PREDICTION OF GREY WATER FOOTPRINT BY USING ARTIFICIAL NEURAL NETWORK AND RANDOM FOREST

RUZIANA BINTI KAMARZAMAN

B. ENG (HONS.) CIVIL ENGINEERING

UNIVERSITI MALAYSIA PAHANG

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name : RUZIANA BINTI KAMARZAMAN

Date of Birth : 27TH FEBRUARY 1996

Title : PREDICTION OF GREY WATER FOOTPRINT BY USING ARTIFICIAL NEURAL NETWORK AND RANDOM FOREST

Academic Session : 2018/2019

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☑	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Universiti Malaysia Pahang reserves the following rights:

1. The Thesis is the Property of Universiti Malaysia Pahang
2. The Library of Universiti Malaysia Pahang has the right to make copies of the thesis for the purpose of research only.
3. The Library has the right to make copies of the thesis for academic exchange.

Certified by:

____
(Student's Signature)

960227-03-5258
New IC/Passport Number
Date: 31ST MAY 2019

____
(Supervisor's Signature)

DR. EDRIYANA BINTI A .AZIZ
Name of Supervisor
Date: 31ST MAY 2019

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

SUPERVISOR'S DECLARATION

I/We* hereby declare that I/We* have checked this thesis/project* and in my/our* opinion, this thesis/project* is adequate in terms of scope and quality for the award of the Bachelor Degree of Civil Engineering

____

(Supervisor's Signature)

Full Name : DR. EDRIYANA BINTI A. AZIZ

Position :

Date :

____

(Co-supervisor's Signature)

Full Name : MOHD SYAZWAN NIZAM BIN MOHD MONI

Position :

Date :

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

____

(Student's Signature)

Full Name : RUZIANA BINTI KAMARZAMAN

ID Number : AA15166

Date : 31ST MAY 2019

PREDICTION OF GREY WATER FOOTPRINT BY USING ARTIFICIAL NEURAL NETWORK AND RANDOM FOREST

RUZIANA BINTI KAMARZAMAN

Thesis submitted in fulfillment of the requirements
for the award of the
Bachelor Degree in Civil Engineering

Faculty of Civil Engineering and Earth Resources

UNIVERSITI MALAYSIA PAHANG

MAY 2019

ACKNOWLEDGEMENTS

Alhamdulillah, all praise and thanks belong to Allah alone because He is the originator of everything.

My deep gratitude goes first to Dr. Edriyana, who expertly guided me in completing my Final Year Project. Your guidance kept me constantly engaged with my research. Without her assistance and dedicated involvement in every step throughout the process, this thesis would have never been accomplished. I would like to thank you very much for your support and understanding over these past one year.

My appreciation also extend to my co-supervisor, Mr. Syazwan Nizam who monitoring all my progress, showing me the right way to do the best research and teach me new knowledge of water footprint. Not to forget my research members, Ainifatihah, Fazlina and Syazwan who cooperated and help each other in completing our research. Thank you for accompanying each other going through hardship of research study.

Most importantly, none of this could have happened without my family. My mother who offered her encouragement through phone calls every week. With their present and financial support, I am really grateful for my sisters Ruziah and Rubi. Above all, thank you again for those who I forget to mention.

ABSTRAK

Loji Rawatan Air (LRA) adalah satu tempat untuk merawat air mentah yang berasal dari bumi seperti sungai, tasik, laut dan air bawah tanah yang akan diagihkan untuk kegunaan masyarakat sekitar. Walaubagaimanapun, Loji Air di Malaysia masih lagi menggunakan proses rawatan air mentah konvensional. Oleh itu, jejak penggunaan air tiada dalam rekod catatan loji rawatan air. Dengan menggunakan pendekatan jejak air (Water Footprint Approach), jejak air kelabu (WFgrey) dinilai untuk mengukur kualiti air mentah. Jejak air (Water footprint) adalah penunjuk penggunaan air bersih yang mengira bukan sahaja air yang digunakan secara terus, tetapi juga air yang digunakn secara tidak terus. Disamping itu, jejak air kelabu (WFgrey) adalah jumlah air bersih yang diperlukan untuk mengasimilasi bahan tercemar di dalam air. Kajian ini fokus dalam mengira jejak air kelabu (WFgrey) di dua Loji Rawatan Air iaitu Loji Rawatan Air Semambu dan Panching. Tempoh yang dikaji bermula dari 2015 hhingga 2017. Terdapat beberapa faktor yang mempengaruhi pengiraan jejak air kelabu (WFgrey) seperti kandungan bahan tercemar didalam air mentah, kadar pelepasan air dan jumlah pengambilan air. Dalam kajian ini, peningkatan jumlah jejak air kelabu (WFgrey) adalah disebabkan oleh kandungan iron dan ammonia yang tinggi dalam sumber air yang kebanyakannya berasal dari perlombongan bauksit di Kuantan. Walaubagaimanapun, jumlah kesuluruhan jejak air kelabu (WFgrey) menunjukkan penurunan yang mana adalah menunjukkan petanda yang baik untuk keberlangsungan sungai. Sebagai rumusan, jumlah jejak air kelabu (WFgrey) dijangka akan menurun pada masa depan. Namun begitu, peluang untuk jejak air kelabu (WFgrey) untuk meningkat masih ada jika sungai tercemar. Kajian ini mencadangkan supaya aktiviti-aktiviti industri berdekatan sungai harus dijalankan mengikut prosedur operasi standard (SOP) dan kilang-kilang juga mesti membersihkan sisa-sisa tercemar terlebih dahulu sebelum dinyahkan ke sungai.

ABSTRACT

Water Treatment Plant (WTP) is a place to treat raw water from earth resources like river, lake, ocean and underground water which will supply to society. However, the water treatment plant in Malaysia still using conventional WTP to treat the raw water. So, the footprint of water usage is not yet in the recorded. By using water footprint (WF) approach, the grey water footprint (WFgrey) is assessed in order to evaluate the raw water quality. Water footprint is the indicator of freshwater use that looks not only at direct water use but also at indirect water use. Meanwhile, grey water footprint is the amount of freshwater needed to assimilate the pollutant. This study focused in calculating grey water footprint of two WTPs which is Semambu WTP and Panching WTP. The study period start from 2015 until 2017. There are factors influenced the calculation of total grey water footprint such as the concentration of pollutant considered, the discharge rate and the amount of water intake. In this study, the increment of total grey water footprint is due to high amount iron and ammonia in water intake which mostly come from bauxite mining in Kuantan. However, the overall grey water footprint trend shows decrement which is good sign for river sustainability. As a conclusion, the total grey water footprint is predicted to decrease in future. But, there is also chance for the grey water to increase if the river is polluted. This study suggested that the industrial activities near the river must be carried out with Standard Operation Procedure and the factory also must treated the effluent before discharge it into river.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

BOD	Biological Oxygen Demand
COD	Chemical Oxygen Demand
Fe^2	Iron
L	Load of pollutant
NH_4	Ammonia

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
JPS	Jabatan Pengairan dan Saliran
MMD	Meteorological Department
MOH	Ministry of Health
PAIP	Random Forest
RF	Pengurusan Air Pahang Berhad
RMSE	Root Mean Square Error
WFgrey	Grey Water Footprint
WTP	Water Treatment Plant Malaysia

CHAPTER 1

INTRODUCTION

1.1 Introduction

Every living thing needs water to survive. No matter where, when and who. It is one of the basic things to continue living. However, people tend to forget about the importance of good quality water's availability when they are flooded with the clean water resource while the other side of the world is thirsting for a drop of clean water. This is a predictable human behaviour. In fact, changing human behaviour is harder than maintaining the water quality itself.

It is important to keep supplying good quality of water resources in a country, especially in a developing country which water demand is high. Apart from providing the daily need of living things for instances drinking, showering and cleaning, water supply also important in maintaining economic activities like agriculture, construction and industries. Plentiful water resources in a dam, lake or river are important to generate energy like electrical energy with lower cost.

Economic activities are production of goods and services which make them available to the consumers. It is human activities which are performed in exchange for money. Rapid economic activities are good in setting up a developing country. Nevertheless, in getting something big, there is price to pay behind it. Water consuming in rapid economic activities is high. This is a big problem for a developed country in maintaining their triumph with limited clean water resources. It is one of the reasons why the places with water scarcity is having problem in expanding their economic activities. For today, not only undeveloped country likes South Africa faces the problems, the developing country also facing the same problem due to water crisis. For instances, Brazil, China, London and Tokyo (Margrit, 2018).

Energy supply is providing the society with comfort to carry on daily life. The applications which can change water into energy are through hydroelectric power, ocean energy and saline water. Hydroelectric power and ocean energy can transform kinetic energy to electric energy. While saline water is an application that transform solar energy to heat and electricity energy (Srinivas, n.d.). Water is a cheap energy producing. However, lacking in water quantity and quality will result expensive charge later on in producing energy.

World population is growing from time to time. Urbanization has become a common thing nowadays and it is one of the causes of water crisis. Demand for good quality of water is high but, the good water availability is decreasing and people are still not aware about it. According to (Wu, Liu, & Chen, 2012), migration from rural area to the cities has resulted water crisis in China. Urbanization is a good process. Despite of that, more water is needed in making sure the prosperity of urbanization. In fact, water might not be existed in future if the society is not conscious about the right way in managing water consuming.

Water footprint is a tool to know the management and consumption of water supply which is used directly or indirectly. It is important for us to know the water cycle so that we can reduce wastage of water due to overuse of the earth resources in human activities.

1.2 Problem statement

Water treatment plant is a place to treat raw water from pollutants and toxins for the usage of society (Alaska Department of Environmental Conservation, 2015). Example of raw water that been treated by water treatment plants are ground water, surface water, ocean water and spring water. There are many stages the raw water should pass through so that the pollutant and toxin can be separated. The last stage of treated water than can be distributed to the end users.

The study is conducted at Panching and Semambu Water Treatment Plants. At the study area, there is large quantity of water which comes from Kuantan river basin. However, the management of the water treatment plants is not yet updated. There is no overall water cycle recorded include rain and evaporation. It is hard for the researchers to track the water cycle due to incomplete data. Malaysia is an equatorial climate which got rain throughout the year. During dry season, the shortage of water results in low water flow and pressure. While during raining season, the rivers basin overflow and resulting flooded in some places. The water treatment plant management must be burdened by this matter because of uneven level of raw water resources in Kuantan river basin.

Currently, the water demand is increasing because of the growing population. It will keep increasing from time to time. It is important to keep track of the water cycle so that the water use can be managed. The awareness among the society also should be raised about the shortage of water supply in upcoming years.

Water footprint (WF) is functioned to upgrade and analyse the water use efficiency and water resources management (Lathuillière, Bulle, & Johnson, 2018). Water footprint generally breaks into three parts which are blue, green and grey water footprint. In this study, water footprint application is used to know the total grey water in Kuantan river basin which is supplied as raw water resources in Panching and Semambu water treatment plants. Grey water footprint is the water that polluted after it is used to produce some productions like the water use after the production of cotton shirts (Ababa et al., 2016). Here, how polluted the raw water resources through water footprint calculation from Kuantan river basin until the final step of filtration and the prediction of grey water footprint trend in the water treatment plants will be analysed.

1.3 Objectives

The objectives of this study are:

i. To calculate total grey water footprint in Semambu and Panching water treatment plant in Kuantan river basin for 2015-2017.

ii. To compare the best one between genetic algorithm and random forest algorithm in grey water footprint forecast prediction.

iii. To predict the trend grey water footprint of Semambu and Panching water treatment plants in Kuantan river basin

1.4 Scope of study

The study area on water footprint in water treatment plants (WTP) is around Kuantan river basin. The focus is only in the water treatment plants and grey water footprint in the water treatment plants which mean the calculation of grey water footprint is within the water treatment process starting from abstraction of raw water to final step filtration before distributing to the society. The main reason of grey water is to know the degree of pollution in the treated water or the total of grey water footprint and to predict the grey water trend in the WTPs. Besides, the algorithms used in calculating the grey water footprint will be compared to choose the best one.

The study area is within Semambu and Panching water treatment plants. Mostly, the raw water resources in water treatment plants in Kuantan are from Kuantan river basin which is the main river in Pahang. Semambu and Paching WTPs are not an exception, their raw water resources also supply from Kuantan river basin. For this study, grey water footprint is a focus. The samples of grey water are taken within the WTPs. This is because to know the total grey water used in the water treatment process before reaching the consumers.

Water footprint is a mostly used alternative to evaluate the water management and consumption (Lathuillière et al., 2018). There are many models that had been created in various researches to calculate and predict the water footprint. The calculation formula of the model is called algorithm such as Genetic Algorithm (Hardwinarto & Aipassa, 2015), Random Forest algorithm (Nolan, Fienen, & Lorenz, 2015), Lavenberg Marquardt Algorithm (Badrzadeh, Sarukkalige, & Jayawardena, 2013) and Regression Tree (Chen et al., 2015). In this study, the frequently used algorithms are chosen instead of creating one. The models used are Artificial Neural Network and Random Forest Algorithm. The grey water foot print trend in Semambu and Panching water treatment plants will be calculate and predict using Artificial Neural Network and Random Forest Algorithm.

CHAPTER 2

LITERATURE REVIEW

2.1 Water Consumption

Generally, the term of water use and water consumption are translated as same meaning. However, in research field, water consumption and water use have different view of study. According to World Resources Institutes, water use is the amount of water extract from it sources to be used to produce a good or to carry an activities (Reig, 2013). For example, the amount of water needs to produce a stack of paper is 100 gallon, 75percent of the water amount is return back to its sources. Still, the water use for the factory to produce a stack of paper is 100 gallon. While water consumption is the water which is taking out from it sources without return it back to its original state (Reig, 2013). As illustration, the water consumption can be a shirt or evaporate in the air which already cannot be recycled for us to reuse. Water consumption can be a measurement on water availability after the human activities.

Calculation of overall water consumption is crucial to estimate the water supply and water demand in water management system planning (Bari, Begum, Nesadurai, & Pereira, 2015). The world population keep increasing day by day and demand on natural resources is proportionally increasing as the population. Everyone will be using water every second of their life. But, most of the societies are not yet realise that they are facing water availability problem. Water is a renewable resource. But, the pace of water consumption is faster than the pace of the water resource been replenished. The United Nation limits 165 litre of water per day per capita (Chung, 2018). This limitation use of water will increase in upcoming years. By calculating the water consumption, the demand of water resources can be fulfilled and the standard of living can be improved.

Next is to identify places with water scarcity. Water stress is a worldwide challenge in this era. There are communities suffer from limited sources of clean water. According to (M.Mekonnen & Hoekstra, 2016) the research find out about 4.0 billion people experience water stress which can be conclude as two third of the world population. Even so, from previous studies, the results stated about 1.7 to 3.1 billion people only facing water scarcity (M.Mekonnen & Hoekstra, 2016). This is because the studies cover at specific basin scale and on annual pattern (M.Mekonnen & Hoekstra, 2016). Measuring the water stress annually will not display which period of time experiences the highest water scarcity because the result will be calculated as the average value (M.Mekonnen & Hoekstra, 2016). At the same time, some developed countries retain abundant of water resources but the water is contaminate due to pollutions which result in high cost to treat the contaminate water (Millock & Nauges, 2010). The pollutions come from industry, agriculture and domestic use (Millock & Nauges, 2010). Water scarcity is not only because of the limited clean water resources but also the polluted water resources. The people should be educated about the awareness of water availability and the country water policies. In this condition, the government should start to solve this at the root cause which is the behaviour of the society itself. Until then, the water management system can run smoothly with the right behaviour of the users. What use of the great water system if the users are not appreciating it. The water resources are not only needed for current days but should be preserved to ensure the well-being of the next generation.

The awareness on water availability in society is still low. In fact, a Malaysian uses about 300 litres of water per day which exceed the water requirement set by the United Nation (UN) that is 165 litres per person per day (Chung, 2018). The amount of water set by the United Nation (UN), 165 litres of water is already enough for domestically use likes washing, showering and drinking. In meantime on another side of world, there are about 844 million people who experience limited excess to clean water (Chung, 2018). Half of water use by Malaysians can overcome the limited access of clean water by the water shortage's countries if there is even water distribution across the world. This is an illustration of uneven water distribution. Calculation of overall water consumption is easier than finding the solution to uneven water distribution. Topography and climate of the places are different. Some places suffer long drought in a year while some places are having rain along the year. It sure needs a big developed innovation to share equal water distribution across the world and it is indeed will be costly.

Figure 2.1 Drought in India caused water scarcity (Mehta & Up, 2018)

The water life cycle is important to maintain the availability of water resources in upcoming years. Virtual water caused the water use will not return 100% to the source. Virtual water is the water that embedded in the product after being process (Water Civilization International Centre, 2018). Water footprint is widely used to identify the amount of fresh water consumed and polluted.

Nowadays, water consumption by the industries is very high. For instance, paper industries which use abundant of water in the process. According to 2015 annual report of China's paper industry stated the water consumption by paper-making factories was 3.355 billion tons in 2015 (China Paper Association, 2015). The liquid of 1 ton equals to 1000 litre. 3.355 billion ton is quiet a big value if it is converted into litre. The recycled water from the paper making process was 71.96% from the water use (China Paper Association, 2015). It is about 28.04% of the water use had become virtual water which evaporated or became part of the product themselves. The industries use a lot of water to operate.

Meanwhile, the agriculture also part of the biggest water user. Actually most of the water consumptions are hidden in the production of the food. The water use behind every food production is invisible to the user. That is called as virtual water, the concept invented by Professor Tony Allan. The virtual water can be understood through the example of one kilogram beef production. In order to produce one kilogram beef, it needs approximately 15400 litre of water. The foods to feed the cow also need water to produce. Not to forget the farmhouse service and slaughtering process which use a lot of water to be run. Every litre of water is counted in order to produce a product. For example, it needs 822 litre of water to produce one kilogram apple, 2500 litre for rice, 6000 litre for pork and 18900 litres for coffee.

