041**
REPTree-M2-V0.001-N7-S5-L-1	1.127	0.947	4.227	2.049
REPTree-M2-V0.001-N9-S7-L-1	1.164	1.004	7.854	2.133

From **Table 4.19**, we found that the RMSE of 1.093 (Drying season) and 4.098 (Rainy season) of the *REPTree-M2-V0.001-N5-S2-L-1* method are less than the RMSE of the REPTree-M2-V0.001-N7-S5-L-1 and REPTree-M2-V0.001-N9-S7-L-1 method. For this reason, the *REPTree-M2-V0.001-N5-S2-L-1* will give the highest precision and credibility than another method when we would like to use it to predict the alum dosage in the coagulation process of water treatment plant. However, we will know the method precision from the real applications.

4.2.1.2 The model building and adjustment from the second data group of the CWTP

The model building and adjustment from the second data group is also used the same 4 methods i.e. Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree.

1) Multilayer Perceptron (MLP) method

We adjust 43 methods by using the MLP algorithms and we got three method that they gave the less RMSE value, they are:

- The 1st method gave the RMSE of 3.3276 (No.43 in Table 4.20) that it is completely built from the *12-hiddenLayers, 0.3-learningRate, 0.2-momentum, 9-seed, 5000-trianingTime,* and *20-validationThreshold* function. This method symbol is named of the "*MLP-L0.3-M0.2-N5000-V0-S9-E20-H12*";
- The 2nd method gave the RMSE of 3.4329 (No.29 in Table 4.20) that it is successfully built from the *4-hiddenLayers, 0.3-learningRate, 0.2-momentum, 3-seed, 3000-trianingTime,* and *20-validationThreshold* function. This method symbol is named of the "*MLP-L0.3-M0.2-N3000-V0-S3-E20-H4*"; and
- The 3rd method gave the RMSE of 3.4453 (No.7 in Table 4.20) and this method is already built from the *a-hiddenLayers, 0.3-learningRate, 0.2-momentum, 0-seed, 10000-trianingTime,* and *20-validationThreshold* function. Its configuration symbol is entitled of the "*MLP-L0.3-M0.2-N3000-V0-S9-E20-H12*".

The adjustment of MLP algorithms results are indicated in **Table 4.20**:

Table 4.20 *The adjustment of each MLP algorithms for the model building from the 2nd data group of the CWTP*

The adjustment of Multilayer Perceptron (MLP) algorithms

No	Hidden layers	Learning rate	Momentum	Seed	Training time	Validation Threshold	Correlation Coefficient	MAE	RMAE
1	a	0.3	0.2	0	500	20	0.8374	2.5416	3.8464
2	a	0.3	0.2	0	1000	20	0.8596	2.3114	3.5825
3	a	0.3	0.2	0	2000	20	0.8696	2.2051	3.4679
4	a	0.3	0.2	0	3000	20	0.8704	2.1944	3.4585
5	a	0.3	0.2	0	4000	20	0.8709	2.1892	3.4537
6	a	0.3	0.2	0	5000	20	0.8712	2.1859	3.4507
7	**a**	**0.3**	**0.2**	**0**	**10000**	**20**	**0.8717**	**2.1791**	**3.4453**
8	i	0.3	0.2	0	500	20	0.8187	2.4480	4.0223
9	i	0.3	0.2	0	1000	20	0.8172	2.4649	4.0425
10	i	0.3	0.2	0	2000	20	0.8352	2.3271	3.8529
11	i	0.3	0.2	0	3000	20	0.8383	2.3131	3.8191
12	i	0.3	0.2	0	4000	20	0.8400	2.3076	3.8018
13	i	0.3	0.2	0	5000	20	0.8412	2.3042	3.7898
14	o	0.3	0.2	0	500	20	0.8029	2.5608	4.1919
15	o	0.3	0.2	0	1000	20	0.8021	2.5657	4.2002
16	o	0.3	0.2	0	2000	20	0.8020	2.5661	4.2009
17	o	0.3	0.2	0	3000	20	0.8020	2.5661	4.2009
18	o	0.3	0.2	0	4000	20	0.8020	2.5661	4.2009
19	o	0.3	0.2	0	5000	20	0.8020	2.5661	4.2009
20	t	0.3	0.2	0	500	20	0.8425	2.4520	3.7804
21	t	0.3	0.2	0	1000	20	0.8585	2.2891	3.5963
22	t	0.3	0.2	0	2000	20	0.8673	2.2044	3.4925
23	t	0.3	0.2	0	3000	20	0.8684	2.1979	3.4787
24	t	0.3	0.2	0	4000	20	0.8687	2.2016	3.4739
25	t	0.3	0.2	0	5000	20	0.8687	2.2067	3.4733
26	4	0.3	0.2	0	500	20	0.8610	2.3923	3.5475
27	4	0.3	0.2	3	1000	20	0.8662	2.3337	3.4872
28	4	0.3	0.2	3	2000	20	0.8702	2.3236	3.4388
29	**4**	**0.3**	**0.2**	**3**	**3000**	**20**	**0.8706**	**2.3259**	**3.4329**
30	4	0.3	0.2	3	4000	20	0.8702	2.3330	3.4371
31	4	0.3	0.2	3	5000	20	0.8698	2.3377	3.4420
32	8	0.3	0.2	6	500	20	0.8071	3.4584	4.7034
33	8	0.3	0.2	6	1000	20	0.816	3.3542	4.5814
34	8	0.3	0.2	6	2000	20	0.8244	3.2643	4.4615
35	8	0.3	0.2	6	3000	20	0.8272	3.2288	4.4131
36	8	0.3	0.2	6	4000	20	0.8289	3.1975	4.3855
37	8	0.3	0.2	6	5000	20	0.8297	3.1808	4.3691
38	12	0.3	0.2	9	500	20	0.8757	2.1378	3.3992
39	12	0.3	0.2	9	1000	20	0.8772	2.1192	3.3798
40	12	0.3	0.2	9	2000	20	0.8794	2.1079	3.3493
41	12	0.3	0.2	9	3000	20	0.8802	2.1033	3.3349
42	12	0.3	0.2	9	4000	20	0.8804	2.1057	3.3287
43	**12**	**0.3**	**0.2**	**9**	**5000**	**20**	**0.8804**	**2.1108**	**3.3276**

Gray row: Good method

From Table 4.20, we found that the *MLP-L0.3-M0.2-N5000-V0-S9-E20-H12* method gave the less RMSE of 3.3276 than another method. Thus, it has the highest precision and credibility than another method when we would like to use it to predict the alum dose and the predictive alum dose will be nearly the actual alum. Besides that, we would like to know the precision of those three methods in the drying and raining season, we analyzed the precision by the RMSE value. The results are indicated in **Table 4.21**.

Table 4.21 *The RMSE and MAE value of 3 MLP methods in the drying and rainy season for the model building from the 2^{nd} data group of the CWTP.*

Multilayer Perceptron (MLP) method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
MLP-L0.3-M0.2-N5000-V0-S9-E20-H12	**1.622**	**1.291**	**8.464**	**2.760**
MLP-L0.3-M0.2-N10000-V0-S0-E20-Ha	1.634	1.302	8.711	2.781
MLP-L0.3-M0.2-N3000-V0-S3-E20-H4	2.038	1.560	8.893	2.823

From **Table 4.21**, we found that the *MLP-L0.3-M0.2-N5000-V0-S9-E20-H12* method gave us the less RMSE in both drying and raining season than another method. In the drying season, it gave the RMSE of 1.622 is less than the RMSE of 8.464 in the rainy season. Thus, it will provide more precision in the drying season than raining season. The model will be accuracy, it is responded the real application that it can predict the alum dosage in the coagulation process or not.

2) M5Rules method

For the model building and adjustment by using M5Rules method is also use "*buildRegressionTree", "unpruned", "minNumInstances (given value=4.0)",* and *"useUnsmoothed"* function. We adjusted eight method by using the M5Rules algorithms and we got three method that they gave the less RMSE value, they are:

- The 1st method gave the RMSE of 3.2239 (No.7 in Table 2.22) that it successfully made from the *buildRegressionTree is "FALSE", debug is "FALSE", unpruned is "FALSE", 4-minNumInstances, and useUnsmoothed is "TRUE"*. This method symbol is entitled of the "*M5Rules-U-M4.0*";
- The 2nd method gave the RMSE of 3.2733 (No.1 in Table 4.22) and this configuration completely built from the *buildRegressionTree is "FALSE", debug is "FALSE", 4-minNumInstances, unpruned is "FALSE", and useUnsmoothed is "FALSE"*. This method symbol is named of the "*M5Rules-M4.0*"; and
- The 3rd method gave the RMSE of 3.5807 (No.8 in Table 4.22) that it completely built from the *buildRegressionTree is "TRUE", debug is "FALSE", unpruned is "FALSE", 4-minNumInstances, and useUnsmoothed is "TRUE"*. This method symbol is named of the "*M5Rules-U-R-M4.30*".

The M5Rules algorithms adjustment results are indicated in **Table 4.22** and each method symbol is indicated in **Table 4.12**:

Table 4.22 *The adjustment of each M5Rules algorithms for the model building from the 2nd data group of the CWTP*

M5Rules algorithms adjustment								
No	Build Regression Tree	Debug	Min Num Instances	unpruned	Use Unsmoothed	Correlation Coefficient	MAE	RMSE
1	**FALSE**	**FALSE**	**4**	**FALSE**	**FALSE**	**0.8837**	**1.9554**	**3.2733**
2	FALSE	FALSE	4	TRUE	FALSE	0.5882	2.1537	7.3445
3	TRUE	FALSE	4	TRUE	FALSE	0.8403	2.2353	3.8497
4	FALSE	FALSE	4	TRUE	TRUE	0.8328	1.8388	3.9418
5	TRUE	FALSE	4	TRUE	TRUE	0.8328	1.8388	3.9418
6	TRUE	FALSE	4	FALSE	FALSE	0.8524	2.1362	3.6529
7	**FALSE**	**FALSE**	**4**	**FALSE**	**TRUE**	**0.8872**	**1.9324**	**3.2239**
8	**TRUE**	**FALSE**	**4**	**FALSE**	**TRUE**	**0.8580**	**2.0393**	**3.5807**

From **Table 4.22**, we found that the *M5Rules-U-M4.0* method has the less RMSE of 3.2239 than another method. For this reason, the M5Rules-U-M4.0 method will give the highest precision than another method when we would like to use it to predict the alum dosage.

Because we would like to know the accuracy of those three methods in the drying and raining season, we divided the 2nd data group out in term of the drying and raining season. The results are indicated in **Table 4.23** as below:

Table 4.23 *The RMSE and MAE value of 3 M5Rules methods in the drying and raining season for the model building from the 2nd data group of the CWTP.*

M5Rules method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
M5Rules-U-M4.0	**1.466**	**1.162**	**5.758**	**2.490**
M5Rules-M4.0	1.467	1.183	6.097	2.597
M5Rules-U-R-M4.0	1.430	1.151	8.342	2.764

From the **Table 4.23**, we found that the RMSE of 1.466 (Drying season) and 5.758 (Raining season) of the *M5Rules-U-M4.0* method is less than another method in both drying and raining season. Therefore, the *M5Rules-U-M4.0 method* has more the accuracy than another method. However, the M5Rules method built from two data group and the M5rules method of the 1st data group is "*M5Rules-N-M4.0*" that it will be more accuracy than another method and the M5Rules method of 2nd data group is "*M5Rules-U-M4.0*" that it will be more accuracy than another method. For this reason, we will summarize which M5Rules method will give the highest precision by the real applications.

3) M5P method

Normally, M5P has six functions i.e. *buildRegressionTree, minNumInstances, unpruned, debug, saveInstances, and useUnsmoothed.* Because the *debug* and *saveInstances* function is not the effective to the model building or didn't have any changeable things when we used them with another function.

Therefore, the model building and adjustment, we especially used the functions like M5Rules method i.e. *buildRegressionTree, minNumInstances, useUnsmoothed,* and *unpruned.*

We adjusted eight methods by using the M5P algorithms and we got three method that they gave the less RMSE, they are:

- The RMSE of the 1st method is 3.1813 (No.1 in Table 4.24) and method is completely built form the *buildRegressionTree is "FALSE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, unpruned is "FALSE", and useUnsmoothed is "FAlSE.* This method symbol is named of the "*M5P-M4.0";*
- The RMSE of the 2nd method is 3.6047 (No.3 in Table 4.24) that it is successfully built from the *buildRegressionTree is "TRUE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, unpruned is "TRUE", and useUnsmoothed is "FALSE"*. This method symbol is "*M5P-N-R-M4.0*"; and
- The RMSE of the 3rd method is 3.2435 (No.7 in Table 4.24) and this method is built from the *buildRegreesionTree is "FALSE", debug is "FALSE", saveInstances is "FALSE", unpruned is "FALSE", 4-minNumInstances, and useUnsmoothed is "TRUE"*. This method symbol is entitled of the "*M5P-U-M4.0*".

The M5P algorithms adjustment results indicated in **Table 4.24** and the method symbol is shown in **Table 4.15**.

Table 4.24 *The adjustment of each M5P algorithms for the model building from the 2nd data group of the CWTP*

M5P algorithms adjustment									
No	Build Regression Tree	Debug	Min Num Instances	Save instances	Unpruned	Use unsmoothed	Correlation Coefficient	MAE	RMSE
1	**FALSE**	**FALSE**	**4**	**FALSE**	**FALSE**	**FLASE**	**0.8898**	**1.9414**	**3.1813**
2	FALSE	FALSE	4	FALSE	TRUE	FALSE	0.6402	2.0114	6.6987
3	**TRUE**	**FALSE**	**4**	**FALSE**	**TRUE**	**FALSE**	**0.8583**	**1.9779**	**3.6047**
4	FALSE	FALSE	4	FALSE	TRUE	TRUE	0.8448	1.8865	3.8138
5	TRUE	FALSE	4	FALSE	TRUE	TRUE	0.8848	1.8865	3.8138
6	TRUE	FALSE	4	FALSE	FALSE	FALSE	0.8498	2.0714	3.6948
7	**FALSE**	**FALSE**	**4**	**FALSE**	**FALSE**	**TRUE**	**0.8854**	**1.9518**	**3.2435**
8	TRUE	FALSE	4	FALSE	FALSE	TRUE	0.8486	2.0315	3.6985

Gray row: Good method

From **Table 4.24**, we found that the RMSE of 1.9414 of the *M5P-M4.0* method (No.1) is less than the RMSE of another method. For this reason, the M5P-M4.0 method is more the precision than another method. On the other hand, if we would like to predict the alum dose by using this method, it will give the predictive alum dosage values are nearly the actual alum dosage.

Because we would like to know the accuracy of those 3 methods i.e. M5P-M4.0, M5P-N-R-M4.0, and M5P-U-M4.0 in the drying and raining season, we divided the 2nd data group out in term of the drying and raining season. The results indicated in **Table 4.25**.

Table 4.25 *The RMSE and MAE value of 3 M5P methods in the drying and rainy season for the model creating of the CWTP*

M5P method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
M5P-M4.0	1.437	1.174	9.964	2.655
M5P-N-R-M4.0	**1.234**	**1.110**	**8.002**	**2.519**
M5P-U-M4.0	1.425	1.150	8.013	2.651

Form **Table 4.25**, the RMSE values of three methods are very nearly in the drying season and the RMSE values of M5P-M4.0 and M5P-U-M4.0 are also very nearly in the raining season. The RMSE of 1.234 (Drying season) and 8.002 (Raining season) of the M5P-N-R-M4.0 method is less than another method. For this reason, the M5P-N-R-M4.0 method has the highest precision than another method. However, we will summarize the precision of those three methods by the real applications.

4) REPTree method

Normally, the REPTree has seven functions i.e. *debug*, *maxDepth*, *minNum*, *minVarianceProp*, *noPruning*, *numFolds,* and *seed*.

We adjusted 12 methods by using REPTree algorithms and we got three methods that they gave the less RMSE value, they are:

- The RMSE of the 1st method is 3.6647 (No.2 in Table 4.26) and this method is successfully built from the *debug is "FALSE", -1-maxDepth, noPruning is "FAlSE", 2-minNum, 0.001minVarianceProp, 3-numFolds,* and *1-seed*. This method symbol is named of the "*REPTree-M2-V0.001-N3-S1-L-1*";
- The RMSE of the 2nd method is 3.6938 (No.6 in Table 4.26) and this method is completely built from the *debug is "FALSE", -1-maxDepth, 2-minNum, 7-numFolds, 0.001-minVarianceProp, noPruning is "FALSE", and 5-seed*. This method symbol is entitled of the "*REPTree-M2-V0.001-N7-S5-L-1*"; and
- The RMSE of the 3rd method is 3.7262 (No.8 in Table 4.26) and this method is successfully built from the *debug is "FALSE", -1-maxDepth, noPruning is "FALSE", 2-minNum, 0.001-minVarianceProp, 9-numFolds, and 7-seed*. This method symbol is named of the "*REPTree-M2-V0.001-N9-S7-L-1*".

The REPTree algorithms adjustment results are indicated in **Table 4.26** and the method symbols shown in **Table 4.18**.

Table 4.26 *REPTree algorithms adjustment for the model building from the 2nd data group of the CWTP*

The adjustment of REPTree algorithms										
No	Debug	Max Depth	Min Num	Min Variance Prop	No Pruning	Num Folds	Seed	Correlation Coefficient	MAE	RMSE
1	FALSE	-1	2.0	0.001	TRUE	3	1	0.8210	1.9261	4.0265
2	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**3**	**1**	**0.8510**	**2.0298**	**3.6647**
3	FALSE	-1	2.0	0.001	TRUE	5	2	0.8210	1.9261	4.0265
4	FALSE	-1	2.0	0.001	FALSE	5	2	0.8236	2.0290	3.9728
5	FALSE	-1	2.0	0.001	TRUE	7	5	0.8210	1.9261	4.0265
6	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**7**	**5**	**0.8484**	**1.9625**	**3.6938**
7	FALSE	-1	2.0	0.001	TRUE	9	7	0.8210	1.9261	4.0265
8	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**9**	**7**	**0.8453**	**1.9802**	**3.7262**
9	FALSE	-1	2.0	0.001	TRUE	20	18	0.8210	1.9261	4.0265
10	FALSE	-1	2.0	0.001	FALSE	20	18	0.8244	2.0279	3.9689
11	FALSE	-1	2.0	0.0001	TRUE	20	18	0.8210	1.9256	4.0265
12	FALSE	-1	2.0	0.0001	FALSE	20	18	0.8244	2.0277	3.9688

Gray row: Good method

Because we would like to know the accuracy of those three REPTree methods in the drying and raining season, we divide the 2nd data group in term of the drying and raining season. The results are indicated in **Table 4.27** as below:

Table 4.27 *The RMSE and MAE value of three REPTree methods in the drying and raining season for the model building from 2nd data group of the CWTP*

REPTree method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
REPTree-M2-V0.001-N3-S1-L-1	1.307	1.020	8.437	2.563
REPTree-M2-V0.001-N7-S5-L-1	1.195	0.999	7.969	2.505
REPTree-M2-V0.001-N9-S7-L-1	**1.088**	**0.931**	**7.626**	**2.281**

From **Table 4.27**, we found that the RMSE of 1.088 (Drying season) and 7.626 (Raining season) of the *REPTree-M2-V0.001-N9-S7-L-1* method is less than another method in the drying and raining season. Thus, it has the precision and credibility than another method. When we would like to use it to predict the alum dosage, it will also give the highest precision. For this reason, it will give the predictive alum dosage values are nearly the actual alum dosage values. However, each method will be accuracy or not, they are upon the real applications.

4.2.2 Dongmarkkaiy Water Treatment Plant (DWTP)

The model for alum dosage prediction by using Weka data mining software of this plant is built from 4 methods i.e. Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree with 2 data groups. *First group*, we substituted all missing values of each parameter by the average value of that parameter, computed by each month, this group has 2,861 records. *Second group*, we cut off the missing value to reduce bias, this group has 2,284 records.

4.2.2.1 The model building for alum dosage prediction by using Weka data mining software from the 1st data group of the DWTP

1) Multilayer Perceptron (MLP)

For the model adjustment and building of this data group by using MLP method, we have got three configurations that they gave the less RMSE value, they are:

- The RMSE of the 1st method is 3.7110 (No.25 in Table 4.28) and this method completely built from the giving the *t-hiddenLayer, 0.3-learningRate, 0.2-momentum, 20-validationThreshold, 0-seed, and 5000-trainingTime.* This method symbol is named of the "*MLP-L0.3-M0.2-N5000-V0-S0-E20-Ht*";
- The RMSE of the 2nd method is 3.7305 (No.29 in Table 4.28) and this method successfully built from the giving the *4-hiddenLayer, 0.3-learningRate, 0.2-momentum, 20-validationThreshold, 3-seed, and 3000-trainingTime.* This method symbol is named of the "*MLP-L0.3-M0.2-N3000-V0-S3-E20-H4*"; and
- The RMSE of the 3rd method is 3.7347 (No.6 in Table 4.28), this method successfully built from the giving the *a-hiddenLayer, 0.3-learningRate, 0.2-momentum, 20-validationThreshold, 0-seed, and 5000-trainingTime.* This method symbol is named of the "*MLP-L0.3-M0.2-N5000-V0-S0-E20-Ha*".

For those three good methods and another MLP algorithms adjustment from the 1st data group is indicated in **Table 4.28**

Table 4.28 *The MLP algorithms adjustment for the model building from the 1st data group of the DWTP*

The adjustment of Multilayer Perceptron (MLP) algorithms									
No	Hidden layers	Learning rate	Momentum	Seed	Training time	Validation Threshold	Correlation Coefficient	MAE	RMAE
1	a	0.3	0.2	0	500	20	0.7963	3.0260	3.9261
2	a	0.3	0.2	0	1000	20	0.8027	2.9658	3.8670
3	a	0.3	0.2	0	2000	20	0.8067	2.9237	3.8278
4	a	0.3	0.2	0	3000	20	0.8121	2.8632	3.7706
5	a	0.3	0.2	0	4000	20	0.8143	2.8358	3.7473
6	**a**	**0.3**	**0.2**	**0**	**5000**	**20**	**0.8155**	**2.8202**	**3.7347**
7	a	0.3	0.2	0	10000	20	0.8179	2.7871	3.7395
8	i	0.3	0.2	0	500	20	0.8035	2.9320	3.8423
9	i	0.3	0.2	0	1000	20	0.8061	2.8853	3.8305
10	i	0.3	0.2	0	2000	20	0.8022	2.8455	3.8742
11	i	0.3	0.2	0	3000	20	0.8020	2.8225	3.8777
12	i	0.3	0.2	0	4000	20	0.8023	2.8075	3.8748
13	i	0.3	0.2	0	5000	20	0.8027	2.7975	3.8716
14	o	0.3	0.2	0	500	20	0.7804	3.0908	4.0413
15	o	0.3	0.2	0	1000	20	0.7818	3.0837	4.0320
16	o	0.3	0.2	0	2000	20	0.7820	3.0820	4.0309
17	o	0.3	0.2	0	3000	20	0.7820	0.0828	4.0309
18	o	0.3	0.2	0	4000	20	0.7820	3.0828	4.0310
19	o	0.3	0.2	0	5000	20	0.7820	3.0828	4.0310
20	t	0.3	0.2	0	500	20	0.8039	2.9240	3.8408
21	t	0.3	0.2	0	1000	20	0.8105	2.8858	3.7874
22	t	0.3	0.2	0	2000	20	0.8149	2.8335	3.7409
23	t	0.3	0.2	0	3000	20	0.8160	2.8092	3.7265
24	t	0.3	0.2	0	4000	20	0.8167	2.7944	3.7188
25	**t**	**0.3**	**0.2**	**0**	**5000**	**20**	**0.8175**	**2.7822**	**3.7110**
26	4	0.3	0.2	3	500	20	0.8115	2.8121	3.7998
27	4	0.3	0.2	3	1000	20	0.8174	2.7603	3.7565
28	4	0.3	0.2	3	2000	20	0.8206	2.7224	3.7387
29	**4**	**0.3**	**0.2**	**3**	**3000**	**20**	**0.8216**	**2.7001**	**3.7305**
30	4	0.3	0.2	3	4000	20	0.8208	2.6959	3.7342
31	4	0.3	0.2	3	5000	20	0.8201	2.6941	3.7390
32	8	0.3	0.2	6	500	20	0.7316	3.2826	4.5560
33	8	0.3	0.2	6	1000	20	0.7401	3.2217	4.4994
34	8	0.3	0.2	6	2000	20	0.7382	3.2114	4.5387
35	8	0.3	0.2	6	3000	20	0.7361	3.1994	4.5781
36	8	0.3	0.2	6	4000	20	0.7240	3.1931	4.5314
37	8	0.3	0.2	6	5000	20	0.7455	3.1827	4.5052
38	12	0.3	0.2	9	500	20	0.6442	3.1948	5.9211
39	12	0.3	0.2	9	1000	20	0.6075	3.1562	6.5296
40	12	0.3	0.2	9	2000	20	0.5506	3.0640	7.6022
41	12	0.3	0.2	9	3000	20	0.5009	3.0713	8.6126
42	12	0.3	0.2	9	4000	20	0.4672	3.0609	9.4084
43	12	0.3	0.2	9	5000	20	0.4427	3.0747	10.0541

Gray row: Good method

Because we would like to know the preciseness of those three methods in the drying and raining season. We divided the 1st data group out in term of drying and raining season. The drying season has 1,389 records and the raining season has 1,472 records. The results are indicated in **Table 4.29**.

Table 4.29 *The RMSE and MAE value of 3 MLP methods in the drying and raining season for the model building from the 1st data group of the DWTP*

Multilayer Perceptron (MLP) method				
Method	**Drying season**		**Rainy season**	
	RMSE	**MAE**	**RMSE**	**MAE**
MLP-L0.3-M0.2-N5000-V0-S0-E20-Ht	4.354	2.386	12.735	3.884
MLP-L0.3-M0.2-N3000-V0-S3-E20-H4	**2.380**	**1.731**	**8.195**	**3.039**
MLP-L0.3-M0.2-N5000-V0-S0-E20-Ha	4.183	2.328	8.337	3.069

For the model adjustment and building from the 1st group of data by using MLP method, the **Table 4.29** indicated that the MLP method by using the *4-hiddenLayer, 0.3-learningRate, 3000-trainingTime, 3-seed, 3000-trainingTime, and 0.2-momentum* or in the form of "*MLP-L0.3-M0.2-N3000-V0-S3-E20-H4*" is the best method because its RMSE of 2.380 (Drying season) and 8.195 (Raining season) is less than another method. Therefore, we will use this method for the testing in the real application that it will be accurate or not.

2) M5Rules method

For the model building and adjustment from the 1st data group by using M5Rules method, we adjust in 8 methods and we got three methods that they gave the less RMSE value, they are indicated in the gray row of **Table 4.30**, they are:

- The 1st method gave the RMSE of 3.2394 (No.1 in Table 4.30) and this method is set from the *buildRegressionTree is "FALSE", debug is "FALSE", unpruned is "FALSE", 4-minNumInstances, and useUnsmoothed is also "FALSE"*. This method symbol is named of the "*M5Rules-M4.0*";
- The 2nd method gave the RMSE of 3.2447 (No.7 in Table 4.30) and this method is set from the *buildRegressionTree is "FALSE", debug is "FALSE", unpruned is "FALSE", 4-minNumInstances, and useUnsmoothed is "TRUE"*. This method symbol is named of the "*M5Rules-U-M4.0*"; and

- The 3rd method gave the RMSE of 3.2602 (No.2 in Table 4.30) and this method is set from the *buildRegressionTree is "FALSE", debug is "FALSE", unpruned is "TRUE", 4-minNumInstances, and useUnsmoothed is "FALSE"*. This method symbol is named of the "*M5Rules-N-M4.0*".

Table 4.30 *Adjustment of each M5Rules algorithms*

M5Rules algorithms adjustment								
No	Build Regression Tree	Debug	Min Num Instances	unpruned	Use Unsmoothed	Correlation Coefficient	MAE	RMSE
1	**FALSE**	**FALSE**	**4**	**FALSE**	**FALSE**	**0.8632**	**2.1966**	**3.2394**
2	**FALSE**	**FALSE**	**4**	**TRUE**	**FALSE**	**0.8628**	**2.1002**	**3.2602**
3	TRUE	FALSE	4	TRUE	FALSE	0.8509	2.2746	3.4003
4	FALSE	FALSE	4	TRUE	TRUE	0.8204	2.0407	3.7996
5	TRUE	FALSE	4	TRUE	TRUE	0.8204	2.0407	3.7996
6	TRUE	FALSE	4	FALSE	FALSE	0.8541	2.2676	3.3485
7	**FALSE**	**FALSE**	**4**	**FALSE**	**TRUE**	**0.8626**	**2.1913**	**3.2447**
8	TRUE	FALSE	4	FALSE	TRUE	0.8578	2.2017	3.2984

Gray row: Good method

Because we would like to know the precision of those three methods that they are adjusted from the M5Rules method in the drying and raining season. Thus, we divided the 1st data group out in term of drying and raining season, the drying season has 1,389 records and the rainy season has 1,472 records. the results are already shown in **Table 4.31**.

Table 4.31 *The RMSE and MAE value of 3 M5Rules methods for the model building from the 1st data group of the DWTP*

M5Rules method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
M5Rules-M4.0	1.979	1.395	7.648	2.822
M5Rules-N-M4.0	**1.630**	**1.163**	**4.967**	**2.124**
M5Rules-U-M4.0	1.997	1.396	7.590	2.804

From Table 4.31, we found that the M5Rules-N-M4.0 method gave the less RMSE than another method. In the drying season, it gave the RMSE of 1.630 is less than the RMSE of 4.967 in the raining season. For this reason, when we would like to use it to predict the alum dosage, the predictive alum dosage in the drying season will be nearly the actual alum dosage than the raining season. Therefore, we will use M5Rules method in term of the *M5Rules-N-M4.0 method* to predict the alum dose in the real application that it will be highest the accuracy or not.

3) M5P method

For the model building from the 1st data group of the DWTP by using this M5P method, we adjusted eight methods of each M5P algorithms and we got three methods that they gave the less RMSE than another method, they are shown in gray row of **Table 4.32**, they are:

- The RMSE of the 1st method is 3.1721 (No.3 in Table 4.32) and this method is set from the *buildRegressionTree is "TRUE", debug is "FALSE", 4-minNumInstances, saveInstances is "FALSE", unpruned is "TRUE", and useUnsmoothed is "FALSE"*. This method symbol is named of "*M5P-N-R-M4.0*";
- The RMSE of the 2nd method is 3.1795 (No.1 in Table 4.32) and this method is completely set from the *buildRegressionTree is "FALSE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, unpruned is "FALSE", and useUnsmoothed is "FALSE"*. This method symbol is named of "*M5P-M4.0*"; and
- The 3rd method gave the RMSE of 3.1950 (No.7 in Table 4.32) and this method is successfully set from the *buildRegressionTree is "FALSE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, unpruned is "FALSE", and useUnsmoothed is "TRUE"*. This method is named of the "*M5P-U-M4.0".*

For each M5P algorithms adjustment indicated in **Table 4.32** as below:

Table 4.32 *Adjustment of each M5P algorithms for the model building from the 1st data group of the DWTP*

The M5P algorithms adjustment									
No	Build Regression Tree	Debug	Min Num Instances	Save instances	Unpruned	Use unsmoothed	Correlation Coefficient	MAE	RMSE
1	**FALSE**	**FALSE**	**4**	**FALSE**	**FALSE**	**FLASE**	**0.8686**	**2.1370**	**3.1795**
2	FALSE	FALSE	4	FALSE	TRUE	FALSE	0.5589	2.1047	7.3625
3	**TRUE**	**FALSE**	**4**	**FALSE**	**TRUE**	**FALSE**	**0.8707**	**2.1049**	**3.1721**
4	FALSE	FALSE	4	FALSE	TRUE	TRUE	0.8255	2.0319	3.7188
5	TRUE	FALSE	4	FALSE	TRUE	TRUE	0.8255	2.0319	3.7188
6	TRUE	FALSE	4	FALSE	FALSE	FALSE	0.8636	2.2151	3.2489
7	**FALSE**	**FALSE**	**4**	**FALSE**	**FALSE**	**TRUE**	**0.8674**	**2.1236**	**3.1950**
8	TRUE	FALSE	4	FALSE	FALSE	TRUE	0.8644	2.1558	3.2264

Gray row: Good method

Because we would like to know the accuracy of those three methods in the drying and raining season, we divided the 1st data group (2,861 records) out in term of the drying and raining season. The drying season has 1,389 records and the raining season has 1,472 records. We explained their precision by the RMSE value, the results are indicated in **Table 4.33.**

Table 4.33 *The RMSE and MAE value of 3 M5P methods in the drying and raining season for the model building from the 1st data group of the DWTP*

M5P method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
M5P-M4.0	1.903	1.356	7.368	2.729
M5P-N-R-M4.0	**1.817**	**1.314**	**6.476**	**2.510**
M5P-U-M4.0	1.894	1.327	7.311	2.711

From **Table 4.33**, we found that the *M5P-N-R-M4.0 method* gave the less RMSE of both drying and raining season than another method. In the drying season, it gave the RMSE of 1. 817 is less than the RMSE of 6.476 in the rainy season. Thus, it gives the highest precision in the drying season. Therefore, when we would like to use it to predict the alum dosage in the water treatment plant, the predictive alum dose is certainly nearly the actual alum dosage than another method. However, we will know the precision of M5P-N-R-M4.0 method when we use it to predict the alum dosage in the real applications.

4) REPTree method

For the model building from the 1st data group by using this REPTree method, we adjusted in 12 methods and we got three methods that they gave the less RMSE than another method. They are shown in the gray row of **Table 4.34**, they are:

- The 1st method gave the RMSE of 3.2903 (No.6 in Table 4.34) and this method is successfully set from the *debug is "FALSE", -1-maxDepth, noPruning is "FALSE", 0.2-minNum, 0.001-minVarianceProp, 7-numFolds, and 5-seed*. This method symbol is named of "*REPTree-M2-V0.001-N7-S5-L-1*";
- The 2nd method gave the RMSE of 3.3120 (No.4 in Table 4.34) that it completely set from the *debug is "FALSE", -1-maxDepth, 0.2-minNum, 5-numFolds, 0.001-minVarianceProp, noPruning is "FALSE", and 2-seed*. This method symbol is named of "*REPTree-M2-V0.001-N5-S2-L-1*"; and

- The 3rd method gave the RMSE of 3.3143 (No.2 in Table 4.34) and this method is already set from the *"FALSE", -1-maxDepth, noPruning is "FALSE", 0.2-minNum, 0.001-minVarianceProp, 3-numFolds, and 1-seed*. This method symbol is named of "*REPTree-M2-V0.001-N3-S1-L-1*".

Table 4.34 *Adjustment of each REPTree algorithms for the model building from the 1st data group of the DWTP*

REPTree algorithms adjustment										
No	Debug	Max Depth	Min Num	Min Variance Prop	No Pruning	Num Folds	Seed	Correlation Coefficient	MAE	RMSE
1	FALSE	-1	2.0	0.001	TRUE	3	1	0.8426	2.0939	3.5159
2	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**3**	**1**	**0.8577**	**2.1440**	**3.3143**
3	FALSE	-1	2.0	0.001	TRUE	5	2	0.8426	2.0939	3.5159
4	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**5**	**2**	**0.8578**	**2.1169**	**3.3120**
5	FALSE	-1	2.0	0.001	TRUE	7	5	0.8426	2.0939	3.5159
6	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**7**	**5**	**0.8590**	**2.1023**	**3.2903**
7	FALSE	-1	2.0	0.001	TRUE	9	7	0.8426	2.0939	3.5159
8	FALSE	-1	2.0	0.001	FALSE	9	7	0.8521	2.1227	3.3746
9	FALSE	-1	2.0	0.001	TRUE	20	18	0.8426	2.0939	3.5159
10	FALSE	-1	2.0	0.001	FALSE	20	18	0.8504	2.1216	3.3964
11	FALSE	-1	2.0	0.0001	TRUE	20	18	0.8426	2.0897	3.5157
12	FALSE	-1	2.0	0.0001	FALSE	20	18	0.8504	2.1192	3.3964

Gray row: Good method

Because we would like to know the accuracy of those three methods. As we explained in previous method, we divided the 1st data group out in term of the drying (1,389 records) and raining (1,472 records) season. We explained that methods precision by the RMSE value, if the RMSE is very low or nearly zero that shown that method has the highest accuracy. The results indicated in **Table 4.35.**

Table 4.35 *The RMSE and MAE value of 3 REPTree methods in the drying and raining season for the model building from the 1st data group of the DWTP*

REPTree method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
REPTree-M2-V0.001-N5-S2-L-1	**1.584**	**1.033**	**5.603**	**2.243**
REPTree-M2-V0.001-N3-S1-L-1	1.846	1.242	6.804	2.626
REPTree-M2-V0.001-N7-S5-L-1	1.592	1.099	5.852	2.255

Table 4.35 shown that the *REPTree-M2-V0.001-N5-S2-L-1* method gave the less RMSE than another method. In the drying season, it gave the RMSE of 1.584 is less than the RMSE of 5.603 in the raining season. For this reason, it has the highest accuracy in the drying season and it will be more precise than another method. Of course, when we would like to use it to predict the alum dose, the predictive alum dose

is certainly nearly the actual alum dose than another method. However, we will know its precision when we bring it to use in the real applications.

4.2.2.2 The model building for alum dosage prediction by using Weka data mining software from the 2nd data group of the DWTP

The model adjustment and building for alum dosage prediction by using Weka data mining software from the 2nd data group of the DWTP also used 4 methods i.e. Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree.

