PREDICTION OF BLUE WATER FOOTPRINT AT SEMAMBU AND PANCHING WATER TREATMENT PLANTS

MUHAMMAD SYAZWAN BIN MAT AKHIR

B. ENG (HONS.) CIVIL ENGINEERING

UNIVERSITI MALAYSIA PAHANG

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name : MUHAMMAD SYAZWAN BIN MAT AKHIR

Date of Birth : 10^{th} FEBRUARY 1996

Title : PREDICTION OF BLUE WATER FOOTPRINT AT SEMAMBU AND PANCHING WATER TREATMENT PLANTS

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)	(Supervisor's Signature)
960210-07-5835	DR. EDRIYANA BINTI A. AZIZ
Date: 31^{sr} MAY 2019	Date: 31^{sr} MAY 2019

THESIS DECLARATION LETTER (OPTIONAL)

Librarian,
Perpustakaan Universiti Malaysia Pahang,
Universiti Malaysia Pahang,
Lebuhraya Tun Razak,
26300, Gambang, Kuantan.

Dear Sir,

CLASSIFICATION OF THESIS AS RESTRICTED

Please be informed that the following thesis is classified as RESTRICTED for a period of three (3) years from the date of this letter. The reasons for this classification are as listed below.

Author's Name
Thesis Title

Reasons (i)

(ii)

(iii)

Thank you.

Yours faithfully,

____
(Supervisor's Signature)

Date:

Stamp:

Universiti Malaysia PAHANG

Engineering • Technology • Creativity

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis 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 A.AZIZ

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 : MUHAMMAD SYAZWAN BIN MAT AKHIR

ID Number : AA15103

Date : 31^{st} MAY 2019

PREDICTION OF BLUE WATER FOOTPRINT AT SEMAMBU AND PANCHING WATER TREATMENT PLANTS

MUHAMMAD SYAZWAN BIN MAT AKHIR

Thesis submitted in fulfillment of the requirements
for the award of the
B. Eng (Hons.) Civil Engineering

Faculty of Civil Engineering & Earth Resources

UNIVERSITI MALAYSIA PAHANG

MAY 2019

ACKNOWLEDGEMENTS

Alhamdullilah, I would like to praise to Allah SWT for giving me this opportunity to complete my thesis. Thanks to Him for giving me a strength and a good health while conducting this research.

I would like to express sincere appreciation and thanks to my supervisor, Dr. Edriyana A. Aziz, for her guidance, advices and encouragement in all aspect through this research. High appreciation for her comments, criticism and willing to share her knowledge within this field. All the knowledge that she gave to me were precious and priceless.

I would also like to express my heartfelt to my beloved parents, Mat Akhir bin Ahmad and Halijah binti Abd. Wahab for the support and encouragement. Grate motivation from them boost up my spirit until I managed to complete this task.

Finally, not to be forgotten, my appreciation will go to En. Syazwan Nizam, lecturers, staffs and all my friends in final year in the Faculty of Civil Engineering and Earth Resources, for their ideas and supports. Thank you so much for all the help, information and braveness through all the difficult time in doing the project and to finish study. Highly appreciate and millions of love for all of you from me.

ABSTRAK

Permasalahan air di dunia sekarang menjadi isu yang membimbangkan disebabkan pembandaran yang semakin pesat. Terdapat banyak penyelidikan berkaitan air telah dilakukan untuk menangani masalah ini seperti (Chang, Chang, Huang, & Kao, 2016; Nguyen-ky et al., 2017; Rosecrans, Nolan, & Gronberg, 2017). Kajian ini tertumpu kepada penilaian jejak air biru (WFblue) di Loji Rawatan Air (LRA) Semambu dan Panching. Kemudian, jumlah jejak air biru akan dimodelkan dan menjalani satu siri latihan untuk meramalkan trend dengan menggunakan 2 algoritma iaitu Rangkaian Neural Buatan (ANN) dan Random Forest (RF). Perbandingan telah dibuat di antara kedua-dua algoritma bagi memilih algoritma terbaik dalam melakukan ramalan trend berkaitan air. Objektif kajian ini adalah: (1) untuk mengira jumlah WFblue di LRA Semambu dan Panching yang terletak di lembah Sungai Kuantan bagi tempoh 2015-2017, (2) untuk membandingkan algoritma terbaik antara ANN dan RF dalam model ramalan WFblue dan (3) untuk meramalkan trend WFblue di LRA Semambu dan Panching. Sehubungan dengan itu, jumlah pengambilan air, penggunaan hujan dan jumlah penyejatan akan diambil kira dalam pengiraan jumlah WFblue di mana WFblue boled ditakrifkan sebagai jumlah penggunaan air dalam rantaian produk. Pada akhir penyelidikan ini, jumlah WFblue telah berjaya dihasilkan. Trend yang diramalkan menunjukkan penurunan dari 2015 hingga 2017 selepas menjalani siri latihan dalam perisian WEKA. Hasil dari kajian ini, pengawasan yang baik mengenai jumlah pengambilan air perlu dilaksanakan dan semua LRA dicadangkan untuk menggunakan penilaian jejak air sebagai pendekatan bagi memastikan kecekapan penggunaan air.

ABSTRACT

Water stress in the world is becoming more alarming issue due to urbanisation. There are a lot of water related researches to address this issue (Chang, Chang, Huang, & Kao, 2016; Nguyen-ky et al., 2017; Rosecrans, Nolan, & Gronberg, 2017). This study focused on blue water footprint (WFblue) assessment in Semambu and Panching water treatment plants (WTPs). Then, the total WFblue will be modelled and undergo a series of training to predict the trend by using 2 algorithms which is Artificial Neural Network (ANN) and Random Forest (RF). In order to choose the best algorithm, comparison has been made between those two algorithms. The objectives of this research are; (1) to calculate the total WFblue in Semambu & Panching WTPs which are located in Kuantan river basin for the 2015-2017 period; (2) to predict the trend of total blue water footprint Semambu & Panching water treatment plants in Kuantan river basin; and, (3) to compare the best algorithm between ANN and RF in WFblue prediction model. Water intake, rainfall utilization and total evaporation will be taken into account in total WFblue calculation where WFblue can be defined as total water consumption within a product chain. at the end result of this research, the total blue water footprint prediction trend has been produced. The predicted trend of WFblue showed a decrement from 2015 until 2017 after undergoes training in WEKA software. From this research, correct monitoring of water intake amount need to be implemented and it is suggested that all WTPs applies water footprint assessment as an approach to ensure the efficiency of water utilization

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

ET_o	Evapotranspiration
WF_{blue}	Blue Water Footprint
WF_{grey}	Grey Water Footprint
WF_{green}	Green Water Footprint
CO_2	Carbon Dioxide
H_2	Hydrogen
NH_3	Ammonia
T_{mean}	Mean Daily Temperature
ρ	Mean Daily Percentage