Figure 2.2 Water footprint for foods

(Source : http://www.eniscuola.net/en/mediateca/water-footprint-of-food/)

Domestic water use is unbelievably high. Drinking, washing, cooking and showering are domestically water use. As stated before, 300 litre of water use by a Malaysian a day (Chung, 2018). For the solutions of the problem, the water management and society behaviours play a big role. Every accomplishment starts with a first small step. In order to raise awareness in the society, it starts from home; the place people first learn things. To reduce domestic water use (Chung, 2018), first is identifying any pipe leakage which can cause the continuous wastage for long period of time. Second is the water used to wash rice or any ingredients at kitchen sink can be recycled to water the plants. Third, stop the habit of opening the water tap while brushing teeth and no long-shower. Forth, use the washing machine when there are big load of laundry. Last but not least, collect the rain water to wash car, water the plants or other domestic uses.

Water policies is a guide for the government to determine the rule and guide in managing the water consumption , water contamination and also in fulfilling the need of clean water resources (The CEO Water Mandate, 2010). It is a good way in raising awareness to expose the society about what their county's water policies are. This is because the society can know how the government manage the limited water resources in order to fulfil the society's need to clean water. Besides, it also can increase the society's knowledge about the current water issues. Indeed, the catastrophic like drought cannot be resisted but, the innovation of new water system can be upgraded.

2.2 Grey Water Footprint

Water life cycle is a model that can be illustrated about the movement of water. The transition state allows water molecules to release and gain heat energy. The water undergoes three phase which are liquid, solid and gas. The water evaporates from stream or ocean by the heat from sun. Then it form a cloud that later produce rain. The rain again occupies the river, ocean, lake or get beneath ground. The cycle keeps repeating and moving every moment. This complex process that allows the movement of water across the earth is permitting the continuity of every living thing. Water is a renewable resources but it is limited due to the increasing of world population demand.

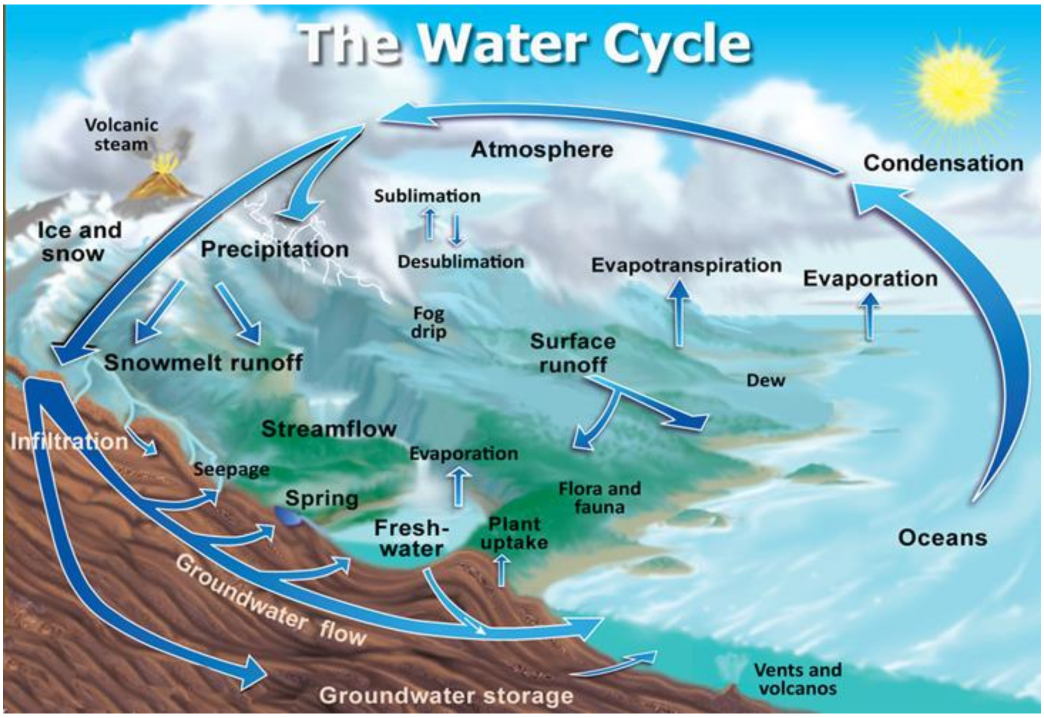


Figure 2.3 Water cycle process

(Source:https://www.usgs.gov/special-topic/water-science-school/science/water-cycle?qt-science_center_objects=0#qt-science_center_objects)

Today, water scarcity is everywhere across the horizon day by day due to unlimited used. To avoid further inefficient water use, a water footprint concept had been developed (Aldaya, Chapagain, Hoekstra, & Mekonnen, 2011). Most of the water consumption is invisible. This is why the human do not realise that they have used abundant of water in their life. Water footprint is a measurement of water use to produce goods in supply chain (Aldaya et al., 2011). There are three type of water footprint which are green water footprint, blue water footprint and grey water footprint. In production process, green water footprint is water retain in the ground. Blue water footprint is the volume of water resources from rain, river or lake use to produce goods. Grey water footprint is the water that had been polluted from the production of goods.

Water footprint is a good concept to relate with water life cycle. The production of goods will demand a big amount of clean water resources parallel to the increasing of population (Haida, Chapagain, Rauch, Riede, & Schneider, 2018). Certainly the volume of wastewater produce will be a big amount also. However, the amount of water take out from it resource to produce a good will not same as the amount of water retained at the end of the production process. The water footprint network can detect the path of the water use. Start from the blue water used to the grey water produced.

The grey water management or waste water management is quiet an issue nowadays. Untreated grey water that been disposed to the water catchment basin like river, lake and ocean will pollute the environments in long term (Pellicer-martínez & Martínez-paz, 2016). It is not easy to maintain the water resources quality with increasing of population. The municipal water waste should be treated in the right way so that the water back to it sources in good state. Water consumption keeps increasing so, to replace the water use, treating the grey water is one of the ways to maintain its continuity. Plus, the society also should take part by reducing daily water use so that wastewater volume can be reduced.

Many countries are still using conventional wastewater treatment system to treat wastewater from industries, agricultures and domestics. In order to fulfil the future's demand, the conventional system already is not relevant due to large amount of wastewater and density of pollutant in the wastewater itself (Hague & Office, 2004). For example, in urban area where the population density is high and space is limited, it needs on-site treatment or collected to be treated elsewhere (Hague & Office, 2004). Indeed, a conventional wastewater treatment requires low cost to be establish but it needs a big space than before and more energy to treat the wastewater in this day. The quality of the treated water also is not guarantee clean even it is already been treated.

That is how the river and ocean, the discharge point of treated wastewater become polluted because of the incomplete wastewater treated.

According to (Haida et al., 2018), grey water footprint is the volume of water needed to assimilate the pollutant so that it can meet the clean water quality standard. A study conducted by (Haida et al., 2018) trying to relate climate change adaptation and capacity development with water footprint concept. This study first cooperates with a school in Austria to provide starting point for WF assessment. The pupil became the change agents by reducing 9% of their WF. To a good water management, it required a shared responsibility between consumers, governments, businesses and investors who play different roles each (Bauwens, Lobe, Segers, & Tsaliki, 2009). This study relate to Climate Change Education in improving the clear figure of climate change. Also, it provides current and future generations to reflect to its obstacles by enhancing the individuals to manage its complexity.

2.3 Algorithms

2.3.1 What is algorithm?

As stated by (Horowitz, Sahni, & Rajasekaren, 1997), the word algorithm is existed because of mathematician Muhammad ibn Musa al-Khwarizmi on 9th century. This word and its concept have been used in computer science as a method to the problem solution pathways. However, the word algorithm is more special than the word process, technique or method due to its complexity. Algorithm in computer science means a finite set of instruction that need to be followed to complete a task (Horowitz et al., 1997). Hence, the algorithm has its own characteristics which allow it functioned at its best. The criteria are the input must be zero or more in clearly supply, the output should be produced at least one, the definiteness which is the clear instruction, the finiteness is when the instruction of the algorithm is ended after few steps of instruction and the last criteria is the effectiveness where every instruction must be simple and basic (Horowitz et al., 1997).

2.3.2 How to choose the best algorithm?

There is a tremendous variety of candidate algorithms in Data Mining to select essential features according to certain criteria during the pre-processing step (Rafael, Parmezan, Diana, & Chung, 2017). This wide availability of algorithms performing the task of Feature Selection creates the difficulty of choosing between the algorithms at hand and the most promising one for a specific problem (Rafael et al., 2017).

In pre-processing or data mining, there are steps to ensure the data provided can be run successfully using selected algorithm. Data Mining is considered one of the most expensive steps because it can consume almost 80% of the entire process (Rafael et al., 2017). The correct planning and preparation of the data is of major importance to ensure the data quality (Rafael et al., 2017). From the practical perspective, the selection for a particular choice algorithm should be based on two main elements: the technical understanding of the chosen algorithm dependent on the computer experts and the knowledge of the domain reliant on the domain specialists in general (Rafael et al., 2017).

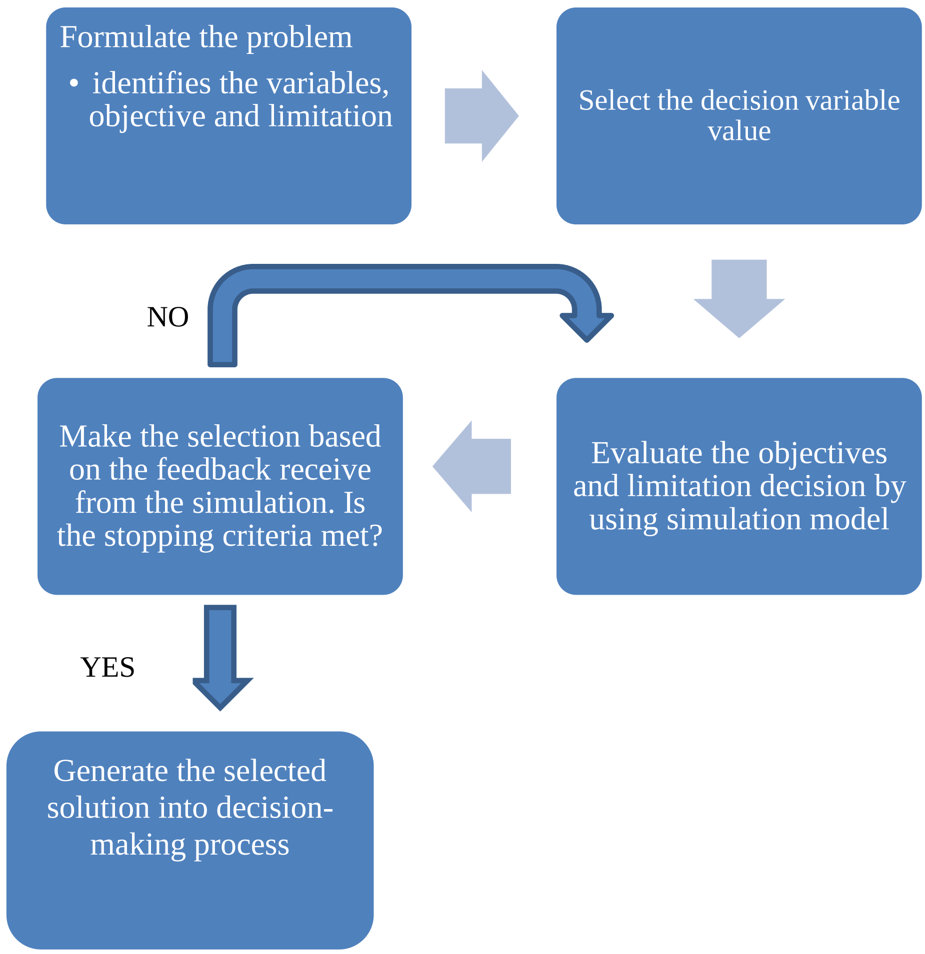
However, the great availability of algorithms can not be used to its maximum since the difficulty of deciding between algorithms with different features increases based on the particularities of the problem, since no algorithm can be considered the best regardless of the problem.

2.3.3 Application of algorithm

(Maier et al., 2014) in their study revealed that evolutionary algorithm (EA) has been widely used in water resources analysis lately. The algorithm application in research areas are calibration of model, water distribution system, management of groundwater, planning and management of river-basin are investigated. Besides, they also measure the obstacles and future route of applying EA in research study (Maier et al., 2014). It is important for the researchers to fully understanding the case study and solutions, the characteristic of the chosen algorithm and performances of the algorithm in recent research. This is to ensure the understanding of the algorithm performances in the study - why the algorithm reacts that way which resulted to good or worse performances. Rather than just simply conclude the algorithm is performing good or bad on specific case study (Maier et al., 2014).

Over two decades, computer science knowledge had been upgraded. The application of algorithms and mathematical formula in water resources system had been popular in research field. The researches had upgraded the algorithms and their application in many areas like calibration of model, water distribution system, management of groundwater, planning and management of river-basin, etc (Maier et al., 2014)

Figure 2.4 Step to optimise the algorithm process (Maier et al., 2014)



2.4 Artificial Neural Network (ANN)

2.4.1 What is ANN?

A nonlinear flexible model is created to imitate biological neural system called as artificial neural networks (ANNs) (Kuan, 2000). Usually, the biological neural system is formed by several layers compose by a big number of neurons that can generate information in parallel (Kuan, 2000). At first stage of ANN introduction, their capability and data generating were minimal caused ANN did not recognised by many people (Kuan, 2000). In line with research field development, ANN has been developed into more complex and flexible structure. Until then, ANN gets more attention in research field. ANN can operate like a brain which can receive input and make a conclusion. They can discover, create and organize information. ANN is not yet discovered to the fullest but they already contribute in upgrading current development like information classification, information interpretation and data forecasting (Staub, Karaman, Kaya, Karapınar, & Güven, 2015). The field as medicine, engineering, finance and manufacturing are using ANN to manage their data (Staub et al., 2015).

Artificial neural network (ANN) is a supervised machine learning (Olawoyin & Chen, 2018). Machine learning is a computational intelligence that simply programmed a model to learn finding unseen pattern (Olawoyin & Chen, 2018). The complex pattern and trend or non-linear data can be detected by ANN. There are many applications using ANN like controlling the machine, prediction of time series, recognition of handwritten and many more (Olawoyin & Chen, 2018).

Before, in forecasting time-series data, physical or conceptual model had been used as a conventional method. But, they had been criticised due to difficulty in prediction application, need many type of data and complex model built. To overcome the difficulty, data driven model had been developed with minimum data needed, faster in processing and higher accuracy in prediction. Most of the data driven model developed only can process linear data forecasting. This limited progress had led to existence of Artificial Neural Network (ANN) to forecast non-linear data. After that, ANN has become a popular method in predicting data time-series.

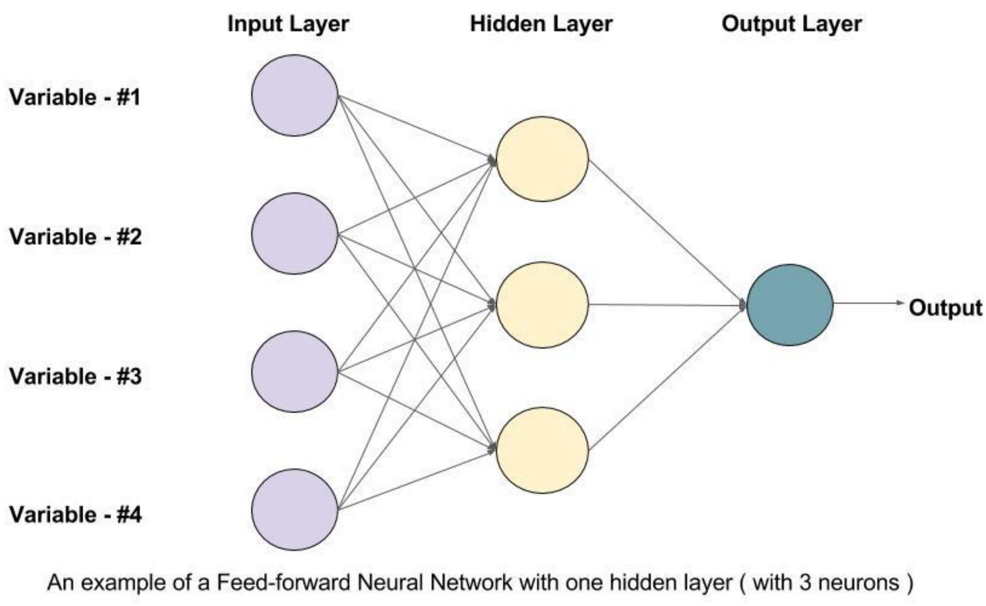


Figure 2.5 Structure of ANN
(Source : https://www.learnopencv.com/understanding-feedforward-neural-networks/)

An ANN is a computer model influenced by the nervous system's basic unit: the neuron. The model simulates how information is processed in animal brains, how it is organized and interconnected (Jimenez-martinez & Alfaro-ponce, 2019). The main features of ANNs are their ability to reproduce complicated nonlinear input-output interactions, apply linear training procedures and adapt themselves to input data (Jimenez-martinez & Alfaro-ponce, 2019). These characteristics are the result of their functionalities that form artificial neuron layers connected by adaptive weights. Training is the method used to adjust the ANN weights.