1) Multilayer Perceptron method

For the model building and adjustment from the 2nd data group of the DWTP by using this MLP method, we completely adjusted in eight methods of each MLP algorithms and we got three methods that they gave the less RMSE value. They are indicated in the gray rows of **Table 4.36**, they are:

- The 1st method gave us the RMSE of 4.0296 (No.43 in table 4.36) and this method is successfully set from the *12-hiddenLayer, 0.3-learningRate, 0.2-momentum, 5000-trainingTime, 20-validationThreshold, and 9-seed.* This method symbol is named of the "*MLP-L0.3-M0.2-N5000-V0-S9-E20-H12*";
- The 2nd method gave us the RMSE of 4.2702 (No.24 in Table 4.36) and this method is completely set from the *t-hiddenLayer, 0.3-learningRate, 0.2-momentum, 4000-trainingTime, 20-validationThreshold, and 0-seed.* This method symbol is named of the "*MLP-L0.3-M0.2-N4000-V0-S0-E20-Ht*"; and
- The 3rd method gave us the RMSE of 4.3335 (No.3 in Table 4.36) and this method is also set from the *a-hiddenLayer, 0.3-learningRate, 2000-trainingTime, momentum of 0.2, 20-validationThreshold, and 0-seed.* This method symbol is named of the "*MLP-L0.3-M0.2-N2000-V0-S9-E20-Ha*".

Each MLP algorithms adjustment results are also indicated in the **Table 4.36.**

Table 4.36 *Adjustment of each MLP algorithm for the model building from the 2nd data group of the DWTP*

Multilayer Perceptron (MLP) algorithms adjustment									
No	Hidden layers	Learning rate	Momentum	Seed	Training time	Validation Threshold	Correlation Coefficient	MAE	RMAE
1	a	0.3	0.2	0	500	20	0.7646	3.3245	4.3935
2	a	0.3	0.2	0	1000	20	0.7697	3.291	4.3577
3	**a**	**0.3**	**0.2**	**0**	**2000**	**20**	**0.7715**	**3.2545**	**4.3335**
4	a	0.3	0.2	0	3000	20	0.7718	3.2479	4.3363
5	a	0.3	0.2	0	4000	20	0.7721	3.2462	4.3366
6	a	0.3	0.2	0	5000	20	0.7719	3.2472	4.3396
7	a	0.3	0.2	0	10000	20	0.7716	3.2472	4.3442
8	i	0.3	0.2	0	500	20	0.7666	3.2977	4.3606
9	i	0.3	0.2	0	1000	20	0.7717	3.2524	4.3147
10	i	0.3	0.2	0	2000	20	0.7503	3.2668	4.5571
11	i	0.3	0.2	0	3000	20	0.7376	3.2760	4.7033
12	i	0.3	0.2	0	4000	20	0.7280	3.2830	4.8153
13	i	0.3	0.2	0	5000	20	0.7202	3.2885	4.9087
14	o	0.3	0.2	0	500	20	0.7526	3.3713	4.4527
15	o	0.3	0.2	0	1000	20	0.7536	3.3622	4.4490
16	o	0.3	0.2	0	2000	20	0.7532	3.3652	4.4551
17	o	0.3	0.2	0	3000	20	0.7531	3.3662	4.4566
18	o	0.3	0.2	0	4000	20	0.7531	3.3664	4.4569
19	o	0.3	0.2	0	5000	20	0.7531	3.3665	4.4570
20	t	0.3	0.2	0	500	20	0.7653	3.2897	4.3660
21	t	0.3	0.2	0	1000	20	0.7704	3.2362	4.3177
22	t	0.3	0.2	0	2000	20	0.7711	3.2223	4.3080
23	t	0.3	0.2	0	3000	20	0.7727	3.2127	4.2918
24	**t**	**0.3**	**0.2**	**0**	**4000**	**20**	**0.7758**	**3.2065**	**4.2702**
25	t	0.3	0.2	0	5000	20	0.7755	3.2086	4.2738
26	4	0.3	0.2	0	500	20	0.7671	3.4338	4.4411
27	4	0.3	0.2	3	1000	20	0.7669	3.4253	4.4542
28	4	0.3	0.2	3	2000	20	0.7656	3.4063	4.4732
29	4	0.3	0.2	3	3000	20	0.7657	3.4067	4.4816
30	4	0.3	0.2	3	4000	20	0.7652	3.4097	4.4925
31	4	0.3	0.2	3	5000	20	0.7648	3.4153	4.5000
32	8	0.3	0.2	6	500	20	0.7406	3.1438	4.6450
33	8	0.3	0.2	6	1000	20	0.7348	3.1201	4.7251
34	8	0.3	0.2	6	2000	20	0.7092	3.1375	5.0692
35	8	0.3	0.2	6	3000	20	0.7041	3.1478	5.1376
36	8	0.3	0.2	6	4000	20	0.7033	3.1363	5.1622
37	8	0.3	0.2	6	5000	20	0.6730	3.1396	5.5361
38	12	0.3	0.2	9	500	20	0.7906	3.0587	4.0924
39	12	0.3	0.2	9	1000	20	0.7824	3.0408	4.1625
40	12	0.3	0.2	9	2000	20	0.7830	2.9760	4.0617
41	12	0.3	0.2	9	3000	20	0.7929	2.9587	4.0786
42	12	0.3	0.2	9	4000	20	0.7977	2.9425	4.0368
43	**12**	**0.3**	**0.2**	**9**	**5000**	**20**	**0.7988**	**2.9376**	**4.0296**

Gray row: Good method

Because we would like to know the preciseness of those three methods in the drying and raining season, we divided the 2nd data group (2,284 records) out in term of the drying and raining season. The drying season has 1,107 records and 1,177 records in the rainy season. We always decide the method precision by the RMSE value, if the RMSE value is very less, that method has the highest accuracy. On the other hand, if the RMSE value is quite high that indicated that method is low accuracy. The RMSE and MAE value of three good MLP methods indicated in Table **4.37**.

Table 4.37 *The RMSE and MAE value of 3 MLP methods in term of the drying and raining season for the model building from the 2nd data group of the DWTP*

MLP method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
MLP-L0.3-M0.2-N5000-V0-S9-E20-H12	**2.433**	**1.729**	**12.586**	**3.774**
MLP-L0.3-M0.2-N4000-V0-S0-E20-Ht	3.034	1.946	13.428	3.818
MLP-L0.3-M0.2-N2000-V0-S0-E20-Ha	2.844	1.815	14.485	3.984

From **Table 4.37**, we found that the RMSE of 2.433 (Drying season) and 12.586 (Raining season) of the MLP-L0.3-M0.2-N5000-V0-S9-E20-H12 method is less than another method in both drying and raining season. For this reason, it has the highest accuracy in the drying season than rainy season and its predictive alum dose results will be nearly the actual alum dose too. However, we will know the method precision when we bring it to predict the alum dosage in the real applications.

2) M5Rules method

For the model building and adjustment from the 2nd data group of the DWTP. We adjusted in eight methods and got three methods that they gave the less RMSE values because we hold the low RMSE is decided the method precision.

Those three methods are already indicated in gray rows of **Table 4.38**, they are:

- The 1st method gave us the RMSE of 3.5671 (No.8 in Table 4.38) and this method is completely set from the *buildRegressionTree is "TRUE", debug is "FALSE", 4-minNumInstances, unpruned is "FALSE", and useUnsmoothed is "TRUE".* This method symbol is named of the "*M5Rules-U-R-M4.0*";
- The 2nd method gave us the RMSE of 3.6229 (No.1 in Table 4.38) that it is successfully set from the *buildRegressionTree is "FALSE", debug is "FALSE", unpruned is "FALSE", 4-minNumInstances, and useUnsmoothed is "FALSE".* This method symbol is named of the "*M5Rules-M4.0*"; and
- The 3rd method gave us the RMSE of 3.6327 (No.7 in Table 4.38) that it is also set from the *buildRegressionTree is "FALSE", debug is "FALSE", unpruned is "FALSE", 4-minNumInstances, and useUnsmoothed is "TRUE".* This method symbol is entitled of the "*M5Rules-U-M4.0*".

***Table 4.38** Adjustment of each M5Rules algorithms for the model building from the 2nd data group of the DWTP*

M5Rules algorithms adjustment								
No	Build Regression Tree	Debug	Min Num Instances	unpruned	Use Unsmoothed	Correlation Coefficient	MAE	RMSE
1	**FALSE**	**FALSE**	**4**	**FALSE**	**FALSE**	**0.8376**	**2.4829**	**3.6229**
2	FALSE	FALSE	4	TRUE	FALSE	0.8258	2.4788	3.7743
3	TRUE	FALSE	4	TRUE	FALSE	0.8293	2.6402	3.7351
4	FALSE	FALSE	4	TRUE	TRUE	0.7888	2.5353	4.2427
5	TRUE	FALSE	4	TRUE	TRUE	0.7888	2.5353	4.2427
6	TRUE	FALSE	4	FALSE	FALSE	0.8361	2.5681	3.6378
7	**FALSE**	**FALSE**	**4**	**FALSE**	**TRUE**	**0.8369**	**2.4827**	**3.6327**
8	**TRUE**	**FALSE**	**4**	**FALSE**	**TRUE**	**0.8430**	**2.494**	**3.5671**

Gray row: Good method

Because we would like to know the preciseness of those three M5Rules methods in the drying and raining season, we divided the 2nd data group out in term of the drying and raining season that the drying season has 1,107 records and 1,177 records in the rainy season. We also determined the methods precision by the RMSE value, if low RMSE means more accuracy (The best RMSE value is nearly zero or zero). The results are indicated in **Table 4.39** as below:

***Table 4.39** The RMSE and MAE value of 3 M5Rules methods in the drying and raining season for the model building from 2nd data group of the DWTP*

M5Rules method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
M5Rules-U-R-M4.0	**2.411**	**1.574**	**8.045**	**2.948**
M5Rules-M4.0	2.780	1.807	8.266	3.010
M5Rules-U-M4.0	2.776	1.803	8.401	3.112

From **Table 4.39**, we found that the M5Rules-U-R-M4.0 method gave us the less RMSE in both drying and raining season. In this case, it gave the RMSE of 2.411 and 8.045 in the drying and rainy season respectively. For these reason, this M5Rules method will give the highest precision in the drying season than rainy season when we would like to use it to predict the alum dosage. However, we will know the exactitude of this method when we bring it to use in the real applications.

3) M5P method

For the model building and adjustment by using this M5P method, we adjusted M5P algorithms in eight methods and we got three methods that they gave the less RMSE because we hold the RMSE value is decided the precision of the model. Those three methods already shown in the gray rows in **Table 4.40**, they are:

- The 1st method has provided us the RMSE of 3.4429 (No.2 in Table 4.40) that it is completely set from the *buildRegression is "FALSE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, useUnsmoothed is "FALSE", and unpruned is "TRUE".* This method symbol is entitled of the "*M5P-N-M4.0*";
- The 2nd method gave us the RMSE of 3.5003 (No.1 in Table 4.40) that it is successfully set from the *buildRegressionTree is "FALSE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, useUnsmoothed is "FALSE", and unpruned is also "FALSE"*. This method symbol is named of the "*M5P-M4.0*"; and
- The 3rd method gave us the RMSE of 3.5207 (No.3 in Table 4.40) that it is already set from the *buildRegressionTree is "TRUE", saveInstances is "FALSE", debug is "FALSE", 4-minNumInstances, useUnsmoothed is "FALSE", and unpruned is "TRUE"*. This method symbol is entitled of the "*M5P-N-R-M4.0*".

Table 4.40 *Adjustment of each M5P algorithms for the model building from 2nd data group of the DWTP*

M5P algorithms adjustment									
No	Build Regression Tree	Debug	Min Num Instances	Save instances	Unpruned	Use unsmoothed	Correlation Coefficient	MAE	RMSE
1	**FALSE**	**FALSE**	**4**	**FALSE**	**FALSE**	**FLASE**	**0.8491**	**2.4325**	**3.5003**
2	**FALSE**	**FALSE**	**4**	**FALSE**	**TRUE**	**FALSE**	**0.8545**	**2.3398**	**3.4429**
3	**TRUE**	**FALSE**	**4**	**FALSE**	**TRUE**	**FALSE**	**0.8490**	**2.4488**	**3.5207**
4	FALSE	FALSE	4	FALSE	TRUE	TRUE	0.8056	2.4578	4.0489
5	TRUE	FALSE	4	FALSE	TRUE	TRUE	0.8056	2.4578	4.0489
6	TRUE	FALSE	4	FALSE	FALSE	FALSE	0.8433	2.5379	3.5788
7	FALSE	FALSE	4	FALSE	FALSE	TRUE	0.8452	2.4455	3.5439
8	TRUE	FALSE	4	FALSE	FALSE	TRUE	0.8458	2.4891	3.5363

Gray row: Good method

Because we would like to know those three M5P methods in the drying and raining season. Thus, we divided the 2nd data group out in term of the drying and raining season. The drying season has 1,107 records and 1,177 records in the raining season. We also decide those methods precisions by the RMSE value. The results are indicated in **Table 4.41**.

Table 4.41 *The RMSE and MAE value of 3 M5P methods in the drying and raining season for the model building of the DWTP*

M5P method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
M5P-M4.0	2.306	1.527	8.266	3.010
M5P-N-M4.0	**1.916**	**1.368**	**6.831**	**2.651**
M5P-N-R-M4.0	2.198	1.517	8.266	3.010

From **Table 4.40**, we found that the *M5P-N-M4.0* method gave us the less RMSE in both drying and raining season than another method. In the drying season, it gave us the RMSE of 1.916 and 6.831 in the raining season. In this case, its RMSE in the drying season is less than the RMSE of raining season. For this reason, it will be more accuracy in the drying season than raining season. However, we will know this method precision when we use it to predict the alum dosage in the real application which it can use in the real water treatment plant or not.

4) REPTree method

For the model building from the 2nd data group of the DWTP by using this REPTree method, we adjusted REPTree algorithms in 8 methods and we got three method that they gave the less RMSE than another method. These three REPTree methods are shown in the gray rows of **Table 4.42**, they are:

- The 1st method gave us the RMSE of 3.5789 (No.2 in Table 4.42) that it is completely set from the *debug is "FALSE", -1-maxDepth, 2-minNum, noPruning is "FALSE", 0.001-minVarainceProp, 3-numFolds, and 1-seed*. This method symbol is named of the "*REPTree-M2-V0.001-N3-S1-L-1*";
- The 2nd method has provided us the RMSE of 3.6257 (No.8 in Table 4.42) that it is successfully set from the *debug is "FALSE", -1-maxDepth, 2-minNum, noPruning is "FALSE", 0.001-minVarianceProp, 9-numFolds, and 7-seed*. This method symbol is entitled of the "*REPTree-M2-V0.001-N9-S7-L-1*"; and
- The 3rd Method gave us the RMSE of 3.6339 (No.6 in Table 4.42) and this method is set from the *debug is "FALSE", -1-maxDepth, 2-minNum, noPruning is "FALSE", 0.001-minVarianceProp, 7-numFolds, and 5-seed*. This method symbol is named of the "*REPTree-M2-V0.001-N7-S5-L-1*".

Table 4.42 *Adjustment of each REPTree algorithms for the model building from 2^{nd} data group of the DWTP*

REPTree algorithms adjustment										
No	Debug	Max Depth	Min Num	Min Variance Prop	No Pruning	Num Folds	Seed	Correlation Coefficient	MAE	RMSE
1	FALSE	-1	2.0	0.001	TRUE	3	1	0.8227	2.4373	3.8376
2	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**3**	**1**	**0.8423**	**2.4167**	**3.5789**
3	FALSE	-1	2.0	0.001	TRUE	5	2	0.8227	2.4373	3.8376
4	FALSE	-1	2.0	0.001	FALSE	5	2	0.8357	2.4801	3.6527
5	FALSE	-1	2.0	0.001	TRUE	7	5	0.8227	2.4373	3.8376
6	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**7**	**5**	**0.8375**	**2.4324**	**3.6339**
7	FALSE	-1	2.0	0.001	TRUE	9	7	0.8227	2.4373	3.8376
8	**FALSE**	**-1**	**2.0**	**0.001**	**FALSE**	**9**	**7**	**0.8388**	**2.4350**	**3.6257**
9	FALSE	-1	2.0	0.001	TRUE	20	18	0.8227	2.4373	3.8376
10	FALSE	-1	2.0	0.001	FALSE	20	18	0.8328	2.4649	3.6841
11	FALSE	-1	2.0	0.0001	TRUE	20	18	0.8227	2.4374	3.8376
12	FALSE	-1	2.0	0.0001	FALSE	20	18	0.8328	2.4649	3.6841

Gray row: Good method

Because we would like to know the precision of those three REPTree methods in the drying and rainy season, we already divided the 2^{nd} data group out in term of drying (1,107 records) and raining (1,177 records) season. We also decide the accuracy of REPTree method by the RMSE value. As we know, if low RMSE value means more accurate. The results are indicated in **Table 4.43**.

Table 4.43 *The RMSE and MAE value of 3 REPTree methods for the model building from the 2^{nd} data group of the DWTP*

REPTree method				
Method	**Drying season**		**Raining season**	
	RMSE	**MAE**	**RMSE**	**MAE**
REPTree-M2-V0.001-N3-S1-L-1	2.030	1.338	7.408	2.803
REPTree-M2-V0.001-N7-S5-L-1	2.072	1.323	7.129	2.659
REPTree-M2-V0.001-N9-S7-L-1	**2.017**	**1.379**	**6.409**	**2.474**

From **Table 4.43**, we found that the *REPTree-M2-V0.001-N9-S7-L-1* method is given us the less RMSE than another method. In this case, it gave the RMSE of 2.017 in the drying season is less than the RMSE of 6.409 in the raining season. For this reason, it is more accurate than another method and has the highest precision in the drying season. However, we will know this method accuracy when we bring it to predict the alum dosage in the real application which it can use in the real plant or not, or it can be an agent the Jar-Test experiment or not.

4.3 Applications

We used Weka data mining software for the models building of alum dosage prediction in the coagulation of water treatment plant processes in Vientiane capital, Lao PDR. In the models building, we used 4 methods to classify the models i.e. Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree with 2 data groups i.e. *first data group*, we substituted all missing values of each parameter by the average value of that parameter, computed by each month and *second data group*, we cut off the missing value to reduce bias.

Because the built models will be precise and credible, they are depended on the real applications. For the real applications, we use the best models built from 4 methods to predict the alum dose by using 3 data sets i.e. *first data set is the original data that used for the model building, second data set is the new data that collected from both of water treatment plants (November 2016-January 2017 or 3 months)* and *third data set is the new data from Jar-Test experiment*.

On the other hand, we used the best model for alum dosage prediction in the Bangkhen Water Treatment Plant (Thailand) to compare the predicted alum dose results of Lao's water treatment plant and Bangkhen water treatment plant in Thailand. We did like this because we would like to know the model can use to alum dosage prediction in another plant or not and it will be precise or not.

4.3.1 Applications of the models from the Chinaimo Water Treatment Plant

4.3.1.1 Alum dosage prediction from original data

The models for alum dosage prediction are built from 4 methods i.e. Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree with 2 data groups. In this case, we use built models predict the old data that we used them to build the models. We use them to predict the original data because we would like to compare the predicted alum dose results with the alum dose result of new data that we didn't use them to create the models.

1) Model group 1

We chose the original data from the record of 900th-1000th or 101 records for alum dosage prediction. In these records has the drying and rainy season and they are started from ***19th March to 27 June 2012***. In the alum dosage prediction, we would like to use the best methods that they built from 4 methods, they are:

- The best Multilayer Perceptron (MLP) method gave the *RMSE of 2.9790* that it is successfully set from the *8-hidden layer, 0.3-learning rate, 6-seed, 3000-training time, 0.2-momentum, and 20-validation threshold*. Its method symbol is entitled of the "*MLP -L0.3 -M0.2 -N3000 -V0 -S6 -E20 -H8*".
- The best M5Rules method gave us the RMSE of 2.9025 and this method is completely set from the *buildRegressionTree is "FALSE", debug is "FALSE", miniNumInstances of 4, unpruned is "TRUE", and useUnsmoothed is "FALSE"*. This method symbol is named of the "*M5Rules-N-M4.0*".
- The best M5P method gave the RMSE of 2.5955 and this method is completely set from the *buildRegressionTree is "FALSE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, unpruned is "TRUE", and useUnsmoothed is "FALSE"*. This method symbol is named "*M5P -N -M4.0"*.
- The best REPTree method gave the RMSE of 3.0137 and this method is completely built from the *2-minNum, -1-maxDepth, 5-numFolds, seed of 2, debug is "FALSE", 0.001-minVarianceProp, and noPruning is "FALSE"*. This method symbol is named of the "REPTree-M2-V0.001-N5-S2-L-1".

The predictive alum dose results are indicated in **Table 4.44**. We analyze the model precision by the *RMSE* as shown in **Table 4.45**.

Table 4.44 *The predictive alum dosage results of model group 1 of the CWTP*

No.	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree	No.	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree
1	12	10	11	11	10	52	20	12	13	14	15
2	10	10	10	10	10	53	20	13	15	15	16
3	12	10	11	11	10	54	25	18	20	20	20
4	10	9	10	10	10	55	25	18	20	20	20
5	10	10	10	10	10	56	20	17	20	20	20
6	10	10	10	10	10	57	25	14	19	18	16
7	10	10	10	11	10	58	25	15	18	19	20
8	10	10	10	10	10	59	25	15	19	19	19
9	10	10	10	10	10	60	25	14	19	19	16
10	10	10	10	10	10	61	20	14	16	16	16
11	10	10	10	10	10	62	12	13	12	13	12
12	10	9	10	10	10	63	20	13	14	13	12
13	10	10	10	11	10	64	20	13	16	14	13
14	10	10	11	11	10	65	20	14	16	16	13
15	10	10	10	10	10	66	25	16	20	17	16
16	10	10	10	10	10	67	25	17	20	18	17
17	10	10	10	10	10	68	25	19	21	20	20
18	12	10	11	11	10	69	15	15	15	15	15
19	10	9	10	10	10	70	25	17	21	21	20
20	12	10	11	11	10	71	20	16	18	19	16
21	10	10	10	10	10	72	25	19	20	18	18
22	10	10	11	11	10	73	20	21	21	20	20
23	10	10	10	10	10	74	25	24	25	24	26
24	10	9	10	10	10	75	25	24	25	24	25
25	10	11	11	11	11	76	25	23	24	23	24
26	10	10	10	10	10	77	20	20	21	21	20
27	10	10	11	10	10	78	20	21	20	20	20
28	10	10	10	10	10	79	20	20	20	20	20
29	10	10	10	10	10	80	20	18	20	18	18
30	10	10	10	10	10	81	20	20	19	17	18
31	10	10	10	10	10	82	20	20	20	20	20
32	12	11	11	11	11	83	20	21	21	20	20
33	12	10	10	10	10	84	20	21	21	20	20
34	12	10	12	11	10	85	30	26	28	28	28
35	12	12	12	13	12	86	25	23	25	23	26
36	12	10	11	11	10	87	20	19	19	19	19
37	10	10	10	10	10	88	20	17	19	19	19
38	10	10	10	10	10	89	20	18	19	17	18
39	10	10	10	10	10	90	20	17	19	19	19
40	10	10	11	11	10	91	20	18	20	17	16
41	10	17	17	19	20	92	15	16	15	15	13
42	10	10	10	10	10	93	15	15	16	14	13
43	10	10	10	10	10	94	15	14	14	13	13
44	12	10	11	11	10	95	15	14	14	14	13
45	12	9	10	10	10	96	20	18	19	17	18
46	12	10	11	10	10	97	20	21	21	20	20
47	12	10	11	11	10	98	20	19	19	20	20
48	12	10	11	11	10	99	15	16	16	17	15
49	12	10	11	11	11	100	15	18	19	17	18
50	15	11	12	11	11	101	15	18	17	18	18
51	20	11	13	14	15						

Table 4.45 *The RMSE and MAE value of 4 methods in the alum dosage prediction from model group 1 of the CWTP*

Method	RMSE	MAE
Multilayer Perceptron (MLP)	7.004	2.390
M5Rules	**3.157**	**1.545**
M5P	4.127	1.854
REPTree	4.998	1.933

From **Table 4.45**, we found that the M5Rules gave the less RMSE of 3.157 and MAE of 1.545 than another method. For this reason, the M5Rules method is more precise and credible than another method. For the predictive alum dose results indicated in **Figure 4.5** as shown below:

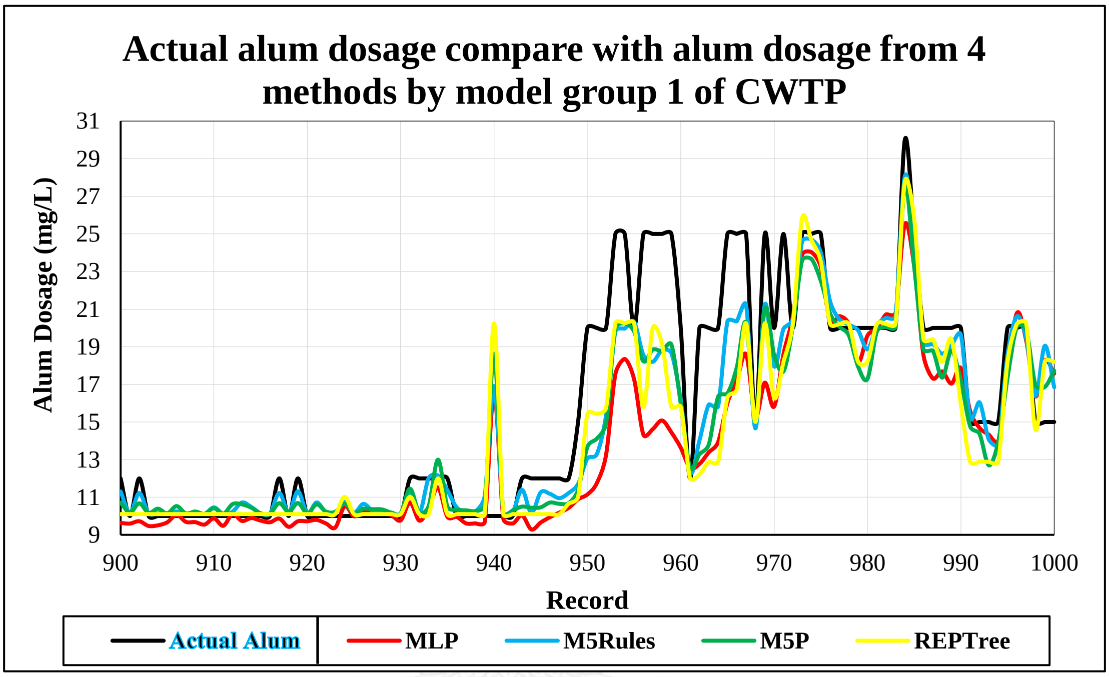


Figure 4.5 *Graph of predictive alum dose from model group 1 of the CWTP*

From Figure 4.5, we found that in the record 900th-950th, the predictive alum dosage values are so nearly or the same actual alum dosage values and from record 973rd-1000th. The predictive alum dosage values are also nearly or same the actual alum dosage value because some points of curves are the same points. When we carefully look at the blue curve (M5Rules), it is nearly the actual alum. Therefore, the M5Rules has the precision and credibility than another method when we compared the RMSE values of 4 methods together.

Thus, we can summarize the method is completely built from the *buildRegressionTree is "FALSE", debug is "FALSE", 4-miniNumInstances, unpruned is "TRUE", and useUnsmoothed is "FALSE"* or we call *M5Rules-N-M4.0* is precise and credible than another method i.e. MLP, M5P, and REPTree.

2) **Model group 2**

We chose the record of 900th-1,000th of 2nd data group for the alum dosage prediction and we also use the best methods that they built from 4 methods, they are:

- The best MLP method gave the RMSE of 3.3276 and this method completely built from the *12-hiddenLayers, 0.3-learningRate, 0.2-momentum, 5000-trianingTime, 9-seed,* and *20-validationThreshold* function. This method symbol is named of the "*MLP-L0.3-M0.2-N5000-V0-S9-E20-H12*".
- The best M5Rules method gave the RMSE of 3.2239 that it successfully made from the *buildRegressionTree is "FALSE", debug is "FALSE", unpruned is "FALSE", 4-minNumInstances, and useUnsmoothed is "TRUE"*. This method symbol is entitled of the "*M5Rules-U-M4.0*".
- The best M5P method gave the RMSE of 3.6047 that it successfully built from the *buildRegressionTree is "TRUE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, unpruned is "TRUE", and useUnsmoothed is "FALSE"*. This method symbol is "*M5P-N-R-M4.0*".
- The best REPTree method gave the RMSE of 3.7262 and this method successfully built from the *debug is "FALSE", -1-maxDepth, noPruning is "FALSE", 2-minNum, 0.001-minVarianceProp, 9-numFolds, and 7-seed*. This method symbol is named of the "*REPTree-M2-V0.001-N9-S7-L-1*".

The predicted alum dose results are indicated in **Table 4.46** and **Figure 4.6**. We analyze the methods precisions by the *RMSE* as shown in **Table 4.47**.

Table 4.46 *The predictive alum dosage from model group 2 of the CWTP*

No.	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree	No.	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree
1	12	12	12	13	12	52	25	23	24	24	24
2	12	10	10	11	11	53	20	18	18	19	17
3	10	10	10	11	10	54	20	17	17	19	17
4	10	10	10	10	10	55	20	18	17	17	17
5	10	10	10	10	10	56	20	17	17	19	17
6	10	10	10	11	10	57	20	18	17	17	17
7	10	17	20	19	16	58	15	16	16	15	16
8	10	10	10	10	10	59	15	15	14	15	12
9	10	10	10	10	10	60	15	14	14	13	12
10	12	11	10	11	10	61	15	14	13	14	12
11	12	10	10	10	10	62	20	17	17	17	17
12	12	10	10	11	11	63	20	20	21	22	24
13	12	11	10	11	11	64	20	19	20	20	17
14	12	11	10	10	10	65	15	16	17	17	17
15	12	11	10	11	11	66	15	18	18	17	17
16	15	11	11	11	11	67	15	17	17	18	17
17	20	12	12	13	17	68	15	17	17	18	17
18	20	12	12	13	17	69	15	16	16	16	16
19	20	14	16	16	17	70	15	17	17	18	17
20	25	17	20	21	22	71	15	17	17	17	17
21	25	18	20	21	22	72	15	15	15	15	12
22	20	17	20	19	22	73	15	15	15	17	17
23	25	14	19	18	15	74	15	14	13	14	12
24	25	15	19	19	19	75	15	14	13	14	13
25	25	15	19	19	19	76	15	15	14	15	12
26	25	15	19	19	17	77	15	15	15	15	19
27	20	14	16	17	15	78	15	16	15	16	16
28	12	13	13	13	12	79	20	19	20	20	21
29	20	13	13	13	16	80	25	21	22	22	24
30	20	13	13	14	12	81	20	20	19	20	20
31	20	14	16	17	15	82	25	21	22	22	24
32	25	16	16	16	19	83	25	21	22	23	24
33	25	17	17	18	17	84	20	19	20	20	21
34	25	18	20	21	25	85	20	18	20	21	22
35	15	15	14	15	13	86	20	18	20	21	21
36	25	17	20	21	22	87	20	18	18	17	17
37	20	16	19	18	19	88	20	18	17	18	17
38	25	18	18	18	17	89	20	16	20	20	25
39	20	20	19	19	20	90	25	21	26	22	24
40	25	23	24	24	24	91	20	16	16	16	19
41	25	23	24	24	24	92	20	17	20	19	17
42	25	23	23	24	24	93	25	20	24	22	24
43	20	20	20	19	18	94	30	25	25	25	24
44	20	20	21	23	24	95	30	22	23	24	24
45	20	20	19	19	20	96	30	23	23	24	24
46	20	18	18	18	17	97	25	22	22	23	24
47	20	19	19	17	17	98	20	22	22	21	24
48	20	20	19	19	20	99	25	23	24	24	24
49	20	20	21	23	24	100	25	24	25	24	24
50	20	20	21	22	24	101	40	37	24	35	39
51	30	25	26	25	24						

Table 4.47 *The RMSE and MAE of 4 methods from model group 2*

Method	RMSE	MAE
Multilayer Perceptron (MLP)	8.384	2.967
M5Rules	7.056	2.615
M5P	**5.421**	**2.480**
REPTree	5.437	2.550

From Table 4.47, we found that the RMSE of 5.421 and MAE of 2.480 of the M5P method are less than another method. For this reason, the M5P method has the highest precision than another method.

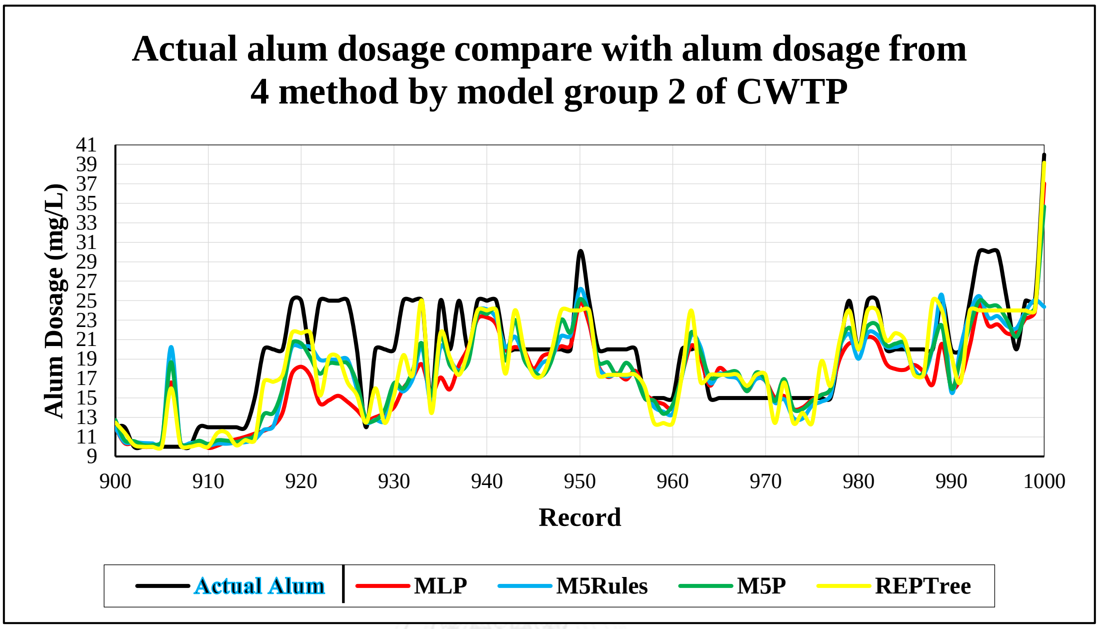


Figure 4.6 *Graph of predictive alum dose from model group 2 of the CWTP*

From the Figure 4.6, we found that the 4 methods curves are nearly the actual alum curve from record of 900th-920th and 932nd-992nd as shown in Figure 4.14 above. But we carefully look at the green curve (M5P), it is so nearly the black curve (Actual alum) than another method. For this reason, the M5P method from model group 2 has the highest precision and credibility than another method.

4.3.1.2 Alum dosage prediction by using new data

For the new data collection, the researchers collected from November 2016 to January 2017 or 92 records. We also use 2 models that they built from 4 methods i.e. MLP, M5Rules, M5P, and REPTree for the alum dosage prediction in this CWTP.

1) **Model group 1**

We synthesize the model precision by using 3 different measurement values as shown in **Table 4.48** and analyze by using RMSE value as shown in **Table 4.49.**

Table 4.48 *Measurement of model precision (Boonnao et al., 2015)*

Measurement	Definition
± 2 mg/L	High precision
± 3 mg/L	Moderate precision
± 5 mg/L	Low precision
$> 5\ and < -5$	Imprecision

- ± 2 mg/L: The predictive alum dosage values are in this value, we hold that those predicted alum dosage values can use in the real water treatment plant.
- ± 3 mg/L: The predictive alum dosage values are in this value, we hold that those predictive alum dosage values can use in the real water treatment plant but use with the Jar-Test experiment.
- ± 5 mg/L: The predictive alum dosage values are in this value can’t use in the real water treatment plant.
- $> 5\ and < -5$ mg/L: The predictive alum dosage values are in this value can’t use in the real water treatment processes.