LIST OF ABBREVIATIONS

WWO	World Water Organization
WF	Water Footprint
WFblue	Blue Water Footprint
ANN	Artificial Neural Network
RF	Random Forest
WFA	Water Footprint Assessment
WFN	Water Footprint Network
WTP	Water Treatment Plant
WWTP	Waste Water Treatment Plant
GIS	Geographical Information Services
FAO’s	Food Agriculture Organization of United Nations
SWAT	Soil and Water Assessment Tool
HRB	Haihe River Basin
BWFI	Blue Water Footprint Index
ULB	Upper Litani Basin
PUNN	Product Unit Neural Network
MLP	Multi-Layer Perceptron
FFBP	Feed Forward Back Propagation
GR	Generalized Regression
RBF	Radial Basis Function
RMSE	Root Mean Square Error
CSV	Common Separated Value

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Water is a fundamental natural resource that plays an important role for humans, animals, plants and environment development (Dur, 2018). Adequate water supply can help in accomplishing duties and responsibilities for many parties. Human needs fresh water for their daily used and agricultural activities where a statistics shows that agricultural sector companies is the largest consumer of water which is 62% followed by 21% of company and 17% of domestics (Lucia, Maiello, & Quintslr, 2018). Besides, water are also important to the environment as a support for the biological process and to stabilize global temperature. Water supply sources come from water catchment areas including rivers, lakes and also reservoirs. It is reported by World Water Organization (WWO) in 2010 that water demand in the world rapidly increasing which has tripled since 1950. 70% of the world is surrounded with water but the amount of fresh water supply is relatively limited compared to saline. However, a good quality of water is difficult to obtain even in the country that be blessed with fresh water resources.

The existing water supply in a region can be exhausted when the population grows drastically. The process of delivering water from the source area to the city can be negatively affect the environment as well as intensive farming activities. Furthermore, when the population in the world increase, the development of a country will increase rapidly and requires adequate water supply for the use (Asare, Zhao, Asante, & Nyarko, 2018) . This is because it is expected that the world's population will all live in the city by 2050. The progress of development in any country needs a wide area which sometimes required to cut off trees and elimination of catchment areas where will cause global temperature rise. The occurrence of this incident will automatically reduce the sources of water supply in a country. Government parties need to plan the future development wisely in order to prevent the exhausting of water supply.

Uncontrolled climate changes will also be a problem in providing enough water supply especially for the management in the water treatment plants. Some countries in the world experience hot and rainy weather throughout the year which can affect in the water purification process. Natural water sources that will be treated at water treatment plants getting lesser in prolonged summer while in the rainy season, water treatment plants may not be able to accommodate a large amount of water at any given time.

Pollution of water that often occurs will make the supply of clean water been disturbed (Udimal, Jincai, Ayamba, & Owusu, 2017) . This happen when unscrupulous parties like in industrial areas that release toxic waste directly into the river. Water treatment plants unable to remove all contaminated substances that mixed with the river water, thus the quality of water to be deliver are not guaranteed. Treating the waste water before releasing it into the water body will give a little bit help to reduce water pollution scale. Moreover, poor management in development activities will also lead to pollution. Poor management here mean when the parties are doing uncontrolled tree cutting, proper waste cleaning need to be done. If the waste just be left on the ground, probability of the substances to enter the river is high which can interfere with the use of water by agricultural and other sectors.

Water footprint (WF) can be used to measure water resource requirements by the consumers for the products and services (Hogeboom, Knook, & Hoekstra, 2018). Water footprint assessment is a process to evaluate the sustainability and efficiency of water consumption and establish which actions should be preferred in order to have sustainable footprint. Water footprint can be classify into three components which is grey, green and blue water footprint. This study focuses on blue water footprint assessment to assess full water utilization in the water treatment plants. Furthermore, water footprint assessment is multi-purpose which can produce wide range of information from different perspectives. From water footprint assessment, the total amount of water consumed within a process can be identified.

1.2 Problem Statement

Water treatment is a process that improves the quality to make it be accepted for any specific use. The end use of the treated raw water may be for industrial water supply, economics and including being return safely to the environment. In regard, a study will be conducted at water treatment plants which most of the source of water comes from Kuantan river basin. Nonetheless, most of the water treatment plants are still managed by using the old method. This method still can be used but ineffective due to some issues. One of the issues are the data about rainfall and evaporation of water are not recorded which can bring problem for the treatment plants. For a country who are in the equator such as Malaysia, the relevant parties will face the problem in retrieving the missing

By 2030, water demand is expected to grow 50% due to the rapidly increasing of the population in the world. The increasing number of population will be mostly in the cities because it will be about 70% of the world's population will live in the cities in 2050, compared to 50% today (UN-Habitat,2016). In this context, the provision of a good quality of water is important for the commercial use of the population. If a good quality of water cannot be issued, any activities that require water usage such as industrial development will be disrupted. This problem will also affect the income source of a country.

The water footprint measures large volume of water resources used to provide goods and services. Water footprint can be classified into three components which is blue, green and grey. These components give a clear picture of the water consumption as the amount of water needed for assimilation of pollutants. Water footprint can be used to observe the level of efficiency that water treatment plants can achieved in purifying water resources (Dur, 2018). This study will focusing on blue water footprint which is an indicator of surface and groundwater needed and also refers to the amount of water used to create a product.

1.3 Objective of Study

There are three objectives for this study based on the problem statement. The objectives are as below:

i. To calculate total blue water footprint in Semambu & Panching water treatment plants in Kuantan river basin for 2015-2017.
ii. To compare the best algorithm between Artificial Neural Network and Random Forest in blue water footprint prediction.
iii. To predict the trend of total blue water footprint in Semambu & Panching water treatment plants at Kuantan river basin.

1.4 Scope of Study

This study mostly focuses on the calculation of water footprint within the process of water treatment. Blue water footprint assessment will be used to assess full water utilization in the water treatment plants. High amount of water resources that will be assessed comes from Kuantan river basin. There are some researches that have assessed blue water footprint either per watershed or river basin (Hoekstra et al, 2012). With respect to that, blue water footprint assessment will only cover at Panching and Semambu water treatment plants starting from water abstraction until the final step before water been supply. This calculation is intended to identify the total amount of water used in the process of distributing to the user. In addition, blue water footprint capacities from the river basin is also affected by runoff from the precipitation and the needs to control flow for ecosystem (Zhuo, Hoekstra, Wu, & Zhao, 2019). The trend of blue water footprint Semambu & Panching water treatment plants at Kuantan river basin will be predicted.