2.4.2 Application of ANN

In 2018, a study on Atrificial Neural Network had been created to forecast the future (Olawoyin & Chen, 2018). At earlier stage of ANN existence, there was only single-layer neural network which had no hidden layer. There was only input neuron connected by weight to process output neuron using Linear Threshold Unit (LTU) activation form.

The single-layer output is defined by

$$t = w_0 + w_1 x_1 + w_2 x_2 \ldots\ldots + w_n x_n$$

Where:

t = threshold

w = weight

x = input

However, the single layer neuron in this single-layer network unable to solve multi-layer neuron which is more complex and only allow for linear calculation only. To solve this limited capability of ANN, the Multilayer Perceptron (MLP) was invented. This is a feed-forward neural network made up from inputs, hidden layers and output (Olawoyin & Chen, 2018).

A study by (Danandeh Mehr, Kahya, Şahin, & Nazemosadat, 2015) explored the prediction on successive-station monthly streamflow using various artificial neural network algorithms. Due to inefficient rain gauge catchment, the monthly stream-flow records were used in this study. A search tool, feed-forward back-propagation (FFBP) neural network algorithm has been used to investigate the best scenario for the river. Then, the selected scenario modelled using generalized regression neural network (GRNN) and radial basic function (RBF). The result showed in order to get right model and good value, one-month-lagged record is enough. RBF also outperform than FFBP and GRNN.

A wavelet neural network and wavelet support vector regression model had been used to predict long-term drought in Awash River Basin in Ethiopia (Belayneh, Adamowski, Khalil, & Ozga-zielinski, 2014). Drought prediction can give a good information to avoid and prepare for disaster in future. The five data driven models were used to estimate the efficiency for long-term (6 and 12 months time) drought forecasting. WA-ANN and WA-SVR models are developed to run the input. The traditional stochastic model (ARIMA) was compared to technique of machine learning like artificial neural network (ANN) and support vector regression (SVM). The parameter is used to compare the performances were using RMSE, MAE and R^2. The result showed WA-ANN models outperformed than other models in predicting standard precipitation index (SPI) of 12 and 24 by 6 and 12 month lead time.

To be exact, this study will use ANN application to forecast the trend of grey water footprint of Semambu and Panching water treatment plant which the water resources come from Kuantan river basin. There are many studies use ANN application to predict the trend of time series data. In order keep in track or more advance movement, this study will make use of ANN mechanism. The reality that many researchers start to interest in ANN and create solution through this mechanism is a good news which will lead to more upgrading development and technologies.

2.5 Random Forest (RF)

2.5.1 What is Random Forest?

Random forest is learning methods of classification, regression and other tasks (Gong, Sun, Shu, & Huang, 2018). The random forests could be a methodology that Breiman has introduced, adds an extra layer of randomness to aggregate bootstrap and is found to perform well compared to several alternative classifiers. It's strong and extremely easy against overfitting (Mustapha & Abdelwahed, 2019). It is a great data mining model is an ensemble algorithm centered on decision trees to reduce the variance of the model while avoiding bias (Chong, Zhu, Luo, & Pan, 2019). It has been shown to be a useful method that is often used in the field of data science. Random forest data sets both observations and training data variables to develop many independent decision trees and take majority vote for classification (Chong et al., 2019)

The advantages of Random Forest is nonlinear parameter relationships do not effect RF performance and RF can handle statistical data with highly skewed or multi-modal predictors as well as categorical predictors with either ordinal or non-ordinary structure (Chong et al., 2019). Next is RF uses boot strapped sampling to develop independent decision trees that can avoid overfitting. Besides, RF can assess the significance of each predictor.

2.5.2 Application of Random Forest

Random forest (RF) algorithm had been used by (Gong et al., 2018) to forecast international roughness index (IRI) asphalt pavement. RF was used successfully in previous machine learning studies. In this study, RF model was developed and more than 11000 samples data is collected from long term pavement performance (LTPP) program. The efficiency of RF was compared with regularized linear regression model. As the outcome showed that the RF outperformed the linear regression model. The factors that gave crucial effect on IRI were transverse cracking, fatigue cracking, rutting, annual average precipitation and service age

CHAPTER 3

METHODOLOGY

3.1 Introduction

At first, this study will be started by doing the preliminary study which is background study about current water scarcity and water quality that happen around the world. The study includes the background study, problem statement, identify objectives and scope, significance of study and literature review. However, the scope of this research is only cover at Kuantan’s water treatment plants. The contents of preliminary study are all written in chapter 1 and chapter 2.

After all the study about water scarcity and water quality in water treatment plant in Kuantan, the water footprint concept is chosen to calculate the total of grey water footprint in panching and semambu water treatment plants. Water footprint is a tool to assess the water quality and grey water is in water footprint’s categories which mean the amount of freshwater needed to assimilate the pollutant.

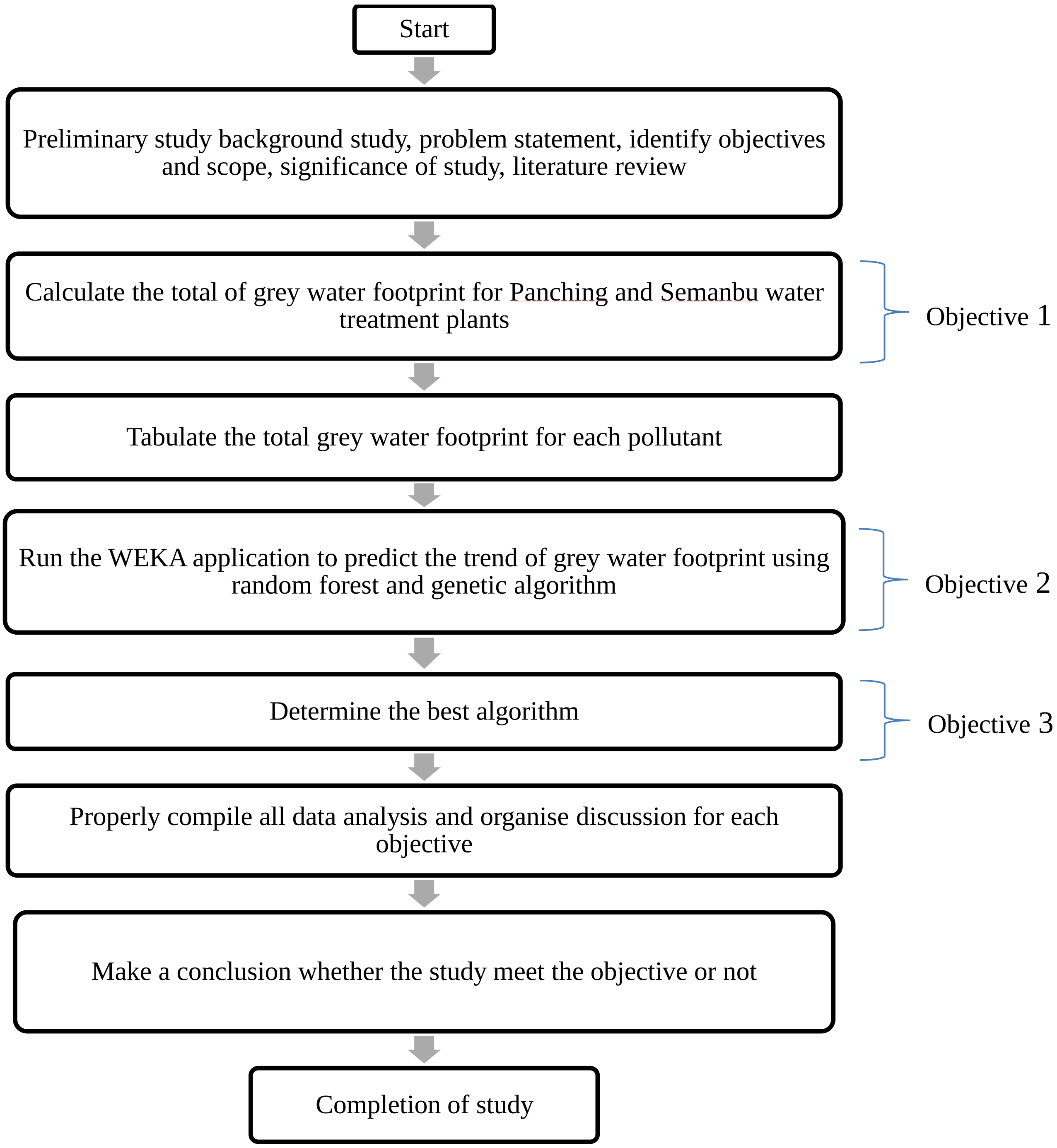
The total of grey water footprint will be calculated in excel. the formula used will be refer in the water footprint manual (Aldaya et al., 2011). The excel software will ease the calculation where calculation will be counted automatically based on the formula that has been set in the excel software. All needed is water quality data to key in excel table. The total of grey water footprint will be tabulated based on the type of pollutants contain in the water resources.

Next, the result of the total grey water footprint will be run in the WEKA software to predict the trend of grey water footprint using Random forest and Artificial Neural Network. WEKA software is a machine learning data processing and contains collection of algorithm for predictive data analysis. This software is chosen because it is a free access software and user-friendly. After the data is run in the software, the result will produced the prediction graph of the grey water footprint trend according to the algorithm used.

The best algorithm will be chosen between Random Forest and Artificial Neural Network. The root mean square error (RMSE) is the best parameter in choosing the best algorithm. The lower RMSE means the error is small. So, the lower RMSE between two algorithm will determined the best algorithm. Next, the trend of grey water footprint will be predicted by using the algorithm involved. the prediction trend is starting from 2018 until 2020.

At last, all the data analysis will be compiled and organised before the discussion is made for each objectives. The discussion is made to elaborate the objectives and results. After the discussion is elaborated, the conclusion is made whether the study meets the objectives or not.

3.2 Flow Chart



3.3 Study Area

This study is covering only at Kuantan river basin, focusing on water treatment plant of Semambu and Panching. The water treatment plants are chosen because there are many cases of water scarcity happen in Kuantan throughout the years until now. This study is also aim to upgrade the water distribution system in Kuantan with good water quality.

Semambu Water Treatment Plant is located at Jalan Pintasan Kuantan, Semambu, 26100 Kuantan, Pahang. It is the first of the biggest water treatment plant in Kuantan with the output of 330 million liters of treated water per day. The water supply is distributed to the west area of Kuantan river which start from Bandar Indera Mahkota to Kemaman border.

The Panching Water Treatment Plant at Felda Panching Timur which was completed in September 2013, has the capacity of 160 million litres daily and currently serves 14,000 acres of industrial area and 140,000 local population. The treated water is distributed along west area of Kuantan river starting from Bukit Rangin to Gambang.

Semambu Water Treatment Plant is the first biggest water treatment plant in Kuantan followed by Panching Water Treatment Plant. Semambu WTP is first built before Panching WTP. Both WTP can fulfilled the water demand around Kuantan by 500,000 population.

Figure 3.1 Location of Water Treatment Plants (Source: Google maps)

3.4 Data Collection

The table below is the type of data used in this study and the department involved:

Table 3.1 **Data collection**

DEPARTMENT	DATA
Ministry of Health (MOH)	Water quality data
Malaysian Meteorological Department (MMD)	Temperature
Jabatan Pengairan dan Saliran (JPS) Pahang	Rainfall Intensity
Pengurusan Air Pahang Berhad (PAIP)	Design plan of WTP

3.5 Site Visit

Site visit is crucial in any research to let the researcher evaluate and experience the site themselves. It is a good way to obtain the desired data as we all know primary data is great to be in the study because it illustrates the originality of the study.

Most of the data used in this study are secondary data. The data are collected through official procedure at the departments relate. However, site visit also is carried out to collect the data which are not available in the department and to get the information about the surrounding area of water treatment plants.

3.6 Water Supply Treatment Process

As stated before, this study is carried out only in water treatment plant of Semambu and Panching. The samples are taken started from the extraction of water resources from nearby water supply until the point of treated water distribution. In water treatment plant, there are many stages that the water will go through too eliminate the pollutant before reaching the consumers. Below is the process in conventional water treatment plant:

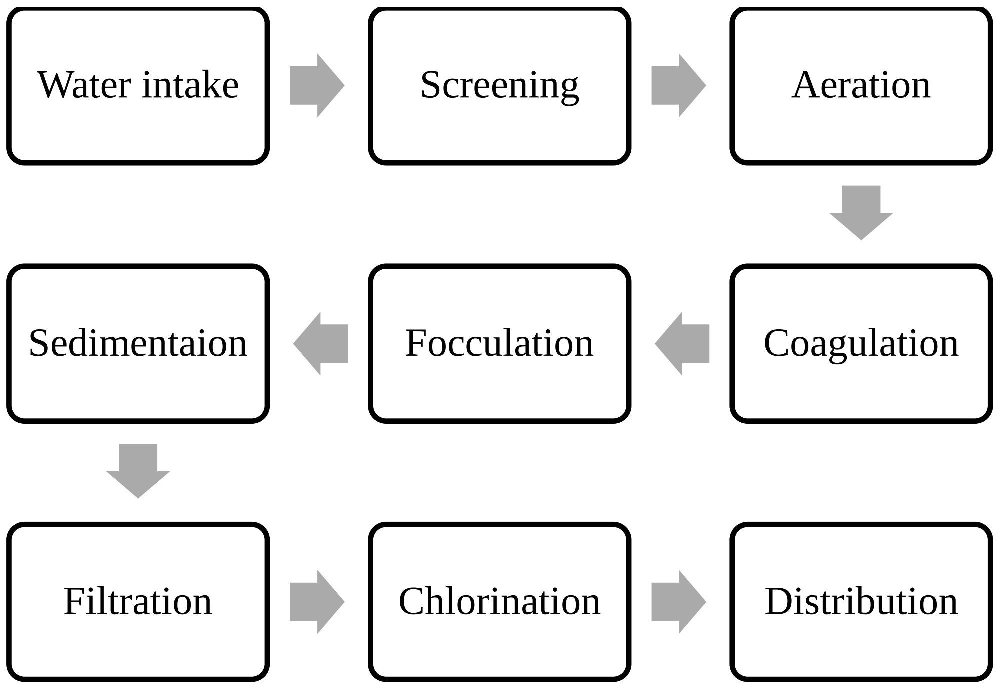


Figure 3.2 Flow process of water treatment plant

The process is starting by taking water from nearby water resources. Next is screening to get rid of big particle from the water. Aeration is a process of movement of water at high velocity to eliminate dissolved gasses and metals. After that, coagulation and flocculation process are carrying on. During coagulation, aluminium sulphate is added to the raw water. When added, the tiny particle of dirt stick together form large and heavy particle which is called floc. This floc particles is easy to remove by settling or filtration. Sedimentation is where the floc particles settled to the bottom (sludge) and filtration is the movement of the water that passes through many layers to remove dirt particles in water. Last stage before distribution is disinfection or chlorination to ensure diseases and bacteria are destroyed before distributing to the consumers.

3.7 Grey Water Footprint Accounting

As for the calculation of grey water footprint, we will be using water footprint manual (Aldaya et al., 2011). The grey water footprint is:

Formula 1:

$$WF_{grey} = \frac{L}{C_{max} - C_{nat}}$$

Where:

$WF_{grey} = grey\ water\ footprint\ (volume/time)$

$L \quad = load\ of\ pollutant\ (mass/time)$

$C_{max} = allowable\ concentration (mass/volume)$

$C_{nat} = natural\ concentration\ (mass/volume)$

Formula 2:

$$Load\ of\ pollutant, L = flow \times concentration\ of\ pollutant \times 86.4$$

Formula 3:

$$C_{nat} = conc.of\ pollutant\ in\ water\ intake - conc.of\ pollutant\ in\ treated\ water$$

3.8 Pre-processing

Pre-processing data is a must step in order for the WEKA Software to run the data process. The first step of pre-processing is treatment of missing data. the missing data is treated by using average method . This method is chosen because the value of one parameter has no much different. So, using average method is not affecting the data much.

Next is data normalisation. This step is crucial to make sure the data is compatible with data mining software. The data must be in the rang of 0 to 1. All the parameters will be normalize so that there is no significance different between parameters values.

Last step in pre-processing is cleaning the data. The data is cleaned by removing the outliers or the significance difference between values. The big values in parameters are not prefered because they can affected the end result.

3.9 Prediction of Grey Water Footprint

After the calculation and pre-processing of total grey water footprint data for each pollutant, the value will be tabulate in excel so that it is systematically organised. Later, the excel document will be convert into Comma Separated Value (CSV) file in order for the data to be run in WEKA software.

WEKA software is a machine learning which can perform data mining task. This software will be used to predict the grey water trend in Semambu and Panching water treatment plants for three years. WEKA contains algorithm which can generate data analysis and prediction for examples, data processing, clustering, classification and regression. For this study, Artificial Neural Network and Random Forest are utilised to choose the best algorithm between the two and predict the trend.