The real predictive alum dosage results are indicated in **Table 4.49** as below:

Table 4.49 *The predictive alum dosage results of 4 methods from model group 1 in the real applications for the CWTP*

No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree	No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree
1	15	14	14	13	13	47	20	17	19	19	20
2	15	14	12	14	13	48	20	15	15	15	15
3	15	14	14	14	20	49	15	13	13	13	12
4	15	14	13	14	13	50	15	13	14	14	13
5	20	14	15	13	13	51	15	13	15	14	12
6	20	15	18	19	12	52	15	12	12	13	12
7	25	19	21	20	20	53	15	12	12	13	12
8	20	15	15	15	13	54	15	12	12	13	12
9	15	14	15	14	13	55	15	11	13	12	12
10	25	28	29	29	33	56	15	11	12	12	12
11	20	24	24	23	26	57	12	11	11	12	12
12	25	19	19	20	20	58	15	11	12	12	12
13	25	19	20	20	20	59	15	11	13	14	15
14	20	17	21	20	25	60	12	10	11	10	10
15	20	16	16	16	16	61	12	11	11	11	11
16	25	19	19	20	15	62	15	11	13	14	15
17	25	18	20	17	18	63	15	12	13	12	12
18	30	31	33	33	33	64	12	10	10	11	10
19	40	43	40	40	33	65	12	11	11	11	10
20	30	36	33	36	33	66	12	10	10	11	10
21	20	25	23	25	25	67	12	10	10	10	10
22	30	20	21	21	20	68	12	11	11	11	11
23	30	19	21	20	20	69	12	11	11	11	11
24	20	17	21	20	20	70	10	10	11	11	10
25	20	17	19	20	25	71	12	10	11	10	10
26	20	17	20	20	20	72	12	11	10	11	11
27	20	17	20	19	20	73	12	11	10	10	10
28	20	16	15	15	13	74	12	11	11	11	10
29	20	15	16	14	13	75	15	12	13	13	12
30	15	14	13	13	13	76	12	11	13	12	12
31	15	14	15	16	20	77	12	11	11	11	10
32	20	16	16	16	16	78	12	11	13	14	15
33	15	14	14	16	16	79	12	11	10	12	12
34	15	14	14	15	14	80	12	11	13	11	10
35	15	12	13	13	12	81	12	12	13	12	12
36	15	12	12	14	15	82	12	10	10	11	10
37	15	12	13	12	12	83	12	11	11	11	11
38	15	12	13	14	15	84	12	12	13	14	15
39	15	13	14	13	12	85	12	12	14	14	15
40	15	13	12	13	12	86	12	12	13	14	15
41	15	13	14	13	12	87	12	12	14	13	15
42	15	12	13	12	12	88	12	12	14	14	15
43	15	12	13	14	15	89	12	11	13	12	11
44	15	12	13	13	15	90	12	13	14	14	15
45	15	13	12	13	12	91	12	12	13	13	15
46	20	16	15	15	12	92	12	11	11	11	11

Table 4.50 *The precision measurement of predictive alum dosage from 4 methods of model group 1 in the real applications for the CWTP*

	MLP		M5Rules		M5P		REPTree	
Measurement	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**
± 2	48	52	63	69	66	71	41	45
± 3	19	21	12	13	8	9	26	28
± 5	15	16	13	14	12	13	11	12
$>5\ and < -5$	10	11	4	4	6	7	14	15

From Table 4.50, we found that the M5P has 66 records are high precision that it is higher than another method. The M5Rules has 63 records are high precision that it is second. However, the M5P has the high precision records are more than M5Rules but it has the low precision records are more than the M5Rules. Thus, we decide the model precision by using RMSE in **Table 4.51.**

Table 4.51 *The RMSE and MAE of the 4 methods from model group 1 in the real applications for the CWTP*

Method	RMSE	MAE
MLP	5.554	2.633
M5Rules	**4.043**	**2.240**
M5P	4.650	2.326
REPTree	7.438	3.041

From Table 4.51, we found that the M5Rules gave the RMSE of 4.043 and MAE of 2.240 are less than another method. Thus, for this reason, the model group 1 built from the M5Rules method has the highest accuracy than another method.

On the other hand, we look at the **Figure 4.7** for the graph of the predictive alum dosage of 4 methods from model group in the real applications.

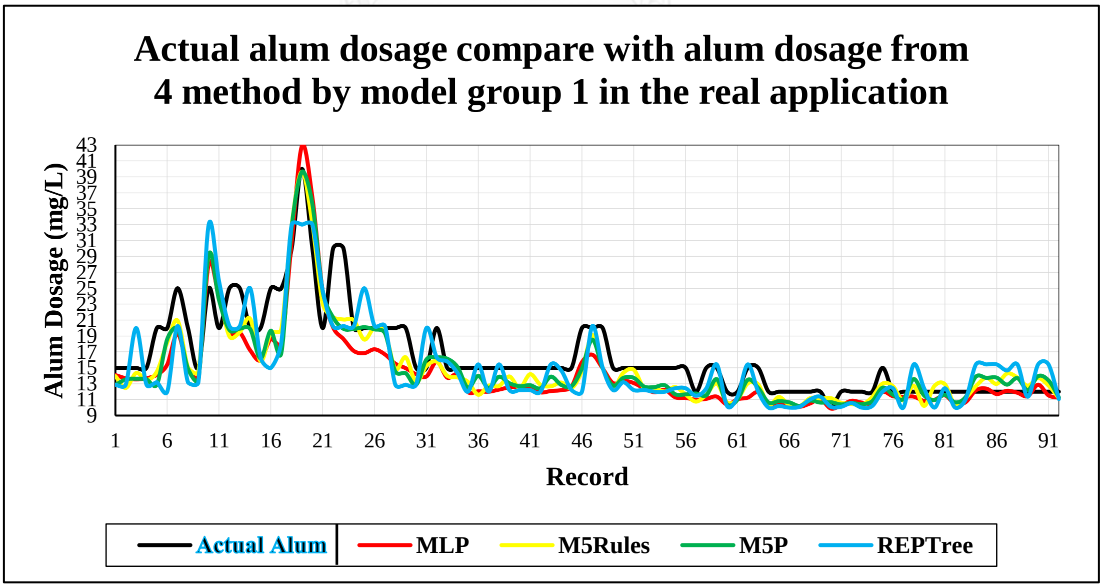


Figure 4.7 *Graph of predictive alum dosage from 4 method of model group 1 in the real application of the CWTP*

From Figure 4.7, we found that the yellow (M5Rules) and green (M5P) curve are nearly the black curve (Actual alum). But we carefully look at the yellow curve, it is so nearly the black curve than another curve. At the 56th-92nd record, the yellow and green curve are very nearly the black curve or they are in the same point with black curve. Therefore, on January 2017, the M5Rules can predict the alum dosage because the predictive alum dosage values are so nearly or same the actual alum dosage values.

2) Model group 2

The predictive alum dosage results are indicated in **Table 4.52** and the precision measurement value of 4 methods shown in **Table 4.53** as below:

Table 4.52 *The predictive alum dosage of 4 methods from model group 2 of the CWTP in the real applications for the CWTP*

No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree	No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree
1	15	14	13	13	12	47	20	17	20	19	21
2	15	14	13	14	12	48	20	15	14	15	13
3	15	14	13	14	19	49	15	13	13	12	13
4	15	14	13	14	12	50	15	14	12	14	13
5	20	14	13	13	13	51	15	13	13	13	16
6	20	16	19	19	19	52	15	13	13	12	13
7	25	19	20	21	21	53	15	12	12	12	12
8	20	15	14	15	13	54	15	13	12	13	13
9	15	14	13	14	13	55	15	12	12	12	11
10	25	27	28	28	24	56	15	12	12	12	11
11	20	23	24	24	24	57	12	12	12	11	11
12	25	19	20	19	18	58	15	12	12	12	11
13	25	19	20	20	17	59	15	12	12	13	13
14	20	17	20	20	17	60	12	11	10	11	11
15	20	16	16	16	19	61	12	12	12	11	10
16	25	18	20	20	17	62	15	12	12	13	15
17	25	18	17	17	17	63	15	12	12	13	12
18	30	30	31	31	35	64	12	11	10	11	11
19	40	42	41	35	39	65	12	11	11	11	11
20	30	36	36	35	36	66	12	11	11	11	11
21	20	24	25	24	24	67	12	11	10	10	10
22	30	20	20	20	24	68	12	11	11	11	11
23	30	18	20	21	21	69	12	12	12	11	10
24	20	17	20	20	17	70	10	10	10	11	11
25	20	17	20	20	17	71	12	11	10	11	10
26	20	17	20	20	22	72	12	12	11	11	11
27	20	17	20	20	21	73	12	11	10	11	11
28	20	16	15	15	12	74	12	11	11	11	11
29	20	15	14	15	12	75	15	12	12	13	12
30	15	14	13	13	12	76	12	12	12	12	11
31	15	14	16	16	17	77	12	12	12	12	11
32	20	16	15	16	16	78	12	12	12	13	14
33	15	14	16	17	15	79	12	11	12	12	11
34	15	17	13	13	14	80	12	11	12	12	11
35	15	12	12	13	12	81	12	12	12	12	11
36	15	13	12	13	17	82	12	11	11	11	11
37	15	12	12	13	12	83	12	11	11	11	11
38	15	13	12	13	17	84	12	13	12	13	17
39	15	13	13	13	14	85	12	13	13	13	14
40	15	13	13	12	13	86	12	12	12	13	17
41	15	13	13	13	13	87	12	13	12	13	14
42	15	12	12	13	11	88	12	12	12	13	14
43	15	13	12	13	17	89	12	12	12	12	11
44	15	13	12	13	14	90	12	13	13	13	17
45	15	13	13	12	13	91	12	12	12	13	14
46	20	16	15	15	13	92	12	12	12	11	11

Table 4.53 *The precision measurement of predictive alum dosage of 4 methods from model group 2 for the CWTP*

	MLP		**M5Rules**		**M5P**		**REPTree**	
Measurement	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**
± 2	56	60	56	60	64	70	54	59
± 3	19	21	18	20	9	10	12	13
$\geq \pm 5$	9	10	10	11	14	15	13	14
$> 5\ and < -5$	8	9	8	9	5	5	13	14

From **Table 4.53**, we found that the M5P gave the 64-high precision records are more than another method. For this reason, 64 records or 70% of all record can use in the real plant or can use in the coagulation process. For the RMSE values indicated in **Table 4.54** as below:

Table 4.54 *The RMSE and MAE value of 4 methods from model group 2 in the real applications for the CWTP*

Method	**RMSE**	**MAE**
MLP	5.088	2.369
M5Rules	4.968	2.353
M5P	**4.695**	**2.321**
REPTree	6.106	2.784

From Table 4.54, we found that the M5P gave the RMSE of 4.695 and MAE of 2.321 are less than another method. When we compared the RMSE and the high precision records, the M5P method is better than another method. Therefore, in this case, the M5P method from model group 2 has the highest precision and credibility for the alum dosage prediction in the real applications.

The predictive alum dosage results are also indicated in **Figure 4.8.** We gave the X-axis is the data record and Y-axis is amount of alum (mg/L).

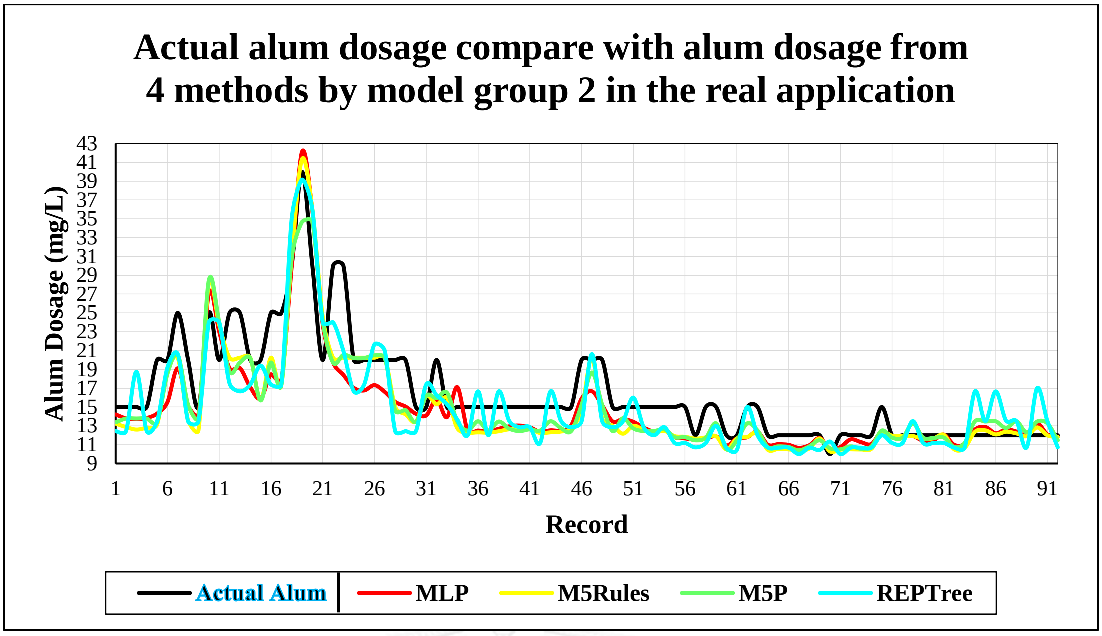


Figure 4.8 *Graph of predictive alum dosage of 4 methods from model group 2 in the real applications of the CWTP*

From Figure 4.8, we found that at the 1st to 30th record of the predictive alum dosage of 4 models is low and moderate precision than high and moderate precision. On the other hand, from 31st to 63nd record of the predictive alum dosage is moderate precision than high and low precision. Besides that, at the 64th to 92nd record, the predicted alum dose of 4 models is high precision than moderate and low precision. For this reason, the model can apply in the real plant for the alum dose finding on January 2017. When we carefully look at the green curve (M5P), it is very nearly the black curve (Actual alum). Thus, the model built from M5P method is high accuracy and credibility than another method.

4.3.1.3 Prediction the laboratory data

The researcher did Jar-Test experiment for bring these data to the model precision testing. The researcher got 72 records from the experiment.

1) Model group 1

Table 4.55 *The predictive alum dosage of 4 methods from model group 1 in the real applications using laboratory data for the CWTP*

No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree	No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree
1	15	12	12	14	15	37	10	14	13	13	13
2	15	12	13	12	12	38	10	14	14	13	13
3	15	12	13	14	15	39	10	14	16	13	13
4	25	19	19	20	20	40	10	14	13	13	13
5	25	19	20	20	20	41	10	14	13	13	13
6	20	17	21	20	25	42	10	14	14	13	13
7	20	16	16	16	16	43	10	14	13	13	13
8	25	19	19	20	15	44	10	15	14	14	13
9	25	18	20	17	18	45	15	16	16	17	15
10	20	22	21	21	21	46	20	17	20	17	17
11	20	22	21	22	20	47	15	16	16	16	16
12	25	26	26	27	28	48	20	16	20	17	16
13	25	25	25	25	25	49	25	24	25	25	25
14	25	24	24	24	26	50	25	21	24	21	24
15	20	21	22	22	22	51	20	21	21	20	20
16	25	24	25	24	26	52	25	22	24	24	25
17	20	23	21	22	24	53	25	23	25	24	25
18	20	23	21	22	20	54	25	23	24	24	25
19	30	28	28	29	33	55	25	22	24	24	25
20	40	37	35	38	33	56	25	25	25	25	25
21	35	34	34	35	33	57	15	12	14	14	14
22	35	31	33	33	33	58	15	13	14	14	14
23	35	32	34	33	33	59	15	13	14	13	15
24	40	43	38	42	33	60	15	13	13	13	12
25	12	12	12	12	11	61	15	13	13	13	12
26	12	12	12	12	11	62	20	20	21	19	20
27	12	13	12	12	12	63	20	20	20	20	20
28	12	12	12	13	12	64	15	19	19	17	17
29	12	12	12	13	12	65	25	18	19	18	18
30	12	12	12	12	11	66	15	16	15	15	16
31	12	14	15	15	16	67	15	16	18	17	16
32	12	14	15	15	16	68	15	17	17	17	15
33	12	13	14	15	16	69	20	17	19	17	17
34	12	14	14	16	16	70	15	17	17	17	15
35	12	14	16	16	16	71	25	24	25	25	26
36	12	14	16	16	16	72	25	24	25	25	26

From Table 4.55, we can count the high, moderate, and low precision record in term of the range and percentage counting. The results indicated in **Table 4.56**.

Table 4.56 *The high, moderate, and low precision of 4 methods from model group 1 in the real alum dosage prediction by using laboratory data for the CWTP*

	MLP		**M5Rules**		**M5P**		**REPTree**	
Measurement	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**
± 2	39	54	**49**	**68**	46	64	39	54
± 3	15	21	8	11	14	20	15	21
± 5	13	18	11	15	10	14	12	17
$> 5\ and > -5$	5	7	4	6	2	3	6	8

From Table 4.56, we found that the M5Rules has 49-high precision records or 68% than another method. On the other hand, we analyzed the methods precisions by using the RMSE value. The results are shown in **Table 4.57** as below:

Table 4.57 *The RMSE and MAE of 4 methods from model group 1 for the alum dosage prediction by using laboratory data for the CWTP*

Method	**RMSE**	**MAE**
MLP	4.415	2.462
M5Rules	**3.159**	**1.916**
M5P	3.438	2.067
REPTree	4.560	2.209

From Table 4.57, the M5Rules gave the less RMSE of 3.159 and MAE of 1.916 than another method. When we compared the RMSE and the precision record counting of the M5Rules, both the RMSE and precision percentage counting value are less than another method. Thus, for this reason, the model group 1 built from the M5Rules method is higher precision and credibility than another method.

The predictive alum dosage also indicated in **Figure 4.9**. We gave the X-axis is the record data and the Y-axis is the amount of alum (mg/L).

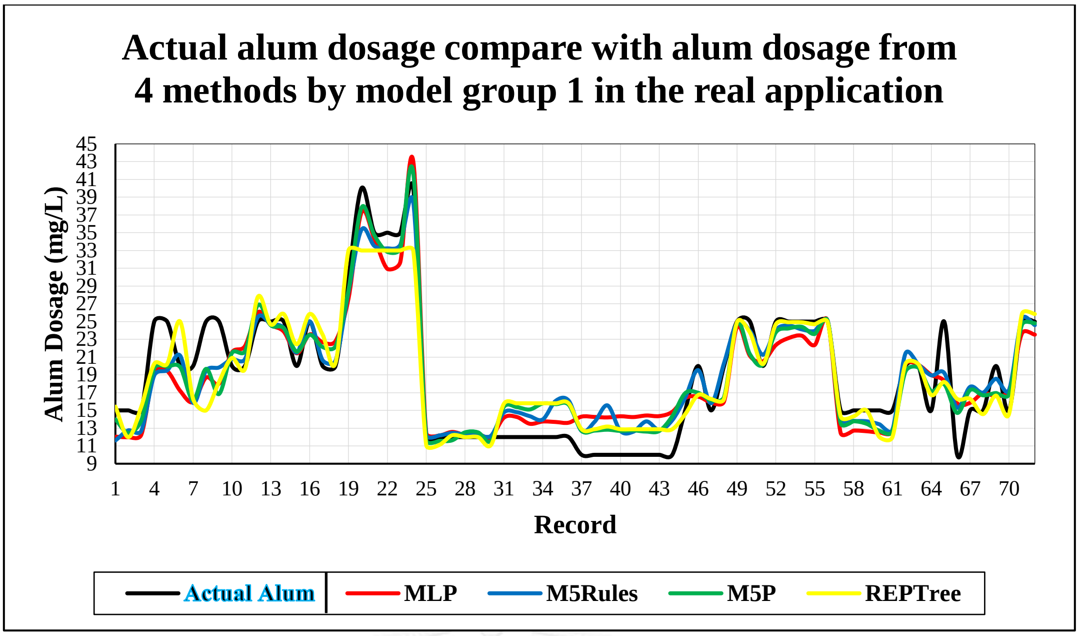


Figure 4.9 *Graph of predictive alum dosage of 4 methods from model group 1 in the real applications of the CWTP by using laboratory data*

From Figure 4.9, we found that the low and moderate precision of 4 methods is in the 1st to 10th, 31st to 44th, and 64th to 67th record. For the high precision of 4 methods is between 11th to 30th, 45th to 63rd, and 68th to 72nd record. When we carefully look at the blue curve (M5Rules), it is very nearly the black curve (Actual alum) and some points are the same. For this reason, the M5Rules method has the highest precision and credibility than another method.

2) Model group 2

Table 4.58 *The predictive alum dosage of 4 methods from model group 2 in the real applications using laboratory data for the CWTP*

No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree	No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree
1	15	13	12	13	17	37	10	14	14	13	12
2	15	12	12	13	12	38	10	14	14	13	12
3	15	13	12	13	17	39	10	14	13	13	13
4	25	19	20	19	18	40	10	14	14	13	12
5	25	19	20	20	17	41	10	14	13	13	12
6	20	17	20	20	17	42	10	14	14	13	12
7	20	16	16	16	19	43	10	14	14	13	12
8	25	18	20	20	17	44	10	15	14	15	12
9	25	18	17	17	17	45	15	16	16	17	17
10	20	21	22	22	24	46	20	17	16	17	17
11	20	22	22	21	24	47	15	16	16	16	16
12	25	25	26	25	24	48	20	16	16	16	19
13	25	24	25	24	24	49	25	24	25	25	24
14	25	23	24	24	24	50	25	21	22	22	24
15	20	21	26	22	24	51	20	21	21	23	24
16	25	23	24	24	24	52	25	22	23	23	24
17	20	22	23	21	24	53	25	23	24	25	24
18	20	22	23	21	24	54	25	23	24	25	24
19	30	27	30	28	24	55	25	22	23	23	24
20	40	37	40	35	39	56	25	25	26	25	24
21	35	34	38	34	35	57	15	13	12	13	15
22	35	30	34	31	35	58	15	13	13	13	14
23	35	31	35	31	35	59	15	13	13	13	15
24	40	42	21	35	39	60	15	13	13	13	12
25	12	13	13	12	12	61	15	13	13	13	12
26	12	12	12	12	12	62	20	20	20	19	20
27	12	13	13	12	12	63	20	20	19	19	20
28	12	12	12	12	13	64	15	19	18	17	17
29	12	13	12	12	13	65	25	18	18	18	17
30	12	12	12	12	11	66	15	16	16	15	12
31	12	14	16	16	17	67	15	16	16	17	12
32	12	14	16	16	15	68	15	17	17	17	17
33	12	14	16	16	15	69	20	17	17	17	17
34	12	14	16	16	15	70	15	17	17	17	17
35	12	14	16	16	15	71	25	23	24	24	24
36	12	14	16	16	15	72	25	23	24	24	24

From Table 4.58, we can count the high, moderate, and low precision record of 4 methods from model group 2 in term of the range and percentage. The results indicated in **Table 4.59** as below:

Table 4.59 *The high, moderate, and low precision record of 4 methods from model group 2 in the real applications using laboratory data for the CWTP*

	MLP		M5Rules		M5P		REPTree	
Measurement	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**
± 2	44	61	38	53	42	58	**45**	**63**
± 3	8	11	12	17	11	15	14	19
± 5	15	21	18	24	16	22	7	10
$> 5\ and < -5$	5	7	4	6	3	4	6	8

From Table 4.59, we found that the REPTree method gave the 45-high precision record or 63% more than another method. For this reason, we should analyze the precision of 4 models by the RMSE value as indicated in **Table 4.60**.

Table 4.60 *The RMSE and MAE value of 4 methods from model group 2 in the real applications using laboratory data for the CWTP*

Method	RMSE	MAE
MLP	**4.621**	2.533
M5Rules	6.575	2.590
M5P	**4.621**	2.533
REPTree	4.731	2.340

From Table 4.60, we found that the RMSE of 4.621of MLP and M5P method are the same and they are less than another method. When we compared the RMSE of new data and lab data of M5P, both are less than another method. Thus, the M5P is also precision and credible than another method.

The results are also indicated in **Figure 4.10** that we gave the X-axis is record and Y-axis is alum dose (mg/L).

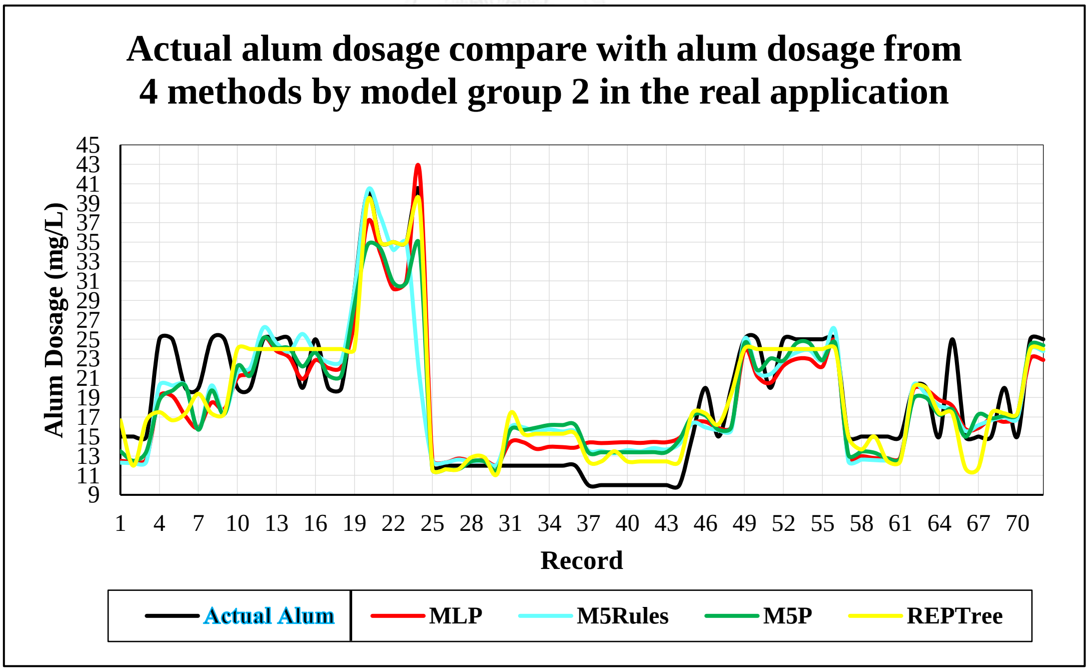


Figure 4.10 *Graph of predictive alum dose of 4 methods from model group 2 in the real application of the CWTP by using laboratory data*

From Figure 4.10, we found that the low and moderate precision at the 1st-9th record, 31st-44th, and 64th-66th of 4 models. For the high precision of 4 models is between the 10th-30th,45th-63rd, and 67th-72nd record.

4.3.2 Application of the models from Dongmarkkaiy water treatment plant (DWTP)

4.3.2.1 Alum dosage prediction from original data

1) **Model group 1**

We got 4 best methods from the model group 1 adjustment and building using 4 methods i.e. Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree. They are:

- The RMSE of the best MLP method is 3.7305 and this method is successfully built from the giving the *4-hiddenLayer, 0.3-learningRate, 0.2-momentum, 3-seed, validationThreshold of 20, and 3000-trainingTime.* This method symbol is named of the "*MLP-L0.3-M0.2-N3000-V0-S3-E20-H4*".
- The best M5Rules method gave the RMSE of 3.2602 and this method is set from the *buildRegressionTree is "FALSE", debug is "FALSE", unpruned is "TRUE", useUnsmoothed is "FALSE", and 4-minNumInstances*. This method symbol is named of the "*M5Rules-N-M4.0*".
- The RMSE of the best M5P method is 3.1721 and this method is already set from the *buildRegressionTree is "TRUE", debug is "FALSE", 4-minNumInstances, saveInstances is "FALSE", unpruned is "TRUE", and useUnsmoothed is "FALSE".* This method symbol is named of "*M5P-N-R-M4.0*".
- The best REPTree method gave the RMSE of 3.3120 (No.4 in Table 4.34) that it completely set from the *debug is "FALSE", -1-maxDepth, 0.2-minNum, 5-numFolds, 0.001-minVarianceProp, noPruning is "FALSE", and 2-seed*. This method symbol is named of "*REPTree-M2-V0.001-N5-S2-L-1*".

For the alum dosage prediction from original data, we chose the 1000th–1100th record from the 1st data group. The results indicated in **Table 4.61** and **Figure 4.11**.

Table 4.61 *Predictive alum dosage results of 4 methods from model group 1 for the DWTP*

No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree	No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree
1000	15	12	15	16	13	1051	5	6	6	6	6
1001	24	26	26	26	27	1052	5	6	7	6	7
1002	15	11	15	16	14	1053	5	6	6	6	7
1003	15	10	13	12	11	1054	5	6	7	6	7
1004	20	14	16	16	15	1055	7	6	7	7	7
1005	15	10	14	12	11	1056	8	6	7	6	7
1006	12	9	12	12	12	1057	7	6	7	7	7
1007	15	8	11	9	10	1058	8	6	8	6	7
1008	12	9	12	12	12	1059	5	6	7	6	7
1009	15	10	15	12	11	1060	5	6	7	6	7
1010	15	8	13	12	13	1061	8	6	8	7	7
1011	15	8	11	10	9	1062	7	6	7	7	7
1012	15	8	14	9	8	1063	5	6	8	7	7
1013	12	9	12	12	12	1064	7	6	7	7	7
1014	10	9	10	10	14	1065	8	6	8	6	7
1015	12	9	12	12	12	1066	8	6	8	6	7
1016	10	8	10	10	14	1067	5	6	6	6	5
1017	10	8	12	14	14	1068	5	6	7	6	7
1018	10	7	10	11	9	1069	5	6	6	6	5
1019	10	9	10	10	14	1070	5	6	7	6	7
1020	12	9	12	12	12	1071	5	6	6	6	5
1021	10	8	10	13	14	1072	8	6	8	6	7
1022	12	9	12	12	12	1073	5	6	7	6	7
1023	10	8	10	10	10	1074	5	6	7	6	7
1024	10	8	10	10	10	1075	5	6	7	6	7
1025	15	18	19	17	15	1076	5	6	6	6	5
1026	10	8	10	10	9	1077	5	6	8	6	7
1027	12	9	12	12	12	1078	5	6	6	6	5
1028	8	7	8	10	7	1079	5	6	7	6	7
1029	12	9	12	12	12	1080	5	6	6	6	5
1030	15	17	17	17	15	1081	5	6	6	6	5
1031	15	16	18	15	16	1082	5	7	6	8	5
1032	8	8	9	10	9	1083	5	6	6	6	5
1033	8	8	9	10	9	1084	5	6	8	6	7
1034	12	9	12	12	12	1085	5	6	6	6	5
1035	10	8	10	10	9	1086	5	6	8	6	7
1036	7	6	7	7	7	1087	5	6	6	6	5
1037	10	8	10	10	10	1088	5	6	8	7	7
1038	10	7	12	12	10	1089	5	6	6	6	5
1039	8	7	8	10	11	1090	5	6	6	6	5
1040	8	7	8	10	7	1091	5	6	6	6	7
1041	7	6	7	7	7	1092	5	6	6	6	5
1042	8	6	7	7	7	1093	5	6	6	6	5
1043	7	6	7	7	7	1094	5	6	6	6	5
1044	8	6	8	7	7	1095	5	6	6	6	7
1045	8	6	8	6	7	1096	5	6	6	6	5
1046	8	6	7	6	7	1097	8	6	8	8	6
1047	8	6	8	6	7	1098	5	6	6	6	5
1048	7	6	7	7	7	1099	5	6	6	6	5
1049	5	6	6	6	6	1100	8	6	8	8	5
1050	7	6	7	7	7						

From Table 4.61, we calculated the RMSE of 4 methods from model group 1. The results are shown in **Table 4.62**.

Table 4.62 *RMSE and MAE of 4 methods from model group1 for the DWTP*

Method	RMSE	MAE
MLP	2.698	1.821
M5Rules	**0.900**	**0.920**
M5P	1.418	1.225
REPTree	2.026	1.346

From Table 4.62, we found that the RMSE of 0.900 and MAE of 0.920 of the M5Rules configuration is less than another method. For this reason, the M5Rules has the highest accuracy and credibility than another method.

The predictive alum dosage of 4 methods from model group 1 in the alum dosage prediction using original data also indicated in **Figure 4.11** as below:

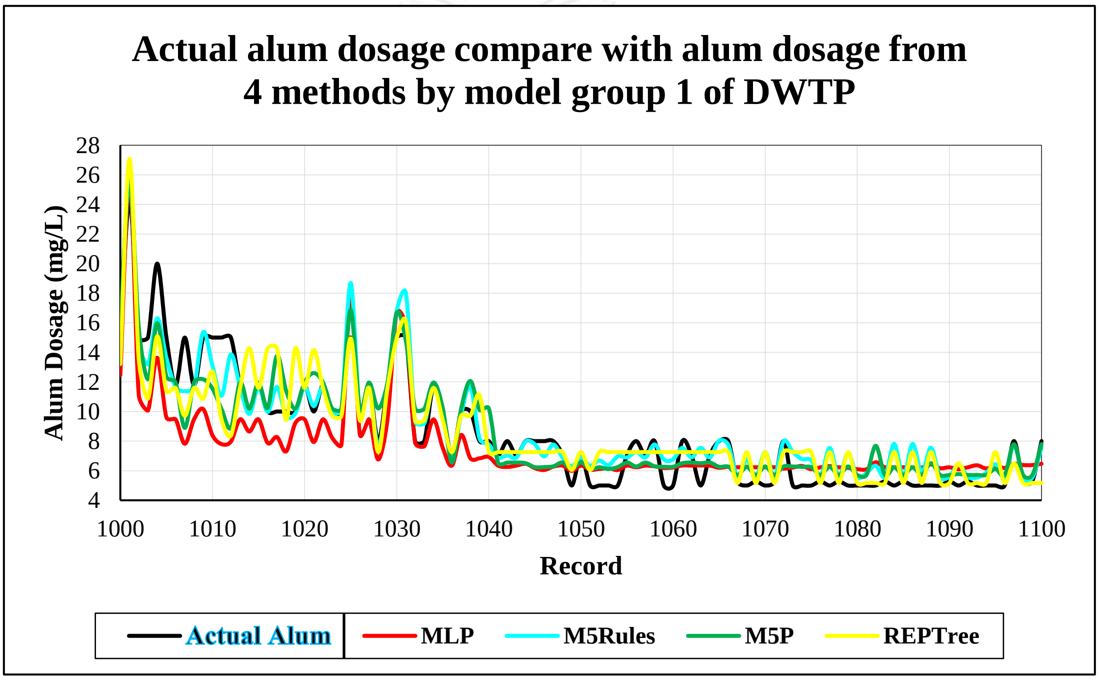


Figure 4.11 *Graph of predictive alum dosage of 4 methods from model group 1 in the alum dosage prediction using old data for the DWTP*

From Figure 4.11, we found that the blue (M5Rules), green (M5P), and yellow (REPTree) curve are nearly the black curve (Actual alum) than the red curve (MLP) at the 1000th-1024th record. But at the 1025th-1100th record, MLP, M5Rules, M5P, and REPTree curve are very nearly the actual alum curve or some points are the same. When we carefully look at the M5Rules curve, it is very nearly the actual alum curve than another method. As the same way, the RMSE of M5Rules is less than another method too.

2) Model group 2

We have gotten the 4 best methods from the model building and adjustment by using 4 methods, they are:

- The best MLP method gave us the RMSE of 4.0296 that it is successfully set from the *12-hiddenLayer, 0.3-learningRate, 5000-trainingTime, 0.2-momentum, validationThreshold of 20, and 9-seed.* This method symbol is named of the "*MLP-L0.3-M0.2-N5000-V0-S9-E20-H12".*
- The best M5Rules method gave us the RMSE of 3.5671 (No.8 in Table 4.37) and this method is completely set from the *buildRegressionTree is "TRUE", debug is "FALSE", 4-minNumInstances, unpruned is "FALSE", and useUnsmoothed is "TRUE".* This method symbol is named of the "*M5Rules-U-R-M4.0*".
- The M5P method has provided us the RMSE of 3.4429 that it is completely set from the *buildRegression is "FALSE", debug is "FALSE", saveInstances is "FALSE", 4-minNumInstances, useUnsmoothed is "FALSE", and unpruned is "TRUE".* This method symbol is entitled of the "*M5P-N-M4.0*".
- The REPTRee method gave us the RMSE of 3.6257 that it is successfully set from the *debug is "FALSE", -1-maxDepth, 2-minNum, noPruning is "FALSE", seed of 7, 0.001-minVarianceProp, and 9-numFolds*. This method symbol is entitled of the "*REPTree-M2-V0.001-N9-S7-L-1*".

The predictive alum dosage results of 4 methods from model group 2 are indicated in **Table 4.63** and **Figure 4.12.**

Table 4.63 *The predictive alum dosage of 4 methods from models group 2 for the DWTP*

No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree	No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree
1000	20	17	27	23	25	1051	15	11	15	14	16
1001	25	17	27	22	25	1052	15	14	15	14	15
1002	30	25	22	24	24	1053	15	22	22	21	18
1003	25	20	27	23	25	1054	15	19	17	16	15
1004	30	21	27	24	24	1055	15	19	17	16	15
1005	25	17	22	20	25	1056	15	20	17	19	15
1006	25	15	17	18	21	1057	10	15	17	16	15
1007	20	11	17	16	15	1058	10	13	15	14	16
1008	20	11	17	17	17	1059	10	11	15	14	16
1009	20	12	14	18	14	1060	10	14	15	14	15
1010	30	25	22	26	25	1061	10	11	15	12	10
1011	30	25	22	27	25	1062	10	10	10	9	10
1012	30	24	22	27	25	1063	10	9	10	8	10
1013	25	26	22	25	24	1064	8	8	10	8	9
1014	25	26	22	25	24	1065	8	8	10	8	9
1015	20	18	17	19	21	1066	8	8	10	8	9
1016	15	14	17	17	15	1067	8	8	10	8	9
1017	15	10	15	14	14	1068	8	8	10	8	9
1018	15	10	15	14	14	1069	8	8	10	8	9
1019	15	12	15	14	15	1070	8	8	10	8	9
1020	25	20	27	23	25	1071	8	7	7	7	9
1021	25	23	22	22	24	1072	8	7	11	8	8
1022	15	20	22	21	18	1073	8	7	7	7	9
1023	20	20	22	21	18	1074	8	7	7	7	8
1024	25	25	22	23	24	1075	8	10	10	8	10
1025	20	25	22	22	24	1076	8	7	7	7	6
1026	25	26	22	26	25	1077	5	7	6	6	6
1027	25	24	22	22	24	1078	5	7	7	7	6
1028	15	20	17	18	15	1079	5	7	7	7	6
1029	15	13	17	16	15	1080	5	7	6	6	6
1030	15	14	15	14	15	1081	5	7	6	6	6
1031	15	15	17	16	15	1082	5	7	7	7	6
1032	15	13	15	14	15	1083	5	7	7	7	6
1033	15	11	15	14	15	1084	5	7	7	7	6
1034	10	12	15	12	10	1085	8	6	6	7	8
1035	10	10	10	9	10	1086	8	6	6	6	6
1036	15	12	15	14	16	1087	5	7	6	6	6
1037	15	10	15	14	14	1088	8	7	7	6	6
1038	15	12	15	14	15	1089	8	6	6	6	6
1039	15	14	15	14	15	1090	8	7	6	6	6
1040	15	15	17	16	15	1091	8	7	6	7	8
1041	10	12	15	12	10	1092	8	7	6	7	8
1042	25	26	22	25	24	1093	8	6	6	6	6
1043	25	24	22	22	24	1094	5	6	6	6	6
1044	30	24	22	27	25	1095	5	6	6	6	6
1045	20	26	22	26	25	1096	5	6	6	6	6
1046	20	26	22	25	24	1097	5	6	6	6	6
1047	25	26	22	25	24	1098	5	7	6	6	6
1048	20	25	22	22	24	1099	5	6	8	7	9
1049	20	24	22	22	24	1100	5	7	7	7	6
1050	15	15	15	15	15						

From Table 4.63, we calculated the RMSE value for the models precision analysis. The results shown in **Table 4.64**.