Blue water footprint that will be assessed will cover for the modelling scope. Most of the studies that have been done have come up with various ways to calculate the blue water footprint that then can predict the trend of total blue water footprint in the water treatment plants. After doing detailed review to the previous researches, two calculation model which also can be called as algorithm are identified that the most being used. Prediction of blue water footprint trend in this research will used this two algorithms which is Artificial Neural Network (Buchtele, Richta, & Chlumecky, 2017) and Random Forest (Z. Wang et al., 2015). Instead of producing new one, those two algorithm will be used in the assessment of blue water footprint. Comparison will be made between Artificial Neural Network and Random Forest algorithms in order to choose the best algorithm in prediction the blue water footprint trend.

1.5 Significance of Study

The study on water footprint by using the blue water footprint assessment is very important to assess full water utilization in Semambu and Panching water treatment plants. By using Random Forest and Artificial Neural Network algorithms for the calculation of total blue water footprint, the amount of water loss can be known for each water treatment plants. Calculation of total blue water footprint in Kuantan river basins will be based on the data obtained from 2015 to 2017 throughout the study. End of this study, the total blue water footprint prediction trend will be produced using those two algorithms. Hence, the effectiveness of both algorithms will be compared in order to choose the best algorithm to be used. The selected algorithm can be widely practice by the government parties and management in order to improvise the quality of water supply in the worldwide.

CHAPTER 2

LITERATURE REVIEW

2.1 Importance of Water Consumption Calculation

Water consumption means water either permanently or temporarily taken from the underground or surface water resources before been distributed for the consumer. There are several mode to differentiate the water usage. The first one is in the stream used. In-stream activities such as hydroelectric power sources and swimming, water are not being used up but the water quality can be downgraded through pollution. The next mode of water usage is the produce of water which including daily used for household, industrial activity, irrigation and agricultural purposes. Most of the water production are consumptions which mean those activity requires the use of water and cannot be returned to the source.

In order to calculate the volume of water consumption, amount of water taken from the sources which can also called water intake need to be measured along with the amount of water returned. Difference between water intake and water returned will identify the correct amount of water consumption.

$$Water\ Intake - Water\ Returned = Water\ Consumption \qquad 2.1$$

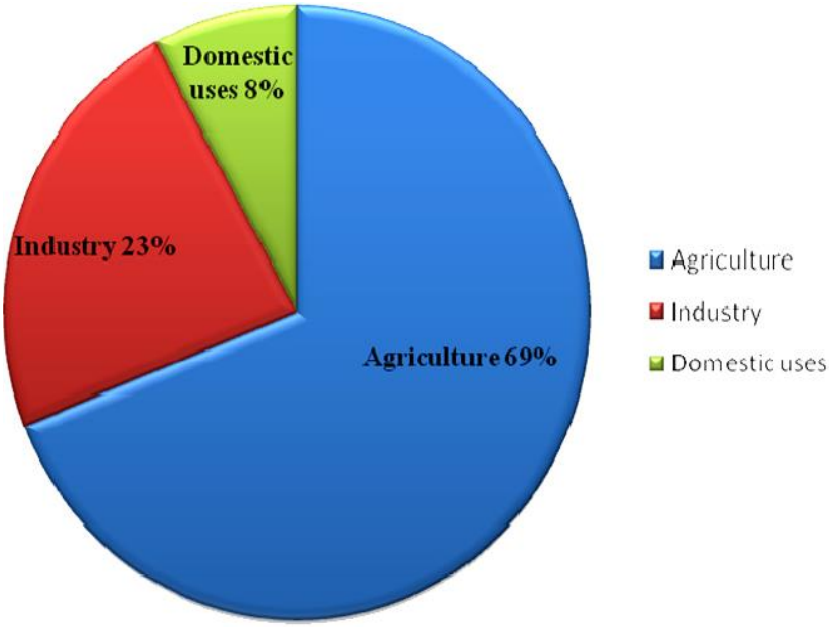


Figure 2.1 Pie Chart of Water Consumption in a country

(Source: https://www.researchgate.net/figure/Total-water-consumption-by-different-sectors - fig1_275964934)

From Figure 2.1, most of the consumers come from the agricultural sector which 69% of water been consumed. 23% of the consumers come from industry sector while another 8% is for the domestic used. Water that needed by the consumer will undergoes some processes of treatment before the good quality of water been distributed. Water from the river basin will be directly go for treatment process in Water Treatment Plant in order to get an acceptable condition for the end-use.

In this research, water footprint will be used to calculate overall water consumption in a WTP. The calculation will only cover within water supply treatment process. Along this treatment process, there are some amount of water that will evaporated due to the surrounding temperature and there will be also addition of water due to the rainfall. These two items will be taken into account in the calculation of total blue water footprint in WTP.

$$Total\ WF_{blue} = Water\ Intake + (Rainfall\ intensity\ x\ Area) + (ET_o\ x\ Area) \quad 2.2$$

2.2 Water Footprint

2.2.1 Water Footprint General

Water footprint can be used in every production and services activities which uses water. Different process possessed its own water footprint. Water footprint concept comes from the Water Footprint Network (WFN) that been created by Dr. Arjen Hoekstra. Water footprint calculation is very important as water been used for almost every important activity in our lives. The amount of water footprint reflects the real demand and occupancy of a person, region or country towards water resources (Liu, Guo, Li, & Zheng, 2019). Water footprint approach will provides useful information for the scientific used of limited water sources.

Water footprint has become as one of the important reference in estimating the impact of products, processes, services and organisation towards water resources (Girolamo, Miscioscia, Politi, & Barca, 2019). Besides, water footprint indicator quantifies the amount of freshwater use as a productive factor which consider direct and indirect used by the producers and also consumers (Zhai et al., 2019). Water footprint also related to the concept of virtual water which overall water consumption throughout the production chain will be taken into account (F. Wang et al., 2019). Furthermore, water footprint approach emphasize the result of observing the entire chain of products by mapping out the magnitude of water consumption.

Water footprint also functions in evaluating water resource requirements by the consumers and depending to the process (Hogeboom et al., 2018). Besides, water footprint assessment is a process measuring the sustainability of water consumption and set up preferable actions in order to have sustainable footprint. Although there are different types of water footprint which is green, grey and blue, this study will only focus on blue water footprint assessment to assess full water utilization in the water treatment plants. Total blue water footprint accounting are involving three parameters which is water intake, rainfall utilisation and total evaporation. This approach will evaluate the total water consumption either water been added into the process or loss during the process.

2.2.2 Blue Water Footprint

Blue water footprint is the fresh water that been taken either from surfaces or ground water resources (Zhuo et al., 2019). The process in acquiring the water can be in several ways. It can be by evaporated from one source of water into a product the being return to another or returned after a period of time. Activities such as irrigation, industry and household used water which each have blue water footprint. Moreover, this assessment can also be used to observe efficiency level of the water treatment plants can achieved in purifying water resources (Dur, 2018).