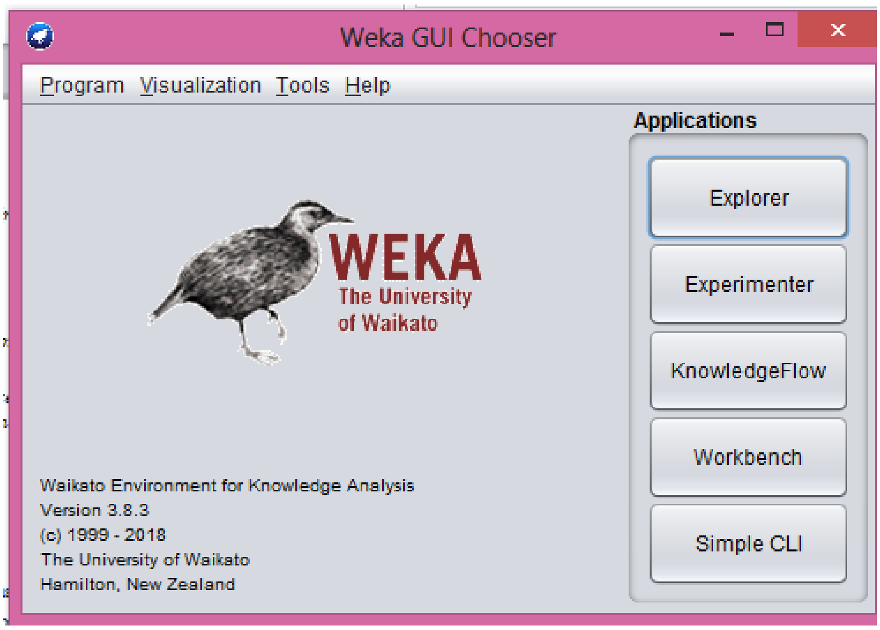


Figure 3.3 WEKA Software interface

Step for data mining in WEKA Software

Step 1:

The data in Excel is converted into Comma Separated Value (CSV) format before inserted into WEKA Software

Step 2:

Normalize the data so that the data range is between 0 to 1

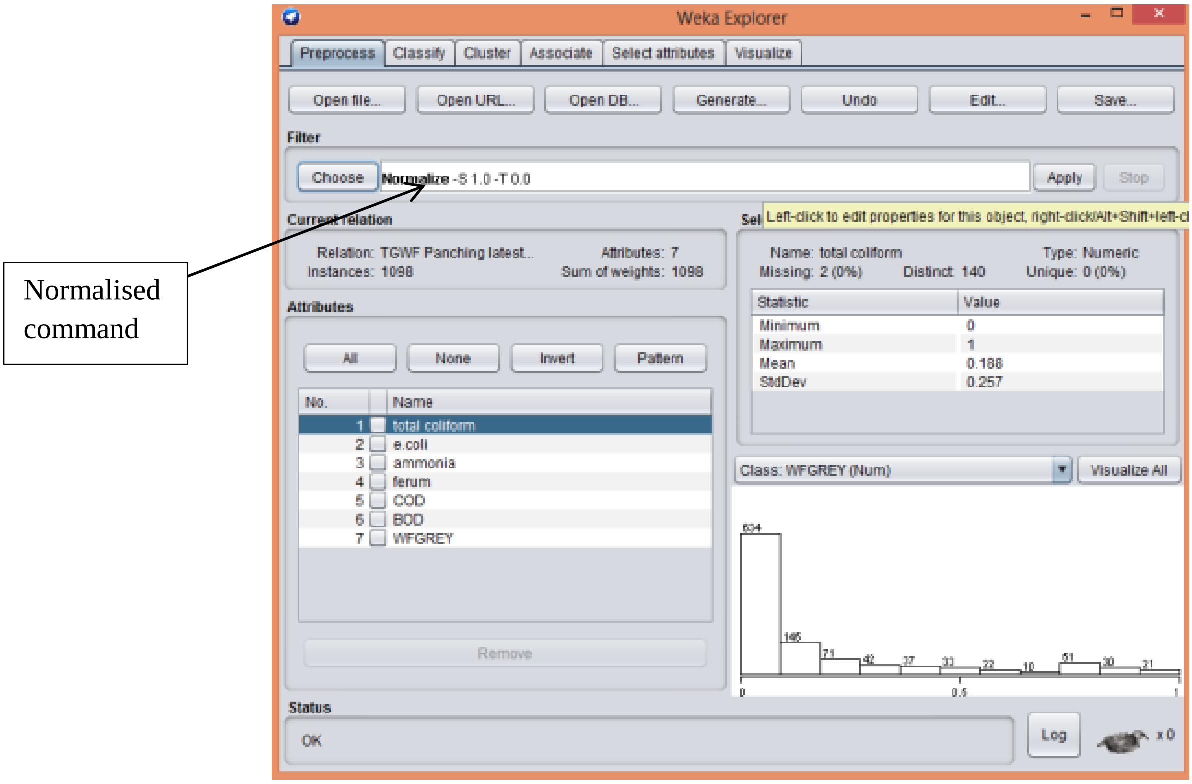


Figure 3.4 WEKA Software

Step 3:

The data then trained by using ANN and Random Forest. The Root Mean Square Error (RMSE) must be near to 0.

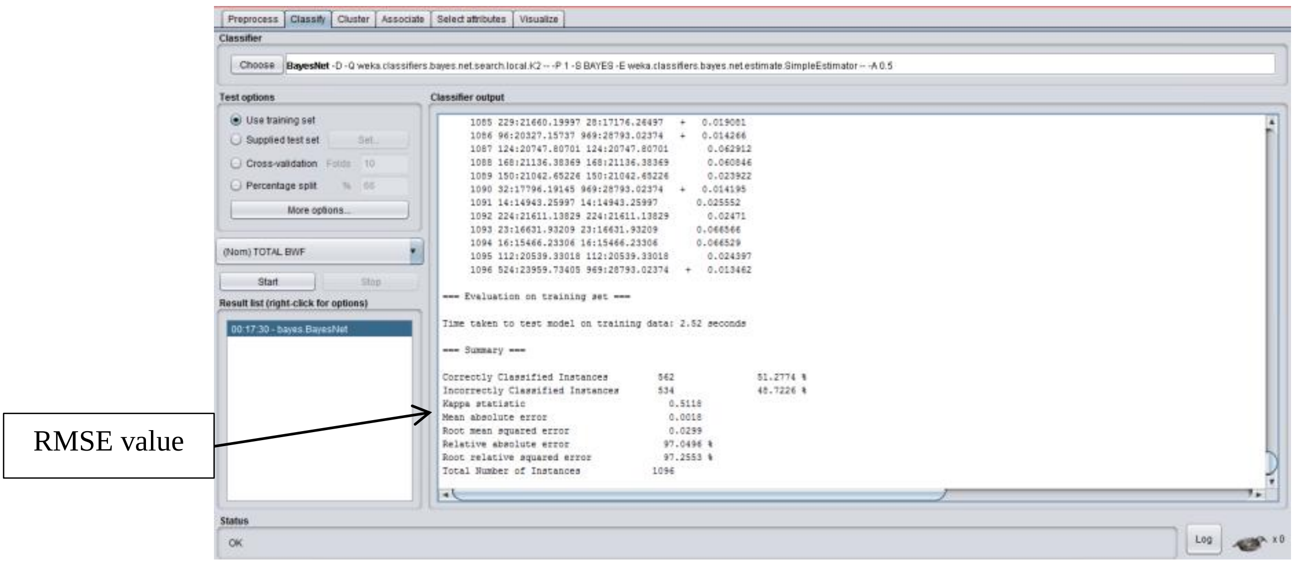


Figure 3.5 WEKA Software

3.10 The Best Algorithm for Prediction

There are two algorithm will be used in predicting the total grey water footprint in this study; Artificial Neural Network and Random Forest.

In this study, the best algorithm is defined by the lowest Root Mean Square Error (RMSE) recorded by WEKA. As stated by its name, RMSE is the indicator of model efficiency (Kumar, Goel, & Bhatia, 2014) by measuring the error of real value with predicted value.

Mostly, in determining the lowest RMSE, the RMSE value must range between zero to one. The nearest the RMSE value to zero, the best the algorithm it is and the lower the error while predicting.

After the best algorithm is selected between ANN and Random Forest, the trend of total grey water footprint will be predicted by using both algorithms. The predicted trend is starting from 2018 until 2020.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

All the results obtained in calculating Grey Water Footprint (WFgrey) will be portrayed in this chapter. The result will be analysed and discussed here. In calculating grey water footprint, there are parameters need to be considered for example load of pollutant in the sample, natural concentration and water quality according to standard of Malaysia river.

Over past few years, the pollution in Kuantan's river became more serious due to many factors for instances, increasing in population, untreated effluent discharge by factories into river, illegal logging and bauxite mining. The chemical concentration in river water had exceeded the allowable chemical concentration. The polluted river can affect the public health if the chemical concentration is not controlled at right amount or simply said if the water is not treated properly before distributing to the public.

4.2 Total Grey Water Footprint

The first objective of this study is to determine the total grey water footprint for two Water Treatment Plants (WTPs) which are Semambu WTP and Panching WTP. Both water treatment plants extract the raw water resources from Kuantan river.

4.2.1 Total Grey Water Footprint of Semambu WTP

4.2.1.1 Total WFgrey of Semambu WTP in 2015

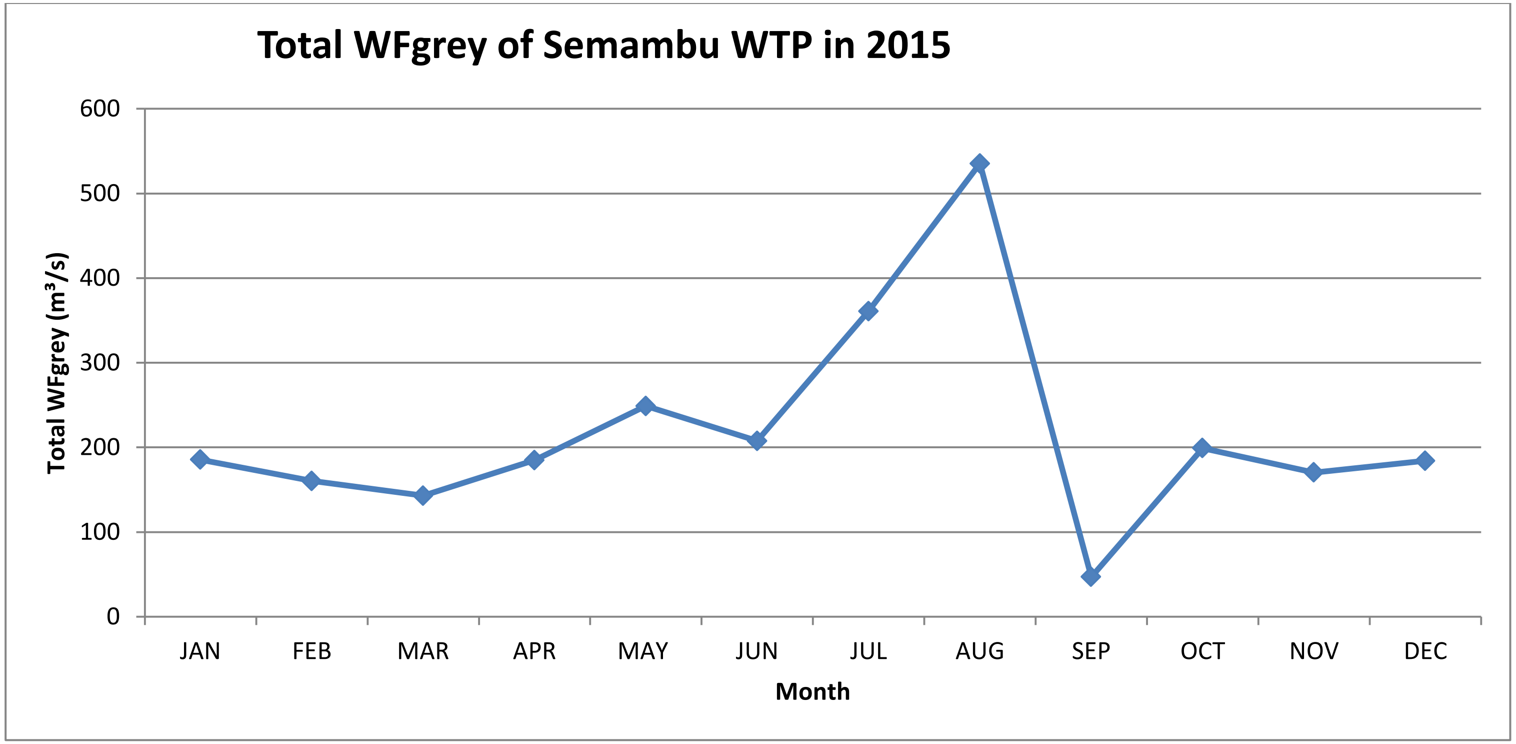


Figure 4.1 Total WFgrey of Semambu WTP in 2015

Table 4.1 The pollutants contribute to total WFgrey of Semambu WTP in 2015

MONTH	Total Coliform	E.coli	Ammonia	Iron	COD	BOD	Total Wfgrey (m^3/s)
JAN	0.008	0.486	4.903	26.389	9.082	25.227	66.093
FEB	0.008	0.082	12.480	83.434	3.636	60.606	160.248
MAR	0.002	0.018	13.817	51.783	11.780	65.444	142.843
APR	0.002	0.026	13.371	126.795	11.807	32.798	184.799
MAY	0.009	0.166	149.774	89.608	2.479	6.885	248.921
JUN	0.018	0.023	13.371	74.673	88.757	30.818	207.661
JUL	0.009	0.052	13.817	259.930	23.046	64.016	360.869
AUG	0.015	0.092	257.034	199.225	15.270	63.627	535.263
SEP	0.004	0.056	0.737	0.824	14.460	30.897	46.978
OCT	0.007	0.114	2.977	124.505	48.393	23.005	199.002
NOV	0.020	0.152	13.371	95.769	30.286	30.779	170.378
DEC	0.007	0.060	13.817	118.349	19.447	32.411	184.089

Figure 4.1 shows the total WFgrey of Semambu WTP in 2015. The WFgrey of Semambu WTP in 2015 is in range 46.98m^3/s to 535.26m^3/s. There is an increment on total WFgrey in August up to 535.26m^3/s. This is due the concentration of ammonia is high than usual. However, in September, there is a sudden decrement until 46.98m^3/s in total WFgrey. As stated in the table 4.1, the decrement is affected by the lowest WFgrey by ammonia and iron in 2015. The concentration of iron in Pahang river comes mostly from bauxite mining (Kusin, Syazwan, Rahman, & Madzin, 2017) and the concentration of ammonia is from decay animal or animal manure. It also can be conclude that the lowest concentration of WFgrey is due to the treated water after affected by pollution. In October, the total WFgrey back to normal range until December. Usually, the WFgrey will be decreased due to high water intake in the end of the year or early year which is caused by the monsoonal season resulting in low concentration in pollutants.

4.2.1.2 Total WFgrey of Semambu WTP in 2016

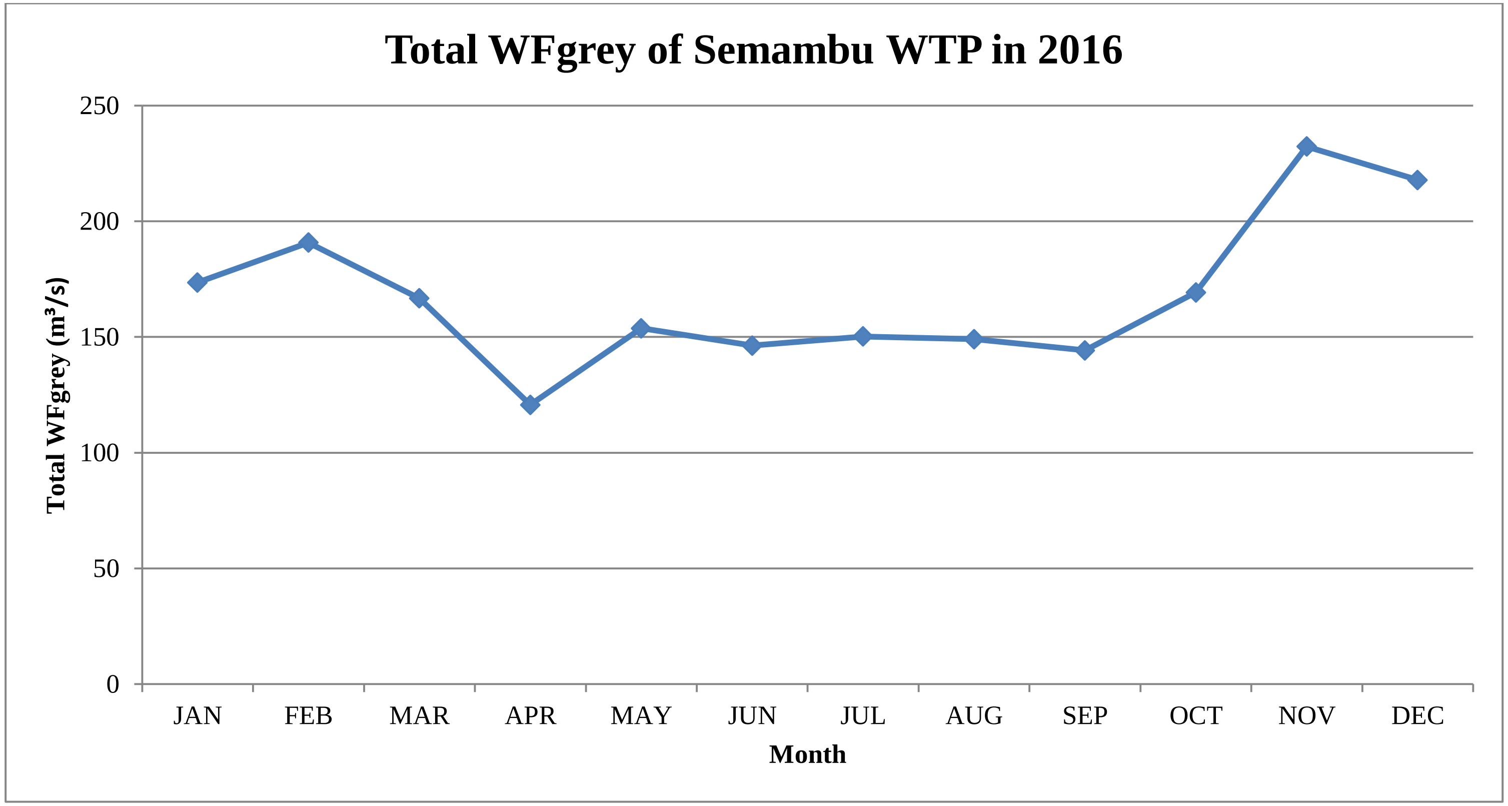


Figure 4.2 Total WFgrey of Semambu WTP in 2016

Table 4.2 The pollutants contribute to total WFgrey of Semambu WTP in 2016

MONTH	Total coliform	E.coli	Ammonia	Iron	COD	BOD	Total Wfgrey (m^3/s)
JAN	0.004	0.046	66.103	44.250	30.929	32.217	173.550
FEB	0.039	0.107	12.925	93.195	22.383	62.174	190.824
MAR	0.001	0.018	13.817	86.101	23.850	42.942	166.729
APR	0.001	0.016	13.371	22.287	22.495	62.485	120.654
MAY	0.004	0.081	13.817	86.040	21.849	31.943	153.733
JUN	0.069	0.661	13.371	70.577	30.187	31.445	146.309
JUL	0.005	0.036	13.817	82.249	24.914	29.242	150.264
AUG	0.066	0.057	13.817	47.945	58.480	28.667	149.032
SEP	0.037	0.050	13.371	46.399	56.594	27.742	144.193
OCT	0.254	0.160	13.817	103.199	25.220	26.603	169.252
NOV	0.058	0.096	13.371	150.305	38.954	29.510	232.295
DEC	0.052	0.066	13.817	74.919	41.839	87.164	217.858

Figure 4.1 shows the total WFgrey of Semamb WTP in 2016. The WFgrey of Semambu WTP in 2016 is in range 120.65m^3/s to 232.29m^3/s. There is a highest increment on total WFgrey in November up to 232.29m^3/s. This is due the high concentration iron than usual in water intake. However, in April there is a decrement until 120.65m^3/s in total WFgrey. This is due to lowest iron concentration recorded for that month. The iron concentration in Pahang river is due to bauxite mining (Kusin et al., 2017). It also can be conclude that the lowest WFgrey in April is due to the treated river water after affected by pollution. From the figure 4.2, it is showed that the total WFgrey start to increased which may causing more freshwater needed to assimilate the pollutant.