Table 4.64 *RMSE and MAE of 4 methods from model group 2 for the DWTP*

Method	RMSE	MAE
MLP	6.318	2.652
M5Rules	5.073	2.456
M5P	3.026	1.889
REPTree	**2.979**	**1.603**

From Table 4.64, we found that the RMSE of 2.979 and MAE of 1.603 of the REPTree method is less than another method. For this reason, the REPTree method from model group 2 is more precise and credible than another method for the DWTP.

The predictive alum dosage results are shown in **Figure 4.12** as below:

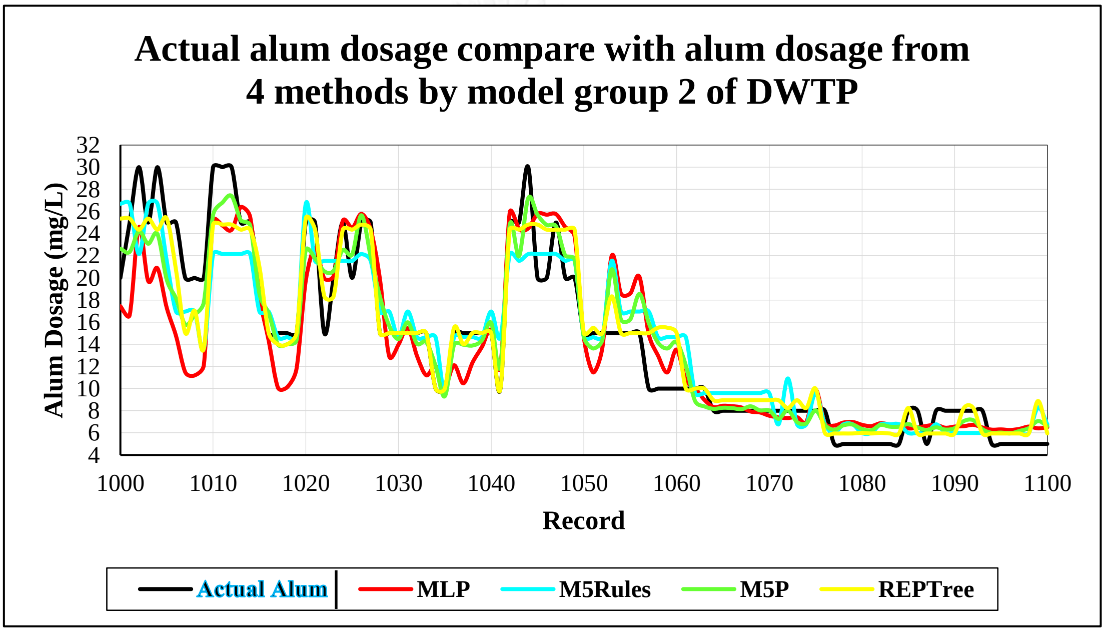


Figure 4.12 *Graph of predictive alum dosage of 4 methods compare with actual alum from model group 2 for the DWTP*

From Figure 4.12, we found that the red curve (MLP) is far from the black curve (Actual alum) than another method at the 1000th-1040th. But the red (MLP), blue (M5Rules), green (M5P), and yellow (REPTree) curve are nearly the black curve from 1041st-1100th record. The 3 curves i.e. blue, green, and yellow curve are nearly the black curve than red curve. On the other hand, when we carefully look at the yellow curve, it is very nearly the black curve than another curve because the RMSE of REPTree is less than another method.

4.3.2.2 Alum dosage prediction in the real application

For the real applications, we collected the new data from November 2016 to January 2017 or 92 records.

1) **Model group 1**

The results indicated in **Figure 4.13** as below:

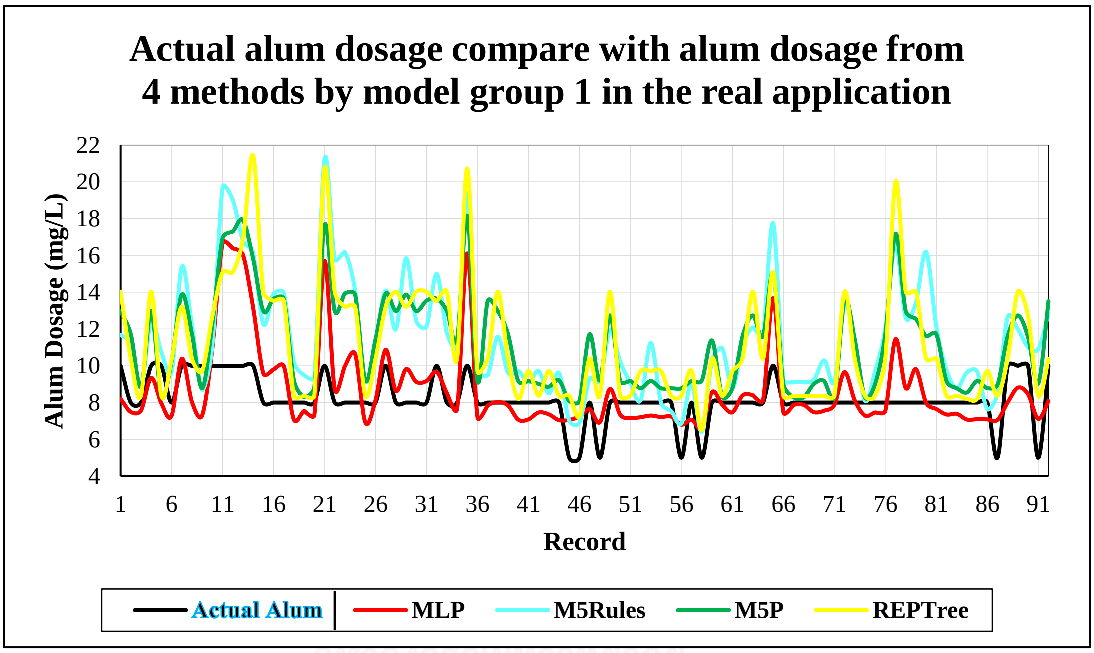


Figure 4.13 *Graph of predictive alum dosage of 4 methods from model group 1 in the real applications for the DWTP*

From Figure 4.13, we found that the red curve (MLP) is nearly the black curve (Actual alum) than another curve. In the 1st-7th, 14th-20th, 22nd-34th, 36th-64th, 66th-76th, and 79th-92nd record, the MLP method gave the predictive alum dosage is high precise than another method. But in the models building from 1st data group, the M5Rules gave the less RMSE, so it should give the predictive alum dosage values are nearly or the same actual alum dose values in the real applications. On the other hand, the MLP gave the highest precision and credibility than another method in the real applications. Therefore, the MLP method from model group 1 has more the accuracy than another method. The predictive alum dosage results of each methods from model group 1 indicated in **Table 4.65**.

Table 4.65 *The predictive alum dosage of 4 methods from model group 1 in the real applications for the DWTP*

No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree	No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree
1	10	8	12	13	14	47	8	8	9	12	10
2	8	7	11	12	10	48	5	7	9	9	8
3	8	8	9	9	8	49	8	9	12	13	14
4	10	9	12	13	14	50	8	7	10	9	8
5	10	8	11	8	8	51	8	7	9	9	8
6	8	7	10	10	10	52	8	7	8	9	10
7	10	10	15	14	13	53	8	7	11	9	10
8	10	8	12	12	10	54	8	7	8	9	10
9	10	7	9	9	10	55	8	7	8	9	8
10	10	11	11	12	13	56	5	7	7	9	8
11	10	17	20	17	15	57	8	7	9	9	10
12	10	16	19	17	15	58	5	7	6	9	6
13	10	16	17	18	17	59	8	9	10	11	10
14	10	13	16	16	21	60	8	8	11	8	8
15	8	10	12	13	14	61	8	7	9	9	10
16	8	10	14	14	14	62	8	8	11	12	10
17	8	10	14	14	14	63	8	8	12	13	14
18	8	7	10	9	8	64	8	8	12	12	10
19	8	8	9	8	8	65	10	14	18	15	15
20	8	7	9	9	8	66	8	7	9	9	8
21	10	16	21	18	21	67	8	8	9	8	8
22	8	9	16	13	14	68	8	8	9	8	8
23	8	10	16	14	13	69	8	7	9	9	8
24	8	11	14	14	13	70	8	8	10	9	8
25	8	7	9	9	8	71	8	8	9	8	8
26	8	8	11	11	10	72	8	10	13	14	14
27	10	11	14	14	13	73	8	8	11	11	10
28	8	9	12	13	14	74	8	7	8	8	8
29	8	10	16	14	13	75	8	7	10	9	8
30	8	9	12	13	14	76	8	8	12	12	10
31	8	9	12	14	14	77	8	11	17	17	20
32	10	10	15	14	14	78	8	9	13	13	14
33	8	9	12	13	14	79	8	10	13	13	14
34	8	8	11	11	10	80	8	8	16	12	10
35	10	16	19	18	21	81	8	8	12	12	10
36	8	7	10	9	10	82	8	7	10	9	8
37	8	8	9	14	10	83	8	7	9	9	8
38	8	8	12	13	14	84	8	7	10	9	8
39	8	8	10	12	10	85	8	7	10	9	8
40	8	7	10	9	8	86	8	7	8	9	10
41	8	7	9	9	10	87	5	7	8	9	8
42	8	7	10	9	8	88	10	8	13	12	10
43	8	7	9	9	10	89	10	9	12	13	14
44	8	7	10	9	8	90	10	8	11	12	13
45	5	7	7	8	8	91	5	7	11	9	8
46	5	7	7	8	7	92	10	8	13	14	10

Table 4.66 *The predictive alum dosage results in term of count in range and percentage of the real applications from model group 1 for the DWTP*

	MLP		**M5Rules**		**M5P**		**REPTree**	
Measurement	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**
± 2	**82**	**89**	48	52	40	43	54	58
± 3	4	4	11	12	8	9	8	9
± 5	1	1	17	19	29	32	11	12
$> 5\ and < -5$	5	6	16	17	15	16	19	21

From Table 4.66, we found that the MLP method gave the high precision record is 82 records or 89 % of 92 records that brought them to predict the alum dosage and it is more than another method. Thus, for the model group 1, the MLP is higher precise and credible than another method.

The RMSE of those methods indicated in **Table 4.67** as below:

Table 4.67 *The RMSE and MAE of 4 method from model group 1 of the real applications*

Methods	**RMSE**	**MAE**
MLP	**1.849**	**1.313**
M5Rules	8.440	3.249
M5P	7.165	3.116
REPTree	8.048	2.981

Of course, when we saw the Figure 4.13, the MLP curve is very nearly the actual alum than another method. Thus, in Table 4.46, the RMSE of 1.849 and MAE of 1.313 of MLP method is less than another method too. Therefore, the MLP method from model group 1 has the highest precision than another method when we used them to predict the alum dosage in the real application.

2) Real application of model group 2

The results indicated in **Figure 4.14** as below:

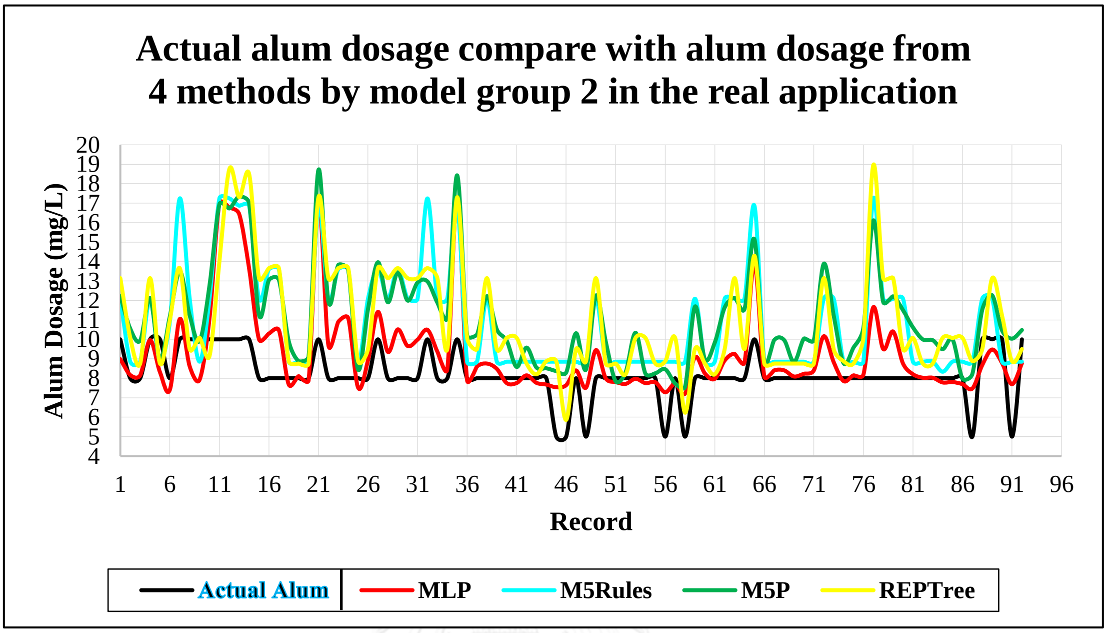


Figure 4.14 *Graph of predictive alum dosage of 4 methods from model group 2 in the real applications of DWTP*

From Figure 4.14, we found that the red curve (MLP) is nearly the black curve (Actual alum) than another method. For this reason, it has higher precision than another method. The MLP has high precision at the 1st-6th, 32nd-76th, and 8th-92nd record. Besides that, the yellow curve (REPTree) is also nearly the black curve. When we carefully look at the curves, we saw the MLP is very nearly the black curve and some points are the same. Therefore, the MLP method of model group 2 gave the predictive alum dosage values are nearly or the same actual alum dose values that the predictive alum dosage results also indicated in **Table 4.68**.

Table 4.68 *The predictive alum dosage results of 4 methods from model group 2 in the real applications of DWTP*

No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree	No	Actual Alum (mg/L)	MLP	M5Rules	M5P	REPTree
1	10	9	12	12	13	47	8	8	9	10	10
2	8	8	9	11	10	48	5	8	9	8	9
3	8	8	9	10	9	49	8	9	12	12	13
4	10	10	12	12	13	50	8	8	9	10	9
5	10	8	9	9	9	51	8	8	9	8	9
6	8	7	11	11	11	52	8	8	9	8	8
7	10	11	17	13	14	53	8	8	9	10	10
8	10	9	12	11	10	54	8	8	9	8	10
9	10	8	9	10	10	55	8	8	9	8	9
10	10	11	12	13	9	56	5	7	9	8	9
11	10	17	17	17	14	57	8	8	9	8	10
12	10	17	17	17	19	58	5	7	9	8	6
13	10	16	17	17	17	59	8	9	12	12	10
14	10	14	17	17	19	60	8	8	9	9	9
15	8	10	12	11	13	61	8	8	9	10	8
16	8	10	14	13	14	62	8	9	12	12	10
17	8	10	14	13	14	63	8	9	12	12	13
18	8	8	9	10	9	64	8	9	12	12	10
19	8	8	9	9	9	65	10	14	17	15	14
20	8	8	9	9	9	66	8	8	9	9	9
21	10	17	17	19	17	67	8	8	9	10	9
22	8	10	12	12	13	68	8	8	9	10	9
23	8	11	14	14	14	69	8	8	9	9	9
24	8	11	14	14	14	70	8	8	9	10	9
25	8	8	9	8	9	71	8	8	9	10	9
26	8	9	12	12	10	72	8	10	12	14	13
27	10	11	14	14	14	73	8	9	12	11	10
28	8	9	12	12	13	74	8	8	9	9	9
29	8	11	14	13	14	75	8	8	9	10	9
30	8	10	12	12	13	76	8	8	9	11	10
31	8	10	12	13	13	77	8	12	17	16	19
32	10	10	17	13	14	78	8	10	12	12	13
33	8	9	12	12	13	79	8	10	12	12	13
34	8	8	12	11	10	80	8	9	12	11	10
35	10	17	17	18	17	81	8	8	9	11	10
36	8	8	9	10	10	82	8	8	9	10	9
37	8	9	9	10	10	83	8	8	9	10	9
38	8	9	12	12	13	84	8	8	8	9	10
39	8	8	9	11	10	85	8	8	9	10	10
40	8	8	9	10	10	86	8	8	9	8	10
41	8	8	9	9	10	87	5	7	9	8	9
42	8	8	9	10	9	88	10	9	12	12	10
43	8	8	9	9	8	89	10	9	12	12	13
44	8	8	9	9	9	90	10	9	9	10	11
45	5	8	9	8	9	91	5	8	9	10	9
46	5	8	9	8	6	92	10	9	9	10	10

Table 4.69 *The predictive alum dosage of 4 methods from model group 2 that count in range and percentage of DWTP*

Measurement	**MLP**		**M5Rules**		**M5P**		**REPTree**	
	Count in range	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**	**Count in range**	**Count in %**
± 2	77	84	49	53	44	48	55	60
± 3	7	8	1	1	18	20	4	4
± 5	3	3	27	30	20	21	22	24
$> 5\ and < -5$	5	5	15	16	10	11	11	12

From Table 4.69, we found that the MLP gave the high precision record of 77 or 84% of 92 records and it is more than another method. The M5P gave the high precision record of 44 or 48% of 92 records that it is less than another method, thus, it is low precise.

The RMSE value used to analyze the model precision. The RMSE results indicated in **Table 4.70** as below:

Table 4.70 *The RMSE and MAE of 4 methods from model group 2 in the real applications*

Methods	**RMSE**	**MAE**
MLP	**2.292**	**1.365**
M5Rules	6.472	2.844
M5P	2.262	2.918
REPTree	6.466	2.774

From Table 4.70, we found that the MLP gave the RMSE of 2.292 and MAE of 1.365 that it is less than another method. Therefore, the MLP method from model group 2 has the highest precision and credibility than another method.

When we compared the MLP method from model group 1 and 2, we found that the MLP method of model group 1 gave the less RMSE of 1.849 than RMSE of 2.292 of MLP method from model group 2. Finally, we can summarize the real applications of DWTP model, the MLP of model group 1 has the highest accuracy and credibility than MLP method from model group 2 and another method from group 1 and 2.

4.4 Alum dosage prediction in Bangkhen water treatment plant

Because we would like to know the precision of the best models of 2 plants i.e. Chinaimo water treatment plant (CWTP) and Dongmarkkaiy water treatment plant (DWTP) which model will give the highest precision when we use them to predict the alum dosage in another plant. Thus, we used them to predict the alum dosage in the Bangkhen water treatment plant. We used data on January to March 2015 for the alum dosage prediction.

We use M5Rules method from model group 1 of CWTP because it is the best method and the MLP method from model group 1 of DWTP because it is also the best method when we use them to predict the alum dose in the real applications. The results indicated in **Figure 4.15.**

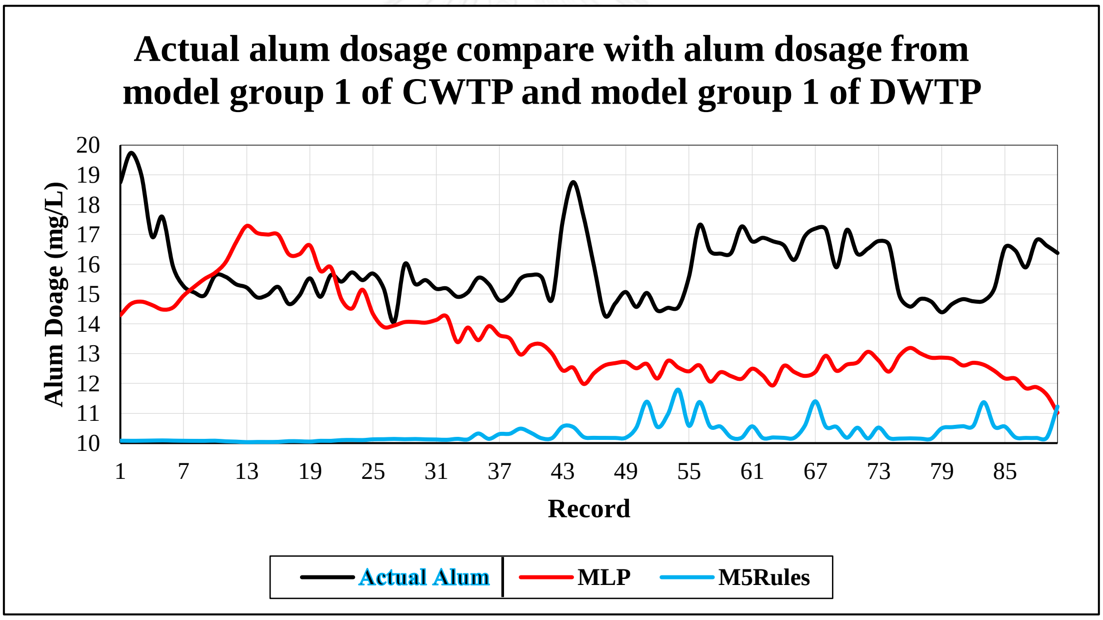


Figure 4.15 *Graph of predictive alum dosage results of 2 models with actual alum dosage in Bangkhen Water Treatment Plant*

From the Figure 4.15, we found that the MLP curve is nearly the actual alum curve than M5Rules curve at the record of 7th – 42nd, 47th – 54th, and 76th-83rd. For these cases, the predictive alum results can use in the real plant because they are in the $\pm 2\ mg/L$ (High precision). On the other hand, For the M5Rules don't have the precision when we saw the graph. Thus, the predictive alum dosage of MLP method is nearly the actual alum dosage than M5Rules method. However, the created model can't use to predict alum dosage in other plant because the natural of water quality characteristic is so different. The predictive alum dosage results indicated in Table 4.71 as below:

Table 4.71 *Predictive alum dosage results of 2 models*

Record	Actual Alum (mg/L)	MLP of Model group 1 of DWTP	M5Rules of Model group 1 of CWTP	Record	Actual Alum (mg/L)	MLP of Model group 1 of DWTP	M5Rules of Model group 1 of CWTP
1	19	14	10	46	16	12	10
2	20	15	10	47	14	13	10
3	19	15	10	48	15	13	10
4	17	15	10	49	15	13	10
5	18	14	10	50	15	13	11
6	16	15	10	51	15	13	11
7	15	15	10	52	14	12	11
8	15	15	10	53	15	13	11
9	15	16	10	54	15	13	12
10	16	16	10	55	16	12	11
11	16	16	10	56	17	13	11
12	15	17	10	57	16	12	11
13	15	17	10	58	16	12	11
14	15	17	10	59	16	12	10
15	15	17	10	60	17	12	10
16	15	17	10	61	17	12	11
17	15	16	10	62	17	12	10
18	15	16	10	63	17	12	10
19	16	17	10	64	17	13	10
20	15	16	10	65	16	12	10
21	16	16	10	66	17	12	11
22	15	15	10	67	17	12	11
23	16	15	10	68	17	13	11
24	15	15	10	69	16	12	11
25	16	14	10	70	17	13	10
26	15	14	10	71	16	13	11
27	14	14	10	72	17	13	10
28	16	14	10	73	17	13	11
29	15	14	10	74	17	12	10
30	15	14	10	75	15	13	10
31	15	14	10	76	15	13	10
32	15	14	10	77	15	13	10
33	15	13	10	78	15	13	10
34	15	14	10	79	14	13	10
35	16	13	10	80	15	13	11
36	15	14	10	81	15	13	11
37	15	14	10	82	15	13	11
38	15	14	10	83	15	13	11
39	16	13	10	84	15	12	11
40	16	13	10	85	17	12	11
41	16	13	10	86	16	12	10
42	15	13	10	87	16	12	10
43	17	12	11	88	17	12	10
44	19	13	11	89	17	12	10
45	18	12	10	90	16	11	11

From Table 4.71, we counted the precision record in term of range and percentage as indicated in **Table 4.72**.

Table 4.72 *Predictive alum dosage of 2 methods from 2 models 2 that count in range and percentage of Bangkhen water treatment plant*

	MLP of model group 1 of DWTP		**M5Rules of model group 1 of CWTP**	
Measurement	**Count in range**	**Count in %**	**Count in range**	**Count in %**
± 2	**54**	**60**	0	0
± 3	6	7	2	2
± 5	28	31	49	54
$> 5\ and < -5$	2	2	39	43

From Table 4.72, we found that the MLP method of model group 1 of DWTP gave more the high precision record of 54 or 60%. On the other hand, the M5Rules method of model group 1 of CWTP don’t have the high precision record, but it has more the low precision. Therefore, the MLP method of model group 1 of DWTP has the highest precision than M5Rules method of model group 1 of CWTP in the alum dosage prediction of Bangkhen water treatment plant as shown in the RMSE and MAE value in Table 4.73 as below:

Table 4.73 *RMSE and MAE of 2 models in the alum dosage prediction in Bangkhen water treatment plant*

Method	**RMSE**	**MAE**
MLP of Model group 1 of DWTP	**4.604**	**2.612**
M5Rules of Model group 1 of CWTP	15.773	5.492

From Table 4.73, the RMSE of 4.604 and MAE of 2.612 of MLP method by model group 1 of DWTP is less than M5Rules method of model group 1 of CWTP. Thus, the model group 1 of DWTP has the highest accuracy than model group 1 of CWTP in the alum dosage prediction in Bangkhen water treatment plant but it can’t not use in the real plant because both models aren’t fixed.

4.5 Application of the model that built from only turbidity

In this case, the researcher would like to compare the model built from three parameters i.e. turbidity, pH, and alkalinity with the model created from only turbidity for alum dosage prediction.

4.5.1 Model testingof Chinaimo Water Treatment Plant (CWTP)

4.5.1.1 Model group 1

The researcher used the model group 1 of CWTP for alum dosage prediction in the real application by using the original, fresh data and laboratory data.

1) Using original data

The 101 records (900th-1000th) were brought to predict the alum dosage. The result indicated in Figure 4.16 as below:

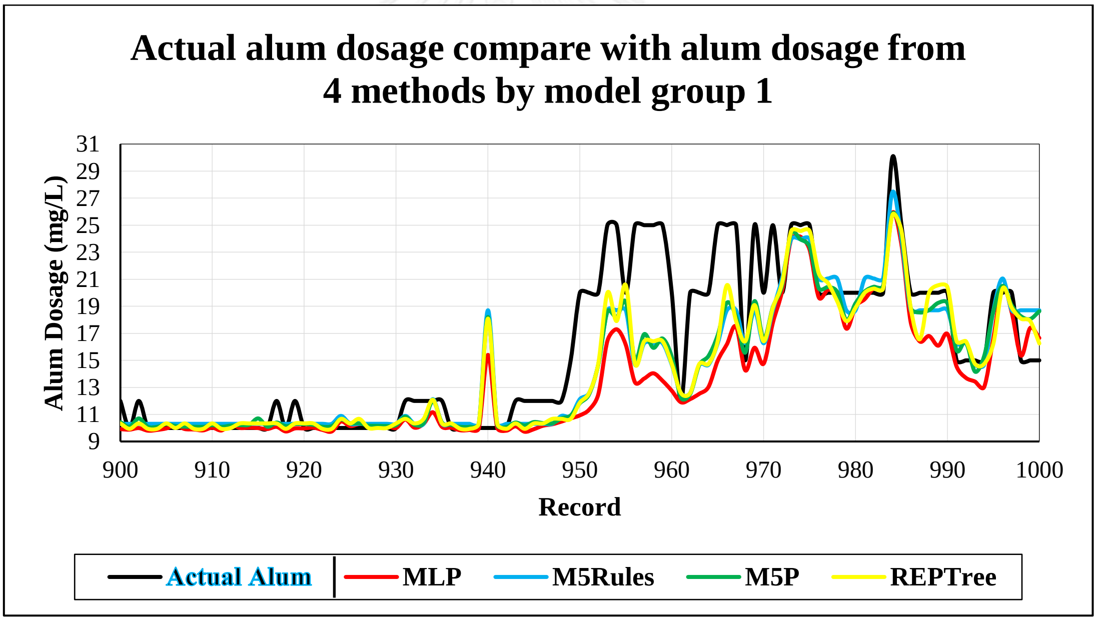


Figure 4.16 *Testing of model group 1 of CWTP that built from only turbidity variable by using original data*

From Figure 4.16, we found that 4 curves of 4 methods are nearly the actual alum curve from 900th-939th and 973rd-998th that mean the predictive alum dosage values are nearly the actual alum dosage. On the other hand, the curves are higher discrepancy values at 940th-973rd. From the graph above is difficult to separate the model accuracy. Therefore, we judged the model precision by using the RMSE value, the result shown in Table 4.74.

Table 4.74 *RMSE and MAE value of 4 methods by model group 1 of CWTP*

	MLP	**M5Rules**	**M5P**	**REPTree**
RMSE	8.376	5.770	8.459	**5.563**
MAE	2.539	2.190	2.539	**2.046**

Table 4.74 indicated that the REPTree method has the higher accuracy than another method because the RMSE and MAE are less than another method too in the model group 1 testing. However, the RMSE value of REPTree is quite high because the best RMSE value is ±2.

2) Using fresh data

The new data, we collected from November 2016 to January 2017 (92 records), the result indicated in Figure 4.17.

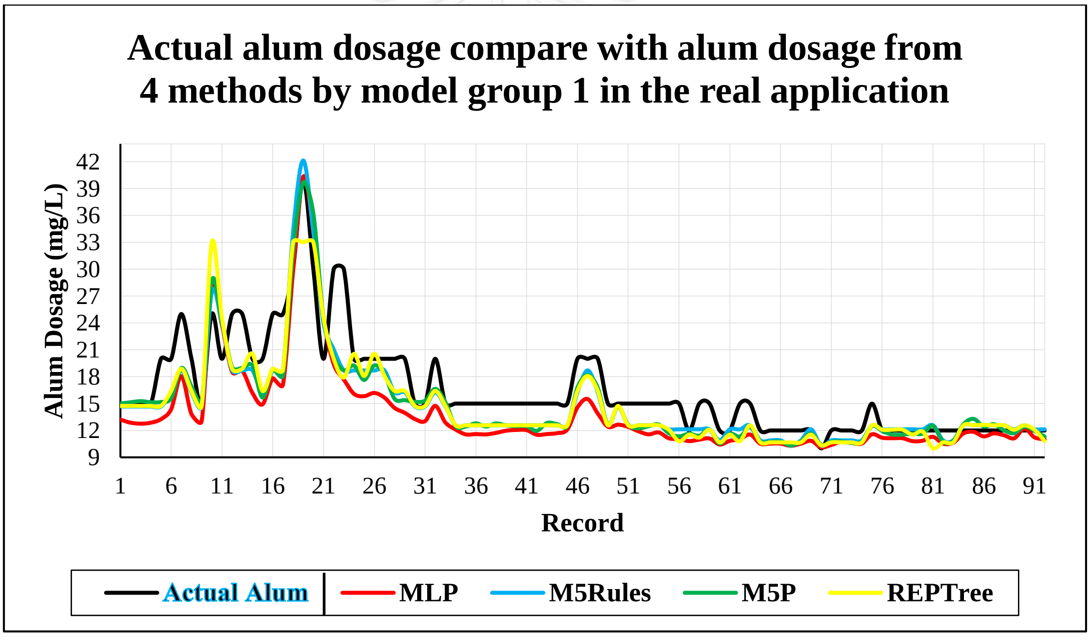


Figure 4.17 *Model group 1 testing that built from only turbidity variable by using fresh data of CWTP*

From Figure 4.17, we found that the 4 curves of 4 methods are nearly the actual alum curve at record of 30th - 92nd. In this case mean the predictive alum dosage values are nearly the actual alum dosage. On the other hand, the predictive alum dosage values are higher discrepancy at data record of 4th - 29th. From the graph is difficult to separate which method is higher accuracy. Therefore, we decided the model accuracy by using RMSE value that indicated in Table 4.75.

Table 4.75 *RMSE and MAE of 4 methods of model group 1 of CWTP in the real testing by using fresh data*

	MLP	**M5Rules**	**M5P**	**REPTree**
RMSE	7.166	**4.373**	4.727	5.206
MAE	3.046	**2.157**	2.247	2.348

Table 4.75 indicated that the M5Rules has the less RMSE and MAE value than another method. In this case, the M5Rules yielded the higher precision than another method in the real application by using fresh data.

3) Laboratory data

The researcher did some Jar-Test experiment for model precision testing. The researcher selected 72 records. The result indicated in Figure 4.18 as below:

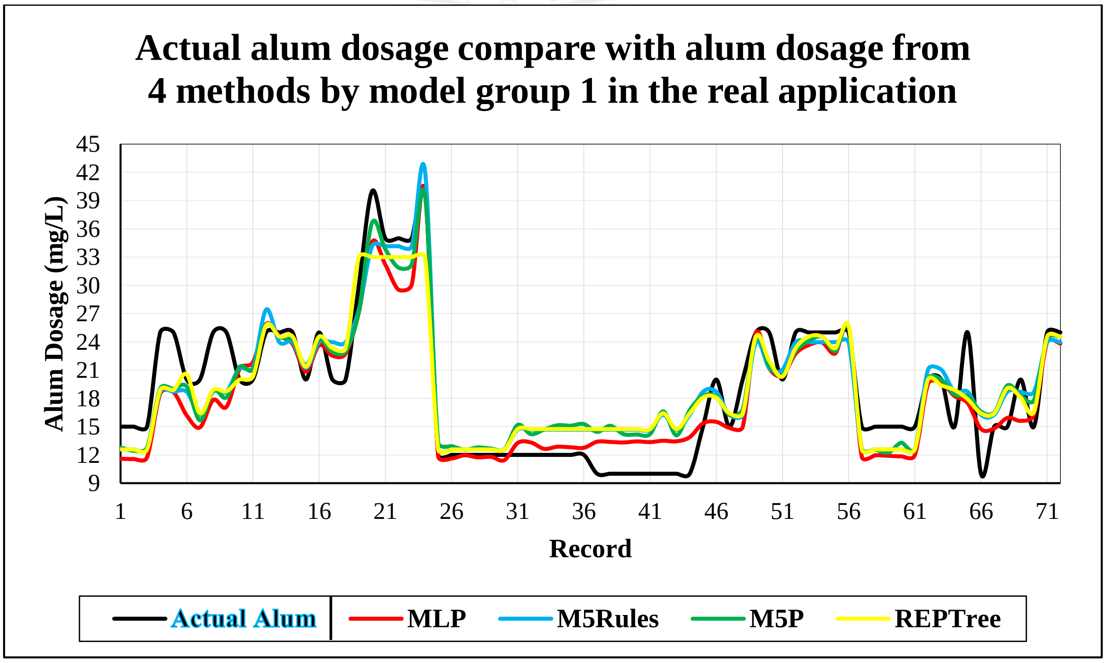


Figure 4.18 *Model group 1 testing of CWTP by using laboratory data*

From Figure 4.18, we found that the 4 curves of 4 methods are nearly the actual alum curve at data record of 8th – 36th and 45th – 64th. In this case mean the predictive alum dosage values are nearly the same actual alum. On the other hand, the predictive alum dosage values have discrepancy values at data record of 1st – 7th and 37th – 44th because the MLP, M5Rules, M5P, and REPTree curve aren't nearly the actual alum curve. From above graph is difficult to separate which method is higher precision. Thus, we judged the model accuracy by using the RMSE as shown in Table 4.76.

Table 4.76 *RMSE and MAE of 4 methods by model group 1 of CWTP for model testing by using laboratory data*

	MLP	M5Rules	M5P	REPTree
RMSE	**5.080**	5.197	**5.036**	5.524
MAE	**2.483**	2.658	**2.574**	2.628

Table 4.76 indicated that the MLP and M5P gave the less RMSE and MAE value than another method. Thus, the MLP and M5P method yielded the higher than another method.

4.5.1.2 Model group 2

1) Using original data

The result indicated in Figure 4.19 as below:

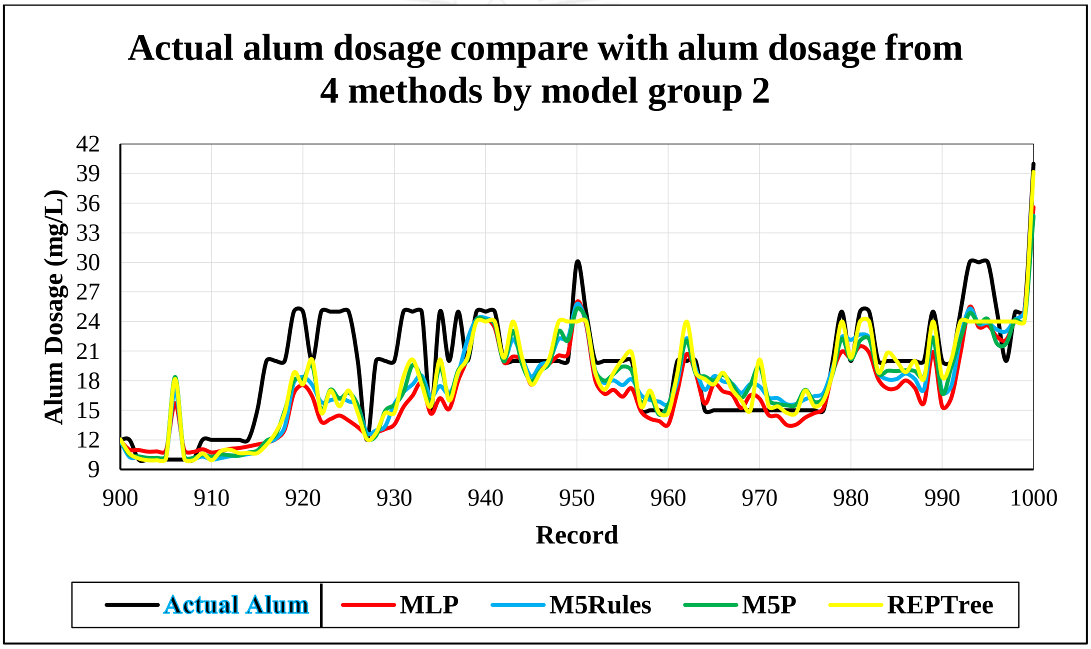


Figure 4.19 *Model group 2 testing of CWTP by using original data*

From Figure 4.19, we found that the predictive alum dosage values are nearly the actual alum dosage values at data record of 900th – 915th and 938th – 992nd. Thus, in this case, the model has the higher precision. On the other hand, at record of 916th to 937th, the model has the less accuracy because the curves aren't nearly the actual alum curve. We will know which method give the highest precision by using RMSE value to be the judge, the results indicated in Table 4.77.