Blue water footprint used as a comprehensive indicator in assessing water resource consumption by considering all direct and indirect process (Ma et al., 2018). This approach is same as the traditional virtual water method which includes an inventory analysis and also consider the environmental impacts toward the entire life cycle of the activities, processes and products. Besides, blue water footprint can revealed the link between water use and the consumption as well as the link between the water management and the global trade (Zhenzhen, Heating, & Wang, 2019). Consumption of water refers to the freshwater withdrawals where the water been evaporated or incorporated into a products and been transferred to different watersheds (Harding, Courtney, & Russo, 2017).

Blue water footprint approach also can be used in quantifying the water scarcity and vulnerability by comparing the ratio of water consumed and water available (Veettil & Mishra, 2016). This information is very useful in identifying water sustainability and to reveal the pattern of geographical hotspot. In addition, blue water footprint accounting also can be used as per unit materials which means the amount of water contributes in producing a materials (Gerbens-leenes, Hoekstra, & Bosman, 2018). This accounting can be classify as blue water footprint where the process along the production also consumed water. Blue water footprint has its own procedures which includes water sustainability assessment and been evaluated from social, economic and an environmental perspective (Civit, Piastrellini, Curadelli, & Pablo, 2018). Therefore, water footprint approach will give clear information that could help in water management and maintain the sustainability.

2.2.3 Blue water footprint application in goods

Previous research has been made by Chapagain & Hoekstra (2011) where a global assessment of the blue, green and grey water footprint of rice by using local data on actual irrigation and high spatial resolution. The environmental effects also been taken into account during this study where the environmental impact of blue water footprint in the production of rice is depends on the location of the water use and the timing of the climate change. Besides, to produce the foods for nation, large quantities of water is required. A study by Hess, Andersson, Mena & Williams (2015) measured blue water footprint of food consumption in the United Kingdom. The study estimated the use of virtual water and global datasets of water scarcity which could help in understanding the potential environmental impacts of alternative diet.

Gush et al. (2019) has performed a study by using Water Footprint Network approach in determining an apple (Malus pumila) orchard growing under the Mediterranean climate condition in South Africa. Blue water footprint and green water footprint were measured through the amount of water involved in transpirations, evaporation, rainfall and irrigation while grey water footprint been determined from fertilizers application. The scale combination of this three water footprint components data been extrapolated to watershed scale by monthly representative of means.

In addition, study also been made for blue and green water footprint accounting in soil water balance by Hoekstra (2019). Comparison between blue and green water footprint in order to distinguish between the consumption of groundwater or surface water versus rainwater. This study allowed for a precise estimation of green and blue water footprints of crop production and for the accurate assessment of irrigation efficiency. Besides, blue water footprint also can assessed the amount of water consumption in a production chain (Gerbens-leenes et al., 2018). An assessment of blue water footprint has been conducted towards five construction materials which is chromium-nickel unalloyed steel, unalloyed steel, Portland cement, Portland composite cement and also soda-lime glass. The total amount of water that been used or loss in the production of these materials been measured and compared to grey water footprint. Grey water footprint will quantify the total amount of fresh water used to assimilate the pollutants.

Zhai et al. (2019) has made a research about water footprint analysis of wheat production. Wheat production also contributes to the global water consumption burden and this study been made to quantify the water consumption throughout the production of wheat in Shandong Province, China from 2009 until 2015. The result showed that largest proportion possessed by grey water footprint, followed by green and blue. The same study has been made by Civit et al. (2018) where comparison between the three water footprint components towards the production of the most relevant variety of grapes and for irrigation system. The result obtained can be used to assist the decision makers about the process of winemaking in the region which contributing to environmental sustainability of the water usage.

Water footprint approach also been used by Xie, Zhang, Wang & Huang (2019) in studying the impacts of shale gas development towards water resources in China. Basically, this study aimed to measure the water intensity in shale gas extraction in China and used water footprint assessment in order to identify the impact of shale gas development towards local water resources. The result showed that heavy amount of wastewater been generated in gas production affects radically affects the amount of water footprint. Furthermore, high amount of water required in the shale gas operation is not affecting the local water supply significantly.

2.2.4 Blue water footprint application in services

Hogeboom et al. (2018) has made an estimation of the blue water footprint of the world's artificial reservoir and attribute it to the aim hydroelectricity generation, residential and industrial water supply, irrigation water supply and flood protection based on their economic value. The blue water footprint in this estimation is the sum of the water footprint of dam construction and the evaporation of water from the reservoir's surface area. Water consumption from artificial reservoirs need to take in account due to the increasing demand or freshwater, increasing water stress levels and continuing dam developments. By estimating the water footprint, substantial variability around the global average can be predicted. Water footprint assesment will provide clear information about the total amount of water being used within the process of producing any products or services

Blue water footprint approach also been applied by Laan, Vahrmeijer, Bristow & Annandale (2017) to investigate sustainability of Steenkoppies Aquifier in South Africa. This research been conducted by comparing the water sufficiency and consumption in a catchment where the information used to develop water footprint framework that gives clear views about hydrology condition of the aquifer. Result from this research indicates that irrigation activities on the Aquifer is unsustainable. This is due to the discrepancies between inflows and outflows water at the catchment.

Dur (2018) has proposed a framework of management assessment regarding blue and grey water footprint to observe the treatment and disposal of wastewater in WWTPs and the efficiency in purifying water resources. This approach illustrated the benefits role of water footprint for WWTPs optimization. In addition, the proposed indicator allows authorities and stakeholders to estimate the margins of quality in the operation activities of WWTP. This research also enable those party to improve freshwater management towards the current demand with the assessment of water resource activity and water cycle impacts.

Wastewater treatment plants (WWTPs) have an important role in protecting received water from untreated discharges (Morera, Corominas, Poch, Aldaya, & Comas, 2016). In the meantime, WWTPs process also give affects to the environment. With respect to that, water footprint estimation is important which it can provide information in evaluating the impact of WWTP with the use of freshwater. In the other words, water footprint assessment also can reduce the fresh water scarcity risk. This assessment are becoming famous because there is about four billion people face severe water scarcity, globally (Owusu-sekyere, Jordaan, & Chouchane, 2017). The calculation of water footprint can be presented as below:

$$WF = WF_{blue} + WF_{grey} + WF_{green} \qquad 2.3$$

Water footprint methodology also allowed the estimation of direct and indirect water consumption that required for the product chain (Casella, Rosa, Salluzzo, & Gisi, 2019). Study that been made deals water footprint estimation in temporary river catchment by combining Geographical Information System (GIS) and Food and

Agriculture Organization of the United Nations (FAO's) water productivity model. Result from this study showed great contribution of green water footprint compared to blue with 686 mm3/year.