4.2.1.3 Total WFgrey of Semambu WTP in 2017

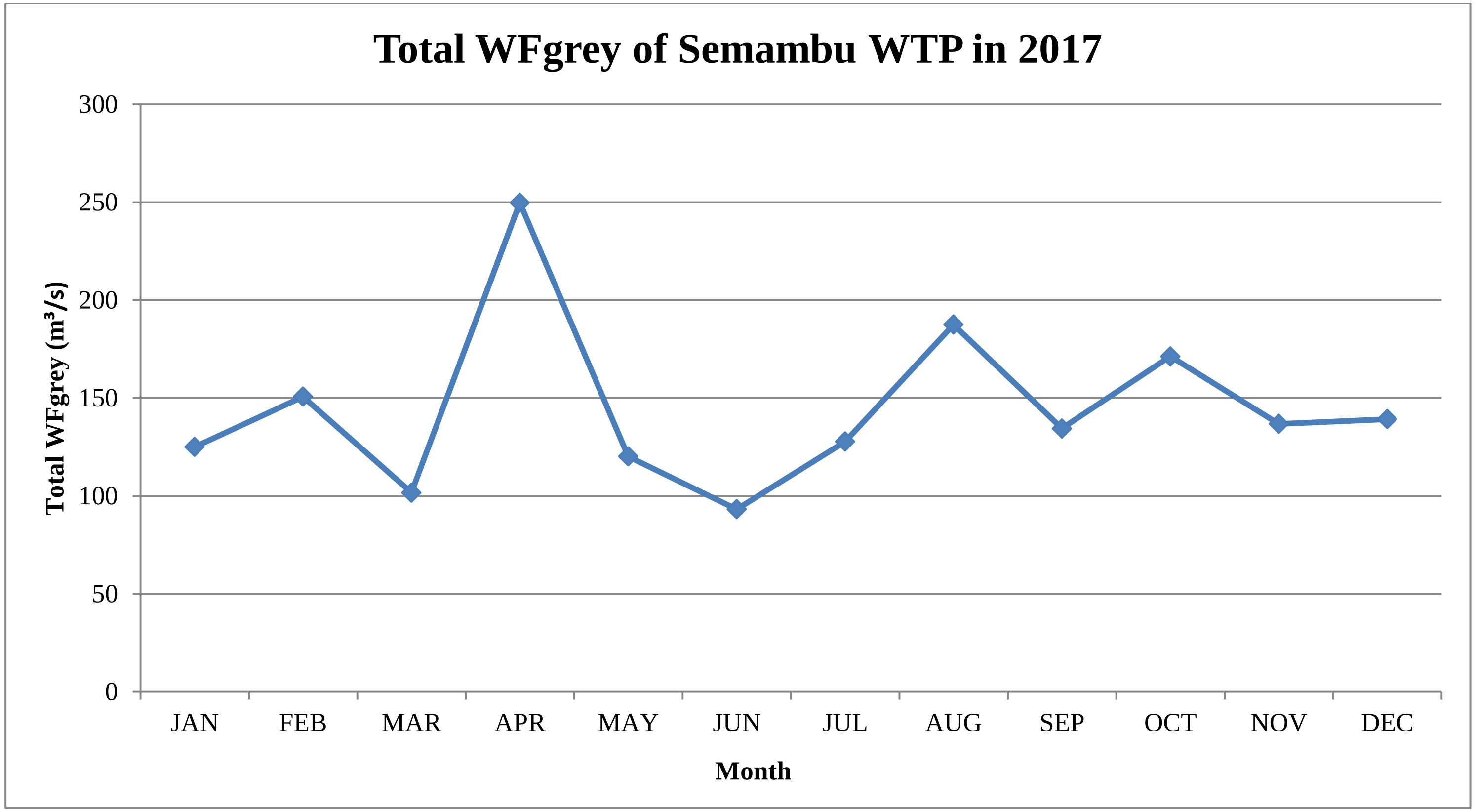


Figure 4.3 Total WFgrey of Semambu WTP in 2017

Table 4.3 The pollutants contribute to total WFgrey of Semambu WTP in 2017

MONTH	Total Coliform	E.coli	Ammonia	Ferum	COD	BOD	Total Wfgrey (m^3/s)
JAN	0.021	0.053	13.817	63.326	24.389	23.381	124.987
FEB	0.009	0.049	12.480	77.769	21.542	38.787	150.635
MAR	0.014	0.015	13.817	47.797	13.630	26.416	101.689
APR	0.012	0.080	107.788	73.427	15.793	52.644	249.744
MAY	0.011	0.068	13.817	40.809	20.056	45.364	120.125
JUN	0.004	0.052	13.371	44.431	11.450	23.854	93.161
JUL	0.009	0.093	13.817	43.160	25.619	45.071	127.768
AUG	0.024	0.092	13.817	106.788	23.850	42.942	187.513
SEP	0.039	0.065	13.371	68.977	2.944	49.060	134.455
OCT	0.032	0.048	13.817	90.570	23.850	42.942	171.258
NOV	0.035	0.059	13.371	79.574	21.288	22.455	136.781
DEC	0.024	0.068	26.874	79.510	9.483	23.243	139.202

Figure 4.3 shows the total WFgrey of Semambu WTP in 2017. The WFgrey of Semambu WTP in 2017 is in range 93.16m^3/s to 249.74m^3/s. There is an increment on total WFgrey in April up to 249.74m^3/s. This is due the high concentration ammonia than usual. The ammonia concentration in river comes from factories that do not treated or fully treated the effluent before discharge into river. It also can be conclude that the concentration of WFgrey lower in May is due to the treated water after affected by pollution.

4.2.1.4 Overall total WFgrey of Semambu WTP

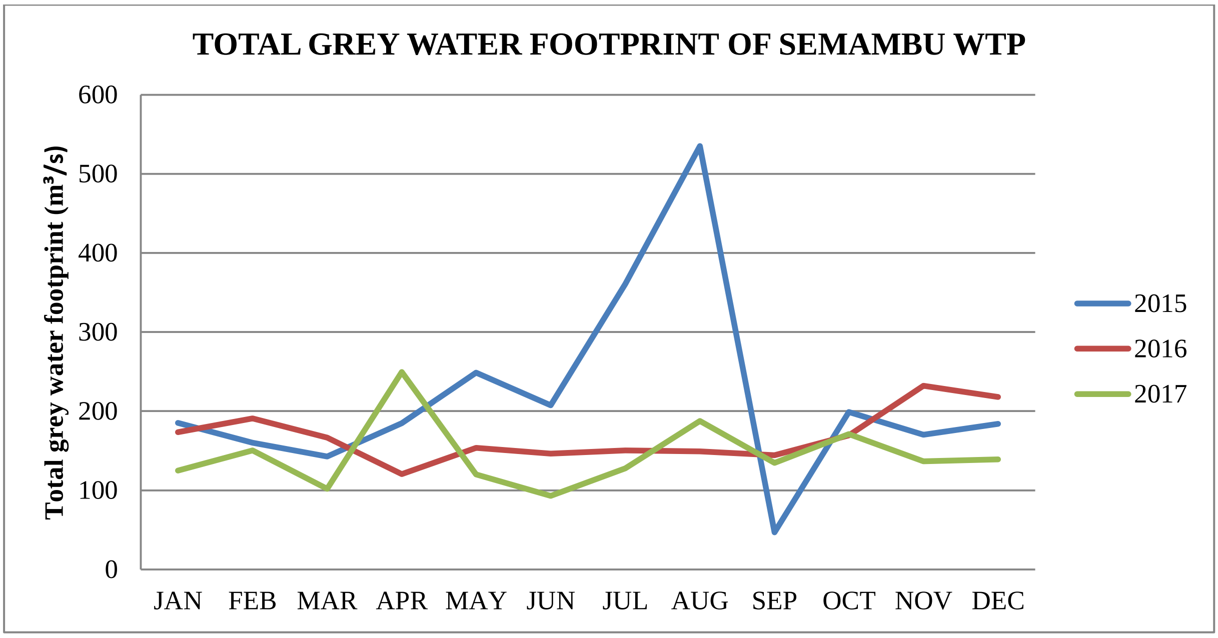


Figure 4.4 Total WFgrey of Semambu WTP

Table 4.4 Total WFgrey in Semambu WTP for each year

Year	**2015**	**2016**	**2017**
Total Wfgrey (m^3/s)	2507.14	2014.69	1737.31

Figure 4.4 shows total WFgrey of Semambu WTP for three years starting from 2015 to 2017. From table 4.4, the total WFgrey is decreasing in 30%. There was significance different of WFgrey value in 2015 where the graph was increasing in August and decreasing in September. The increasing of total WFgrey was affected by high concentration of iron (Fe^2) from bauxite mining in water intake (Kusin et al., 2017). The decreasing value of total WFgrey can be concluded as excessive river treat to eliminate the iron (Fe^2) concentration in the river. However, throughout 3 year time, the total WFgrey trend is up and down pattern in the range $51m^3/s$ to $176m^3/s$.

4.2.2 Total Grey Water Footprint of Panching WTP

4.2.2.1 Total WFgrey of Semambu WTP 2015

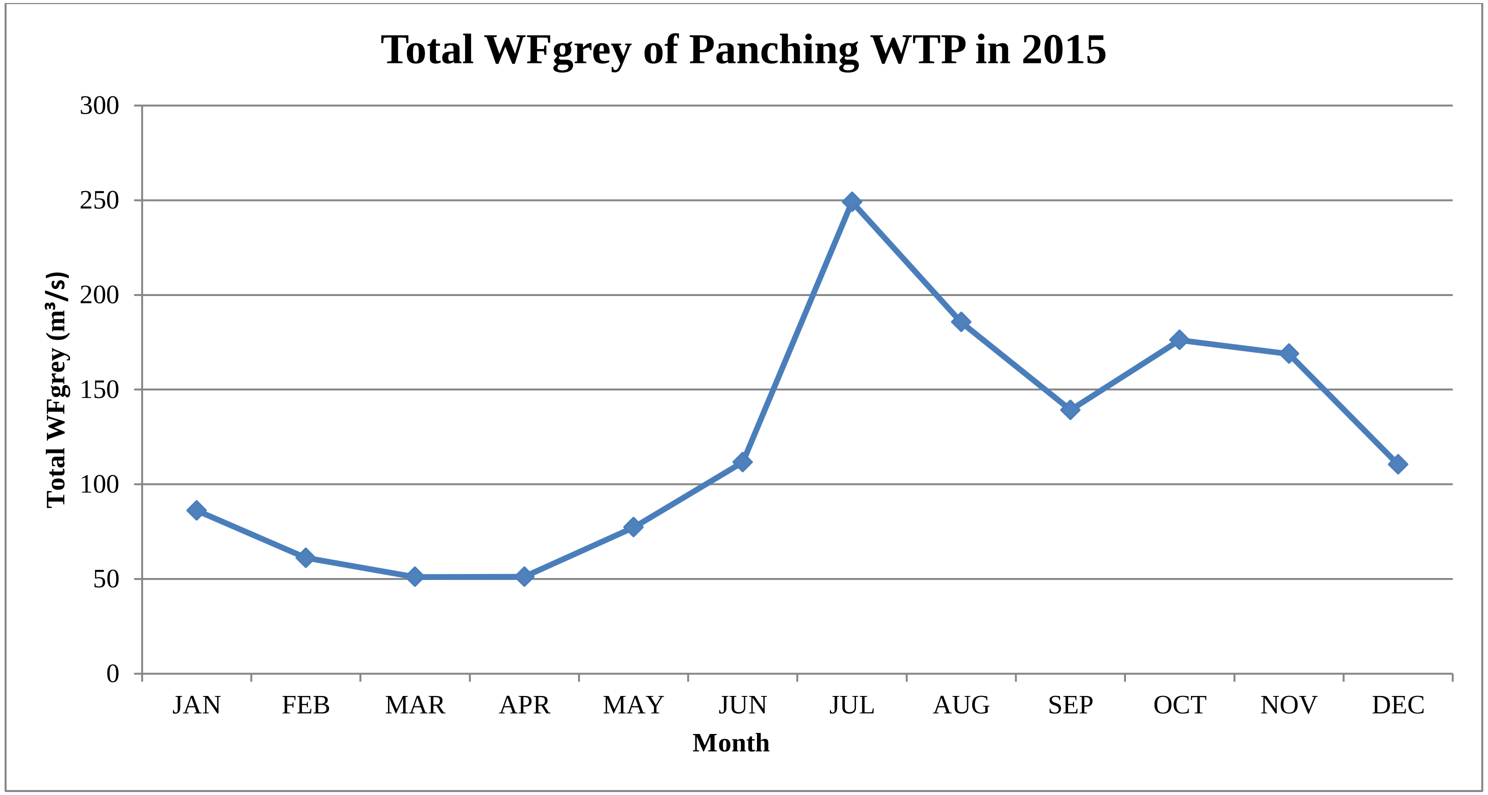


Figure 4.5 Total WFgrey of Panching WTP in 2015

Table 4.5 The pollutants effect WFgrey of Panching WTP in 2015

MONTH	Total Coliform	E.coli	Ammonia	Iron	COD	BOD	Total Wfgrey (m^3/s)
JAN	0.011	0.111	21.195	7.718	8.152	48.913	86.100
FEB	0.022	0.194	19.144	20.083	1.228	20.470	61.142
MAR	0.029	0.064	7.116	22.324	1.219	20.313	51.065
APR	0.018	0.089	1.953	32.619	6.190	10.317	51.187
MAY	0.042	0.084	21.195	32.985	6.086	16.905	77.296
JUN	0.038	0.028	20.512	53.705	21.258	16.105	111.645
JUL	0.019	0.529	21.195	178.719	12.870	35.749	249.080
AUG	0.045	0.439	21.195	115.665	12.773	35.479	185.596
SEP	0.020	0.102	20.512	95.716	6.205	16.680	139.234
OCT	0.046	0.166	21.195	108.598	12.223	33.953	176.181
NOV	0.029	0.232	84.925	60.919	6.314	16.443	168.862
DEC	0.104	0.410	21.195	55.207	11.375	22.152	110.443

Figure 4.5 shows the total WFgrey of Panching WTP in 2015. The WFgrey of Panching WTP in 2015 is in range 120.65m^3/s to 232.29m^3/s. There is an increment on total WFgrey in July up to 232.29m^3/s. From table 4.5, this is due to the high concentration iron than usual which comes from bauxite mining in Kuantan. In October to November, the total WFgrey start to drop from 17618 m^3/s to 110.44 m^3/s. This is due to monsoonal season which decreased the concentration of pollutant in water intake resulted in low WFgrey.