Table 4.77 *RMSE and MAE value of 4 methods testing by model group 2 by using original data of CWTP*

	MLP	M5Rules	M5P	REPTree
RMSE	9.002	7.174	6.703	**6.680**
MAE	3.047	2.849	2.676	**2.522**

Table 4.77 indicated that the REPTree method of model group 2 has the less RMSE and MAE value than another method. Thus, it yielded the higher precision than another method in the model testing by using original data.

2) Using fresh data

The result shown in Figure 4.20 as below:

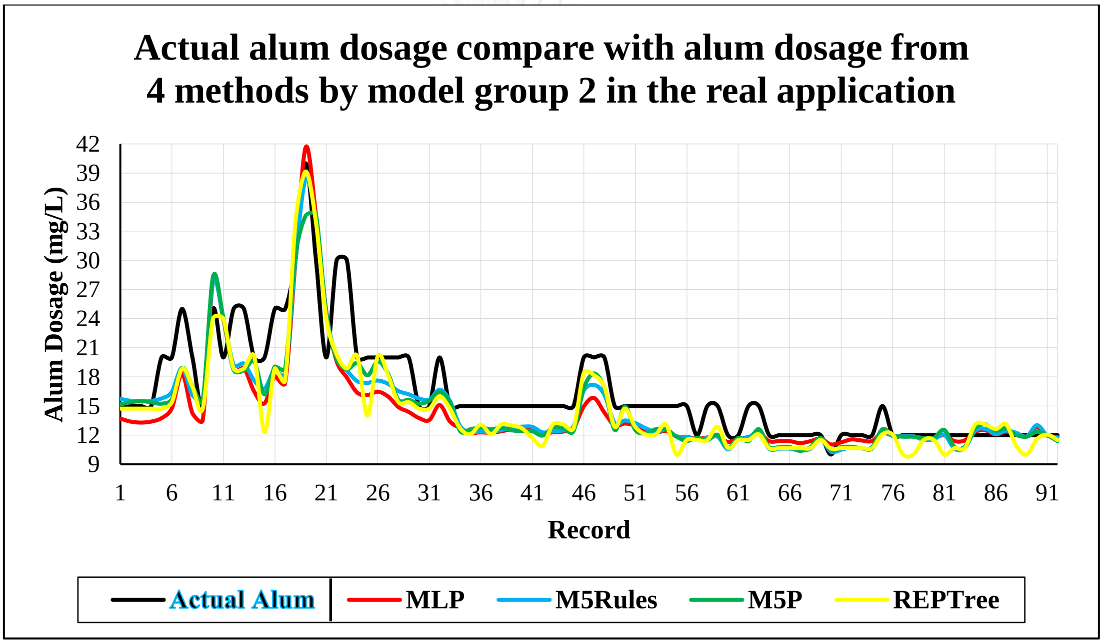


Figure 4.20 *Model group 2 testing by using original data of CWTP*

From Figure 4.20, we found that the 4 curves of 4 methods are nearly the actual alum dosage at data record of 30th – 92nd. Thus, the predictive alum dosage values in this case are also higher accuracy. On the other hand, the predictive alum dosage values are higher discrepancy at data record of 4th – 29th. We will judge which method is going to give the higher accuracy by using RMSE value as indicated in Table 4.78.

Table 4.78 *RMSE and MAE value of 4 methods testing by model group 2 using fresh data of CWTP*

	MLP	M5Rules	M5P	REPTree
RMSE	5.975	**4.523**	4.664	5.179
MAE	2.562	**2.257**	2.222	2.386

Table 4.78 indicated the M5Rules method of model group 2 has the less RMSE and MAE value than another method. Thus, it yielded the highest precision than another method in the real application using fresh data.

3) Using laboratory data

The result indicated in Figure 4.21 as shown below:

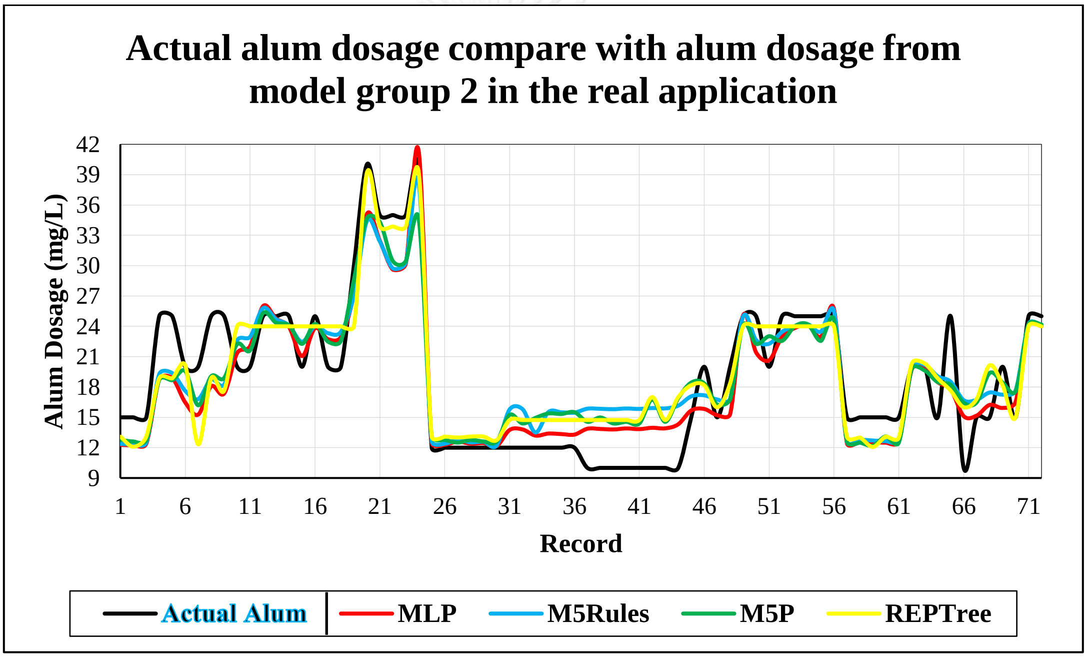


Figure 4.21 *Model group 2 testing by using laboratory data of CWTP*

From Figure 4.21, we found that 4 curves of 4 methods are nearly the actual alum curve at data record of 9th – 36th and 44th – 63rd. Thus, in this case, the predictive alum dosage values are nearly the actual alum dosage values. On the other hand, the predictive alum dosage values are discrepancy at data record of 1st – 8th and 37th – 43rd. When we look carefully at the graph, we are difficult to separate which method give the higher precision. Therefore, we judged the model accuracy by using RMSE value, the results indicated in Table 4.79.

***Table 4.79** RMSE and MAE value of 4 methods testing by model group 2 by using laboratory data of CWTP*

	MLP	M5Rules	M5P	REPTree
RMSE	**4.888**	6.096	5.640	5.937
MAE	**2.493**	2.896	2.777	2.754

Table 4.79 indicated that the MLP method yielded the highest accuracy than another method because it gave the less RMSE and MAE.

4.5.2 Model testing of Dongmarkkaiy Water Treatment Plant (DWTP)

4.5.2.1 Model group 1 of DWTP

The researcher used model group 1 of DWTP to predict alum dosage by using original and fresh data.

1) Using original data

We selected data record of 1000th – 1100th for model testing. The results indicated in Figure 4.22 as below:

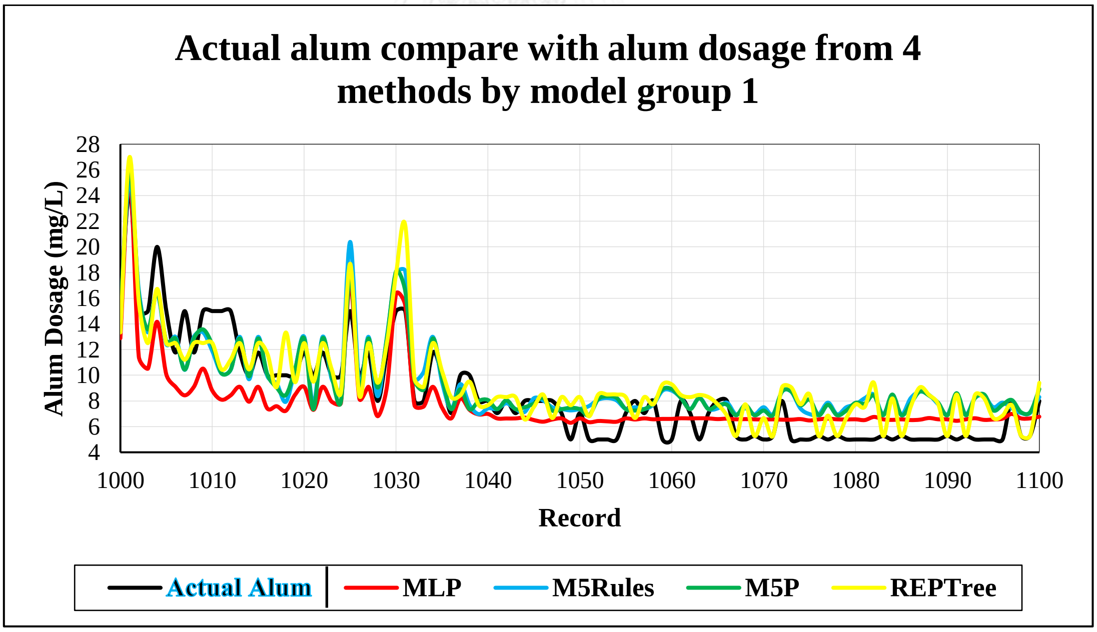


***Figure 4.22** Model group 1 of DWTP testing using original data*

From Figure 4.22, the M5Rules, M5P, and REPTree curve are nearly the actual alum curve than MLP curve. In this case, the predictive alum dosage values are also higher accuracy. Because the graph above is difficult to separate which method yield highest the precision. Therefore, we judged the model precision by using RMSE value as indicated in Table 4.80.

Table 4.80 *RMSE and MAE of 4 methods of model group 1 testing by using original data of DWTP*

	MLP	M5Rules	M5P	REPTree
RMSE	2.666	2.437	**2.278**	2.458
MAE	1.889	1.792	**1.723**	1.729

Table 4.80 indicated that the M5P method yielded the highest precision than another method because it has less the RMSE and MAE value than another method too. Therefore, in the model group 1 testing using original data, the M5P method has the highest accuracy.

2) Using fresh data

We collected the new data from November 2016 to January 2107 (92 records) for model testing. The results shown in Figure 4.23 as below:

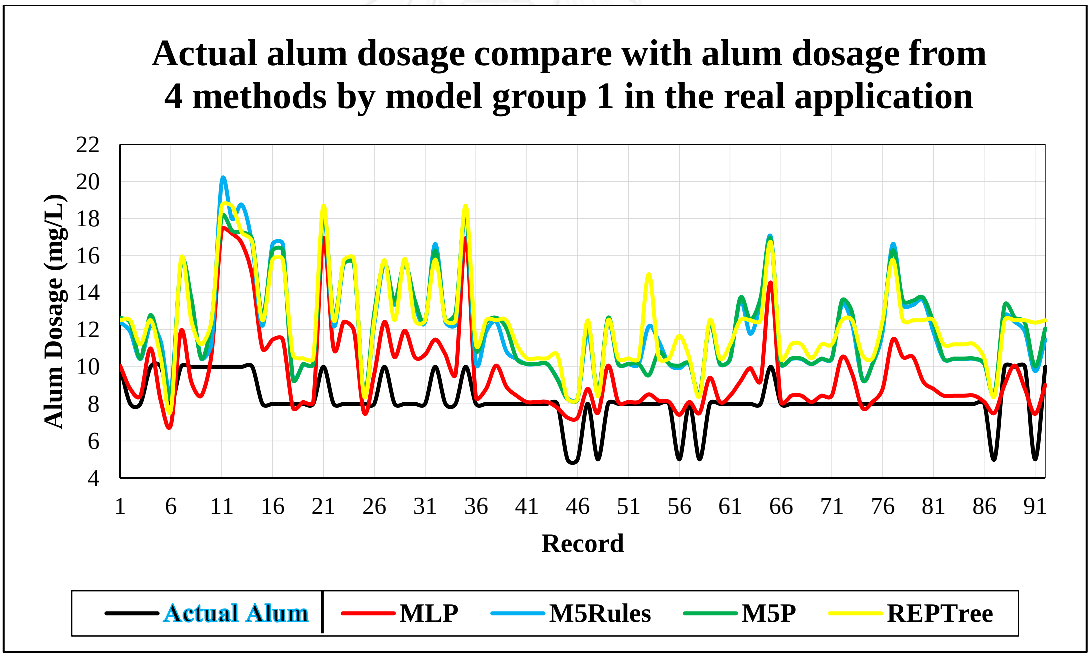


Figure 4.23 *Model group 1 testing by using fresh data of DWTP*

From Figure 4.23, we found that the MLP curve is nearly the actual alum than another method. Thus, the predictive alum dosage values from MLP method are higher accuracy than another method too. For M5Rules, M5P, and REPTree curve aren't nearly the actual alum curve. We judged the model precision by using RMSE value as indicated in Table 4.81.

Table 4.81 *RMSE and MAE of 4 methods of model group 1 testing by using fresh data from DWTP*

	MLP	M5Rules	M5P	REPTree
RMSE	**2.935**	9.878	9.933	10.402
MAE	**1.668**	3.874	3.943	4.110

Table 4.81 indicated the MLP method yielded the highest precision and credibility than another method because it has less RMSE and MAE value. *Therefore, the MLP method of model group 1 of DWTP is accuracy than another method for alum dosage prediction in the real application.*

4.5.2.2 Model group 2 of DWTP

1) Using original data

The researcher selected the data record of 1000th – 1100th for model testing. The results indicated in Figure 4.24 as below:

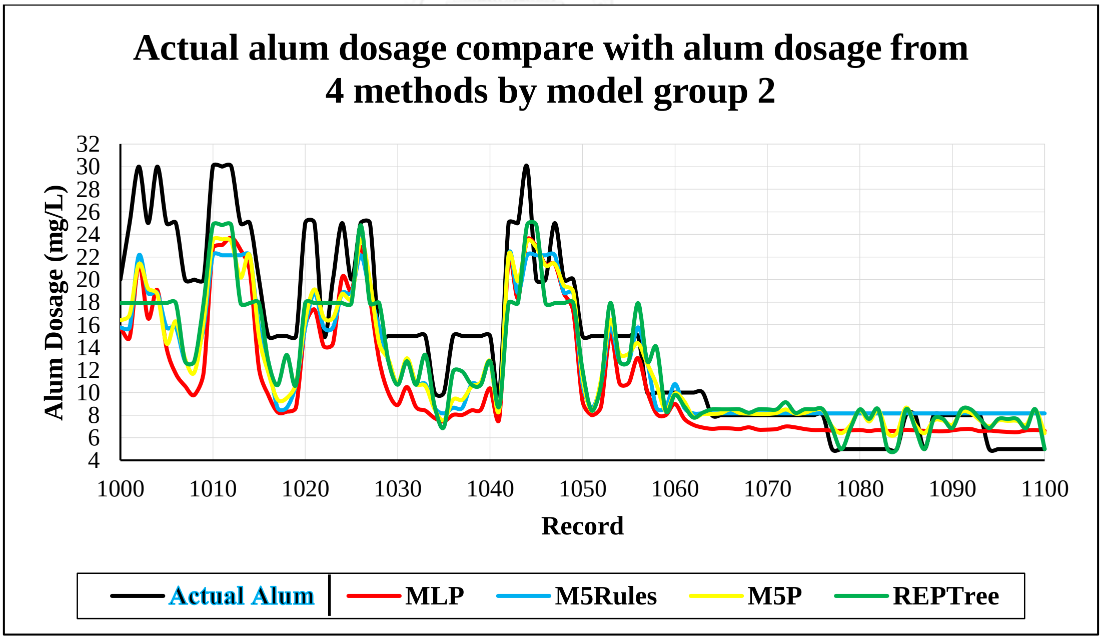


Figure 4.24 *Model group 2 testing by using original of DWTP*

From Figure 4.24, we found that the M5Rules, M5P, and REPTree curve are nearly the actual alum curve than MLP curve. Thus, the predictive alum dosage values from those 3 methods (M5Rules, M5P, and REPTree) are nearly the actual alum dosage values too. When we look at the graph is difficult to separate which method give the highest precision. Therefore, we judged the model precision by using RMSE value as shown in Table 4.82.

Table 4.82 *RMSE and MAE value of 4 methods of model group 2 testing by using original data of DWTP*

	MLP	**M5Rules**	**M5P**	**REPTree**
RMSE	11.715	8.719	**7.785**	8.023
MAE	3.814	3.204	**2.989**	3.039

Table 4.82 indicated that the M5P method yielded the highest precision for alum dosage prediction by using original data because it has the less RMSE and MAE value than another method.

2) ***Using fresh data***

We selected 92 records (Collected from November 2016 to January 2017) for model group 2 testing. The results shown in Figure 4.25 as below:

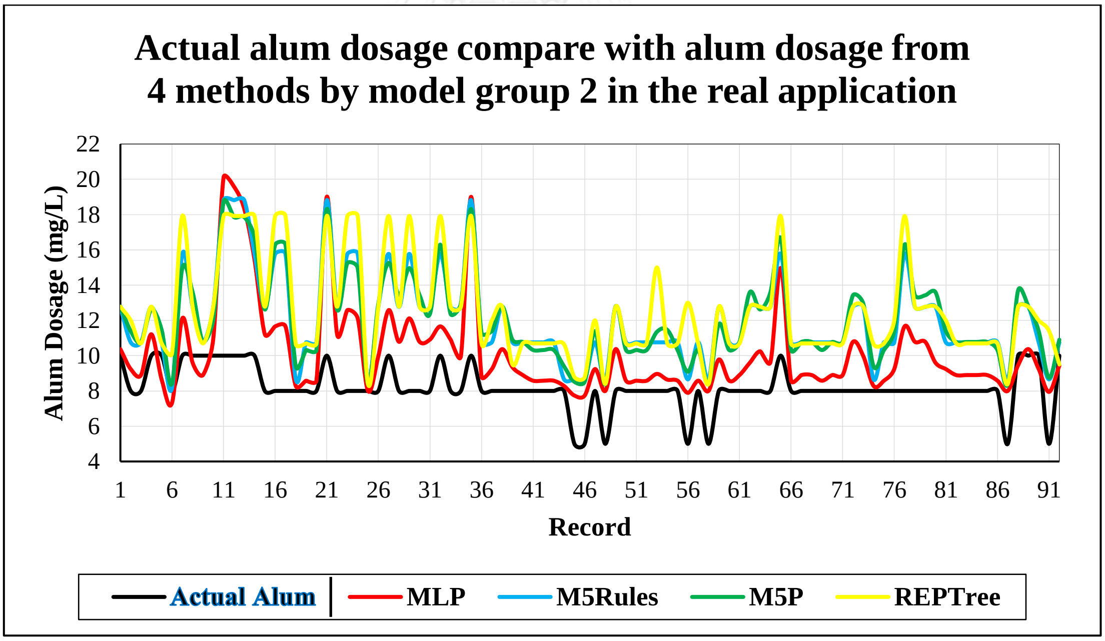


Figure 4.25 *Model group 2 testing by using fresh data of DWTP*

From Figure 4.25, we found that the MLP curve is nearly the actual alum curve than another method. Thus, it has higher precision than another method too. For the model precision judgement, we held the RMSE value as shown in Table 4.83.

Table 4.83 *RMSE and MAE value of 4 methods of model group 2 testing by using fresh data of DWTP*

	MLP	M5Rules	M5P	REPTree
RMSE	**4.328**	9.550	9.627	12.124
MAE	**2.069**	3.852	3.902	4.311

Table 4.83 indicated that the MLP method yielded the highest precision than another method because it has the less RMSE value. Therefore, the MLP method of model group 2 of DWTP has the highest accuracy and credibility for alum dosage prediction than another method in the real application.

4.6 Comparison between models built from three parameters and models built from only one parameter

The researcher built the models from three parameters i.e. turbidity, pH, and alkalinity and the models completely built from only one parameter as turbidity. In this case, we compared those models together about its application for alum dosage prediction in the real plants by using original, fresh, and laboratory data for model precision testing.

4.6.1 Chinaimo Water Treatment Plant (CWTP)

For model testing, we selected original, fresh, and laboratory data for alum dosage prediction. In this case, we compared the models together by using RMSE and MAE value to judge the model precision. The results indicated in Table 4.84 as below:

Table 4.84 *Models testing comparison by using original data of CWTP*

	Models testing by using original data															
	Models built from three parameters								**Models built from only one parameter**							
	Model Group 1				**Model Group 2**				**Model Group 1**				**Model Group 2**			
	MLP	**M5Rules**	**M5P**	**REPTree**	**MLP**	**M5Rules**	**M5P**	**REPTree**	**MLP**	**M5Rules**	**M5P**	**REPTree**	**MLP**	**M5Rules**	**M5P**	**REPTree**
RMSE	7.004	**3.157**	4.127	4.998	8.384	7.056	**5.421**	5.437	8.376	5.770	8.459	**5.563**	9.002	7.174	6.703	**6.680**
MAE	2.390	**1.545**	1.854	1.933	2.967	2.615	**2.480**	2.550	2.539	2.190	2.539	**2.046**	3.047	2.849	2.676	**2.522**

Table 4.84 indicated that the M5Rules of model group 1 that it built from three parameters has the less RMSE and MAE than another method and models built from only one parameter. Therefore, it yielded the highest precision than another model for models testing by using original data.

For models testing by using fresh data is indicated in Table 4.85 as below:

Table 4.85 *Models testing comparison by using fresh data of CWTP*

	Models testing by using fresh data															
	Models built from three parameters								Models built from three parameters							
	Model Group 1				Model Group 2				Model Group 1				Model Group 2			
	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree
RMSE	5.554	**4.043**	4.650	7.438	5.088	4.968	**4.693**	6.106	7.166	**4.373**	4.727	5.206	5.975	**4.523**	4.664	5.179
MAE	2.633	**2.240**	2.326	3.041	2.369	2.353	**2.325**	2.784	3.046	**2.157**	2.247	2.348	2.562	**2.257**	2.222	2.386

Table 4.85 indicated that the M5Rules of model group 1 that it built from three parameters has the less RMSE than another method. In this case, it yielded the highest precision than another method and 4 methods of the models that built from only one parameter. Therefore, we summarized the M5Rules method by model group 1 that it built from three parameters has the highest accuracy and credibility than another method for alum dosage prediction in the real application.

On the other hand, we tested the model precision of CWTP by using laboratory data. The results shown in Table 4.86. From Table 4.86, we found that the M5Rules method of model group 1 that it built from three parameters has the less RMSE and MAE value than another method and 4 methods of models that built from only one parameter.

Therefore, we finally summarized the M5Rules by model group 1 that it built from three parameters yielded the highest precision and credibility than another method because when we used it to predict the alum dosage by using new data, it had the less RMSE than another method too.

***Table 4.86** Models testing comparison by using laboratory data of CWTP*

	Models testing by using laboratory data															
	Models built from three parameters								Models built from three parameters							
	Model Group 1				Model Group 2				Model Group 1				Model Group 2			
	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree
RMSE	4.415	**3.159**	3.438	4.560	**4.621**	6.575	**4.621**	4.731	5.080	5.197	**5.036**	5.524	**4.888**	6.096	5.640	5.937
MAE	2.462	**1.916**	2.067	2.209	**2.533**	2.590	**2.533**	2.340	2.483	2.658	**2.574**	2.628	**2.493**	2.896	2.777	2.754

4.6.2 Dongmarkkaiy Water Treatment Plant (DWTP)

For models testing in this plant, we selected original and fresh data. The results indicated in Table 4.87 and 4.88 as below:

***Table 4.87** Models testing comparison by using original data of DWTP*

	Models testing by using original data															
	Models built from three parameters								Models built from three parameters							
	Model Group 1				Model Group 2				Model Group 1				Model Group 2			
	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree
RMSE	2.698	**0.900**	1.418	2.026	6.318	5.073	3.026	**2.979**	2.666	2.437	**2.278**	2.458	11.715	8.719	7.785	**8.023**
MAE	1.821	**0.920**	1.225	1.346	2.652	2.456	1.889	**1.603**	1.889	1.792	**1.723**	1.729	3.814	3.204	2.989	**3.039**

Table 4.87 indicated that the M5Rules method by model group 1 that it created from three parameters has the less RMSE and MAE than another method and 4 methods of models that built from only one parameter. In this case, it yielded the highest accuracy for alum dosage prediction by using original data.

For models testing by using fresh data indicated in Table 4.88 as below:

Table 4.88 *Models testing comparison by using fresh data of DWTP*

	Models testing by using fresh data															
	Models built from three parameters								Models built from three parameters							
	Model Group 1				Model Group 2				Model Group 1				Model Group 2			
	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree	MLP	M5Rules	M5P	REPTree
RMSE	**1.849**	8.440	7.165	8.048	**2.262**	6.472	6.262	6.466	**2.935**	9.878	9.933	10.402	**4.328**	9.550	9.627	12.124
MAE	**1.313**	3.249	3.116	2.982	**1.365**	2.844	2.918	2.774	**1.668**	3.875	3.943	3.110	**2.069**	3.852	3.902	4.311

Table 4.88 indicated that the MLP method by model group 1 that it built from three parameters has the less RMSE and MAE value than another method and 4 methods of models that built from only one parameter. In this case, it yielded the highest accuracy and credibility than another method too.

Finally, we summarized the MLP method of model group 1 that it created from three parameters yielded the highest precision than another method and 4 methods of models that created from only one parameter. Therefore, we held the MLP method of model group 1 for alum dosage prediction in the real plant of DWTP.

CHAPTER V
CONCLUSION

5.1 5D World Map (5DWM) System

The physical and chemical characteristic of water quality was displayed by 5DWM in term of green color that this color doesn’t have a definition, only visualization of water sampling point. On the other hand, we found that the nitrate of 6 sampling points is over the Lao’s surface water quality standard. Besides that, the potassium of 1st sampling (Mekong River) and 6th (Nam Lik River) is also over the standard. Because we found the nitrate and potassium are over the standard due to the upstream area of those sampling points has a lot of the chemical fertilizers using for the agricultural. However, the water quality in 3 rivers is still good as shown in the heavy metals aren’t over the Lao’s surface water quality standard.

In 1st sampling point (Mekong River), we analyzed the water quality because at this point is the location of the Chinaimo Water Treatment Plant (CWTP) and 2nd sampling point is the location of the Dongmarkkaiy Water Treatment Plant (DWTP). After we analyzed the water quality of both sampling points, the water quality is good condition. Thus, both water treatment plants have the good water supply quality.

5.2 Alum dosage prediction using Weka data mining software

The model building for alum dosage prediction by using 4 methods i.e. Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree in the Weka data mining software with 2 data groups. The *first data group*, we substituted all missing values of each parameter by the average value of that parameter, computed by each month. The *second data group*, we cut off the missing value to reduce bias. We can summarize the results as below:

5.2.1 Chinaimo Water Treatment Plant (CWTP)

The model building for alum dosage prediction of this plant, we have gotten the M5Rules method gave the less RMSE of 0.957 in the drying season and RMSE of 3.197

in the rainy season than another method that it built from the model group 1. This method built from the *Smoothed Linear model* that adjusted from the *buildRegressionTree is "FALSE", debug is "FALSE", 4-miniNumInstances, unpruned is "TRUE", and useUnsmoothed is "FALSE"*. When we used M5Rules method to predict the alum dosage in the real application, it gave the less RMSE of 4.043 than another method. Therefore, we would like to use the model to predict the alum dosage in the water treatment plant, we should choose the model that substituted all missing values of each parameter by the average values of that parameter and use M5Rules method that adjusts the model give the less RMSE.

5.2.2 Dongmarkkaiy water treatment plant (DWTP)

For the model building for the alum dosage prediction of this plant, we have gotten the REPTree method gave the less RMSE of 1.584 (Drying season) and RMSE of 5.603 (Rainy season) than another method that it built from the model group 1. This method adjusted from the *debug is "FALSE", -1-maxDepth, 0.2-minNum, 5-numFolds, 0.001-minVarianceProp, noPruning is "FALSE", and 2-seed*. On the other hand, when we used in the real applications, we found that the REPTree gave the RMSE of 8.048 that it is higher than the RMSE of 1.849 of MLP method that the MLP method successfully adjusted from the *4-hiddenLayer, 0.3-learningRate, 0.2-momentum, seed of 3, 20-validationThreshold, and 3000-trainingTime.* Thus, we will decide the model precision from only the real applications. In this case, if we would like to use the model to predict the alum dosage in the real water treatment plant, we should choose the model group 1 by using the MLP method that adjusts the model give the less RMSE.

5.2.3 Using the best methods for the alum dosage prediction in another plant

We brought the best method of model group 1 of the Chinaimo and Dongmarkkaiy water treatment plant. The results indicated that the Multilayer Perceptron of model group 1 of DWTP had the higher precision than the M5Rules method of model group 1 of CWTP because the MLP gave the RMSE of 4.604, it is less than the RMSE of 15.773 of M5Rules or about 4 times. Therefore, the MLP has the higher precision than M5Rules for the alum dosage prediction at the BangKhen water treatment plant.

Besides that, we can summarize the RMSE value that we got from the model building didn’t effect to the model precision. The model precision measurement will judge from the real applications that we find the RMSE from the predictive alum dosage, we will know the model can use in the real water treatment plant or not.

The created model can’t use in other plant that use different river to be the raw water source because the natural of water quality characteristic of each river is so different. Therefore, the model building must to create from the data of water quality characteristic of that plant.

The models built from three parameters i.e. turbidity, pH, and alkalinity had the highest accuracy than the models built from only one parameter as turbidity. In this research, we found that the main important parameter is turbidity in the model building for alum dosage prediction.

Moreover, we summarized the model will give the high precision in the drying season than rainy season because, in the drying season, all parameters are quite stable such as turbidity and this case, we predicted the alum dosage by using the data in drying season that the predictive alum dosage values are very nearly the actual alum dosage values.

Finally, the model building by Weka Data Mining Software has limitation because it is completely software, some functions can’t adjust by user, some functions haven’t effect to the model building that mean they can’t not improve the model to be good, and we can’t set the new functions in this software.

5.3 Recommendation

If have the researchers would like to study on this topic in the future, we have some recommendations as below:

- 5DWM is a good system in the water quality visualization but it continues development such as giving the meaning of the color.
- In the future, the water treatment plant should apply the automatic alum sensor.
- The model building is only used the Multilayer Perceptron (MLP), M5Rules, M5P, and REPTree. Another method choosing for the model building will increase the opportunity to look for the higher precision of the model.
- Enhance alum dosage prediction by using another software i.e. Scikit-learn, Python, etc. because the Weka data mining software has the limitations because it is completed software that we can’t create the new functions in the software.

REFERENCES

Amirtharajah, A., and Mills, K. M. (1982). Rapid-mix design for mechanisms of alum coagulation. *Journal (American Water Works Association)*, 210-216.

Bae, H., Choi, D.-W., Lee, S.-T., Kim, Y., and Kim, S. (2004). *Application of data mining and artificial modeling for coagulant dosage of water treatment plants corresponding to water quality*. Paper presented at the Industrial Electronics Society, 2004. IECON 2004. 30th Annual Conference of IEEE.

Bagheri, M., Mirbagheri, S., Ehteshami, M., and Bagheri, Z. (2015). Modeling of a sequencing batch reactor treating municipal wastewater using multi-layer perceptron and radial basis function artificial neural networks. *Process Safety and Environmental Protection, 93*, 111-123.

Barker, M., Nickels, M., and Mayfield, H. (2003). Optimizing Drinking Water Treatment Using Neural Networks.

Boonnao, N., Taychamekiatchai, R., and Chawakitchareon, P. (2015). *Prediction of Alum Dosage using in Water Supply Production Process by Weka 3 program.* (Master), Chulalongkorn University.

Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., and Scuse, D. (2016). *WEKA Manual for Version 3-6-14*. University of Waikato, Hamilton, New Zealand.

Charutragulchai, P. (2006). *Alum dosage using in water supply process prediction by decision Tree Forest Method and Genetic Programming.* (Master), Thammasart University.

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, 292*, 83-93.

Chun, M. G., Kwak, K. C., and Ryu, J. W. (1999). *Application of ANFIS for coagulant dosing process in a water purification plant.* Paper presented at the Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE'99. 1999 IEEE International.

Crittenden, J. C., Trussell, R. R., Hand, D. W., Howe, K. J., and Tchobanoglous, G. (2012). *MWH's water treatment: principles and design*: John Wiley & Sons.

Daphne, L. H. X., Utomo, H. D., and Kenneth, L. Z. H. (2011). Correlation between turbidity and total suspended solids in Singapore rivers. *Journal of Water Sustainability, 1*(3), 313-322.

Du, H. (2010). *Data mining techniques and applications: an introduction*: Cengage Learning.

Engelhardt, T. L. (2010). Coagulation, flocculation and clarification of drinking water. *Drinking water sector, Hach Company*.

Gagnon, C., Grandjean, B. P., and Thibault, J. (1997). Modelling of coagulant dosage in a water treatment plant. *Artificial Intelligence in Engineering, 11*(4), 401-404.

García-Laencina, P. J., Sancho-Gómez, J.-L., and Figueiras-Vidal, A. R. (2013). Classifying patterns with missing values using Multi-Task Learning perceptrons. *Expert Systems with Applications, 40*(4), 1333-1341.

Hannouche, A., Chebbo, G., Ruban, G., Tassin, B., Lemaire, B., and Joannis, C. (2011). Relationship between turbidity and total suspended solids concentration within a combined sewer system. *Water Science and Technology, 64*(12), 2445-2452.

Hendricks, D. W. (2006). *Water treatment unit processes: physical and chemical*: CRC press.

Kalmegh, S. (2015). Analysis of WEKA data mining algorithm REPTree, Simple CART and RandomTree for classification of Indian news. *International Journal of Innovative Science, Engineering and Technology, 2*(2), 438-446.

Kawamura, S. (2000). *Integrated design and operation of water treatment facilities*: John Wiley & Sons.

Kiyoki, Y., Chen, X., Sasaki, S., and Koopipat, C. (2016). Multi-Dimensional Semantic Computing with Spatial-Temporal and Semantic Axes for Multi-spectrum Images in Environment Analysis. *Information Modelling and Knowledge Bases XXVII, 280*, 14.

Kiyoki, Y., Kitagawa, T., and Hayama, T. (1994). A metadatabase system for semantic image search by a mathematical model of meaning. *ACM Sigmod Record, 23*(4), 34-41.

Kiyoki, Y., Sasaki, S., Trang, N. N., and Diep, N. T. N. (2012). Cross-cultural multimedia computing with impression-based semantic spaces *Conceptual Modelling and Its Theoretical Foundations* (pp. 316-328): Springer.

Maier, H. R., Morgan, N., and Chow, C. W. (2004). Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. *Environmental Modelling & Software, 19*(5), 485-494.

Nahm, E.-S., Lee, S.-B., Woo, K.-B., Lee, B.-K., and Shin, S.-K. (1996). *Development of an optimum control software package for coagulant dosing process in water purification system.* Paper presented at the SICE'96. Proceedings of the 35th SICE Annual Conference. International Session Papers.

Pizzi, N. G. (2011). *Water treatment operator handbook*: American Water Works Association.

Quinlan, J. R. (1992). *Learning with continuous classes.* Paper presented at the 5th Australian joint conference on artificial intelligence.

Sasaki, S., Takahashi, Y., and Kiyoki, Y. (2010). The 4D World Map System with Semantic and Spatio-temporal Analyzers. *Information Modelling and Knowledge Bases, 21*, 1-18.

Scholz, M. (2015). *Wetlands for water pollution control*: Elsevier.

Stehlé, J., Barrat, A., and Bianconi, G. (2010). Dynamical and bursty interactions in social networks. *Physical review E, 81*(3), 035-101.

Valentin, N., Denoeux, T., and Fotoohi, F. (1999). Modelling of coagulant dosage in a water treatment plant. *Proceedings of EANN’99*, 165-170.

Veesommai, C., Kiyoki, Y., Sasaki, S., and Chawakitchareon, P. (2016). Wide-Area River-Water Quality Analysis and Visualization with 5D World Map System. *Information Modelling and Knowledge Bases XXVII, 280*, 31-41.

Wang, Y., and Witten, I. H. (1996). Induction of model trees for predicting continuous classes.