Despite of using water footprint at water treatment plants, Cai, Liu & Zhang (2019) performed a study at urban household consumption in China which also contributed to pollutant emission and water consumption. In this study, water footprint been used to calculate the effect of urban household consumption towards water resources. Blue water footprint and grey water footprint were represented by water quantity and water quality respectively. This study came out with a result of water footprint per capita of urban household consumption decreasing from 1992 to 2012 due to the increasing technology innovations. Besides, most of the total water footprint were contributed by food consumption at Chinese urban households.

2.2.5 Blue water footprint application towards water scarcity

The previous study by Veettil & Mishra (2016) is about the quantitative assessment of water security by using blue and green water footprints. The water footprints approach can improve water resources management from local up to regional scale. By considering about climatic and anthropogenic factors, an integrated modelling framework has been developed to identify variability of blue and green water availability along with water security quantification at a river basin. The model proposed is helpful in providing a clear picture of the water security within the watershed and to investigate waters stress regions within the river basin. Besides, Soil and Water Assessment Tool (SWAT) been applied to measure the availability of fresh water (blue and green water) in Savannah, USA (Veettil & Mishra, 2016). Water footprint assessment provide very useful information in understanding water consumption.

Li, Xu, Wang & Tan (2018) performed a study using water footprint accounting to analyse dual scale water stress which is water quantity and quality. This study been conducted at Haihe River Basin (HRB), China to assess natural water availability and optimizing allocation among several jurisdictions in order to improve watershed sustainability. In this study, Blue Water Footprint Index (BWFI) and Grey Water Footprint Capacity Coefficient (K) been produced to ensure the evaluation of water scarcity can be done comprehensively. Result from this study showed that most of the cities that covered by HRB are suffering extreme water scarcity.

A research about reducing water scarcity through water footprint reduction in agriculture has been conducted by Nouri, Stokvis, Galindo, Blatchford & Hoekstra (2019). The research aimed to assess the possible method to reduce water scarcity by the alleviation of water footprint in the production of crop by the application of soil mulching and drip irrigation. The global water footprint assessment (WFA) been used in this research to assess the blue and green water footprint of ten crops at the Upper Litani Basin (ULB), Lebanon. This research produced a result that shows the crop production sensitivity is more to climate rather than soil type. Besides, the blue water saving from mulching combined with drip irrigation and mulching only has been estimated with a value 8.3 million m^3/year and 6.3 million m^3/year respectively.

Novoa et al. (2019) has conducted a research about water footprint variability in order to improve water management in Chile. This research been conducted at Cachapoal River agricultural basin under different climate variability throughout the year which is dry, wet and normal. The water footprint result provide the information needed in the assessment of water consumption by considering the agricultural multiple variables and also production. Moreover, the application of the results lead indicators used to deeply understand the process flow and improve water management allocation plans.

Water footprint assessment also has been done by Xu, Li, Wang, Cai & Yue (2018) for optimal industrial water utilization and allocation in Dalian City, China. Overall blue and grey water footprints been evaluated to set up the water allocation models. This allocation model was not only focused on physical water but also revealed the water flows either imported or exported throughout the production process. External water footprint showed the greater value which 72.58% of the overall water footprint. Furthermore, the water allocation model showed that the water allocation plan meet the requirements for blue and grey water footprints in the industry. This two water footprint assessment came out with a useful information that can be applied for the water allocation plan. Therefore, this model is applicable in future water management for sustainable water utilization.

2.2.6 Concluding statement

In this study, blue water footprint will be used as a tool same as in Veettil & Mishra (2016), (Morera et al., 2016) and Dur (2018) in order to enhance water utilisation. In addition, this study will be focussing on water treatment plants and total blue water footprint will be accounted. The data of total blue water footprint will be used to obtain the prediction trend for the future year. Therefore, this study will enable stakeholders or authorities in ensuring the water treatment plant sustainability in the future.

2.3 Algorithm

2.3.1 Algorithm General

Algorithm is a set of instructions for solving some problems by following step by step. Typically, algorithms been executed by computers but human also have algorithm as well. Algorithms can perform data processing, calculation and automatic reasoning tasks. Besides, algorithm is an effective alternative that can express within a limited time and space which can be defined in different formal language in measuring a function (Tauer, Date, Nagi, & Sudit, 2019). Algorithm is also used to operate data in multiple ways where it can be by inserting new data sets, finding a particular item or classifying an item.

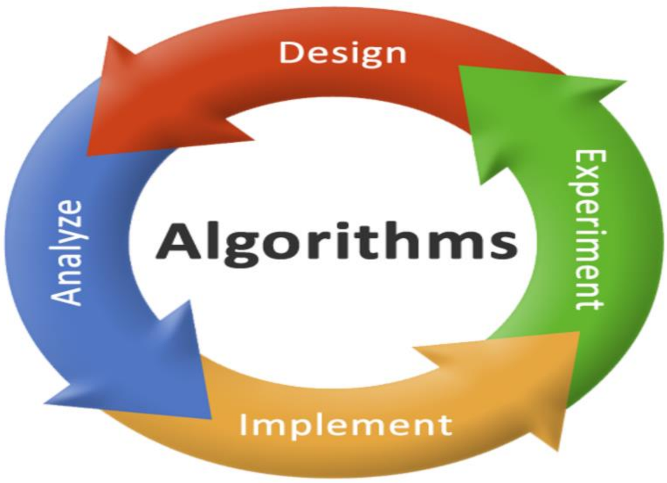


Figure 2.2 Algorithm Procedure

(Source: https://www.quora.com/What-is-an-algorithm-What-are-the-advantages-and-disadvantages-of-it-What-are-its-characteristics)

Figure 2.2 shows the process on how algorithm been used. In using an algorithm, experiment need to be done in order to get the data sets to be inserted as the input. Perfect design of the model is also important so the expected results can be generated. All the variables prepared should be analysed before implementing the output finding process. Algorithm functions provide various advantages. First, algorithm provides solution to any given task which it is able to understand. Algorithm does not depend on any programming language and easy to be handled by the users. By using algorithm, the given tasks is broken down into smaller steps hence it is easier to be converted into real program.

Algorithms are often associated with computer science. Computers used algorithms to get detailed instructions for carrying out any manipulation (Villacampa, Navarro-gonzález, Compañ-rosique, & Satorre-cuerda, 2019). There are various algorithms that can be used in accomplishing any given tasks by entering an appropriate data into the system. In addition, algorithms been widely used throughout every information technology sectors. A searching algorithm, as example it takes few keywords as the input data, then it searches relevant data and produce the output or also can be called results.