4.2.2.2 Total WFgrey of Semambu WTP in 2016

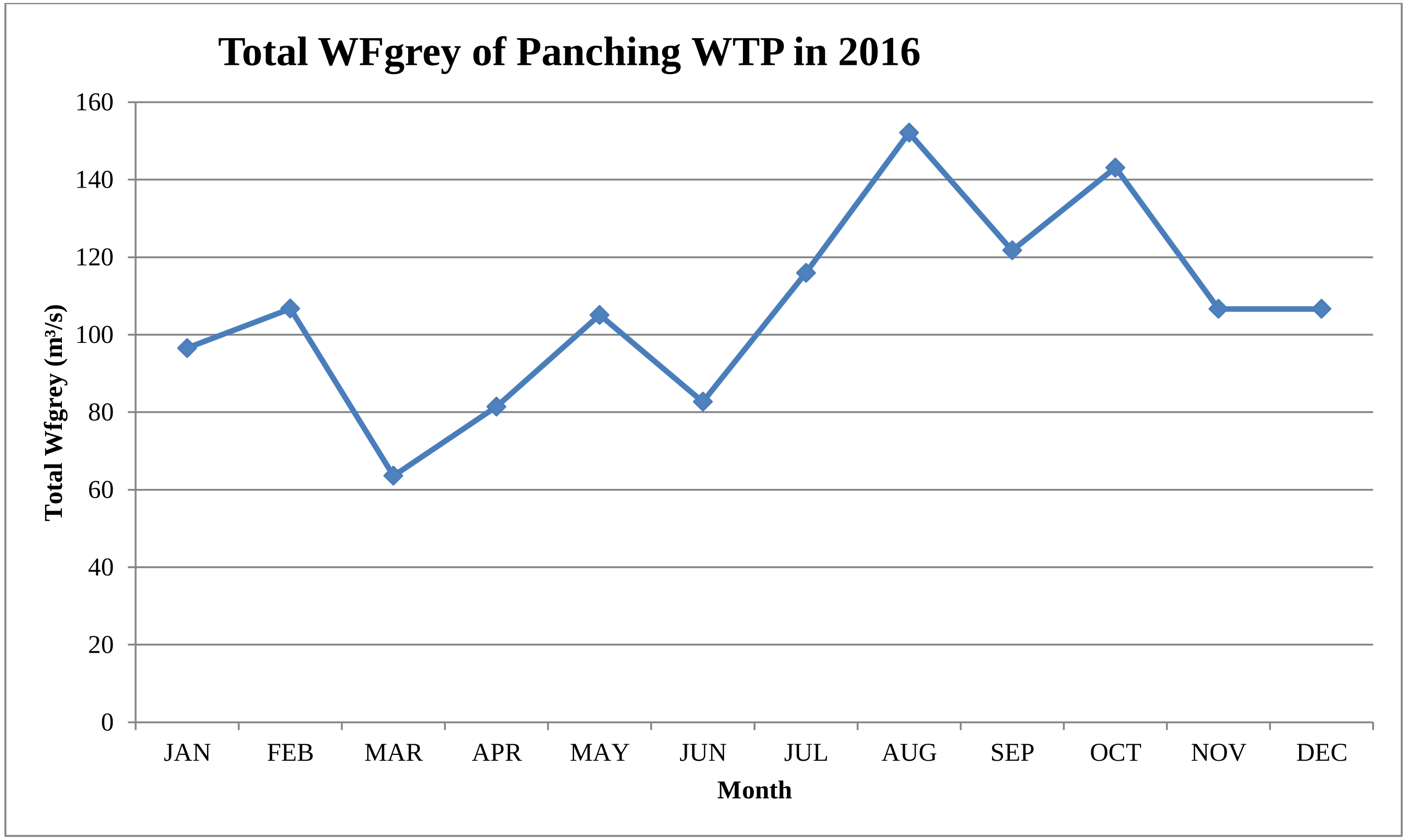


Figure 4.6 Total WFgrey of Panching WTP in 2016

Table 4.6 The pollutants effect total WFgrey of Panching WTP in 2016

MONTH	Total Coliform	E.coli	Ammonia	Iron	COD	BOD	Total Wfgrey (m^3/s)
JAN	0.006	0.029	24.690	49.304	6.104	16.410	96.543
FEB	0.022	0.202	19.828	62.035	9.355	15.286	106.728
MAR	0.022	0.147	5.774	39.937	1.725	15.972	63.577
APR	0.022	0.041	20.512	43.749	1.669	15.456	81.450
MAY	0.063	0.165	32.427	44.725	11.021	16.698	105.100
JUN	0.007	0.075	20.512	42.047	1.954	18.088	82.682
JUL	0.016	0.244	21.195	56.005	18.358	20.130	115.948
AUG	0.060	0.131	21.195	72.231	21.514	36.966	152.098
SEP	0.046	1.047	7.173	66.431	28.679	18.384	121.760
OCT	0.019	0.038	21.195	93.181	10.763	17.938	143.135
NOV	0.046	0.149	20.512	53.426	17.499	15.034	106.665
DEC	0.007	0.042	21.195	44.759	15.675	45.044	126.723

Figure 4.6 shows the total WFgrey of Panching WTP in 2016. The WFgrey of Panching WTP in 2016 is in range 63.57m^3/s to 152.10m^3/s. There is an increment on total WFgrey in August up to 152.1m^3/s. From table 4.6, the WFgrey of each pollutant is in normal range in August. However, it can be conclude that the volume of water intake is low which increase the concentration of pollutant and next resulted to high total WFgrey. As usual, at the end of the year, starting from October to December, the total WFgrey atart to decreased due to high water intake which resulted in low concentration of pollutant.

4.2.2.3 Total WFgrey of Semambu WTP in 2017

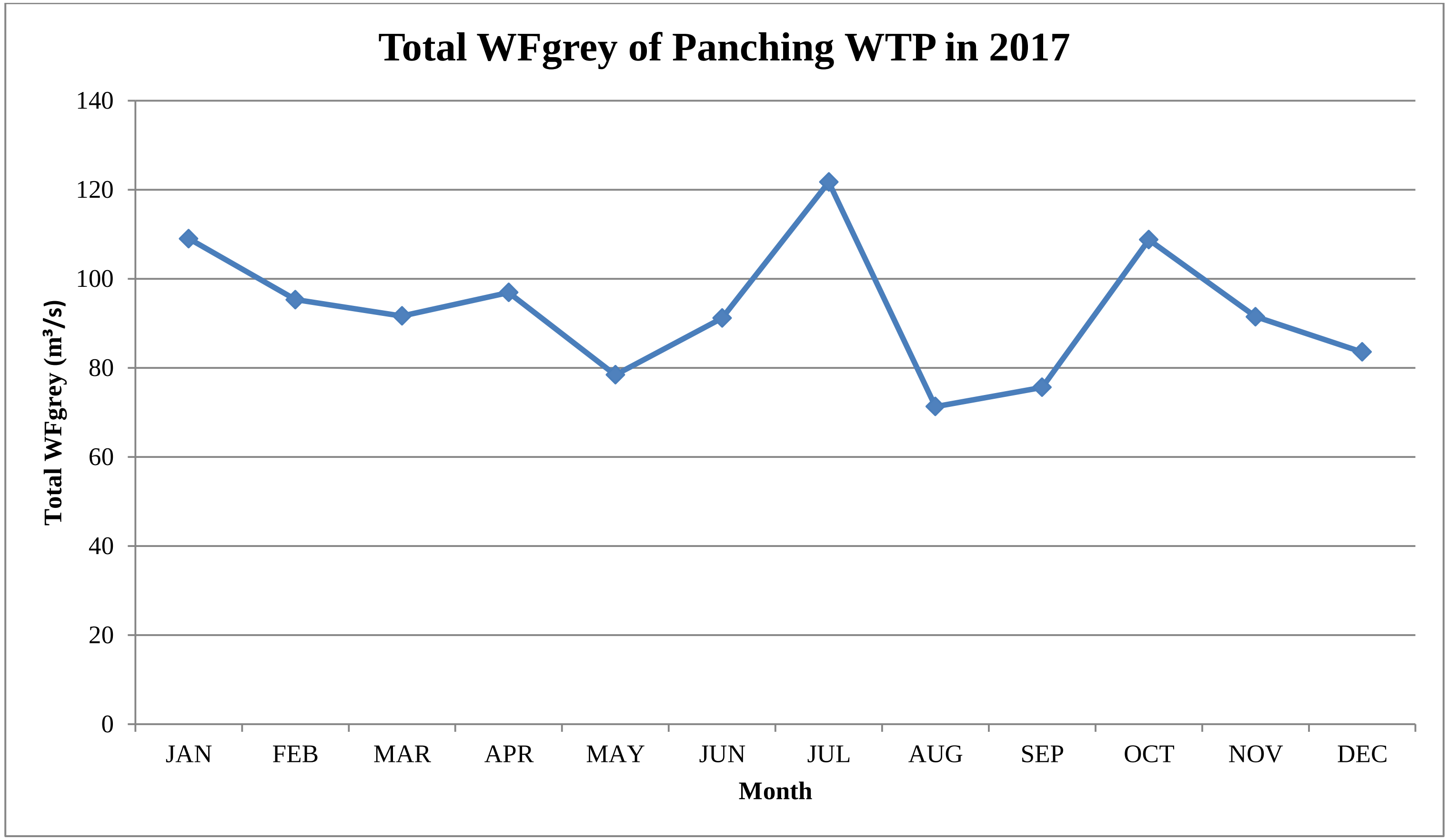


Figure 4.7 Total WFgrey of Panching WTP in 2017

Table 4.7 The pollutants effect total WFgrey of Panching WTP in 2017

MONTH	Total Coliform	E.coli	Ammonia	Iron	COD	BOD	Total Wfgrey (m^3/s)
JAN	0.032	0.051	21.195	65.504	6.026	16.199	109.008
FEB	0.013	0.056	19.144	50.538	10.711	14.876	95.338
MAR	0.015	0.063	21.195	42.272	11.769	16.346	91.661
APR	0.013	0.095	51.740	22.687	8.402	14.004	96.941
MAY	0.005	0.041	21.195	27.301	4.762	25.135	78.440
JUN	0.030	0.119	20.512	35.724	9.215	25.597	91.197
JUL	0.065	0.163	21.195	57.485	17.552	25.265	121.725
AUG	0.012	0.035	1.493	37.735	18.238	13.816	71.329
SEP	0.034	0.081	20.512	32.361	9.202	13.453	75.642
OCT	0.083	1.239	21.195	52.748	11.375	22.152	108.792
NOV	0.060	0.277	4.043	36.987	24.384	25.739	91.491
DEC	0.094	0.169	6.820	42.973	11.375	22.152	83.583

Figure 4.7 shows the total WFgrey of Panching WTP in 2017. The WFgrey of Panching WTP in 2017 is in range 71.33m^3/s to 121.72m^3/s. There is an increment on total WFgrey in July up to 121.73m^3/s. As mentioned in table 4.7, the increment is due to high WFgrey because of E.Coli. The concentration of E.Coli in July is high which resulted to high WFgrey The concentration of e.coli comes from recent sewage or animal waste like dead animal or manure. As usual, at the end of the year, starting from October to December, the total WFgrey atart to decreased due to high water intake which resulted in low concentration of pollutant

4.2.2.4 Overall total WFgrey of Semambu WTP

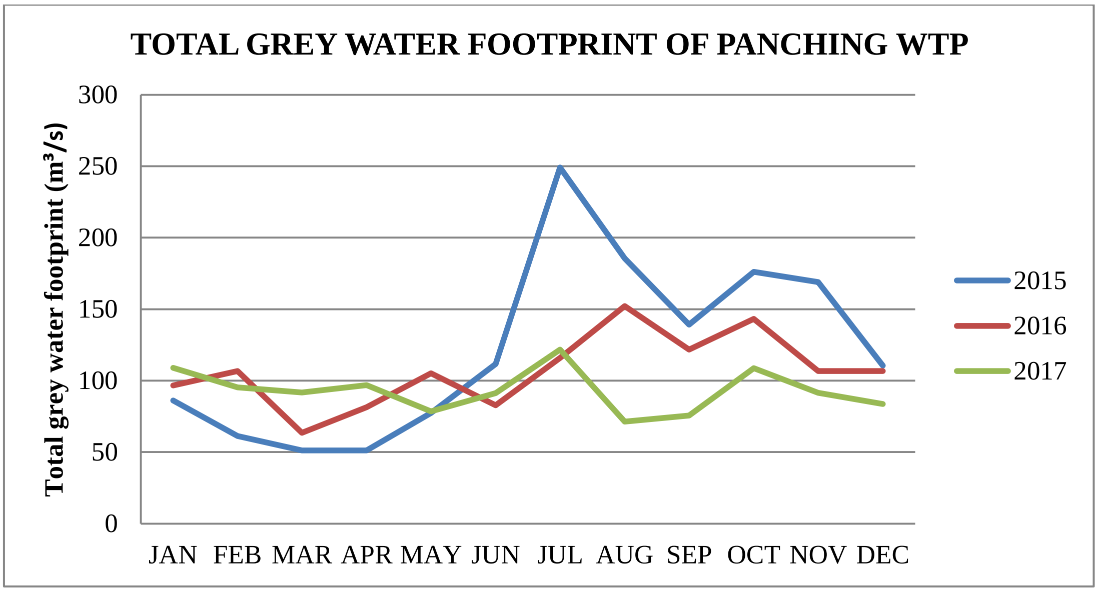


Figure 4.8 Total WFgrey of Panching WTP

Table 4.8 `Total WFgrey in Panching WTP for each year

Year	2015	2016	2017
Total Wfgrey (m^3/s)	1467.83	1302.40	1115.14

Figure 4.8 shows total WFgrey of Panching WTP in the period of 2015 to 2017. There was sudden increment in total WFgrey value up to 249m^3/s in July 2015. This was because of the same reason stated in Semambu WTP which was the increase of iron concentration in river body from bauxite mining (Kusin et al., 2017). At the end of the each year starting from October to December, the total WFgrey start to drop due to monsoonal season. Monsoonal season caused high volume of water intake which resulted in low concentration of pollutant in water body. Along with the growth of development, population and industry, the water body become more polluted with time. It resulted more freshwater needed to assimilate the pollutant inside the polluted water body.

4.3 The Best Algorithm

There are two types of algorithm applied in this study which are Artificial Neural Network (ANN) and Random Forest (RF). Algorithm is a step of task that must be followed to complete a task. By using WEKA software, the selected algorithm is chosen to train the data set. After the training session, the lowest RMSE will be chosen to test the data set. The lowest RMSE selected portrayed the lowest error processed.

Based on the RMSE value generated by the algorithms, ANN has the lowest RMSE value for both Panching and Semambu WTP with 0.0070 and 0.0165 respectively. ANN is known for its effectiveness in Artificial Intelligence (AI). As stated in chapter 2, many study involved with ANN had succeed in performing the objectives. It is proven, for this study, the best algorithm application in carrying prediction is Artificial Neural Network.

The predicted value for one WTP is nearly same even they are trained with different algorithms. This is due to the same data set trained as input. The output is different jus because it is trained by different application.

4.3.1 Root Means Square (RMSE)

The prediction tool is carried on using WEKA software. WEKA software is chosen to predict the trend of total WFgrey because it is user-friendly, open access and easier than MATLAB software. The actual value of total grey water footprint for three years is tabulated in one excel sheet. The excel file then will be converted into CSV format in order for the WEKA software to train the data and produce predicted value.

Root mean square error (RMSE) value shows how well the model performed. It works by measuring the predicted value with the actual value. The higher the RMSE value, the higher the error. In fact, lowest RMSE is more preferable. The application used will display the RMSE value after input is trained in WEKA software. For Artificial Neural Network (ANN), there will be 20 neurons trained. But, only one neuron with lowest RMSE will be chosen to predict the total WFgrey trend. Different in kind for Random Forest (RF) algorithm, it only built one neuron. There is no chance to choose the lowest algorithm for RF algorithm.

4.3.2 RMSE by Artificial Neural Network

Table below shows the value of RMSE produce by 20 hidden neurons in ANN for both WTP.

Table 4.9 RMSE value for Semambu WTP by ANN

No of Neuron	Root Mean Square Error
1	0.1131
2	0.0236
3	0.0272
4	0.0282
5	0.0214
6	0.0168
7	0.0196
8	0.0169
9	0.0187
10	0.0272
11	0.0165
12	0.0173
13	0.0193
14	0.0220
15	0.0175
16	0.0197
17	0.0187
18	0.0190
19	0.0235
20	0.0195

Based on Table 4.9 on the hidden neuron trained for Semambu WTP, the lowest RMSE is a training set with 11 neurons. The RMSE value is 0.0165. Therefore, the prediction value will be using the chosen neuron to predict the total grey water footprint for Semambu WTP.

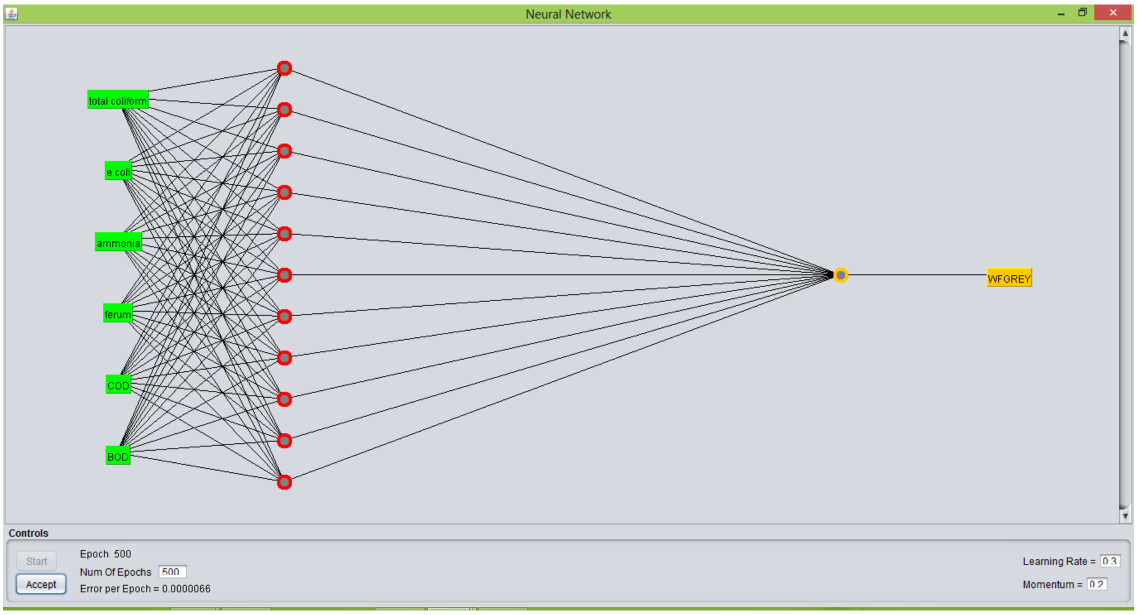


Figure 4.9 Neurons framework by ANN for total WFgrey of Semambu WTP

Based on the lowest RMSE generated by ANN for total WFgrey of Semambu WTP, the eleventh neuron is chosen. Figure 18 above shows how the nodes and neurons of 11 neurons interact. Each input parameters are connected with every nodes by neuron. The red coloured circle is called the nodes. The nodes will generate the information received and transmit it to another nodes. The first layer of nodes (input) is called input layer and the last layer is called output layer. In between the input and output layer is called hidden layer constructed by the nodes.