APPENDIX

Appendix A

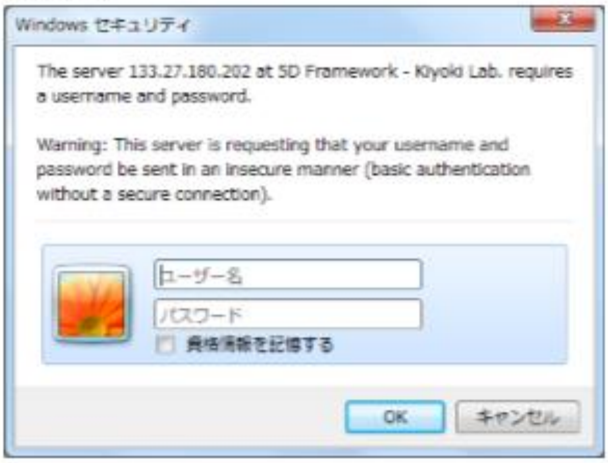
5D World Map (5DWM) System

Environmental Knowledgebase Creation with 5D World Map System

2013.6.9 Yasushi Kiyoki, Shiori Sasaki
(Last updated: 2014.06.17)

Access to 5D World Map System (Environmental Multimedia Data Sharing and Visualization System):
http://133.27.180.202/cake221/

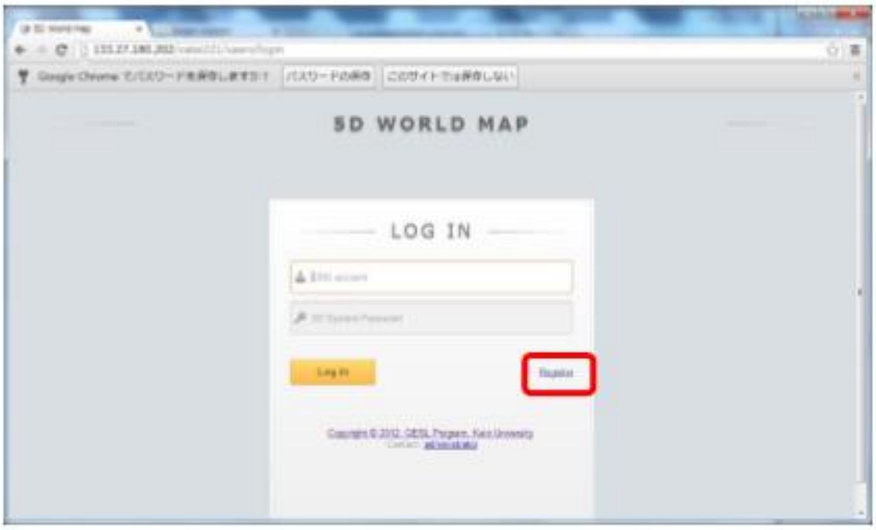
1. Login: Please use a Web browser such as Chrome, Firefox, Safari etc. except Internet Explorer. Please input User name: ******* and Password: ******* on this window. (******* are announced individually.)

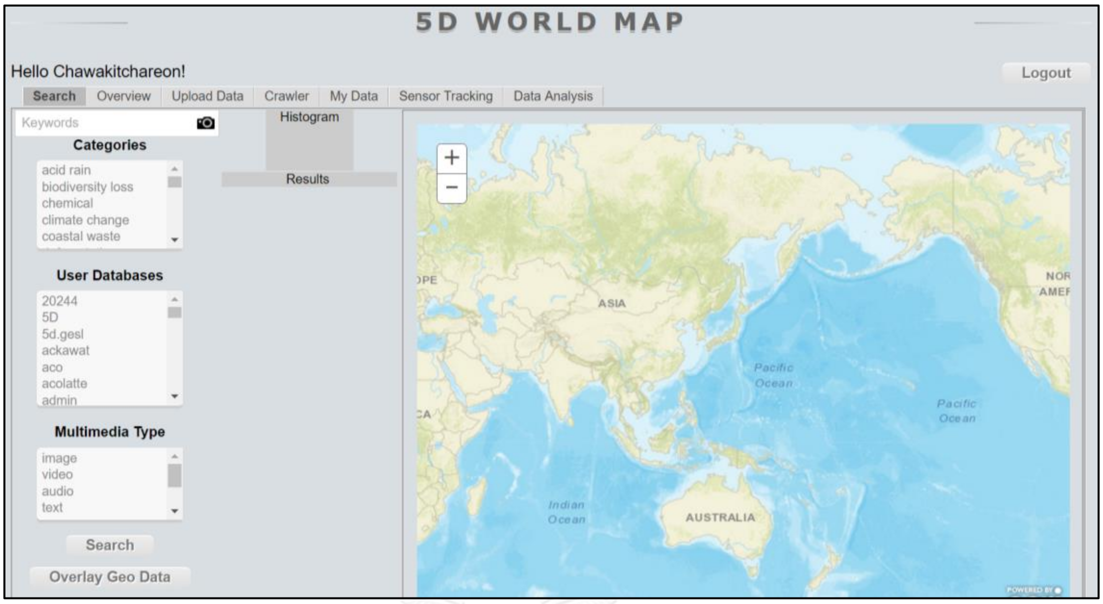


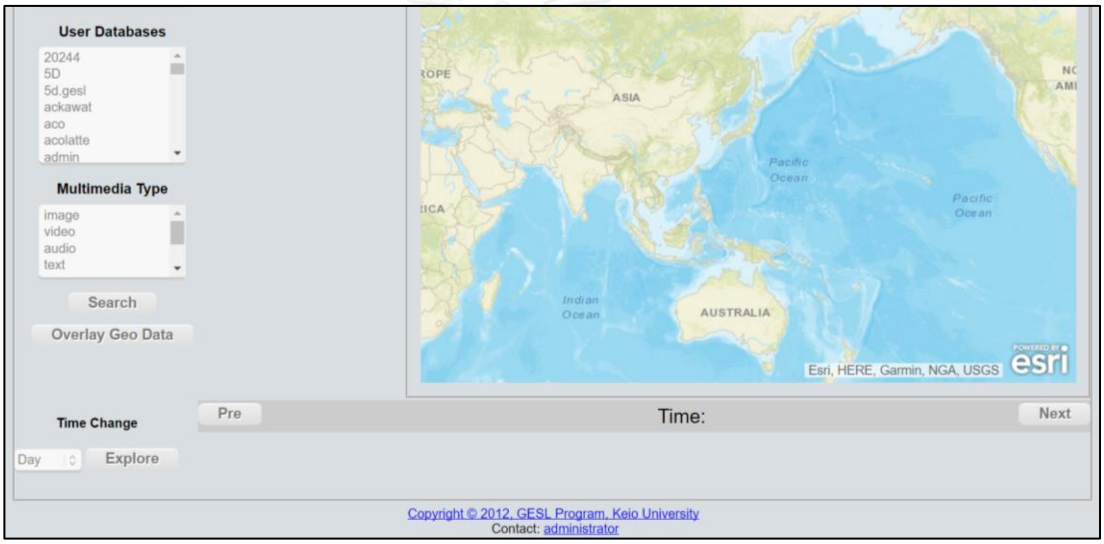
The server 133.27.180.202 at 5D Framework - Kiyoki Lab. requires a username and password.

Warning: This server is requesting that your username and password be sent in an insecure manner (basic authentication without a secure connection).

Register your ID and Password from "Register" button, and log in to the system.

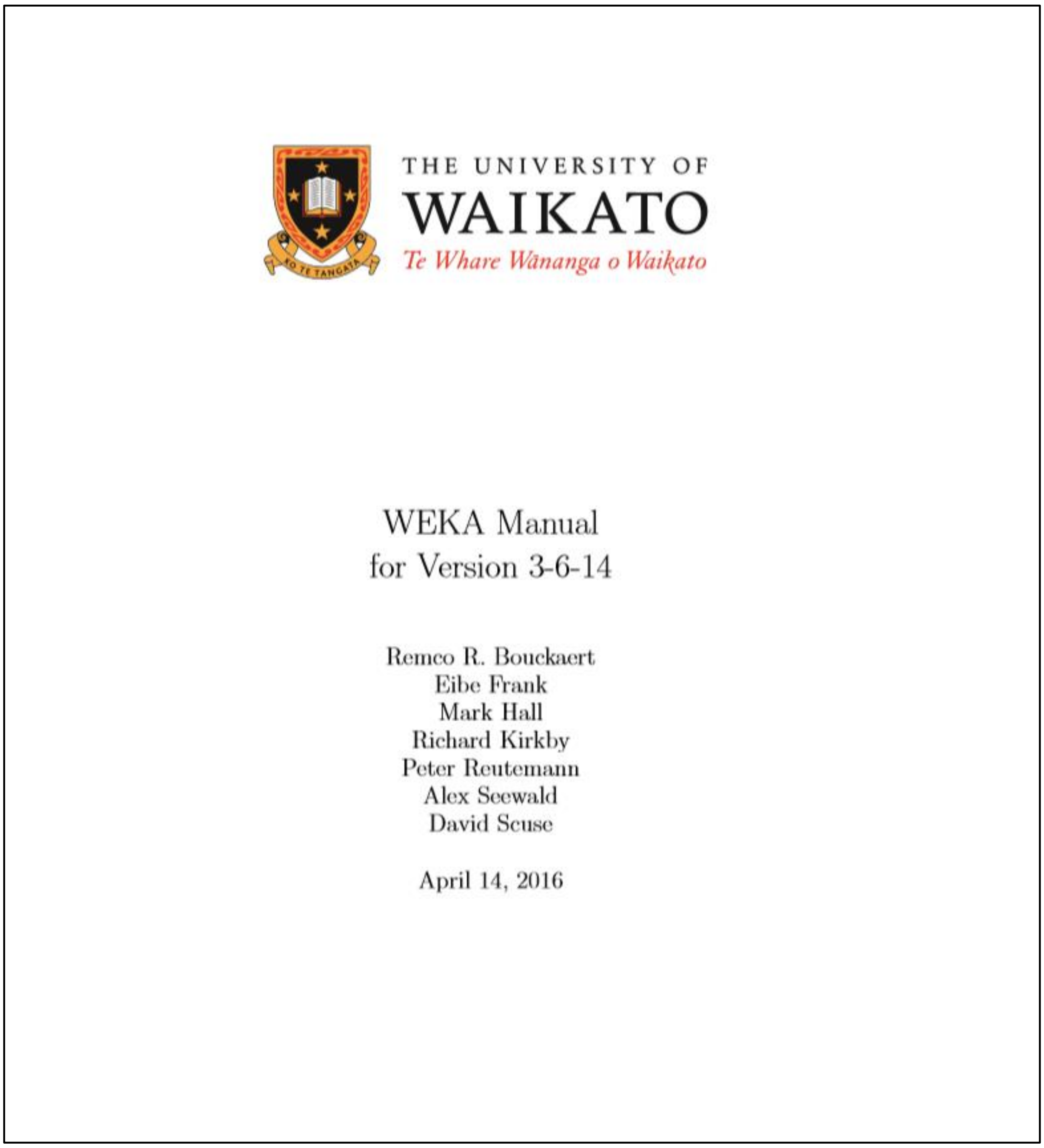



5D WORLD MAP
Hello Chawakitchareon!
Logout
Search
Overview
Upload Data
Crawler
My Data
Sensor Tracking
Data Analysis
Keywords
Histogram
Results
Categories
acid rain
biodiversity loss
chemical
climate change
coastal waste
User Databases
20244
5D
5d.gesl
ackawat
aco
acolatte
admin
Multimedia Type
image
video
audio
text
Search
Overlay Geo Data
ASIA
Pacific Ocean
Pacific Ocean
Indian Ocean
AUSTRALIA


User Databases
20244
5D
5d.gesl
ackawat
aco
acolatte
admin
Multimedia Type
image
video
audio
text
Search
Overlay Geo Data
ASIA
Pacific Ocean
Pacific Ocean
Indian Ocean
AUSTRALIA
Esri, HERE, Garmin, NGA, USGS
esri
Time Change
Pre
Time:
Next
Day
Explore
Copyright © 2012, GESL Program, Keio University
Contact: administrator

Appendix B

Weka Data Mining Software (Weka Manual)



THE UNIVERSITY OF WAIKATO

Te Whare Wānanga o Waikato

WEKA Manual
for Version 3-6-14

Remco R. Bouckaert
Eibe Frank
Mark Hall
Richard Kirkby
Peter Reutemann
Alex Seewald
David Scuse

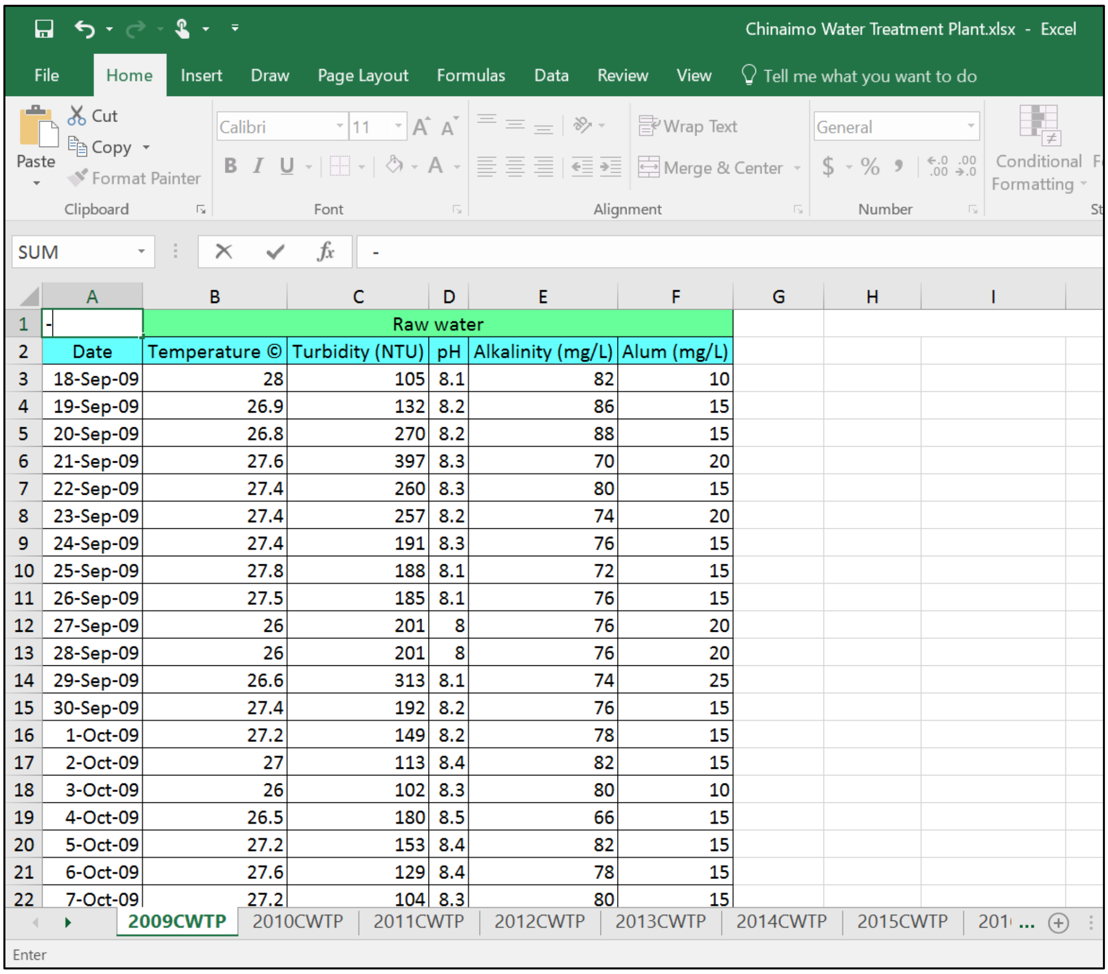
April 14, 2016

Appendix C

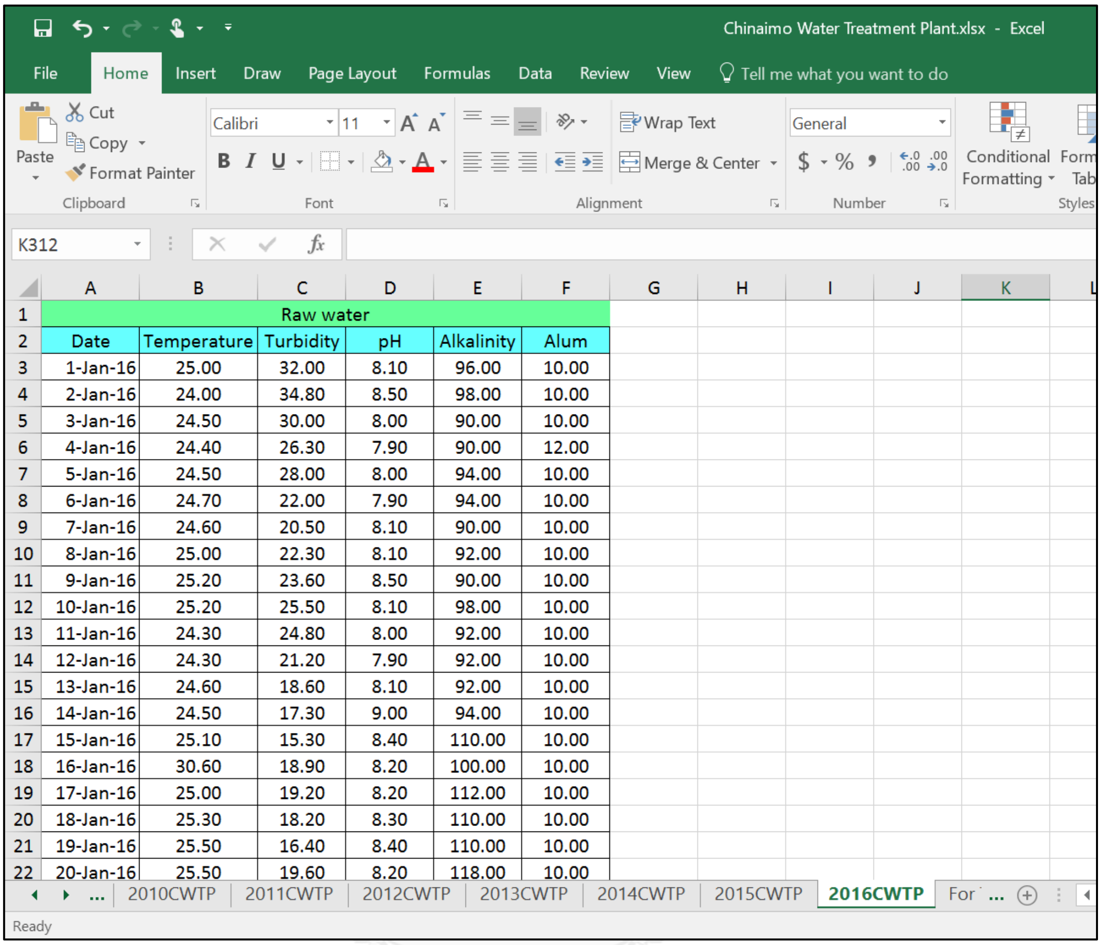
Data information from the Chinaimo Water Treatment Plant (CWTP)

Appendix C.1

Raw data obtained from CWTP 2009-2016 (2,038 records)



	A	B	C	D	E	F
1	-	Raw water				
2	Date	Temperature ©	Turbidity (NTU)	pH	Alkalinity (mg/L)	Alum (mg/L)
3	18-Sep-09	28	105	8.1	82	10
4	19-Sep-09	26.9	132	8.2	86	15
5	20-Sep-09	26.8	270	8.2	88	15
6	21-Sep-09	27.6	397	8.3	70	20
7	22-Sep-09	27.4	260	8.3	80	15
8	23-Sep-09	27.4	257	8.2	74	20
9	24-Sep-09	27.4	191	8.3	76	15
10	25-Sep-09	27.8	188	8.1	72	15
11	26-Sep-09	27.5	185	8.1	76	15
12	27-Sep-09	26	201	8	76	20
13	28-Sep-09	26	201	8	76	20
14	29-Sep-09	26.6	313	8.1	74	25
15	30-Sep-09	27.4	192	8.2	76	15
16	1-Oct-09	27.2	149	8.2	78	15
17	2-Oct-09	27	113	8.4	82	15
18	3-Oct-09	26	102	8.3	80	10
19	4-Oct-09	26.5	180	8.5	66	15
20	5-Oct-09	27.2	153	8.4	82	15
21	6-Oct-09	27.6	129	8.4	78	15
22	7-Oct-09	27.2	104	8.3	80	15

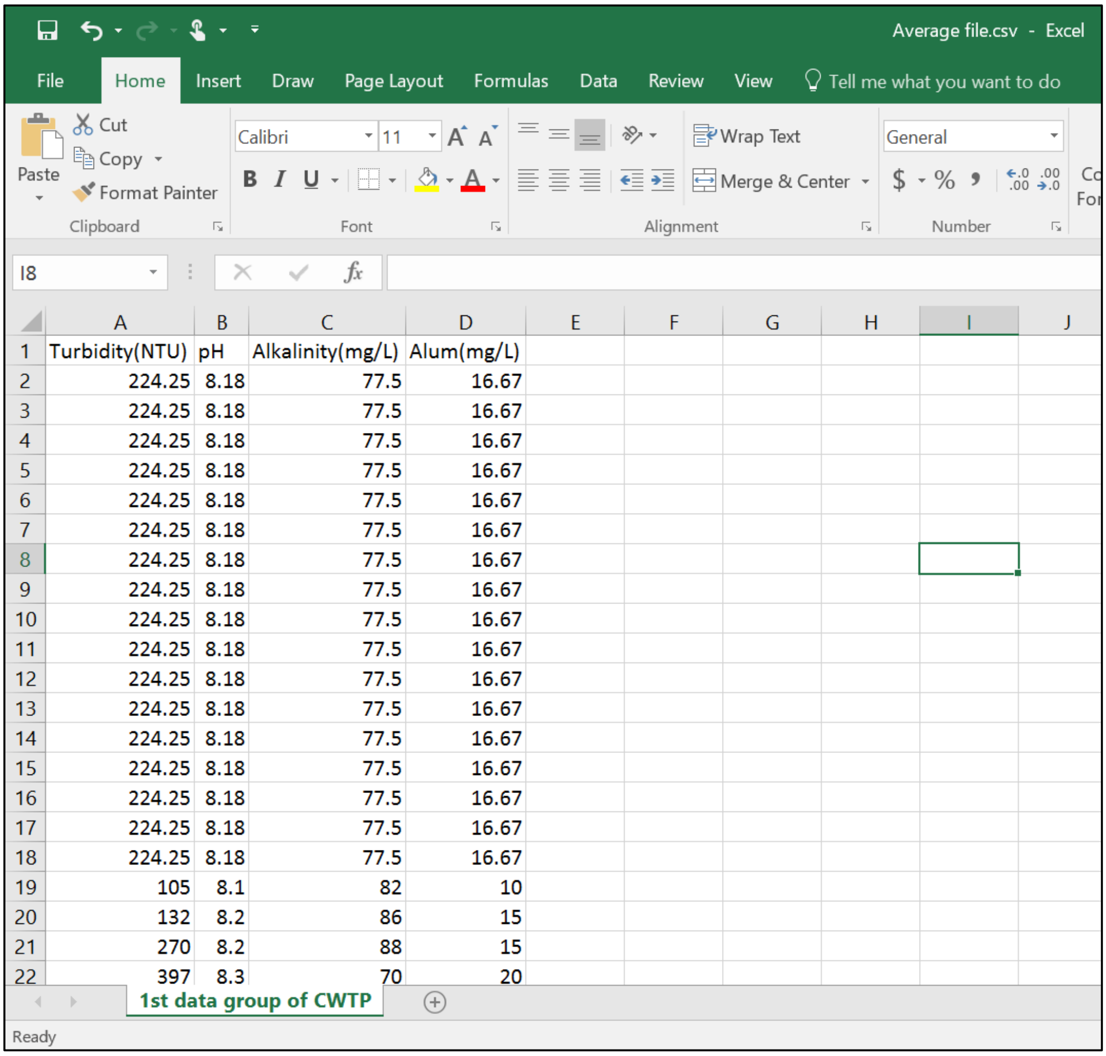


Raw water

Date	Temperature	Turbidity	pH	Alkalinity	Alum
1-Jan-16	25.00	32.00	8.10	96.00	10.00
2-Jan-16	24.00	34.80	8.50	98.00	10.00
3-Jan-16	24.50	30.00	8.00	90.00	10.00
4-Jan-16	24.40	26.30	7.90	90.00	12.00
5-Jan-16	24.50	28.00	8.00	94.00	10.00
6-Jan-16	24.70	22.00	7.90	94.00	10.00
7-Jan-16	24.60	20.50	8.10	90.00	10.00
8-Jan-16	25.00	22.30	8.10	92.00	10.00
9-Jan-16	25.20	23.60	8.50	90.00	10.00
10-Jan-16	25.20	25.50	8.10	98.00	10.00
11-Jan-16	24.30	24.80	8.00	92.00	10.00
12-Jan-16	24.30	21.20	7.90	92.00	10.00
13-Jan-16	24.60	18.60	8.10	92.00	10.00
14-Jan-16	24.50	17.30	9.00	94.00	10.00
15-Jan-16	25.10	15.30	8.40	110.00	10.00
16-Jan-16	30.60	18.90	8.20	100.00	10.00
17-Jan-16	25.00	19.20	8.20	112.00	10.00
18-Jan-16	25.30	18.20	8.30	110.00	10.00
19-Jan-16	25.50	16.40	8.40	110.00	10.00
20-Jan-16	25.50	19.60	8.20	118.00	10.00

Appendix C.2

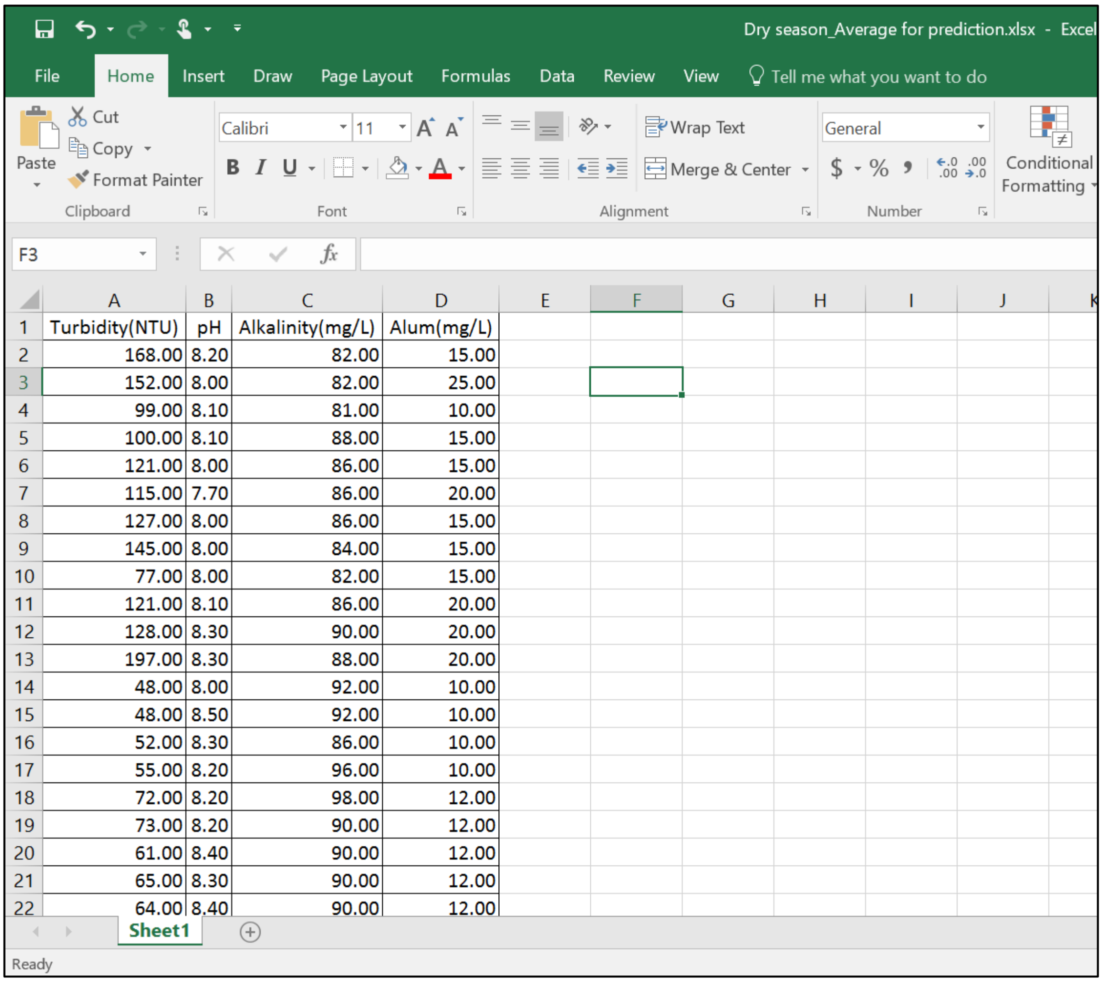
Data preparation from CWTP 2009-2016 for 1st data group (2,069 records)



	A	B	C	D
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	224.25	8.18	77.5	16.67
3	224.25	8.18	77.5	16.67
4	224.25	8.18	77.5	16.67
5	224.25	8.18	77.5	16.67
6	224.25	8.18	77.5	16.67
7	224.25	8.18	77.5	16.67
8	224.25	8.18	77.5	16.67
9	224.25	8.18	77.5	16.67
10	224.25	8.18	77.5	16.67
11	224.25	8.18	77.5	16.67
12	224.25	8.18	77.5	16.67
13	224.25	8.18	77.5	16.67
14	224.25	8.18	77.5	16.67
15	224.25	8.18	77.5	16.67
16	224.25	8.18	77.5	16.67
17	224.25	8.18	77.5	16.67
18	224.25	8.18	77.5	16.67
19	105	8.1	82	10
20	132	8.2	86	15
21	270	8.2	88	15
22	397	8.3	70	20

Appendix C.2.1

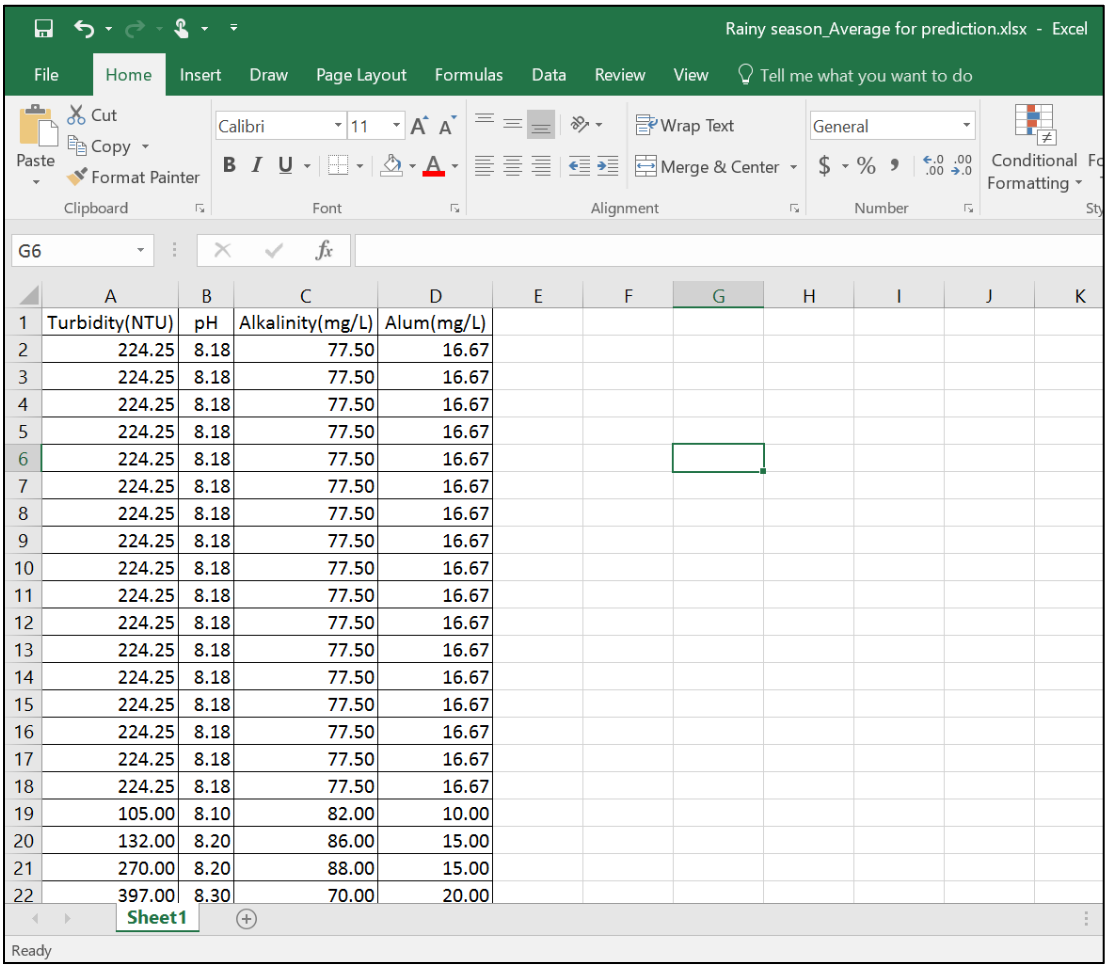
Data preparation from CWTP 2009-2016 for 1st data group in the drying season (1,058 records)



	A	B	C	D
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	168.00	8.20	82.00	15.00
3	152.00	8.00	82.00	25.00
4	99.00	8.10	81.00	10.00
5	100.00	8.10	88.00	15.00
6	121.00	8.00	86.00	15.00
7	115.00	7.70	86.00	20.00
8	127.00	8.00	86.00	15.00
9	145.00	8.00	84.00	15.00
10	77.00	8.00	82.00	15.00
11	121.00	8.10	86.00	20.00
12	128.00	8.30	90.00	20.00
13	197.00	8.30	88.00	20.00
14	48.00	8.00	92.00	10.00
15	48.00	8.50	92.00	10.00
16	52.00	8.30	86.00	10.00
17	55.00	8.20	96.00	10.00
18	72.00	8.20	98.00	12.00
19	73.00	8.20	90.00	12.00
20	61.00	8.40	90.00	12.00
21	65.00	8.30	90.00	12.00
22	64.00	8.40	90.00	12.00

Appendix C.2.2

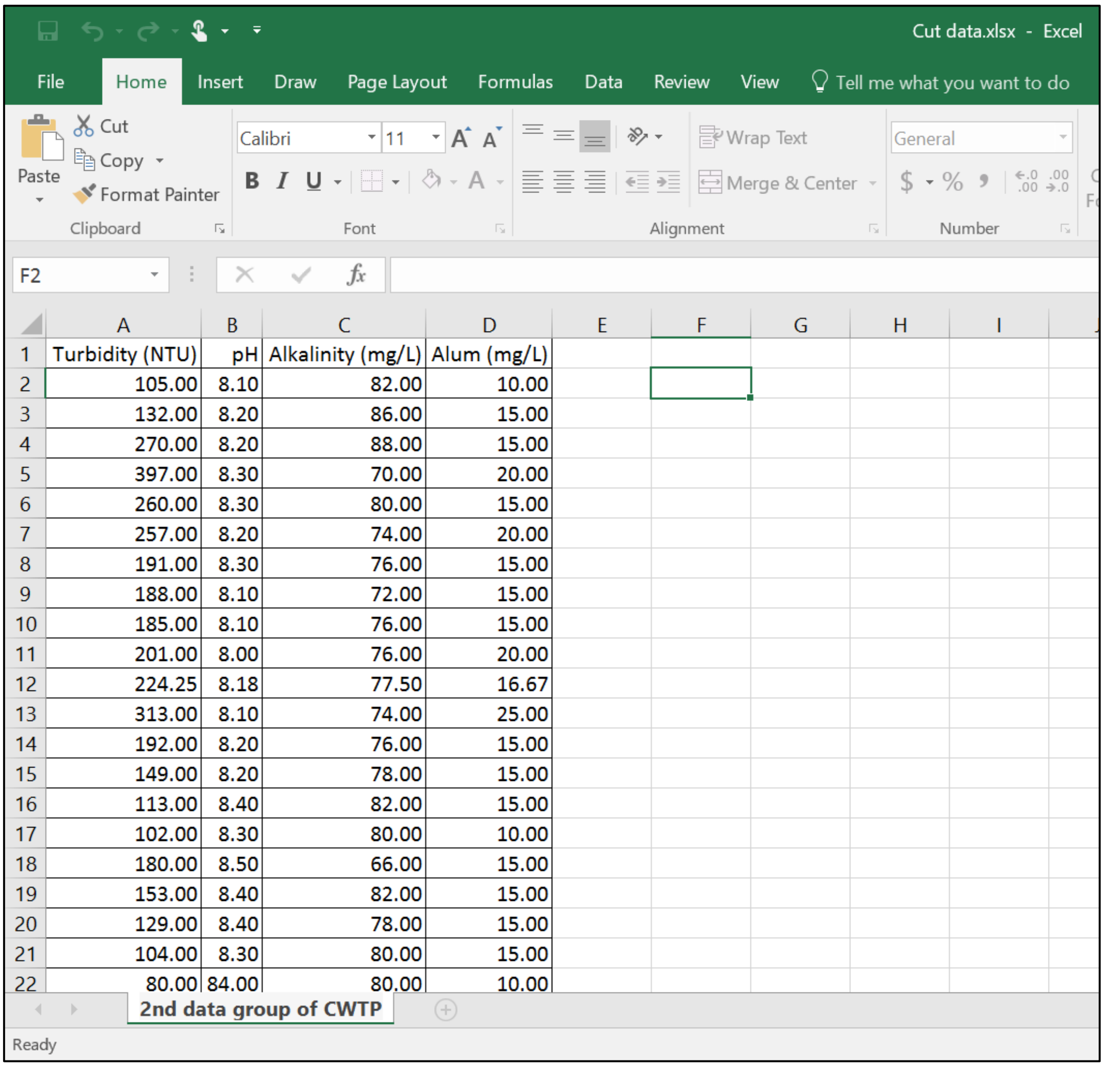
Data preparation from CWTP 2009-2016 for 1st data group in the raining season (1.011 records)



	A	B	C	D
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	224.25	8.18	77.50	16.67
3	224.25	8.18	77.50	16.67
4	224.25	8.18	77.50	16.67
5	224.25	8.18	77.50	16.67
6	224.25	8.18	77.50	16.67
7	224.25	8.18	77.50	16.67
8	224.25	8.18	77.50	16.67
9	224.25	8.18	77.50	16.67
10	224.25	8.18	77.50	16.67
11	224.25	8.18	77.50	16.67
12	224.25	8.18	77.50	16.67
13	224.25	8.18	77.50	16.67
14	224.25	8.18	77.50	16.67
15	224.25	8.18	77.50	16.67
16	224.25	8.18	77.50	16.67
17	224.25	8.18	77.50	16.67
18	224.25	8.18	77.50	16.67
19	105.00	8.10	82.00	10.00
20	132.00	8.20	86.00	15.00
21	270.00	8.20	88.00	15.00
22	397.00	8.30	70.00	20.00