2.3.2 Choosing algorithm towards different roles

In choosing an algorithm, there are some factors that need to be considered in order to produce the desire outputs. There is no algorithm who fits to solve all kinds of problems (Lindauer, Rijn, & Kotthoff, 2019). Some problems are too specific and require a suitable approach. Other problems that are very open can be solved using trial and error approach. Besides, choosing a machine learning algorithm is less related to the technical aspects but more to do with the decisions.

The quality or applicability of any Machine Learning prediction methods are highly depend on the size and precision of the data (Ali, Muhammad, Brahme, Skiba, & Inal, 2019). Data that been collected or prepared need to be understand in order to choose the suitable algorithms. Certain algorithms need to work with small data sets while some algorithms require multiple samples. It is very important to know about the data. Next step is by categorizing the problems which can be separated into two process which is categorized by input and categorized by output. This categorization will help in deciding which algorithms need to be chosen towards the problem.

2.4 Artificial Neural Network (ANN) Algorithm

2.4.1 ANN General

ANN is one of the modern artificial intelligence that can solve non-linear functions, prediction, system identification, modelling, forecasting, data sorting and simulation comprehensively (Đozi & Uro, 2019). ANN functioning as a connecting system which inspired by the biological neural networks that form animal brain. Besides, ANN being widely used due to its clear model with a good performance and easy to be implemented. There is no obvious definition and dictate connection between the input and output variables. ANN modify the real data from the past and adapt the model before producing the exact outputs that suitable to the actual data. ANN is also a mathematical structure that is capable in identifying relationships between the input data sets toward its output. This machine learning provides a framework for various machine learning algorithms which could process any complex data sets.

ANN is also a computational model that inspired by the natural unit of the nervous system. This model stimulated the organization connection in the form of information processed in the animals brain (Jimenez-martinez & Alfaro-ponce, 2019). ANN provides a new approach to design the algorithms without knowing the internal mechanism of a system (Hou, Yuan, Ma, & Sun, 2019). A good training data of ANN will lead to good estimation of outputs even the information are unavailable. ANN can trained large numbers of input combinations in alike (Poort, Ramdin, Kranendonk, & Vlugt, 2019).

In addition, ANN is based on a group of connected data called artificial neurons. Every single neuron has its own internal value which called activation value (Poort et al., 2019). The value been transmitted as a signal from one neuron to another. Signal between the connections of the neurons is a real number and the output is calculated by non-linear function of the sums of its early data sets. The receiver neuron can process the further signal to another connected neurons.

Artificial neurons are group into layers which different layers which each layers may perform different kinds of changes on their inputs. Signals sometimes are not directly delivered from the input layer to the output layer on the first transformation, but it might be after several times. ANN generate the results by using a statistical parameters based on the data sets and correlation. The aim of the ANN is to solve any tasks in the same way that a human brain would. In the brain, the process of gaining knowledge or learning something from a specific task happened through experiences and does the same process till the required objective been achieved (Ali et al., 2019).

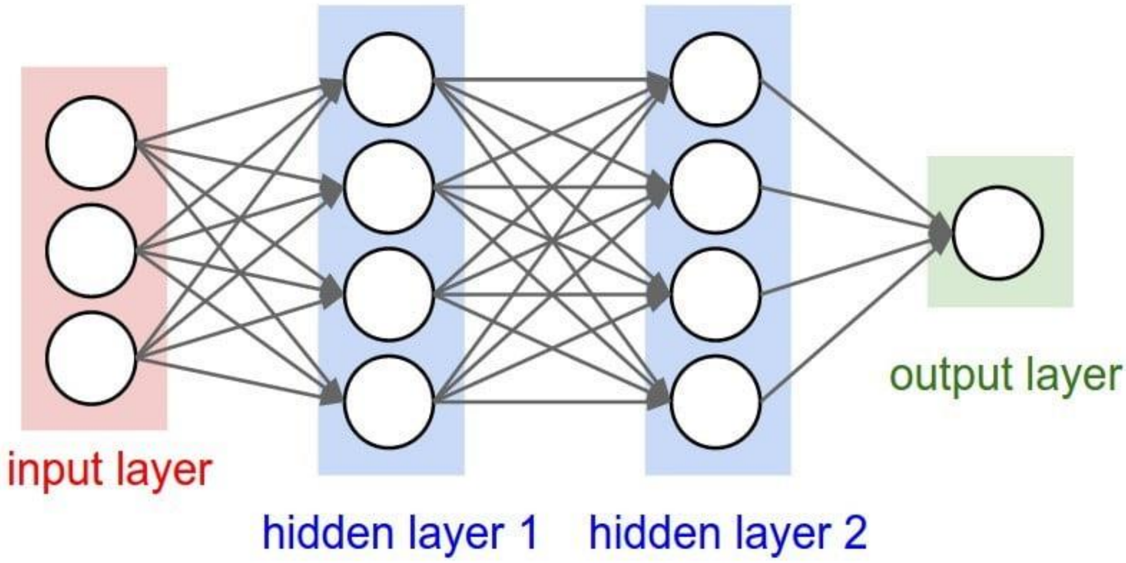


Figure 2.3 Basic idea on how ANN works

(Source:https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/)

Figure 2.3 shows on how deep ANN learns with a factory line analogy. After the data sets are input, all the data will travel to the conveyer belt which subsequent layers will extract different set of features. The first layer will recognize the exact item of the input and analyze one of the features of the input. The next layer will then identify every single features of the item for example if it is an object, the layer will recognize every features in order including the image edges, textures and shapes. After undergoes all the layer, researcher who trained the network can name the output. In addition, back propagation can be used after the getting the output in order to correct any fault that been made. Once the network produced the output, the network itself can perform classification task by its own without any humans help. This is due to the learning styles of this algorithm that learned from previous data, choose and apply the possible solutions to achieve the required objective.

2.4.2 Previous application of ANN algorithm

A study by Piotrowski, Napiorkowski, Napiorkowski & Osuch (2015) showed a comparison between few types of data-driven neural networks and nearest neighbour approach in short time of stream water temperature for two natural catchments which is mountainous and lowlands with different climate zone. Calibration of each neural networks has been independently made about 100 times and the median, mean and standard deviation been used in comparing the networks. Multi-Layer Perceptron (MLP) is one of the most popular ANN that consist a group of nodes which inserted into the input, hidden and output layers (Haykin, 1999). A single hidden layer sometimes sufficient in approximating a continuous function but there is no any rules regarding this issue. Next, Product Units Neural Networks (PUNNs) is opposed to other neural networks where this networks used fewer variables to be optimized. PUNNs is claimed to be the difficult network to be trained but the convergence of the gradient-based for this network has been proved (Zhang et al., 2008).