Table 4.10 RMSE value for Panching WTP by ANN

No of Neuron	Root Mean Square Error
1	0.0481
2	0.0462
3	0.0070
4	0.0099
5	0.0177
6	0.0190
7	0.0187
8	0.0190
9	0.0200
10	0.0182
11	0.0180
12	0.0179
13	0.0179
14	0.0190
15	0.0176
16	0.0182
17	0.0178
18	0.0187
19	0.0190
20	0.0180

Based on the Table 4.10 above, the lowest RMSE value for Panching WTP is 3 layers of hidden neurons obtained from training set which is 0.0070. The neuron will be chosen to predict the total WFgrey in Panching for three years period and will be compared with the actual value.

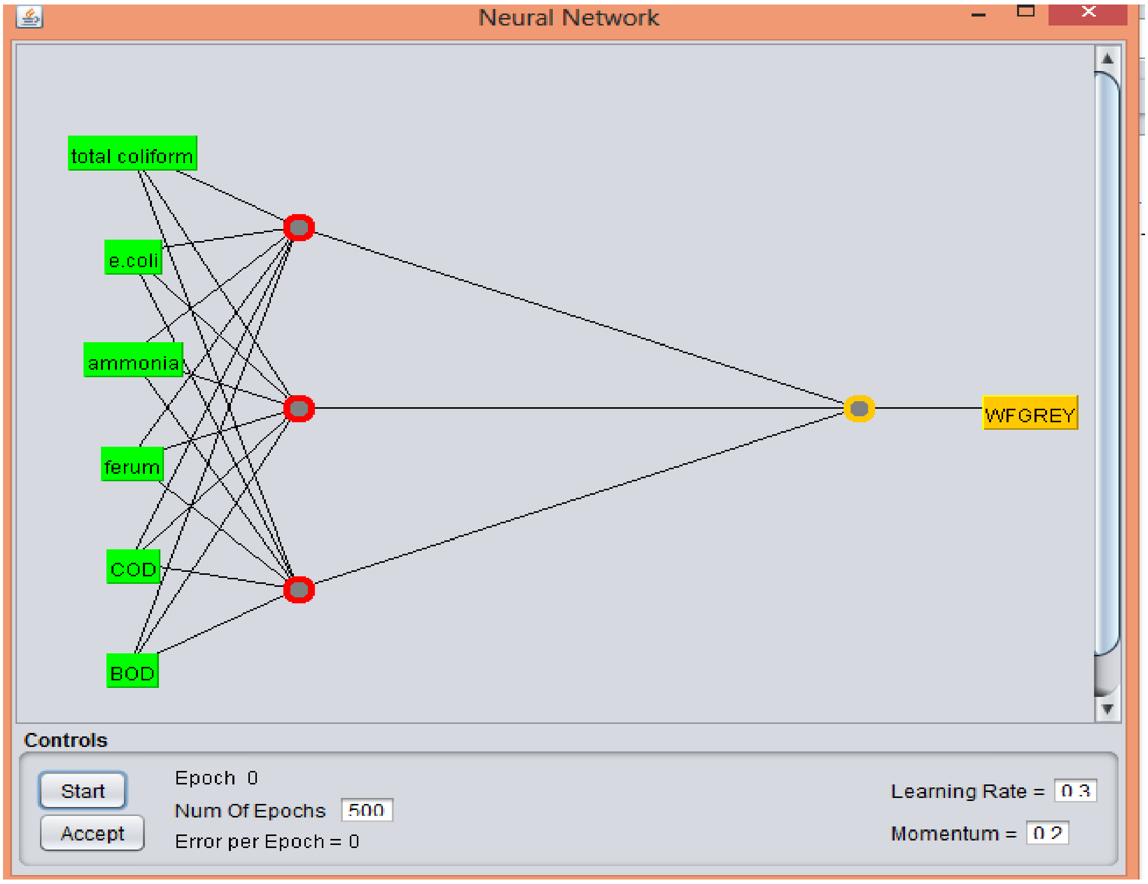


Figure 4.10 Neuron framework by ANN for total WFgrey of Panching WTP

Based on the lowest RMSE generated by ANN for total WFgrey of Panching WTP, the third neuron is chosen. Figure 19 above shows how the nodes and neurons of 3 neurons interact. Each input parameters are connected with every nodes by neuron. The red coloured circle is called the nodes. The nodes will generate the information received and transmit it to another nodes. The first layer of nodes (input) is called input layer and the last layer is called output layer. In between the input and output layer is called hidden layer constructed by the nodes.

4.4 Prediction of Grey Water Footprint

4.4.1 Comparison Between Predicted Value and Actual Value of WFgrey

The RMSE values obtained from ANN and RF algorithm are the average value of RMSE of WFgrey in three years period. For Semambu WTP, the RMSE value for ANN and RF algorithm are 0.0165 and 0.0216 respectively. Meanwhile, for Panching WTP, the RMSE value of ANN and RF algorithm are 0.0070 and 0.0190 respectively.

4.4.1.1 Comparison between predicted value with actual value of WFgrey in Semambu WTP by ANN algorithm

Table 4.11 Actual and predicted value for total WFgrey of Semambu WTP by ANN

Month	Actual	Predicted	Month	Actual	Predicted
Jan-15	185.39	185.20	**Jul-16**	150.25	149.90
Feb-15	160.24	160.02	**Aug-16**	149.03	148.95
Mar-15	142.85	142.41	**Sep-16**	144.19	144.11
Apr-15	184.81	184.71	**Oct-16**	169.25	169.03
May-15	248.91	248.85	**Nov-16**	232.30	232.36
Jun-15	207.65	207.53	**Dec-16**	217.85	217.80
Jul-15	360.88	359.89	**Jan-17**	124.99	124.59
Aug-15	535.27	534.44	**Feb-17**	150.64	150.33
Sep-15	46.97	48.01	**Mar-17**	101.69	101.41
Oct-15	225.35	225.47	**Apr-17**	249.74	249.88
Nov-15	170.38	170.00	**May-17**	120.13	119.60
Dec-15	184.09	183.86	**Jun-17**	93.16	92.86
Jan-16	173.54	172.27	**Jul-17**	127.77	127.23
Feb-16	190.82	190.72	**Aug-17**	187.51	187.26
Mar-16	166.72	166.44	**Sep-17**	134.45	133.86
Apr-16	120.66	120.24	**Oct-17**	171.26	170.92
May-16	153.73	153.31	**Nov-17**	136.79	136.34
Jun-16	146.30	145.91	**Dec-17**	139.20	138.36

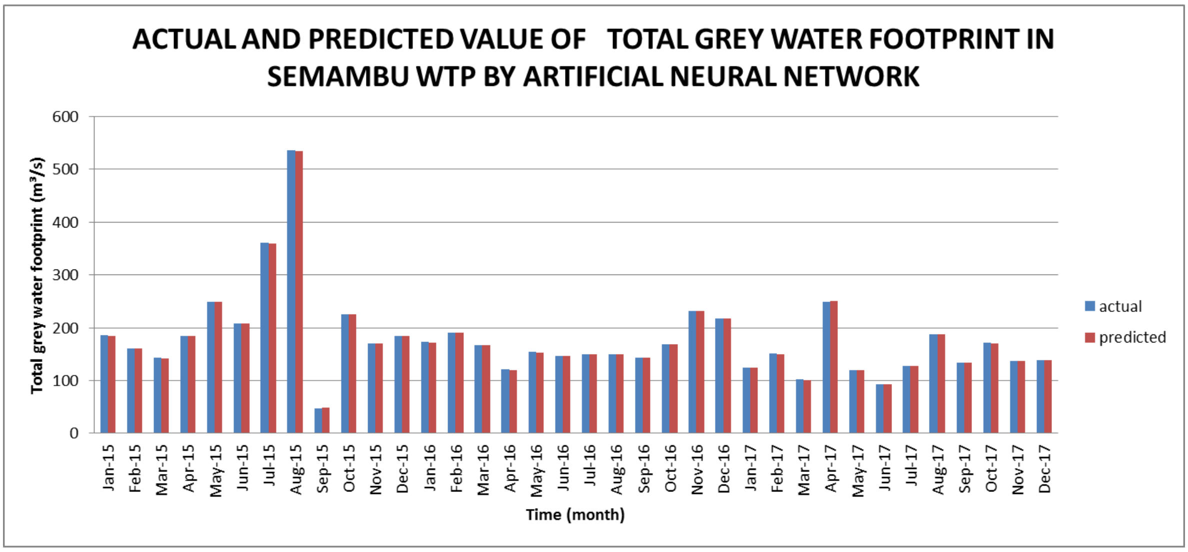


Figure 4.11 Actual and predicted value of WFgrey in Semambu WTP by ANN

Figure 4.11 shows the bar graph comparison between actual and predicted value of WFgrey in Semambu WTP by ANN. There is no significance difference between actual and predicted value due to lowest RMSE chosen which also mean as lowest error between predicted and actual value. The RMSE value for this case is 0.0165.

The highest value of WFgrey for actual value is 535.27m³/s and for predicted value is 534.44m³/s. There is not much difference in both values. Besides that, the lowest of WFgrey value is one month after the highest WFgrey value recorded. The actual and predicted values are 46.97m³/s and 48.01m³/s. Not include the highest and lowest value of WFgrey in Semambu WTP, the range of WFgrey in Semambu WTP is in range 92.86m³/s to 360.88m³/s.

4.4.1.2 Comparison between predicted value with actual value of WFgrey in Panching WTP by ANN algorithm

Table 4.12 Actual and predicted value of total WFgrey in Panching WTP by ANN

Month	Actual	Predicted	Month	Actual	Predicted
Jan-15	86.10	85.89	**Jul-16**	115.94	115.79
Feb-15	61.15	60.99	**Aug-16**	152.11	151.78
Mar-15	51.06	51.04	**Sep-16**	121.76	121.63
Apr-15	51.19	51.11	**Oct-16**	143.14	143.06
May-15	77.30	77.06	**Nov-16**	106.67	106.52
Jun-15	111.65	111.54	**Dec-16**	126.72	126.49
Jul-15	249.09	248.77	**Jan-17**	109.01	108.70
Aug-15	185.60	185.59	**Feb-17**	95.34	95.11
Sep-15	139.24	139.16	**Mar-17**	91.66	91.41
Oct-15	176.17	176.16	**Apr-17**	96.94	96.90
Nov-15	168.86	168.67	**May-17**	78.44	78.17
Dec-15	110.44	110.22	**Jun-17**	91.19	90.93
Jan-16	96.55	96.25	**Jul-17**	121.73	121.57
Feb-16	106.73	106.49	**Aug-17**	71.32	71.08
Mar-16	63.58	63.30	**Sep-17**	75.64	75.46
Apr-16	81.44	81.16	**Oct-17**	108.79	108.54
May-16	105.09	104.94	**Nov-17**	91.50	91.35
Jun-16	82.69	82.37	**Dec-17**	83.58	83.25

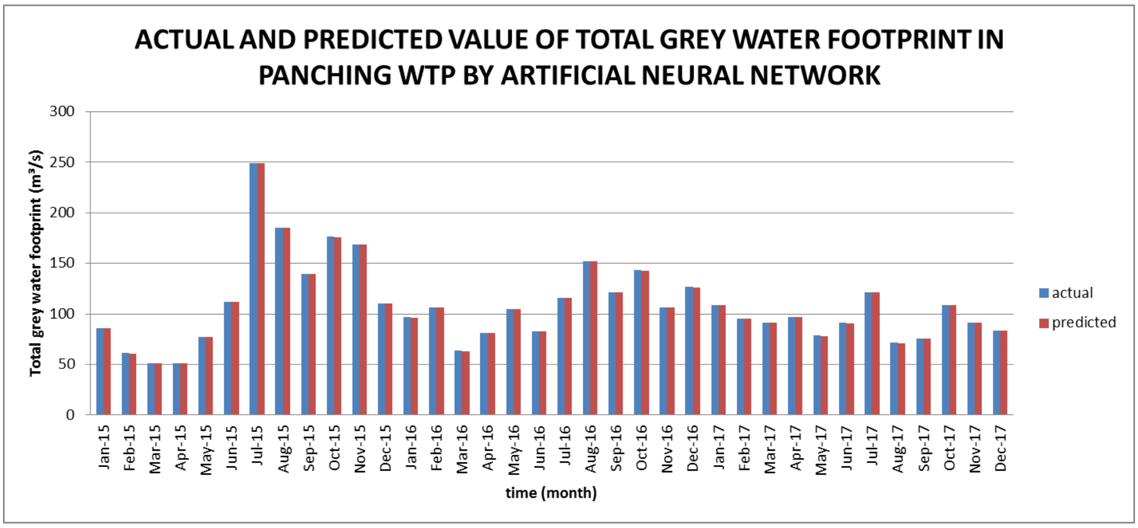


Figure 4.12 Actual and predicted value of WFgrey in Panching by ANN

Figure 4.12 shows the bar graph comparison between actual and predicted value of WFgrey in Panching WTP by ANN. There is no significance difference between actual and predicted value due to lowest RMSE chosen which also mean as lowest error between predicted and actual value. The RMSE value for this case is 0.007.

The highest value of WFgrey for actual value is 249.09m³/s and for predicted value is 248.77m³/s. There is not much difference in both values. Besides that, the lowest of WFgrey value is in March 2015. The actual and predicted values are 51.06m³/s and 51.04m³/s.

4.4.1.3 Comparison between predicted value with actual value of WFgrey in Semambu WTP by Random Forest (RF) algorithm

Table 4.13 Actual and predicted value for Semambu WTP by Random Forest

Month	Actual	Predicted	Month	Actual	Predicted
Jan-15	185.39	185.43	**Jul-16**	150.25	151.05
Feb-15	160.24	160.11	**Aug-16**	149.03	149.58
Mar-15	142.85	141.80	**Sep-16**	144.19	144.68
Apr-15	184.81	184.47	**Oct-16**	169.25	169.23
May-15	248.91	248.93	**Nov-16**	232.30	232.30
Jun-15	207.65	208.53	**Dec-16**	217.85	217.01
Jul-15	360.88	360.87	**Jan-17**	124.99	124.75
Aug-15	535.27	535.28	**Feb-17**	150.64	151.58
Sep-15	46.97	46.98	**Mar-17**	101.69	101.38
Oct-15	225.35	225.34	**Apr-17**	249.74	249.75
Nov-15	170.38	170.00	**May-17**	120.13	120.71
Dec-15	184.09	184.95	**Jun-17**	93.16	93.52
Jan-16	173.54	173.53	**Jul-17**	127.77	127.43
Feb-16	190.82	190.68	**Aug-17**	187.51	187.01
Mar-16	166.72	167.50	**Sep-17**	134.45	135.47
Apr-16	120.66	120.64	**Oct-17**	171.26	170.06
May-16	153.73	152.75	**Nov-17**	136.79	136.06
Jun-16	146.30	145.54	**Dec-17**	139.20	139.98

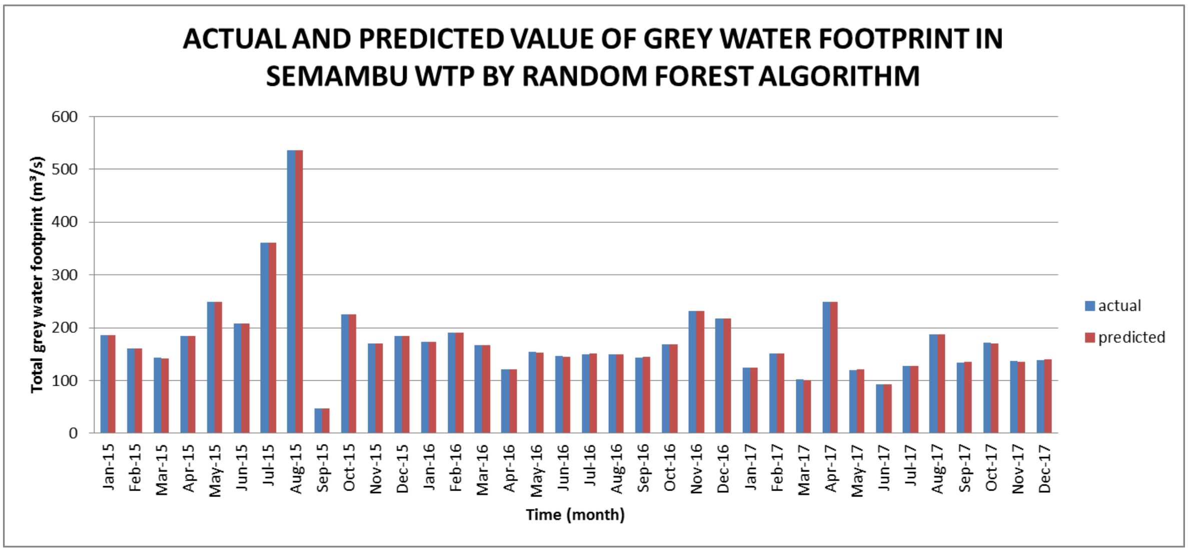


Figure 4.13 Actual and predicted value of WFgrey in Semambu WTP by RF

Figure 4.13 shows the bar graph comparison between actual and predicted value of WFgrey in Semambu by Random Forest. There is no significance difference between actual and predicted value due to lowest RMSE chosen which also mean as lowest error between predicted and actual value. The RMSE value for this case is 0.0216.