Appendix C.3

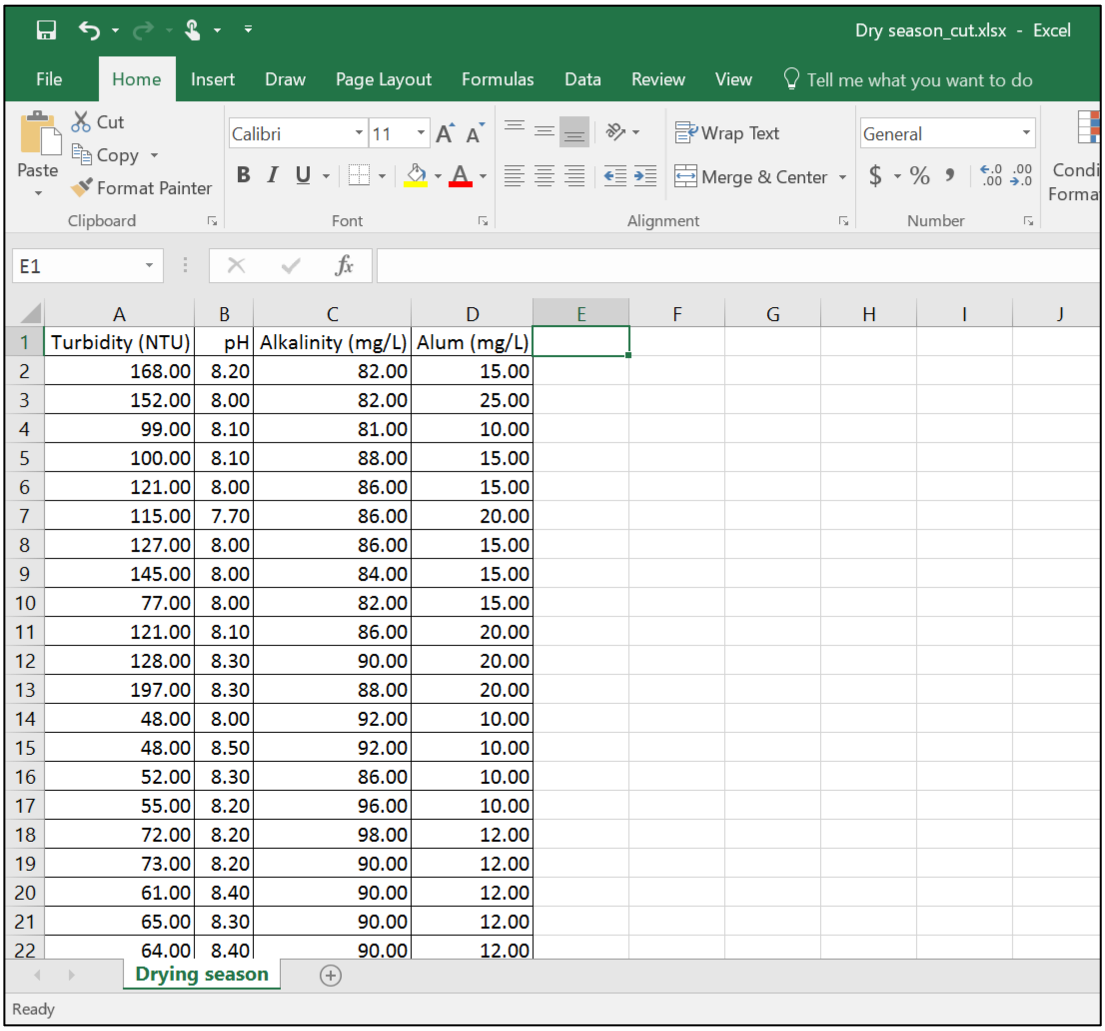
Data preparation from CWTP 2009-2016 for 2nd data group (2,022 records)



	A	B	C	D
1	Turbidity (NTU)	pH	Alkalinity (mg/L)	Alum (mg/L)
2	105.00	8.10	82.00	10.00
3	132.00	8.20	86.00	15.00
4	270.00	8.20	88.00	15.00
5	397.00	8.30	70.00	20.00
6	260.00	8.30	80.00	15.00
7	257.00	8.20	74.00	20.00
8	191.00	8.30	76.00	15.00
9	188.00	8.10	72.00	15.00
10	185.00	8.10	76.00	15.00
11	201.00	8.00	76.00	20.00
12	224.25	8.18	77.50	16.67
13	313.00	8.10	74.00	25.00
14	192.00	8.20	76.00	15.00
15	149.00	8.20	78.00	15.00
16	113.00	8.40	82.00	15.00
17	102.00	8.30	80.00	10.00
18	180.00	8.50	66.00	15.00
19	153.00	8.40	82.00	15.00
20	129.00	8.40	78.00	15.00
21	104.00	8.30	80.00	15.00
22	80.00	84.00	80.00	10.00

Appendix C.3.1

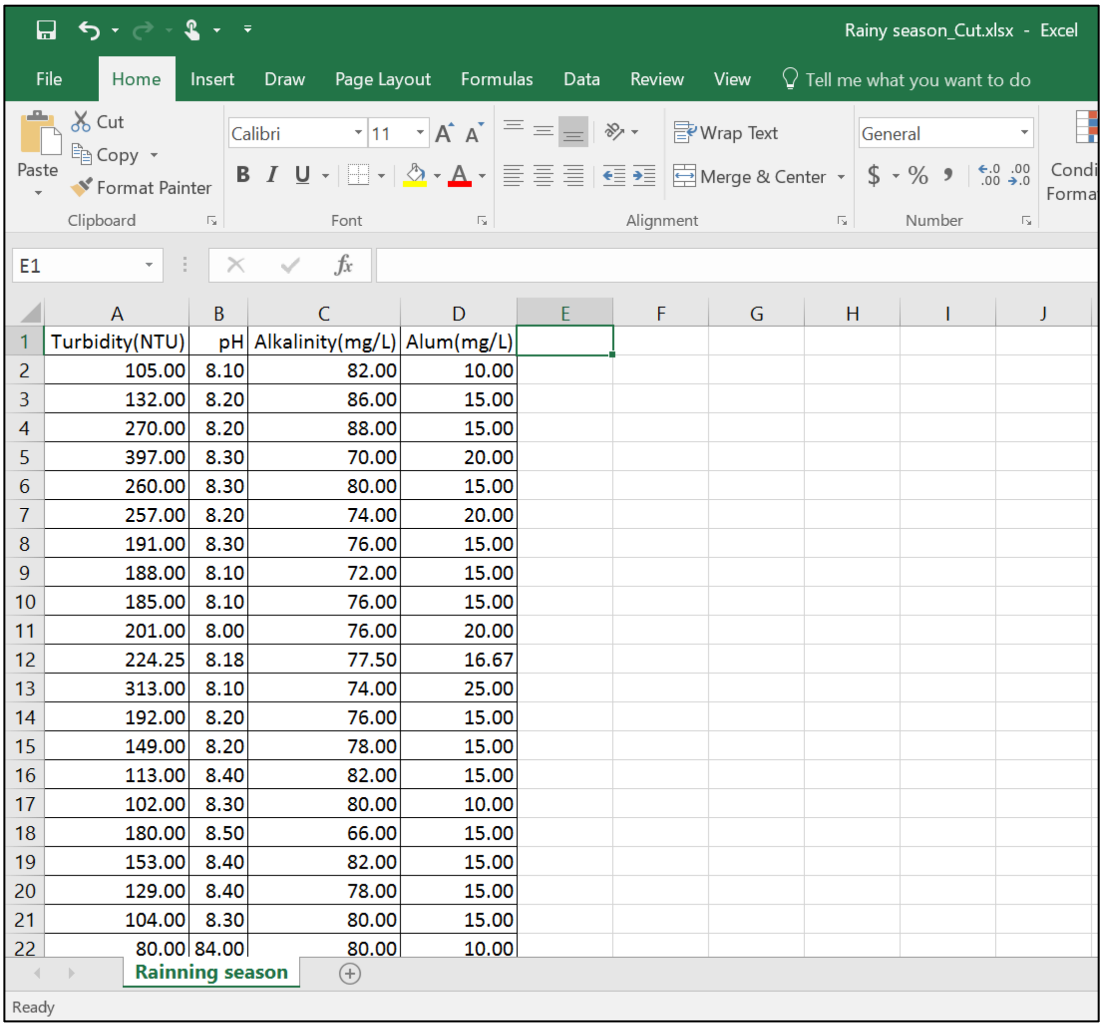
Data preparation from CWTP 2009-2016 for 2nd data group in the drying season (1,038 records)



	A	B	C	D
1	Turbidity (NTU)	pH	Alkalinity (mg/L)	Alum (mg/L)
2	168.00	8.20	82.00	15.00
3	152.00	8.00	82.00	25.00
4	99.00	8.10	81.00	10.00
5	100.00	8.10	88.00	15.00
6	121.00	8.00	86.00	15.00
7	115.00	7.70	86.00	20.00
8	127.00	8.00	86.00	15.00
9	145.00	8.00	84.00	15.00
10	77.00	8.00	82.00	15.00
11	121.00	8.10	86.00	20.00
12	128.00	8.30	90.00	20.00
13	197.00	8.30	88.00	20.00
14	48.00	8.00	92.00	10.00
15	48.00	8.50	92.00	10.00
16	52.00	8.30	86.00	10.00
17	55.00	8.20	96.00	10.00
18	72.00	8.20	98.00	12.00
19	73.00	8.20	90.00	12.00
20	61.00	8.40	90.00	12.00
21	65.00	8.30	90.00	12.00
22	64.00	8.40	90.00	12.00

Appendix C.3.2

Data preparation from CWTP 2009-2016 for 2nd data group in the raining season (983 records)



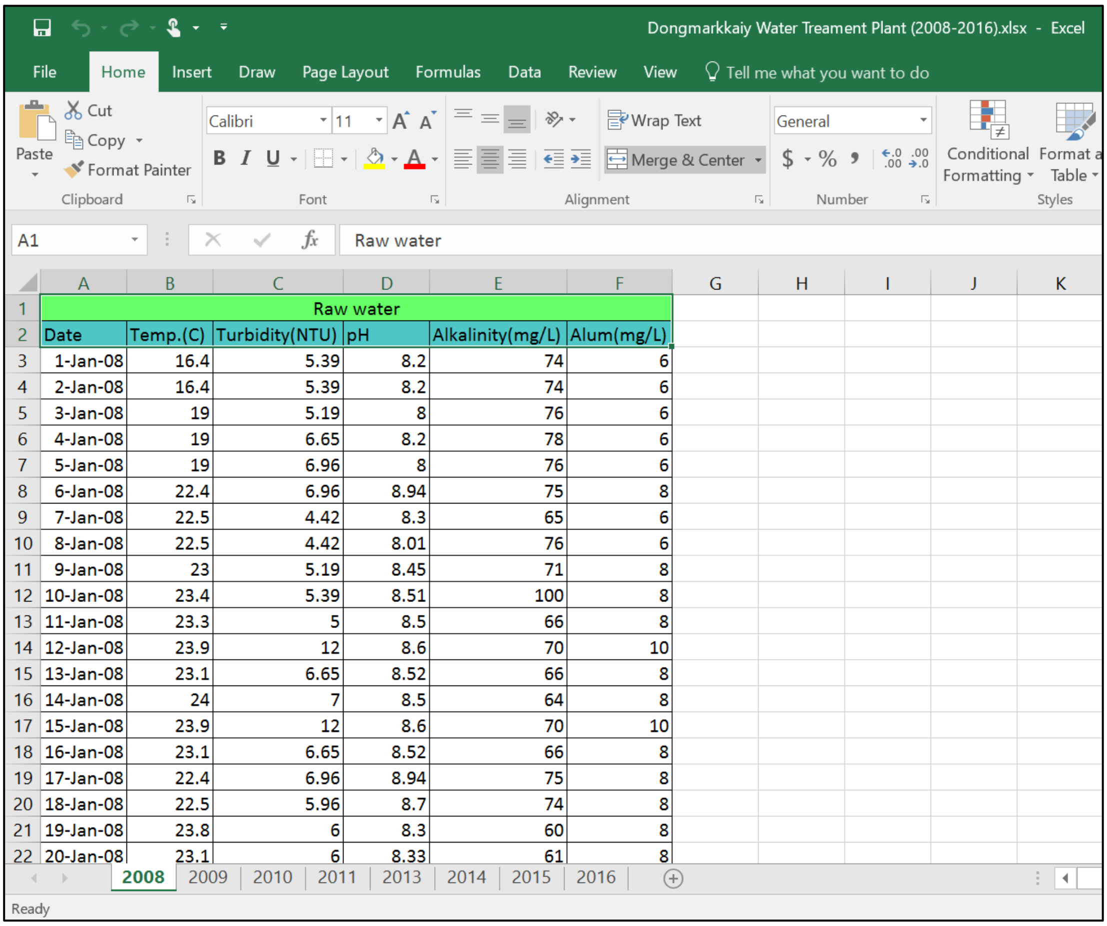
	A	B	C	D
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	105.00	8.10	82.00	10.00
3	132.00	8.20	86.00	15.00
4	270.00	8.20	88.00	15.00
5	397.00	8.30	70.00	20.00
6	260.00	8.30	80.00	15.00
7	257.00	8.20	74.00	20.00
8	191.00	8.30	76.00	15.00
9	188.00	8.10	72.00	15.00
10	185.00	8.10	76.00	15.00
11	201.00	8.00	76.00	20.00
12	224.25	8.18	77.50	16.67
13	313.00	8.10	74.00	25.00
14	192.00	8.20	76.00	15.00
15	149.00	8.20	78.00	15.00
16	113.00	8.40	82.00	15.00
17	102.00	8.30	80.00	10.00
18	180.00	8.50	66.00	15.00
19	153.00	8.40	82.00	15.00
20	129.00	8.40	78.00	15.00
21	104.00	8.30	80.00	15.00
22	80.00	84.00	80.00	10.00

Appendix D

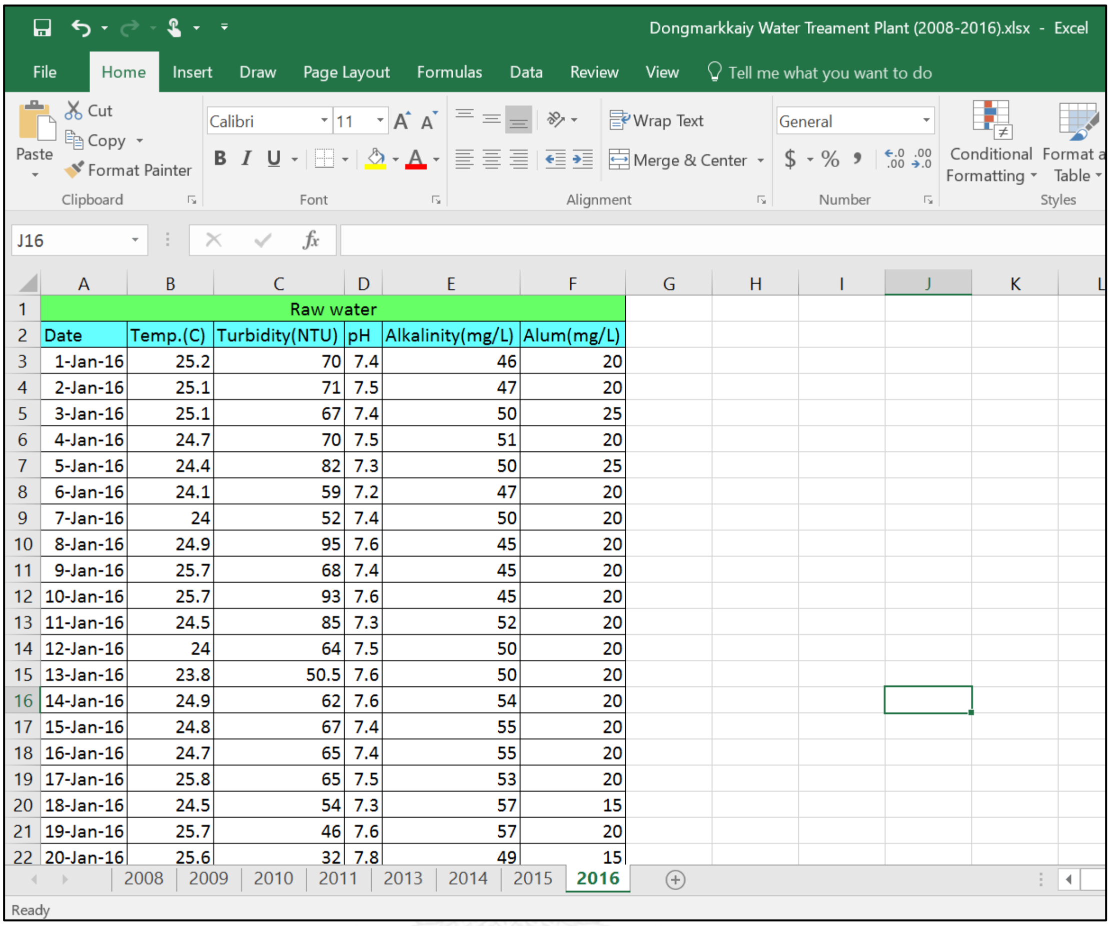
Data information from Dongmarkkaiy Water Treatment Plant (DWTP)

Appendix D.1

Raw data obtained from DWTP 2008-2016 (Total 2,802 records)



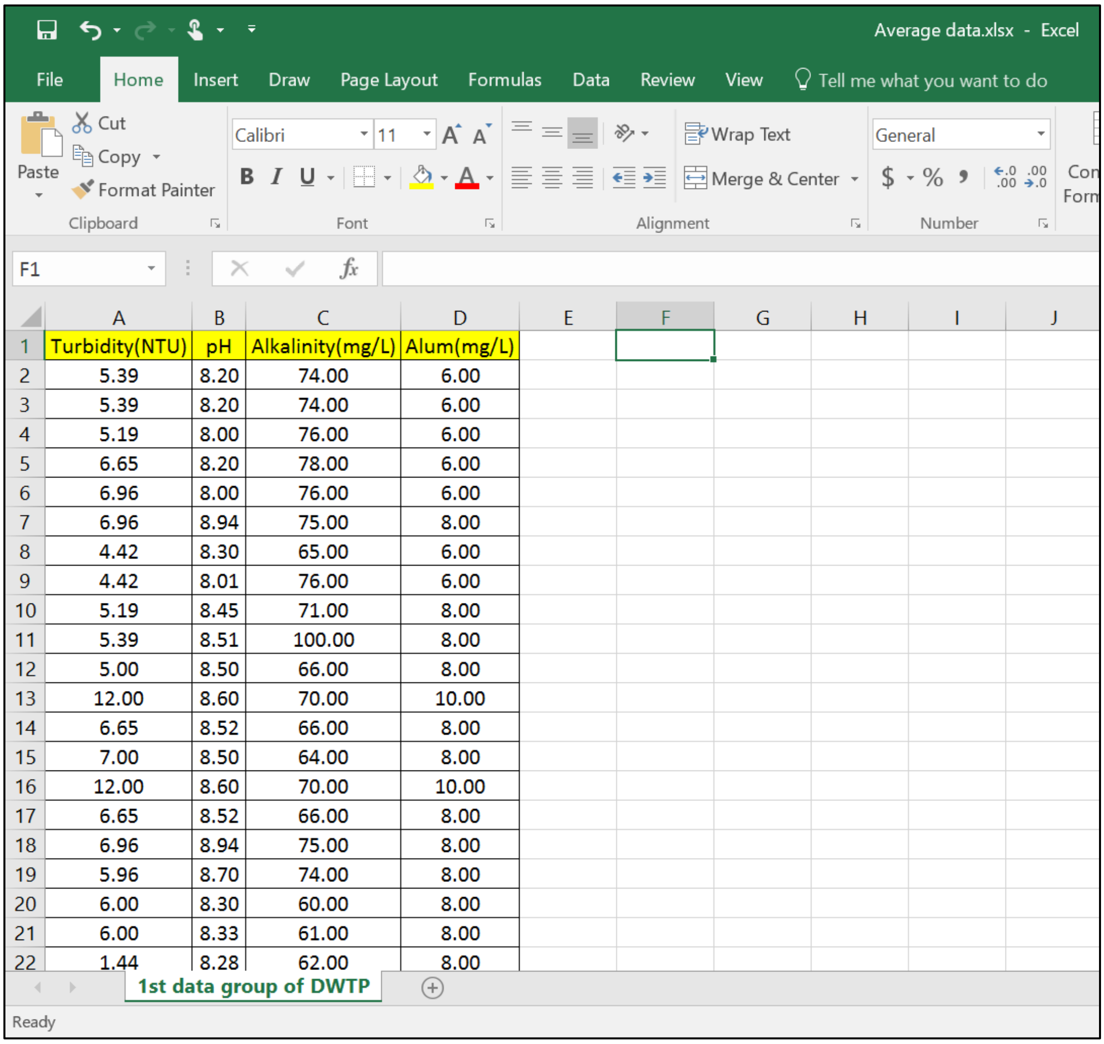
Raw water					
Date	Temp.(C)	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
1-Jan-08	16.4	5.39	8.2	74	6
2-Jan-08	16.4	5.39	8.2	74	6
3-Jan-08	19	5.19	8	76	6
4-Jan-08	19	6.65	8.2	78	6
5-Jan-08	19	6.96	8	76	6
6-Jan-08	22.4	6.96	8.94	75	8
7-Jan-08	22.5	4.42	8.3	65	6
8-Jan-08	22.5	4.42	8.01	76	6
9-Jan-08	23	5.19	8.45	71	8
10-Jan-08	23.4	5.39	8.51	100	8
11-Jan-08	23.3	5	8.5	66	8
12-Jan-08	23.9	12	8.6	70	10
13-Jan-08	23.1	6.65	8.52	66	8
14-Jan-08	24	7	8.5	64	8
15-Jan-08	23.9	12	8.6	70	10
16-Jan-08	23.1	6.65	8.52	66	8
17-Jan-08	22.4	6.96	8.94	75	8
18-Jan-08	22.5	5.96	8.7	74	8
19-Jan-08	23.8	6	8.3	60	8
20-Jan-08	23.1	6	8.33	61	8



Raw water					
Date	Temp.(C)	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
1-Jan-16	25.2	70	7.4	46	20
2-Jan-16	25.1	71	7.5	47	20
3-Jan-16	25.1	67	7.4	50	25
4-Jan-16	24.7	70	7.5	51	20
5-Jan-16	24.4	82	7.3	50	25
6-Jan-16	24.1	59	7.2	47	20
7-Jan-16	24	52	7.4	50	20
8-Jan-16	24.9	95	7.6	45	20
9-Jan-16	25.7	68	7.4	45	20
10-Jan-16	25.7	93	7.6	45	20
11-Jan-16	24.5	85	7.3	52	20
12-Jan-16	24	64	7.5	50	20
13-Jan-16	23.8	50.5	7.6	50	20
14-Jan-16	24.9	62	7.6	54	20
15-Jan-16	24.8	67	7.4	55	20
16-Jan-16	24.7	65	7.4	55	20
17-Jan-16	25.8	65	7.5	53	20
18-Jan-16	24.5	54	7.3	57	15
19-Jan-16	25.7	46	7.6	57	20
20-Jan-16	25.6	32	7.8	49	15

Appendix D.2

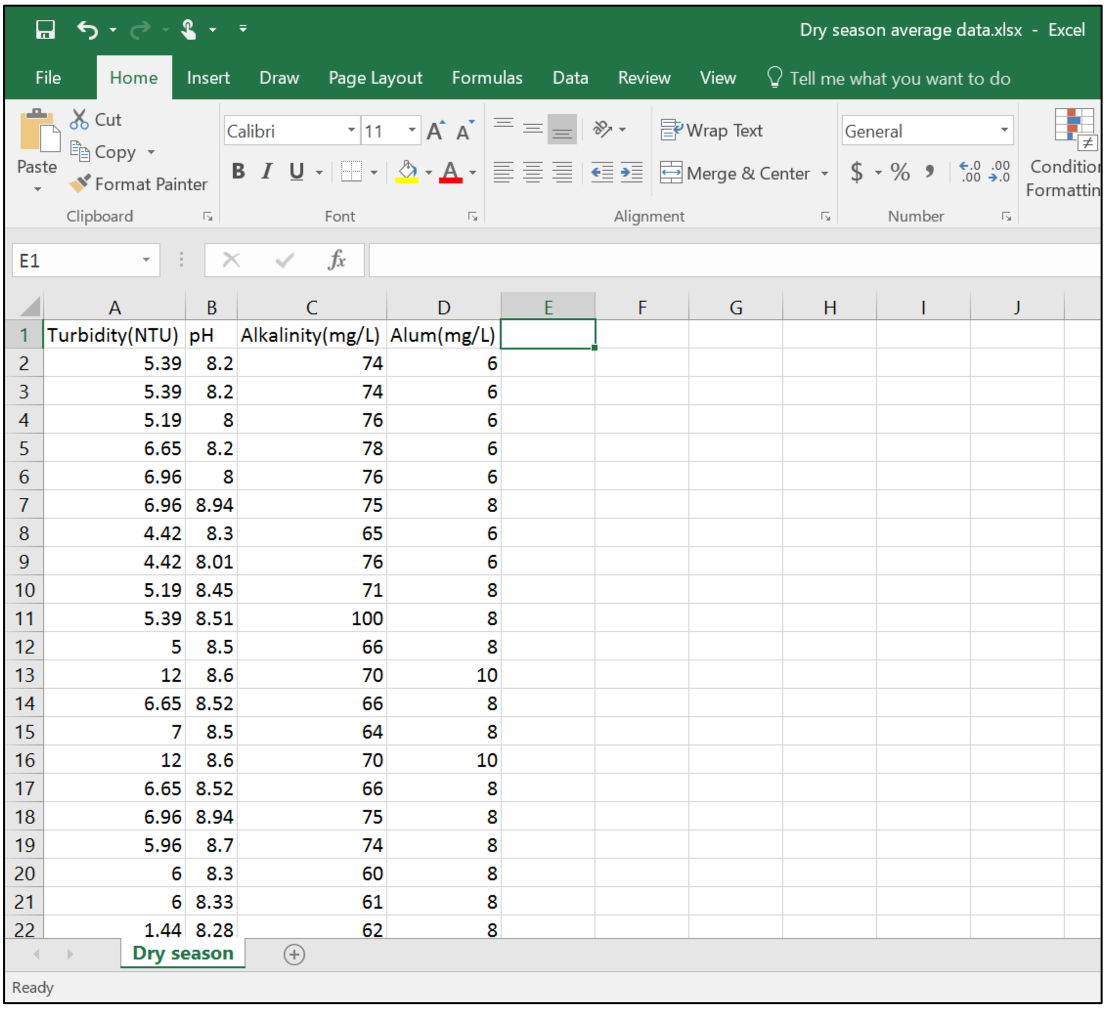
Data preparation from DWTP 2008-2016 for 1st data group (Total 2,861 records)



	A	B	C	D
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	5.39	8.20	74.00	6.00
3	5.39	8.20	74.00	6.00
4	5.19	8.00	76.00	6.00
5	6.65	8.20	78.00	6.00
6	6.96	8.00	76.00	6.00
7	6.96	8.94	75.00	8.00
8	4.42	8.30	65.00	6.00
9	4.42	8.01	76.00	6.00
10	5.19	8.45	71.00	8.00
11	5.39	8.51	100.00	8.00
12	5.00	8.50	66.00	8.00
13	12.00	8.60	70.00	10.00
14	6.65	8.52	66.00	8.00
15	7.00	8.50	64.00	8.00
16	12.00	8.60	70.00	10.00
17	6.65	8.52	66.00	8.00
18	6.96	8.94	75.00	8.00
19	5.96	8.70	74.00	8.00
20	6.00	8.30	60.00	8.00
21	6.00	8.33	61.00	8.00
22	1.44	8.28	62.00	8.00

Appendix D.2.1

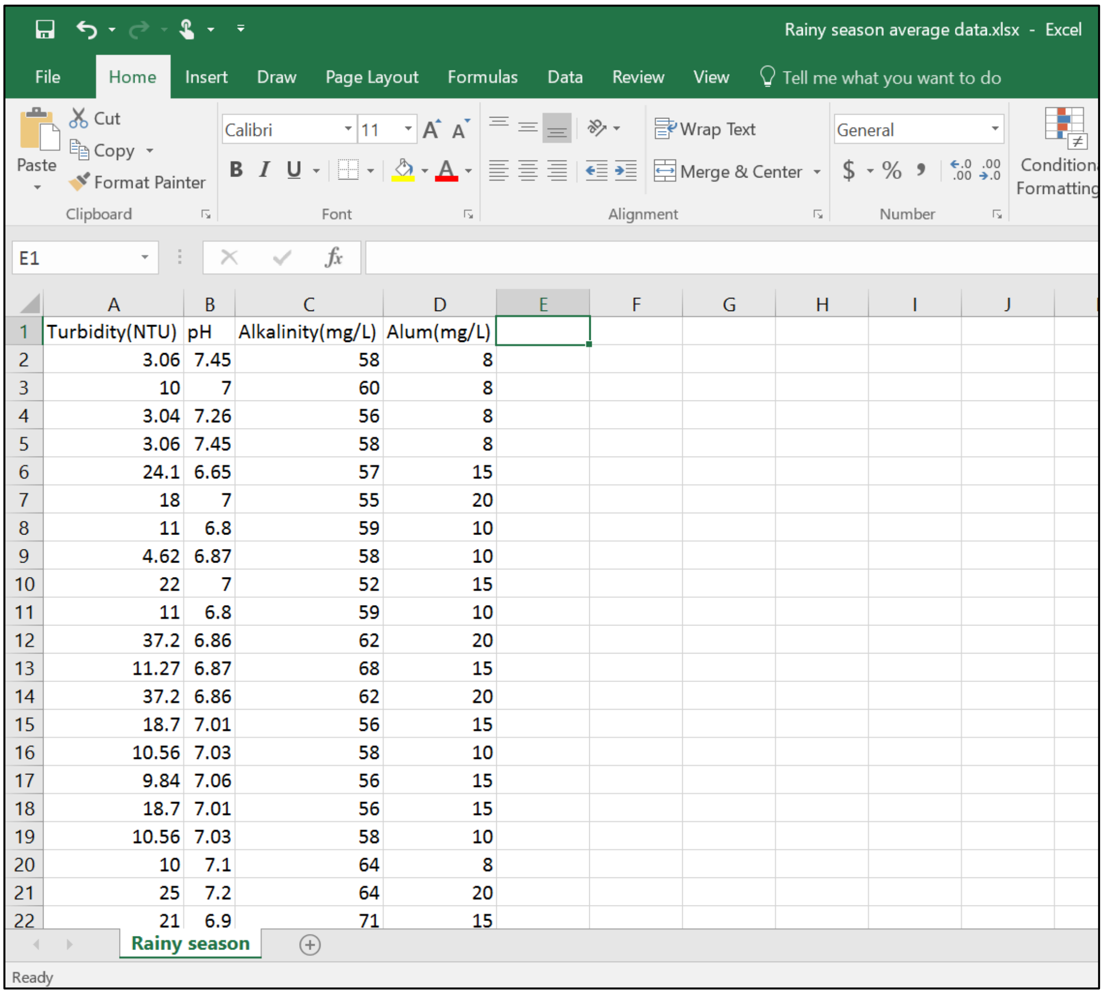
Data preparation from DWTP 2008-2016 for 1st data group in the drying season (1,389 records)



	A	B	C	D
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	5.39	8.2	74	6
3	5.39	8.2	74	6
4	5.19	8	76	6
5	6.65	8.2	78	6
6	6.96	8	76	6
7	6.96	8.94	75	8
8	4.42	8.3	65	6
9	4.42	8.01	76	6
10	5.19	8.45	71	8
11	5.39	8.51	100	8
12	5	8.5	66	8
13	12	8.6	70	10
14	6.65	8.52	66	8
15	7	8.5	64	8
16	12	8.6	70	10
17	6.65	8.52	66	8
18	6.96	8.94	75	8
19	5.96	8.7	74	8
20	6	8.3	60	8
21	6	8.33	61	8
22	1.44	8.28	62	8

Appendix D.2.2

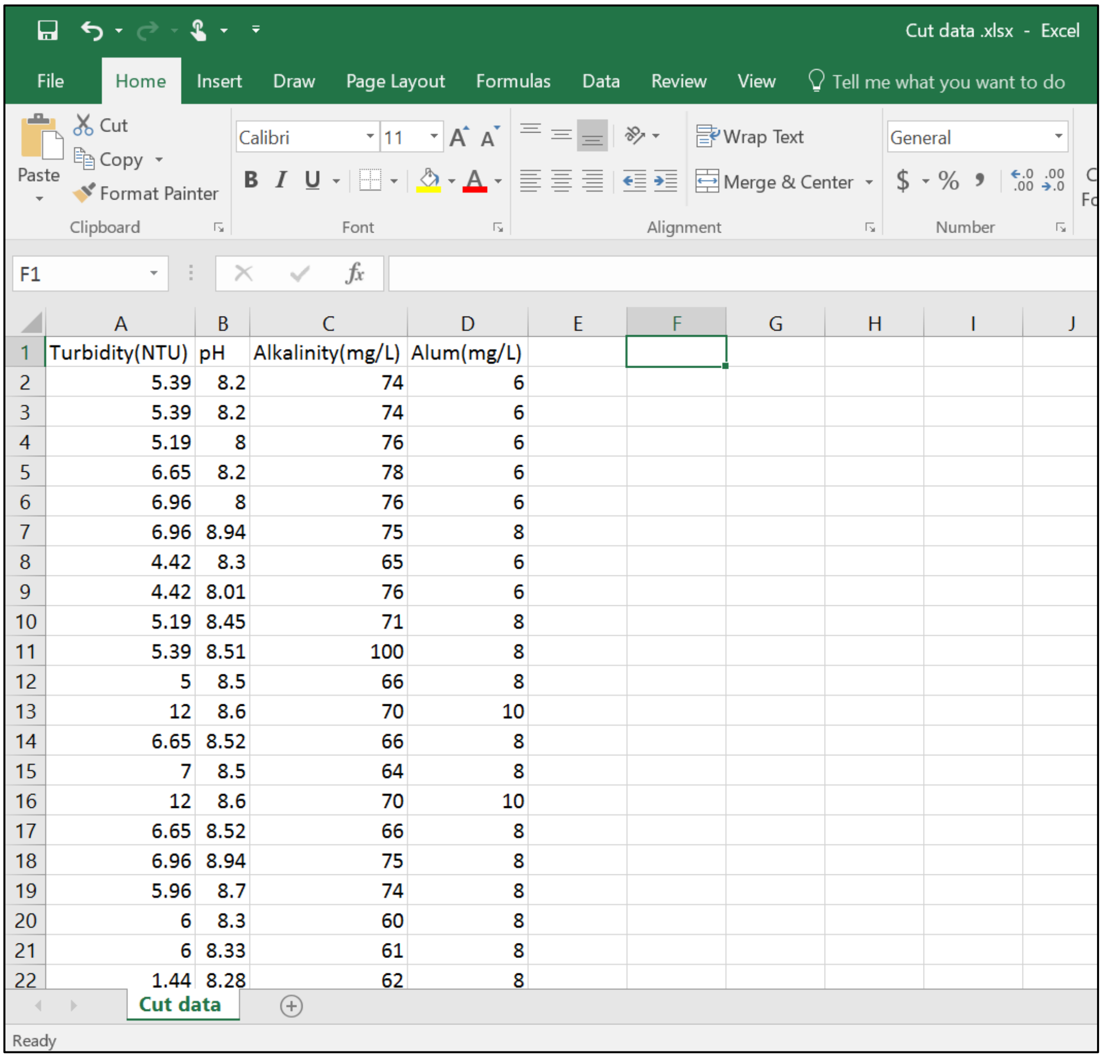
Data preparation from DWTP 2008-2016 for 1st data group in the drying season (1,472 records)



	A	B	C	D
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	3.06	7.45	58	8
3	10	7	60	8
4	3.04	7.26	56	8
5	3.06	7.45	58	8
6	24.1	6.65	57	15
7	18	7	55	20
8	11	6.8	59	10
9	4.62	6.87	58	10
10	22	7	52	15
11	11	6.8	59	10
12	37.2	6.86	62	20
13	11.27	6.87	68	15
14	37.2	6.86	62	20
15	18.7	7.01	56	15
16	10.56	7.03	58	10
17	9.84	7.06	56	15
18	18.7	7.01	56	15
19	10.56	7.03	58	10
20	10	7.1	64	8
21	25	7.2	64	20
22	21	6.9	71	15