Safari, Aksoy & Mohammadi (2016) compared three different ANN techniques which is feed-forward back propagation (FFBP), generalized regression (GR) and radial basis function (RBF) for modelling incipient deposition of sediment in rigid boundary channels. This research was conducted using six parameters that been taken from laboratory experiments. Parameters used are flow discharge, channel bed slope, hydraulic radius, flow depth, median size of sediment particles and relative specific mass of sediment. In this research, by comparing between the three developed ANN models, FFBP is found better to other ANN and all regression model. Performance of any ANN models can be related to the amount of variables taken input and the relative particle size in regression model. This research concluded that ANN model and regression that appropriately set up can be successfully used for any estimation.

A previous study by Ahmad & Simonovic (2005) focused on the estimation of the trend of flood hydrograph by using ANN approach.. This research used ANN in predicting the timing, peak flow and shape of runoff hydrograph based on causal meteorological variables. Five different variables were used in developing runoff hydrograph in Manitoba, Canada which is antecedent precipitation index, winter precipitation, melt index, spring precipitation and timing. A feed-forward ANN is trained on the previous data using back percolation algorithm, In order to produce the desired output, the selection of convenient input variables is plays an important role. Nevertheless, the proposed ANN based hydrograph estimation technique is a useful technique for watershed simulation techniques where limited topographic data is available while time and full understanding of physical process of watershed is a constraint.

ANN is an alternative that been used to generate a non-linear mapping between data sets of a model. Study by Ramı, Cleofe & Jesus (2005) used ANN approach in order to generate site specific quantitative forecast for daily rainfall. In the other words, ANN recognizes the hidden patterns of data naturally and then train itself as well as validating with the its own existing knowledge (Banerjee, Singh, Chatttopadhyay, Chandra, & Singh, 2011). ANN able to understand the relationships between the parameter attributes and beneficial in defining problem by not requiring specific solution. Besides, ANN also can trained and produce the output without knowing the input parameters. This algorithm will train the variables based on the previous exerienced and apply the same process to solve the problems.

2.5 Random Forest (RF) Algorithm

2.5.1 RF General

Random Forest (RF) is a flexible and easy machine learning algorithm to be used. RF is one of the most used algorithms due to its simplicity and the fact stated that RF can be used in either classification or regression tasks. The name itself can show how this algorithm works. It creates forest and make it somehow random. The ‘forest’ is a group of ‘Decision Trees’. In a simple way, RF builds various decision trees and combine them together to get more precise and accurate prediction.

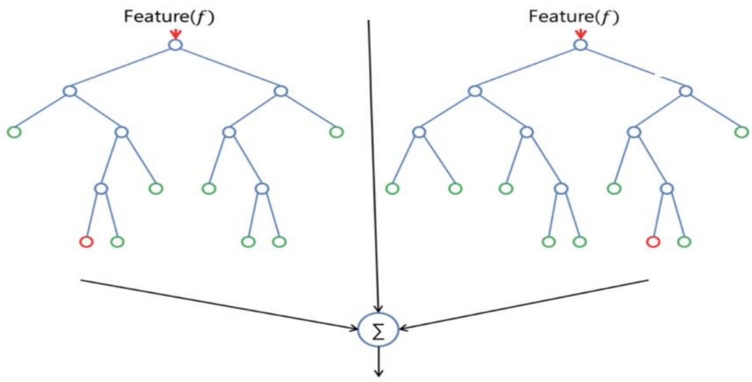


Figure 2.4 How Random Forest look alike

(Source: https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd)

RF adds randomness to the model as well as growing the trees. In the process of searching the important parameters, RF also classify the best feature among a random set of features. This classification will produce comprehensive results in a better model. In addition, only certain features is taken for consideration by the algorithm in splitting a node. RF is slightly difference to another algorithms because it build and choose decision while another algorithms illustrated decision.

2.5.2 Previous application of RF

A study by Chen (2018) explored detection and diagnosis of PV arrays faults by using RF algorithm. This study only takes the real time operating voltage and string currents of PV arrays as the features. In order to optimize the variables of RF, grid-search method is being used by minimizing the error estimation as well as to improve the fault diagnosis model. Comprehensive fault experiments was carried out to obtain sufficient fault data samples. This fault diagnosis model has been successfully integrated in a software called Matlab.

RF algorithm been chosen in a study to estimate biomass in wheat. Wheat biomass can be measure using appropriate spectral vegetation indices but the accuracy of the estimation are not stable (Zhou, Zhu, Dong, & Guo, 2016). Previous study presented are more focusing on developing vegetation indices however limited study exist on modelling algorithm. This study successfully carried out which resulting RF model generates more accurate estimation compared to Support Vector Regression (SVR). RF algorithm provides a handy exploration and prediction in estimating a large scale of wheat biomass in Southern China.

A comparison has been made between RF algorithm and 7 other algorithms in logging regression modelling. The other algorithms are squared linear regression, support vector regression, regression tree, artificial neural networks, gradient descent boosted trees and k nearest neighbour regression. RF algorithm showed a strong learning abilities, robust and feasibility of the hypothesis (Ao, Li, Zhu, Ali, & Yang, 2019). Through this study, the excellence of RF for logging regression modelling is proved. Another study for RF algorithm was carried out for micro kinetic modelling and the computationally integration of micro-kinetics into reaction engineering model. RF can be used to identify the new data sets while keeping the high prediction and low computational load (Partopour, Paffenroth, & Dixon, 2018). This study asserted that RF can be used to identify features of any mechanism over high range of reacting conditions.