The highest value of WFgrey for actual value is 535.27m³/s and for predicted value is 534.44m³/s. There is not much difference in both values. Besides that, the lowest of WFgrey value is one month after the highest WFgrey value recorded. The actual and predicted values are 46.97m³/s and 48.01m³/s. Not include the highest and lowest value of WFgrey in Semambu WTP, the range of WFgrey in Semambu WTP is in range 92.86m³/s to 360.88m³/s.

4.4.1.4 Comparison between predicted value with actual value of WFgrey in Panching WTP by Random Forest (RF) algorithm

Table 4.14 Actual and predicted value of Panching WTP by Random Forest

Month	Actual	Predicted	Month	Actual	Predicted
Jan-15	86.10	86.09	Jul-16	115.94	115.72
Feb-15	61.15	61.14	Aug-16	152.11	152.17
Mar-15	51.06	51.47	Sep-16	121.76	121.78
Apr-15	51.19	50.81	Oct-16	143.14	143.44
May-15	77.30	77.76	Nov-16	106.67	106.52
Jun-15	111.65	111.83	Dec-16	126.72	126.73
Jul-15	249.09	249.07	Jan-17	109.01	109.06
Aug-15	185.60	185.58	Feb-17	95.34	95.34
Sep-15	139.24	138.92	Mar-17	91.66	92.02
Oct-15	176.17	176.19	Apr-17	96.94	96.93
Nov-15	168.86	168.87	May-17	78.44	78.04
Dec-15	110.44	109.90	Jun-17	91.19	90.99
Jan-16	96.55	96.53	Jul-17	121.73	121.73
Feb-16	106.73	106.77	Aug-17	71.32	71.33
Mar-16	63.58	63.57	Sep-17	75.64	75.57
Apr-16	81.44	81.72	Oct-17	108.79	109.42
May-16	105.09	105.15	Nov-17	91.50	91.34
Jun-16	82.69	82.36	Dec-17	83.58	83.64

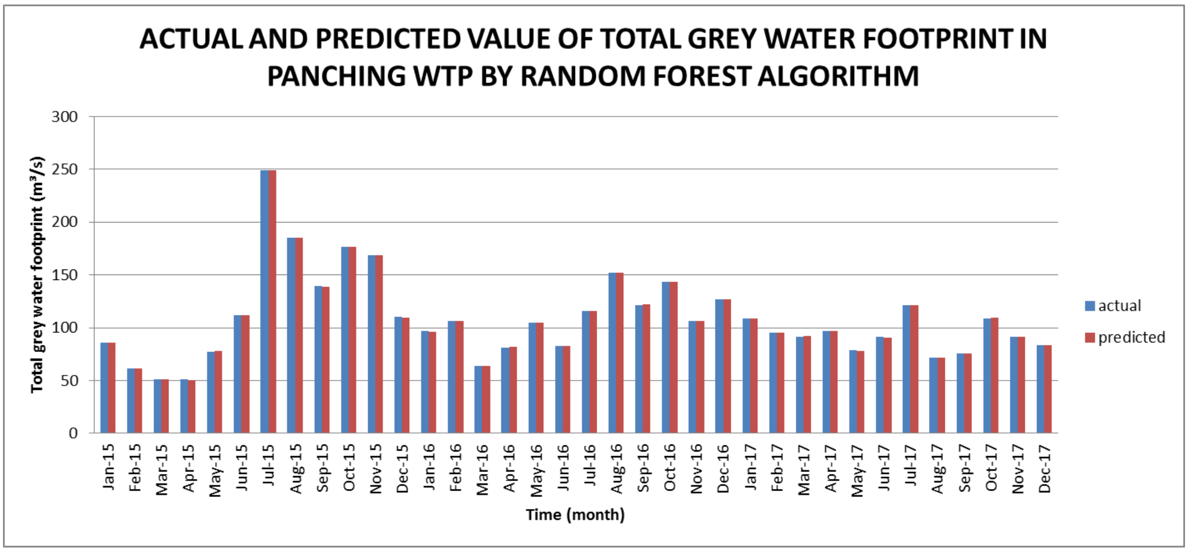


Figure 4.14 Actual and predicted value of WFgrey in Panching WTP by RF

Figure 4.12 shows the bar graph comparison between actual and predicted value of WFgrey in Panching WTP by ANN. There is no significance difference between actual and predicted value due to lowest RMSE chosen which also mean as lowest error between predicted and actual value. The RMSE value for this case is 0.019.

The highest value of WFgrey for actual value is 249.09m^3/s and for predicted value is 248.77m^3/s. There is not much difference in both values. Besides that, the lowest of WFgrey value is in March 2015. The actual and predicted values are 51.06m^3/s and 51.04m^3/s.

CHAPTER 5

CONCLUSION

5.1 Conclusion

Finally, it can be concluded that this study had achieved all the objectives stated successfully. The first objective is achieved by calculating the total Grey Water Footprint (WFgrey) of Semambu and Panching WTP. All the parameter considered are clearly stated and categorised in table before it is substituted into formula. The total WFgrey is tabulated in excel for period of three years starting from 2015 to 2017 so that it is easy to carry on the next objective. Mostly the total WFgrey is increase due to high concentration of iron and ammonia in water intake which result in high amount of water needed to assimilate the pollutant. Bauxite mining in Kuantan is an issue that causes pollution in river which contributing to high concentration of ammonia and iron (Kusin et al., 2017). This case happened in August 2015 for Semambu WTP and July 2015 for Panchig WTP. Overall, the total WFgrey of both WTP shows decrement over time.

The second objective achieved is by comparing the best algorithm used through RMSE value. It found that the best algorithm in predicting total WFgrey is Artificial Neural Network (ANN). This is parallel with previous studies which show the effectiveness of ANN in computer learning. ANN is confirmed to be the best algorithm through Root Mean Square Error (RMSE) value. The lowest RMSE is by ANN at Panching WTP with value of 0.007.

The last objective achieved is the prediction trend of WTP by algorithm training. The prediction trend produced by the algorithm involved and is compared with actual value of total grey water footprint. The trend of both WTP by using ANN shows decrement. However, the trend of total WFgrey by using Random Forest shows increment. If the trend is compared with the actual value of total WFgrey, the trend should be decreased.

5.2 Recommendation

Since the total WFgrey is showing decrement over time, there are few recommendation suggested in order to maintain the sustainability of the water supply. Water treatment plant is a special unit responsible to treat the raw water before distributing among society.

To distribute good quality of water resources, the water treatment plants should consider Water Footprint as an approach to enhance the water supply management. This is because water footprint accounted overall water consumption. As this study focused on WFgrey accounting, the actual amount of freshwater used in the process of treating water is now able to be known. The volume of water wastage in every stage can be reduced and controlled accordingly. Based on the prediction model, the WFgrey is expected to increase in the future, therefore it is important to ensure the river is not polluted especially when it is used for the abstraction.

Grey water footprint is the amount of water needed to assimilate the pollutant so that the water resources are safe for use. River water pollution comes from many sources. For example, the untreated effluent discharge from factories, bauxite mining, illegal logging and uneducated society who throwing the trash into river. The more polluted the river, the higher WFgrey will be. So, the authorities should play their part in this matter to maintain the cleanliness of river. Every party who littered the river should be charge a fine. The society also must be aware about river pollution and its effect. This is no more a small issue in future due to limited clean water resource with increasing in population.

Artificial intelligence is a wide subject to be discussed. A lot of computer learning had been developed in studying water footprint throughout technological era nowadays. It is indeed not an easy technique that can be learnt in a single night. There are many algorithms that had been created to study on water footprint for example, Artificial Neural Network, Random Forest, Genetic Algorithm, Levenbeg Marquad and many more. Algorithm also can be created a new one if there is expert in computer learning and water footprint. However, to make it easy, using the previous algorithm to study the water footprint approach using artificial intelligence is more preferable.

In predicting process of total grey water footprint, it is recommended to use WEKA Software because it is easier than MATLAB. Besides that, with the user-friendly interface and key icon, this software is a fast-learning software for new beginner. It also free access where no payment needed to installed.

REFERENCES

Ababa, A., Adeel, Z., Affairs, S., Bachewe, F., Hirvonen, K., Minten, B., … Workshop, D. (2016). *Progress on Drinking Water , Sanitation and Hygiene* (Vol. 000). https://doi.org/10.1371/journal.pone.0164800

Alaska Department of Environmental Conservation. (2015). Chapter 4: Introduction to Water Treatment. *Introduction to Small Water Systems*, 95–145. https://doi.org/10.1016/B978-1-85617-705-4.00014-9

Aldaya, M. M., Chapagain, A. K., Hoekstra, A. Y., & Mekonnen, M. M. (2011). *The Water Footprint Assessment Manual*. *Earthscan*. https://doi.org/978-1-84971-279-8

Badrzadeh, H., Sarukkalige, R., & Jayawardena, A. W. (2013). Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting, *507*, 75–85.

Bari, A., Begum, R. A., Nesadurai, N., & Pereira, J. J. (2015). Water Consumption Patterns in Greater Kuala Lumpur : Potential for Reduction, (August 2017). https://doi.org/10.3233/AJW-150001

Bauwens, J., Lobe, B., Segers, K., & Tsaliki, L. (2009). *A Shared Responsibility*. *Journal of Children and Media* (Vol. 3). https://doi.org/10.1080/17482790903233325

Belayneh, A., Adamowski, J., Khalil, B., & Ozga-zielinski, B. (2014). Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models, *508*, 418–429.

Chen, Y., Yuan, W., Xia, J., Fisher, J. B., Dong, W., Zhang, X., … Feng, J. (2015). Using Bayesian model averaging to estimate terrestrial evapotranspiration in China. *Journal of Hydrology*, *528*, 537–549. https://doi.org/10.1016/j.jhydrol.2015.06.059

China Paper Association. (2015). 2015 Annual Report of Activities. *Epfl*, 99–111.

Chong, D., Zhu, N., Luo, W., & Pan, X. (2019). Human thermal risk prediction in indoor hyperthermal environments based on random forest. *Sustainable Cities and Society*, *49*(May), 101595. https://doi.org/10.1016/j.scs.2019.101595

Chung, E. (2018). This Is How Much Water Malaysians Waste Every Day, 1–8. Retrieved from https://says.com/my/lifestyle/world-water-day-malaysians-waste-water

Danandeh Mehr, A., Kahya, E., Şahin, A., & Nazemosadat, M. J. (2015). Successive-station monthly streamflow prediction using different artificial neural network algorithms. *International Journal of Environmental Science and Technology*, *12*(7), 2191–2200.

https://doi.org/10.1007/s13762-014-0613-0

Gong, H., Sun, Y., Shu, X., & Huang, B. (2018). Use of random forests regression for predicting IRI of asphalt pavements, *189*, 890–897.

Hague, T., & Office, O. (2004). Unep / Gpa Unep / Ietc Unesco-Ihe. *Environmental Technology*.

Haida, C., Chapagain, A. K., Rauch, W., Riede, M., & Schneider, K. (2018). From water footprint to climate change adaptation: Capacity development with teenagers to save water. *Land Use Policy*, *80*(July 2017), 456–463. https://doi.org/10.1016/j.landusepol.2018.02.043

Hardwinarto, S., & Aipassa, M. (2015). Rainfall Monthly Prediction Based on Artificial Neural Network : A Case Study in Tenggarong Station , East Kalimantan - Indonesia. *Procedia - Procedia Computer Science*, *59*(Iccsci), 142–151. https://doi.org/10.1016/j.procs.2015.07.528

Horowitz, E., Sahni, S., & Rajasekaren, S. (1997). Fundamentals of Computer Algorithms, 2ND Edition.

Jimenez-martinez, M., & Alfaro-ponce, M. (2019). Fatigue damage e ff ect approach by arti fi cial neural network, *124*(2901), 42–47. https://doi.org/10.1016/j.ijfatigue.2019.02.043

Kuan, C.-M. (2000). Artificial Neural Networks, 14.

Kumar, A., Goel, N. K., & Bhatia, K. K. S. (2014). Improving real time flood forecasting using fuzzy inference system, *509*, 25–41.

Kusin, F. M., Syazwan, M., Rahman, A., & Madzin, Z. (2017). The occurrence and potential ecological risk assessment of bauxite mine-impacted water and sediments in Kuantan , Pahang , Malaysia. *Environmental Science and Pollution Research*, 1306–1321. https://doi.org/10.1007/s11356-016-7814-7

Lathuillière, M. J., Bulle, C., & Johnson, M. S. (2018). A contribution to harmonize water footprint assessments, *53*(October), 252–264.

M.Mekonnen, M., & Hoekstra, Y. A. (2016). Four Billion People Experience Water Scarcity. *American Association for the Advancement of Science*, *2*(February), 1–7. https://doi.org/10.1016/j.acra.2014.09.014

Maier, H. R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L. S., Cunha, M. C., … Barbour, E. J. (2014). Environmental Modelling & Software Evolutionary algorithms and other metaheuristics in water resources : Current status , research challenges and future directions * , **. *Environmental Modelling and Software, 62*, 271–299. https://doi.org/10.1016/j.envsoft.2014.09.013

Margrit, A. (2018). Ini 11 Kota Dunia yang Terancam Kekurangan Air. *Kabar24*. Retrieved from http://kabar24.bisnis.com/read/20180211/19/737209/ini-11-kota-dunia-yang-terancam-kekurangan-air

Millock, K., & Nauges, C. (2010). Household Adoption of Water-Efficient Equipment : The Role of Socio-economic Factors , Environmental To cite this version : HAL Id : halshs-00492291 Centre d ' Economie de la Sorbonne Documents de Travail du.

Mustapha, H., & Abdelwahed, N. (2019). ScienceDirect ScienceDirect Investigating the use of random forest in software effort estimation Investigating the use of random forest in software effort estimation. *Procedia Computer Science*, *148*, 343–352. https://doi.org/10.1016/j.procs.2019.01.042

Nolan, B. T., Fienen, M. N., & Lorenz, D. L. (2015). A statistical learning framework for groundwater nitrate models of the Central Valley , California , USA, *531*, 902–911.

Olawoyin, A., & Chen, Y. (2018). Predicting the Future with Artificial Neural Network. *Procedia Computer Science*, *140*, 383–392. https://doi.org/10.1016/j.procs.2018.10.300

Pellicer-martínez, F., & Martínez-paz, J. M. (2016). Grey water footprint assessment at the river basin level : Accounting method and case study in the Segura River Basin , Spain, *60*, 1173–1183.

Rafael, A., Parmezan, S., Diana, H., & Chung, F. (2017). Metalearning for choosing feature selection algorithms in data mining : Proposal of a new framework. *Expert Systems With Applications*, *75*, 1–24. https://doi.org/10.1016/j.eswa.2017.01.013

Reig, P. (2013). What's the Difference Between Water Use and Water Consumption?, 11–13. Retrieved from http://www.wri.org/blog/2013/03/what's-difference-between-water-use-and-water-consumption

Srinivas, H. (n.d.). Water as an energy sources.

Staub, S., Karaman, E., Kaya, S., Karapınar, H., & Güven, E. (2015). Artificial Neural Network and Agility. *Procedia - Social and Behavioral Sciences*, *195*, 1477–1485. https://doi.org/10.1016/j.sbspro.2015.06.448

The CEO Water Mandate. (2010). Guide to Responsible Business Engagement with Water Policy November 2010. *Water Policy*, (November), 112.

Water Civilization International Centre. (2018). Good water, water to “eat”. What is virtual water? *Drops of Water 4*.

Wu, Y., Liu, S., & Chen, J. (2012). Urbanization eases water crisis in China, *2*, 142–144.

APPENDICES

APPENDIX A

Table 5.1 Total of grey water footprint

Total Wfgrey (m^3/s)	2015	2016	2017
Semambu	2507.14	2014.69	1737.31
Panching	1467.83	1302.40	1115.14

APPENDIX B

Table 5.2 Root means square error for Semambu and Panching WTP

RMSE \ WTP	Semambu	Panching
Artificial Neural Network	0.0165	0.0216
Random Forest	0.0070	0.0190

APPENDIX C

Table 5.3 Prediction of grey water footprint of Semambu WTP

Algorithm	Artificial Neural Network (ANN)			Random Forest (RF)		
Year	2015	2016	2017	2015	2016	2017
Actual (m^3/s)	2652.785	2014.637	1737.330	2652.785	2014.637	1737.330
Predicted (m^3/s)	2650.391	2011.047	1732.633	2652.695	2014.469	1737.711
Percentage (%)	-0.0902	-0.1782	-0.2704	-0.0034	-0.0083	0.0219
Overall (%)	-0.1668			0.0019		

APPENDIX D

Table 5.4 Prediction of grey water footprint of Panching WTP

Algorithm	Artificial Neural Network (ANN)			Random Forest (RF)		
Year	2015	2016	2017	2015	2016	2017
Actual (m^3/s)	1467.834	1302.404	1115.141	1467.834	1302.404	1115.141
Predicted (m^3/s)	1466.199	1299.771	1112.473	1467.617	1302.445	1115.409
Percentage (%)	-0.1114	-0.2022	-0.2393	-0.0148	0.0031	0.0240
Overall (%)	-0.1785			0.0024		