Appendix D.3

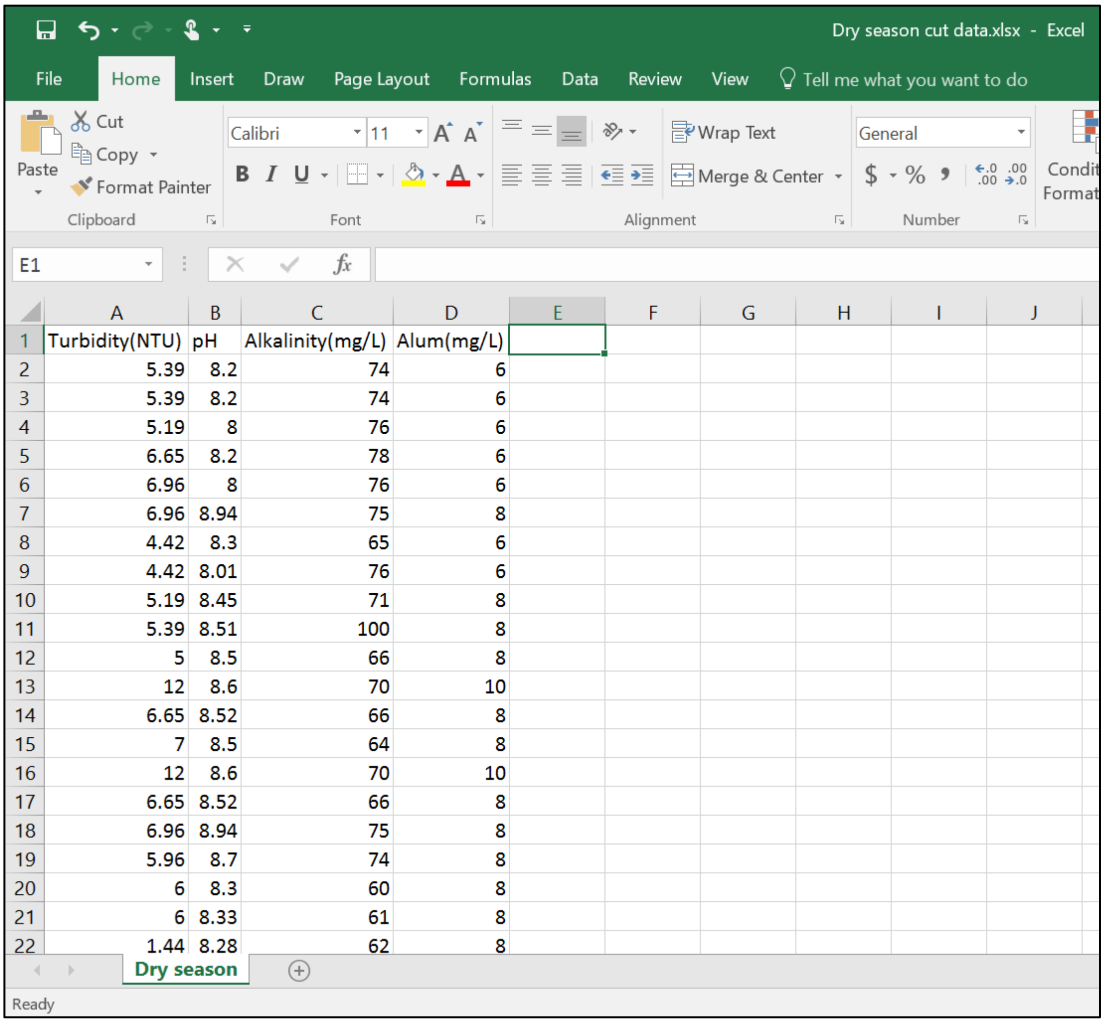
Data preparation from DWTP 2008-2016 for 2nd data group (2,284 records)



	A	B	C	D
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	5.39	8.2	74	6
3	5.39	8.2	74	6
4	5.19	8	76	6
5	6.65	8.2	78	6
6	6.96	8	76	6
7	6.96	8.94	75	8
8	4.42	8.3	65	6
9	4.42	8.01	76	6
10	5.19	8.45	71	8
11	5.39	8.51	100	8
12	5	8.5	66	8
13	12	8.6	70	10
14	6.65	8.52	66	8
15	7	8.5	64	8
16	12	8.6	70	10
17	6.65	8.52	66	8
18	6.96	8.94	75	8
19	5.96	8.7	74	8
20	6	8.3	60	8
21	6	8.33	61	8
22	1.44	8.28	62	8

Appendix D.3.1

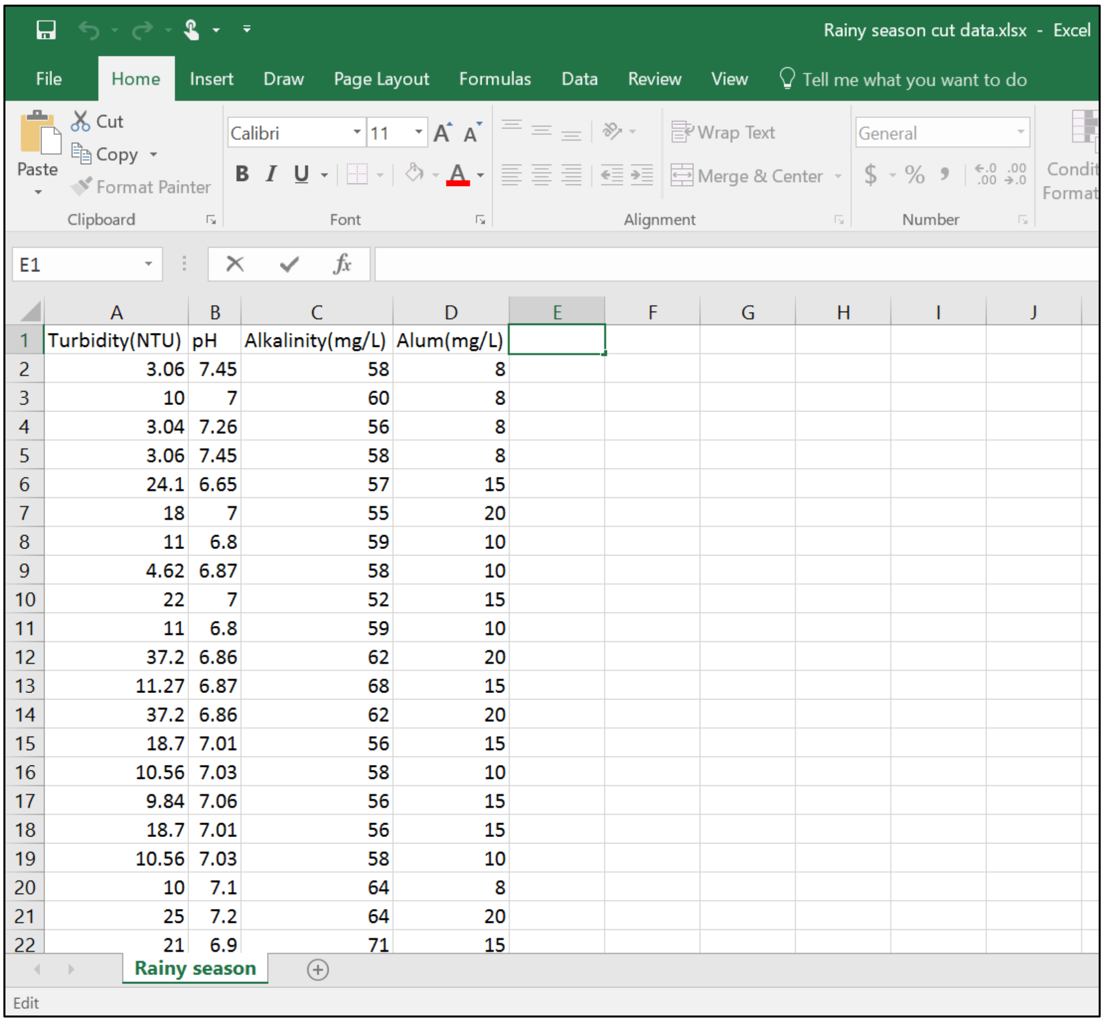
Data preparation from DWTP 2008-2016 for 2nd data group in the drying season (1,107 records)



	A	B	C	D
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	5.39	8.2	74	6
3	5.39	8.2	74	6
4	5.19	8	76	6
5	6.65	8.2	78	6
6	6.96	8	76	6
7	6.96	8.94	75	8
8	4.42	8.3	65	6
9	4.42	8.01	76	6
10	5.19	8.45	71	8
11	5.39	8.51	100	8
12	5	8.5	66	8
13	12	8.6	70	10
14	6.65	8.52	66	8
15	7	8.5	64	8
16	12	8.6	70	10
17	6.65	8.52	66	8
18	6.96	8.94	75	8
19	5.96	8.7	74	8
20	6	8.3	60	8
21	6	8.33	61	8
22	1.44	8.28	62	8

Appendix D.3.2

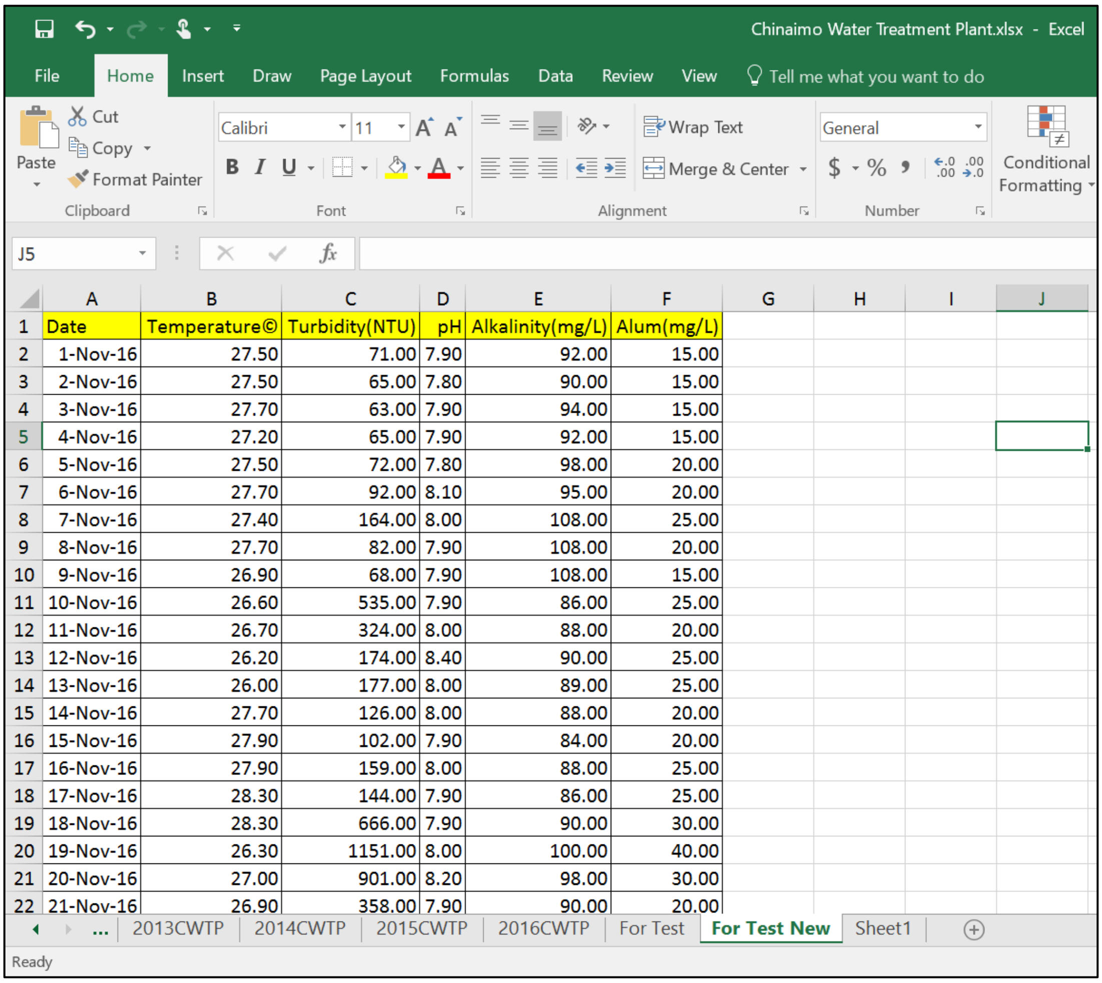
Data preparation from DWTP 2008-2016 for 2nd data group in the raining season (1,177 records)



Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
3.06	7.45	58	8
10	7	60	8
3.04	7.26	56	8
3.06	7.45	58	8
24.1	6.65	57	15
18	7	55	20
11	6.8	59	10
4.62	6.87	58	10
22	7	52	15
11	6.8	59	10
37.2	6.86	62	20
11.27	6.87	68	15
37.2	6.86	62	20
18.7	7.01	56	15
10.56	7.03	58	10
9.84	7.06	56	15
18.7	7.01	56	15
10.56	7.03	58	10
10	7.1	64	8
25	7.2	64	20
21	6.9	71	15

Appendix E

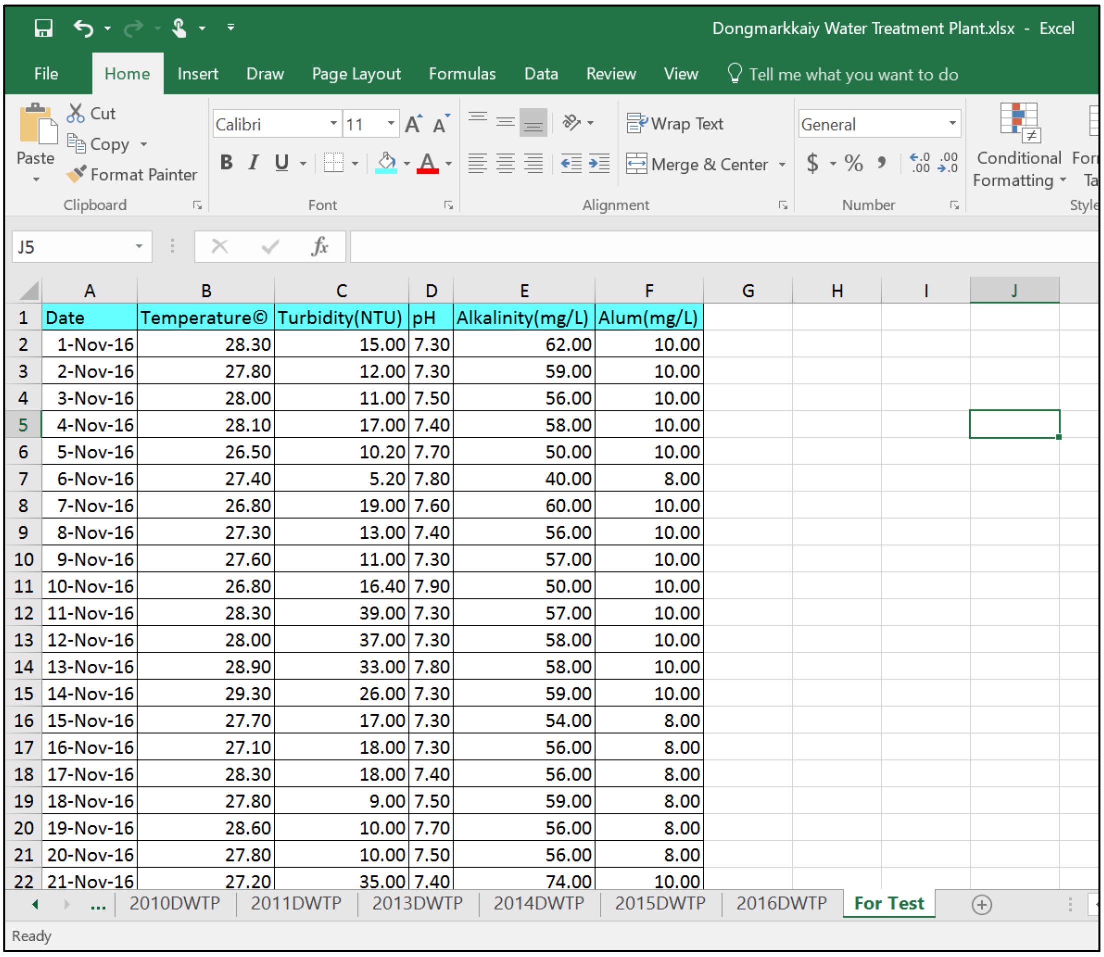
Raw data information obtained from CWTP on November 2016 to January 2017 for the real applications of the model



	A	B	C	D	E	F
1	Date	Temperature©	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	1-Nov-16	27.50	71.00	7.90	92.00	15.00
3	2-Nov-16	27.50	65.00	7.80	90.00	15.00
4	3-Nov-16	27.70	63.00	7.90	94.00	15.00
5	4-Nov-16	27.20	65.00	7.90	92.00	15.00
6	5-Nov-16	27.50	72.00	7.80	98.00	20.00
7	6-Nov-16	27.70	92.00	8.10	95.00	20.00
8	7-Nov-16	27.40	164.00	8.00	108.00	25.00
9	8-Nov-16	27.70	82.00	7.90	108.00	20.00
10	9-Nov-16	26.90	68.00	7.90	108.00	15.00
11	10-Nov-16	26.60	535.00	7.90	86.00	25.00
12	11-Nov-16	26.70	324.00	8.00	88.00	20.00
13	12-Nov-16	26.20	174.00	8.40	90.00	25.00
14	13-Nov-16	26.00	177.00	8.00	89.00	25.00
15	14-Nov-16	27.70	126.00	8.00	88.00	20.00
16	15-Nov-16	27.90	102.00	7.90	84.00	20.00
17	16-Nov-16	27.90	159.00	8.00	88.00	25.00
18	17-Nov-16	28.30	144.00	7.90	86.00	25.00
19	18-Nov-16	28.30	666.00	7.90	90.00	30.00
20	19-Nov-16	26.30	1151.00	8.00	100.00	40.00
21	20-Nov-16	27.00	901.00	8.20	98.00	30.00
22	21-Nov-16	26.90	358.00	7.90	90.00	20.00

Appendix F

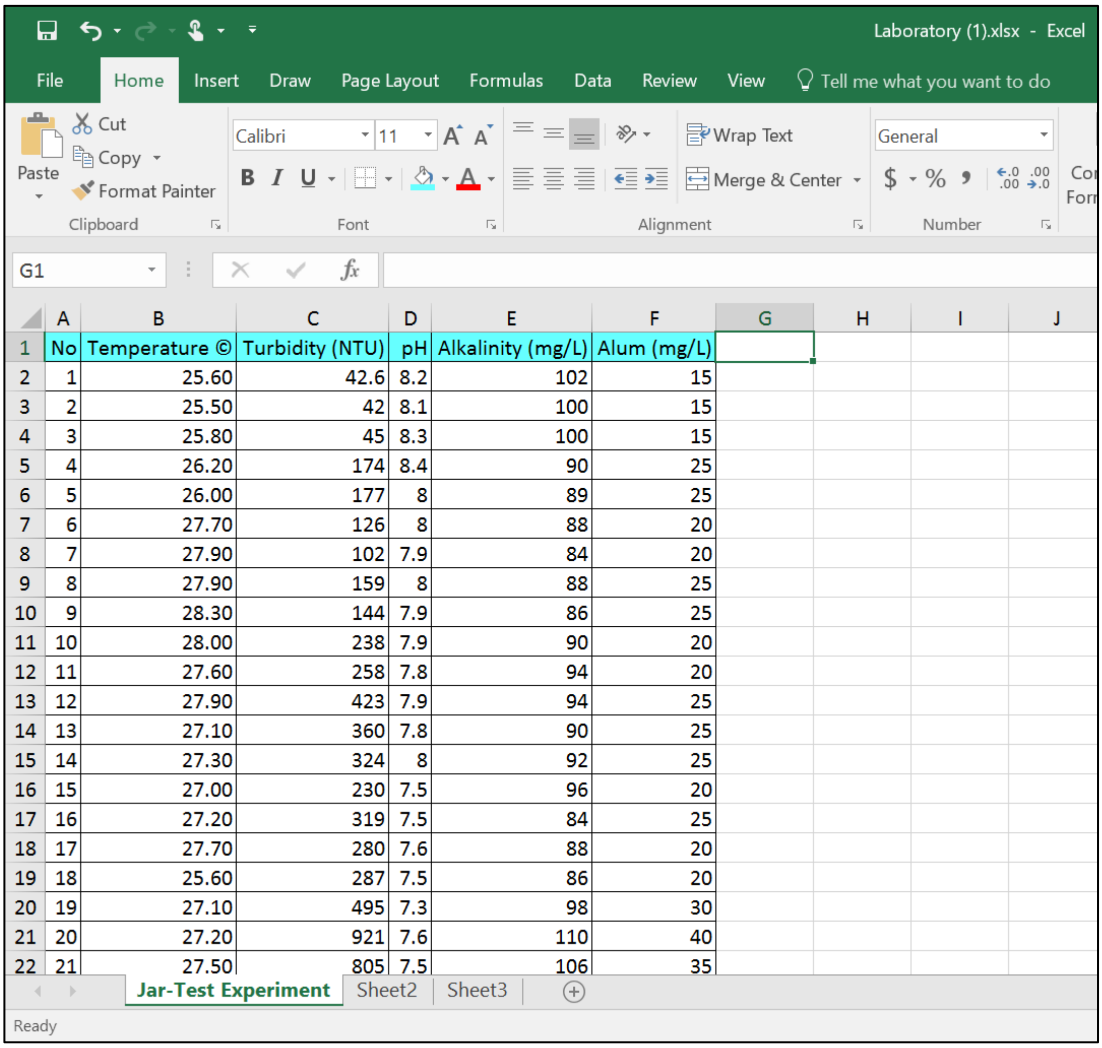
Raw data information obtained from DWTP on November 2016 to January 2017 for the real applications of the model



	A	B	C	D	E	F
1	Date	Temperature©	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	1-Nov-16	28.30	15.00	7.30	62.00	10.00
3	2-Nov-16	27.80	12.00	7.30	59.00	10.00
4	3-Nov-16	28.00	11.00	7.50	56.00	10.00
5	4-Nov-16	28.10	17.00	7.40	58.00	10.00
6	5-Nov-16	26.50	10.20	7.70	50.00	10.00
7	6-Nov-16	27.40	5.20	7.80	40.00	8.00
8	7-Nov-16	26.80	19.00	7.60	60.00	10.00
9	8-Nov-16	27.30	13.00	7.40	56.00	10.00
10	9-Nov-16	27.60	11.00	7.30	57.00	10.00
11	10-Nov-16	26.80	16.40	7.90	50.00	10.00
12	11-Nov-16	28.30	39.00	7.30	57.00	10.00
13	12-Nov-16	28.00	37.00	7.30	58.00	10.00
14	13-Nov-16	28.90	33.00	7.80	58.00	10.00
15	14-Nov-16	29.30	26.00	7.30	59.00	10.00
16	15-Nov-16	27.70	17.00	7.30	54.00	8.00
17	16-Nov-16	27.10	18.00	7.30	56.00	8.00
18	17-Nov-16	28.30	18.00	7.40	56.00	8.00
19	18-Nov-16	27.80	9.00	7.50	59.00	8.00
20	19-Nov-16	28.60	10.00	7.70	56.00	8.00
21	20-Nov-16	27.80	10.00	7.50	56.00	8.00
22	21-Nov-16	27.20	35.00	7.40	74.00	10.00

Appendix G

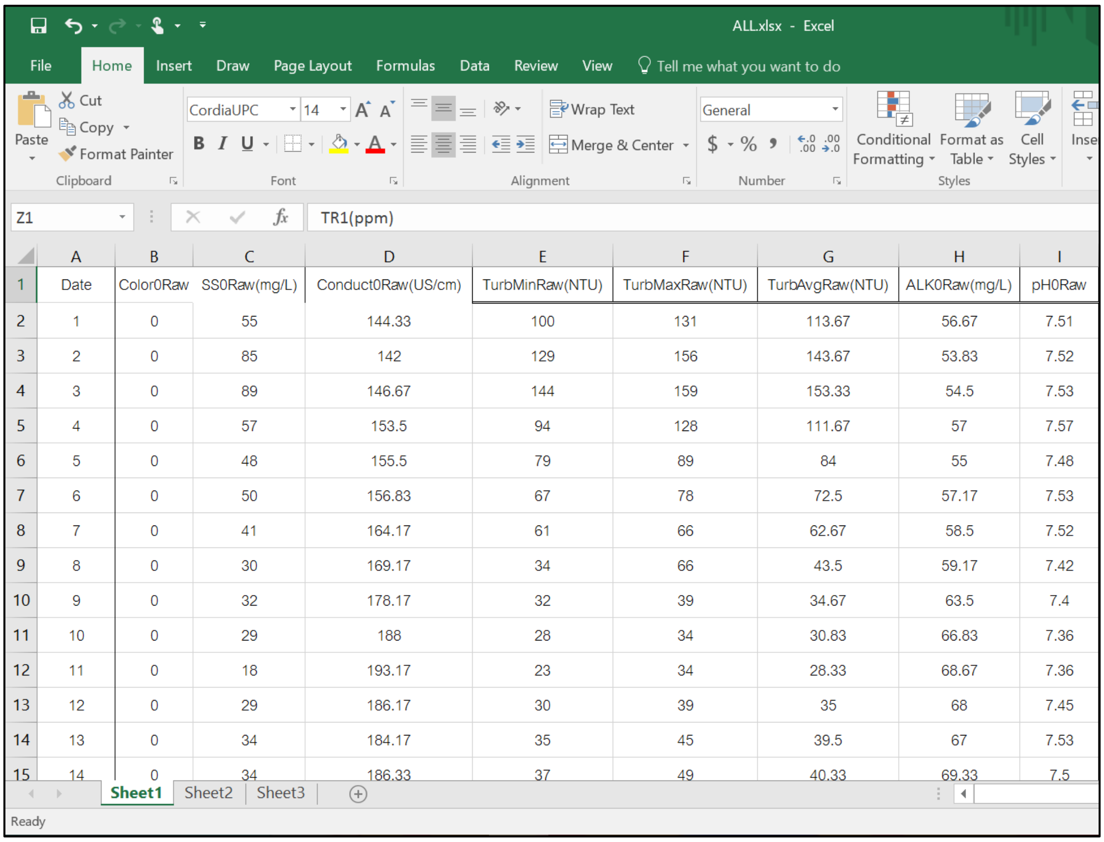
Data of the Jar-Test experiment at the CWTP (72 records)



No	Temperature ©	Turbidity (NTU)	pH	Alkalinity (mg/L)	Alum (mg/L)
1	25.60	42.6	8.2	102	15
2	25.50	42	8.1	100	15
3	25.80	45	8.3	100	15
4	26.20	174	8.4	90	25
5	26.00	177	8	89	25
6	27.70	126	8	88	20
7	27.90	102	7.9	84	20
8	27.90	159	8	88	25
9	28.30	144	7.9	86	25
10	28.00	238	7.9	90	20
11	27.60	258	7.8	94	20
12	27.90	423	7.9	94	25
13	27.10	360	7.8	90	25
14	27.30	324	8	92	25
15	27.00	230	7.5	96	20
16	27.20	319	7.5	84	25
17	27.70	280	7.6	88	20
18	25.60	287	7.5	86	20
19	27.10	495	7.3	98	30
20	27.20	921	7.6	110	40
21	27.50	805	7.5	106	35

Appendix H

Data information from BangKhen Water Treatment Plant 2004-2015 (3,965 records)

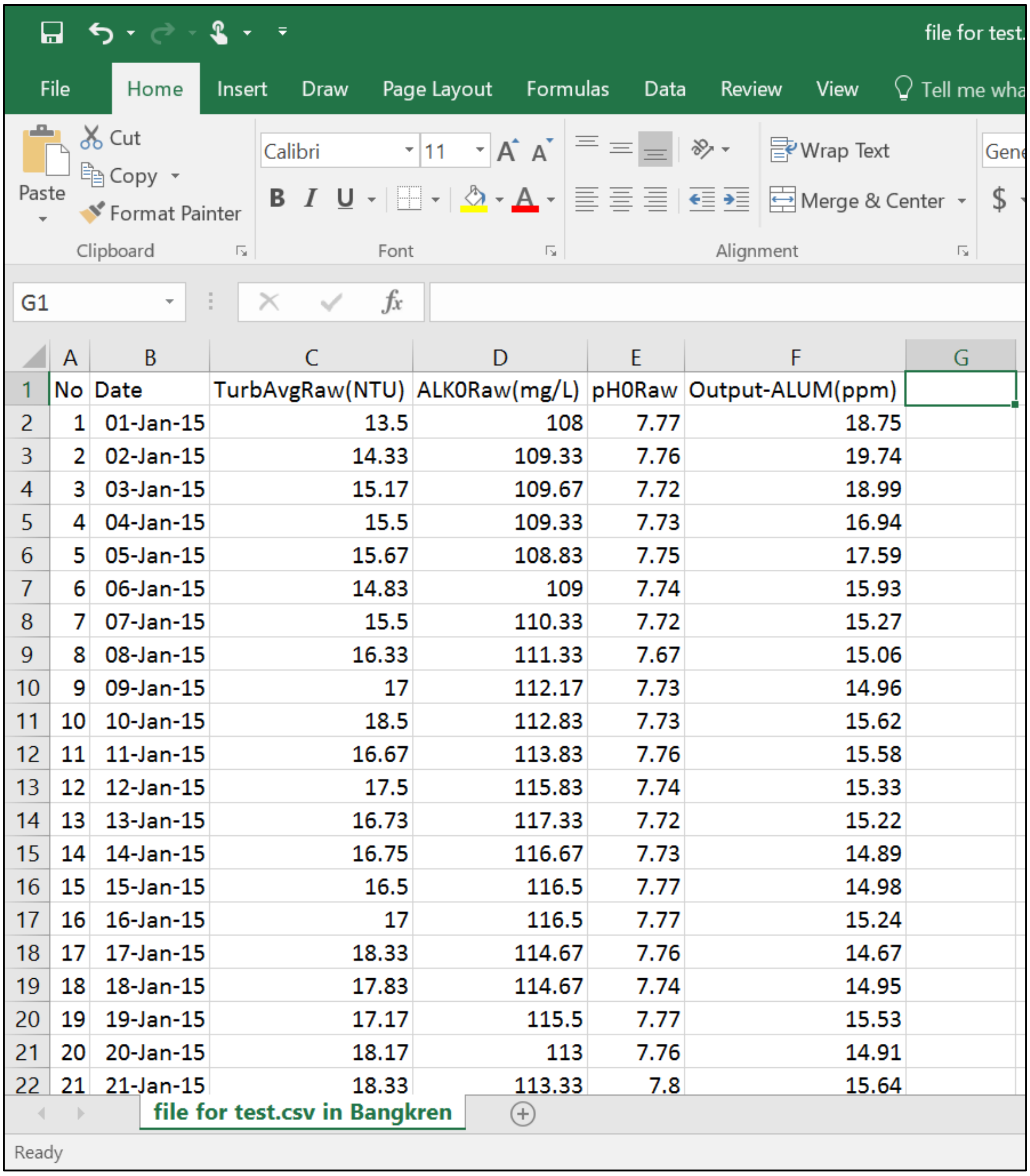


Z1 TR1(ppm)

	A	B	C	D	E	F	G	H	I
1	Date	Color0Raw	SS0Raw(mg/L)	Conduct0Raw(US/cm)	TurbMinRaw(NTU)	TurbMaxRaw(NTU)	TurbAvgRaw(NTU)	ALK0Raw(mg/L)	pH0Raw
2	1	0	55	144.33	100	131	113.67	56.67	7.51
3	2	0	85	142	129	156	143.67	53.83	7.52
4	3	0	89	146.67	144	159	153.33	54.5	7.53
5	4	0	57	153.5	94	128	111.67	57	7.57
6	5	0	48	155.5	79	89	84	55	7.48
7	6	0	50	156.83	67	78	72.5	57.17	7.53
8	7	0	41	164.17	61	66	62.67	58.5	7.52
9	8	0	30	169.17	34	66	43.5	59.17	7.42
10	9	0	32	178.17	32	39	34.67	63.5	7.4
11	10	0	29	188	28	34	30.83	66.83	7.36
12	11	0	18	193.17	23	34	28.33	68.67	7.36
13	12	0	29	186.17	30	39	35	68	7.45
14	13	0	34	184.17	35	45	39.5	67	7.53
15	14	0	34	186.33	37	49	40.33	69.33	7.5

Appendix H.1

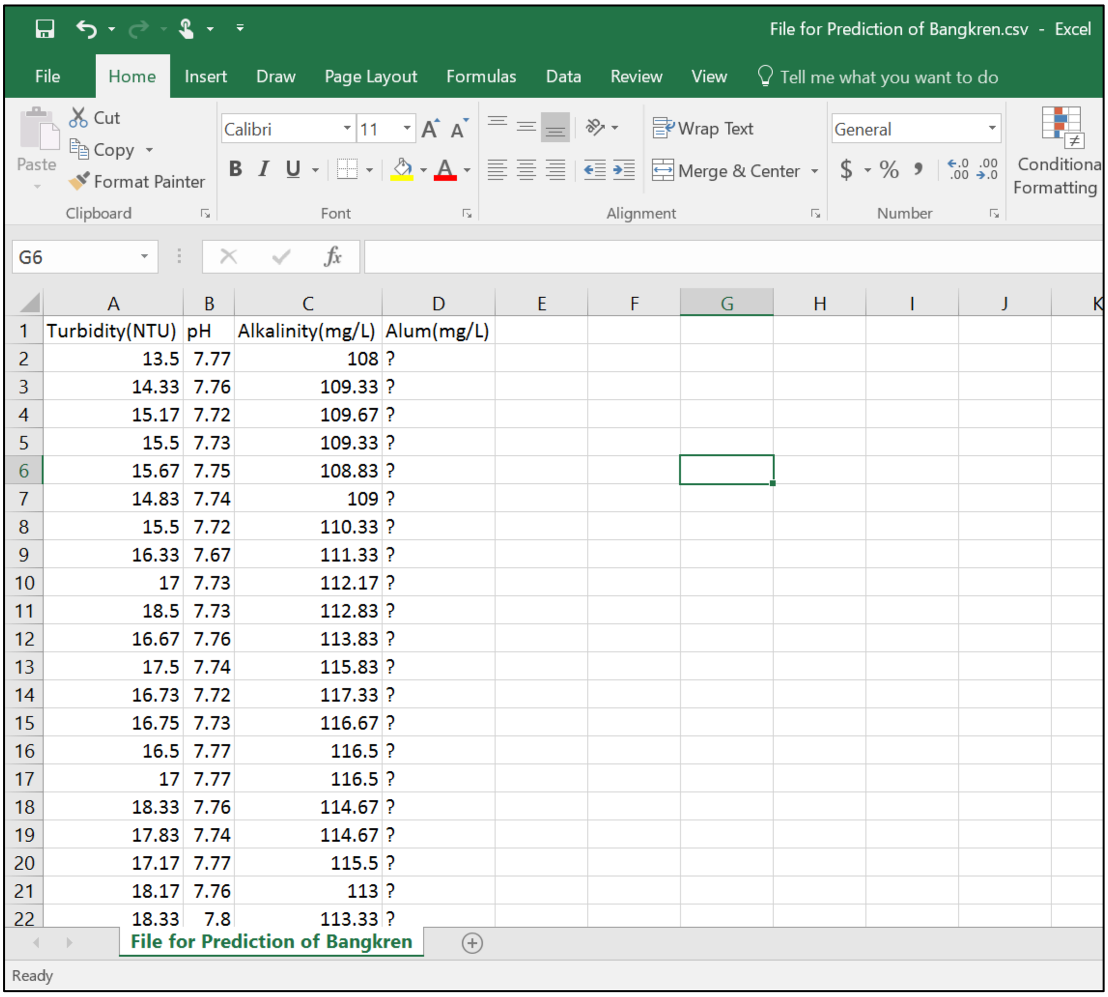
Raw data obtained from BangKhen Water Treatment Plant for the model testing on January, February, and March 2015 (90 records)



	A	B	C	D	E	F	G
1	No	Date	TurbAvgRaw(NTU)	ALK0Raw(mg/L)	pH0Raw	Output-ALUM(ppm)	
2	1	01-Jan-15	13.5	108	7.77	18.75	
3	2	02-Jan-15	14.33	109.33	7.76	19.74	
4	3	03-Jan-15	15.17	109.67	7.72	18.99	
5	4	04-Jan-15	15.5	109.33	7.73	16.94	
6	5	05-Jan-15	15.67	108.83	7.75	17.59	
7	6	06-Jan-15	14.83	109	7.74	15.93	
8	7	07-Jan-15	15.5	110.33	7.72	15.27	
9	8	08-Jan-15	16.33	111.33	7.67	15.06	
10	9	09-Jan-15	17	112.17	7.73	14.96	
11	10	10-Jan-15	18.5	112.83	7.73	15.62	
12	11	11-Jan-15	16.67	113.83	7.76	15.58	
13	12	12-Jan-15	17.5	115.83	7.74	15.33	
14	13	13-Jan-15	16.73	117.33	7.72	15.22	
15	14	14-Jan-15	16.75	116.67	7.73	14.89	
16	15	15-Jan-15	16.5	116.5	7.77	14.98	
17	16	16-Jan-15	17	116.5	7.77	15.24	
18	17	17-Jan-15	18.33	114.67	7.76	14.67	
19	18	18-Jan-15	17.83	114.67	7.74	14.95	
20	19	19-Jan-15	17.17	115.5	7.77	15.53	
21	20	20-Jan-15	18.17	113	7.76	14.91	
22	21	21-Jan-15	18.33	113.33	7.8	15.64	

Appendix H.2

Data preparation from BangKhen Water Treatment Plant for the model testing (January, February, and March 2015 (90 records))



	A	B	C	D
1	Turbidity(NTU)	pH	Alkalinity(mg/L)	Alum(mg/L)
2	13.5	7.77	108	?
3	14.33	7.76	109.33	?
4	15.17	7.72	109.67	?
5	15.5	7.73	109.33	?
6	15.67	7.75	108.83	?
7	14.83	7.74	109	?
8	15.5	7.72	110.33	?
9	16.33	7.67	111.33	?
10	17	7.73	112.17	?
11	18.5	7.73	112.83	?
12	16.67	7.76	113.83	?
13	17.5	7.74	115.83	?
14	16.73	7.72	117.33	?
15	16.75	7.73	116.67	?
16	16.5	7.77	116.5	?
17	17	7.77	116.5	?
18	18.33	7.76	114.67	?
19	17.83	7.74	114.67	?
20	17.17	7.77	115.5	?
21	18.17	7.76	113	?
22	18.33	7.8	113.33	?

VITA

Mr. Khoumkham Ladsavong was born on 27 February 1992, at Vientiane Province, Lao PDR. I graduated the secondary school level in the academic year 2009 at the Keokou Secondary School, Vientiane Province, Lao PDR. I was graduated the Bachelor's degree at the National University of Laos in the academic year 2015, my major is environmental engineering and I have attended to study in the Master's degree in the academic year 2015 at Environmental Engineering Department, Faculty of Engineering, Chulalongkorn University, Thailand.