CHAPTER 3

METHODOLOGY

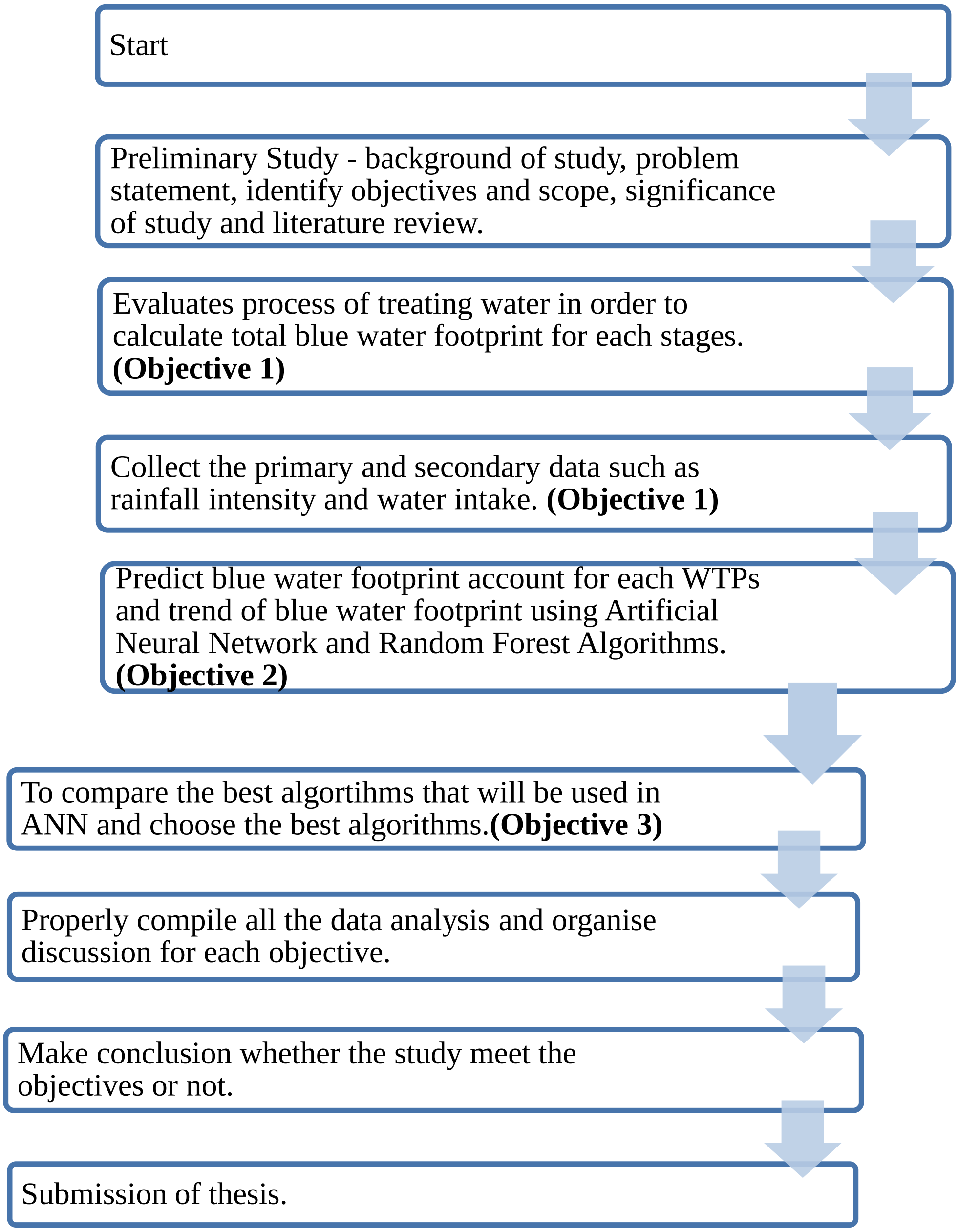
3.1 Introduction

Global Water Footprint Standard has been used and applied worldwide covering many sections which includes detailed instruction and proper guidance. This standard provides high level quantification and tough analytics in performing Water Footprint Assessment (WFA) to maintain water sustainability journey. WFA is important in order to calculate the water resource requirements by the consumers for the products and services (Hogeboom et al., 2018). This assessment is focused on the method to calculate the total blue water footprint which also can be called total water consumption to provide goods and services.

Blue water footprint calculation will only cover in Water Treatment Plant starting from the water intake until the storage process of the water which before the water being distribute to the consumer. Total amount of water evaporated and rainfalls intensity will be taken into account for total water consumption calculation. Blue water footprint here can be defined as the total summation of the water consumed in every stages of water supply treatment process and formed the blue water footprint formula. In this study, the prediction of blue water footprint trend will using WEKA software and two chosen algorithms which is Artificial Neural Network and Random Forest as the training algorithms.

$$WF_{blue} = Total\ Water\ Intake + (Rainfall\ Intensity\ x\ Area) + (ET_o\ x\ Area) \qquad 3.1$$

3.2 Flow of Study



3.1 Flow of Study

3.3 Study Area

Kuantan river basin is located in the district of Kuantan which at the north eastern end of Pahang State in Malaysia. It is one of the most important river basins in Pahang that covers for 1630 km2 area of catchment which started from reserved forest in Mukim Ulu Kuantan, Kuantan Town up to the South China Sea. Furthermore, it also consists a numbers of important rivers which flow to the industrial area, rural area and also agricultural.

This study will measure the total blue water footprint for two main water treatment plants in Kuantan which is Panching and Semambu Water Treatment Plants. Both WTPs water intake are freshly from Kuantan river basin.

Figure 3.1 Location of Study

(Source: https://www.google.com/maps)

Panching Water Treatment Plant
(3.86° N, 103.18° E)

Semambu Water Treatment Plant
(3.87° N, 103.33° E)

Semambu WTP is located 18 km from Sungai Kuantan while Panching WTP is also 18 km from the same water sources. Panching WTP will cover for Gambang, Jaya Gading and Panching. Meanwhile, Semambu WTP covers for the area up to the North of Pahang. For Kuantan's Town area will be cover by Bukit Ubi WTP.

3.4 Data Collection

During this research, some departments are indirectly involved in the data collection process. Table 3.1 below shows the data collected with the sources.

Table 3.1 Data collection and departments involved

DATA	SOURCE
Primary data : - **Area of WTP**	- Panching Water Treatment Plant - Semambu Water Treatment Plant
Secondary data : - **Water Intake**	- Pengurusan Air Pahang Berhad (PAIP)
- **Rainfall intensity**	- Jabatan Pengairan dan Saliran Negeri Pahang (JPS)
- **Temperature**	- Jabatan Meteorologi Malaysia (MET)

3.5 Site Visit

In this study, primary data such as Area of WTP are required in order to calculate the total blue water footprint. This kind of data can be obtained directly from the location of study. Proper site visit has been done in order to collect the data for the assessment. Besides, by performing the site visit, the clear view of the water treatment condition can be experienced. Ton of information starting from the water intake going to the stage by stage of water supply treatment process are smoothly gained.

Both water treatment plants still using conventional water treatment process which some stages of the treatment process tanks still exposed to the surface without any cover. This issue will led the rainfall to enter the tanks as well as the water from the tanks evaporated to the environment. Hence total blue water footprint calculation is important to identify the missing information in order to know total amount of water that been consumed along the process.

3.6 Water Supply Treatment Process (WSTP)

Water treatment is any type of process in order to improve the water quality to make it acceptable for every single end-use. The treated water that been distributed can be either direct or indirectly used by the consumer as example for drinking, irrigation, industrial water supply and many other uses. Water treatment removes undesirable contaminants in the water and also fix the concentration so that the water becomes suitable and acceptable for its wish end-use. As been mentioned before, both water treatment plants chosen still using conventional water treatment process which the treatment process undergoes stage by stage start from water intake, aeration, coagulation, flocculation, sedimentation, filtration, disinfection and water distribution. This research will only cover from water intake until water storage before the water being distribute to the consumer.

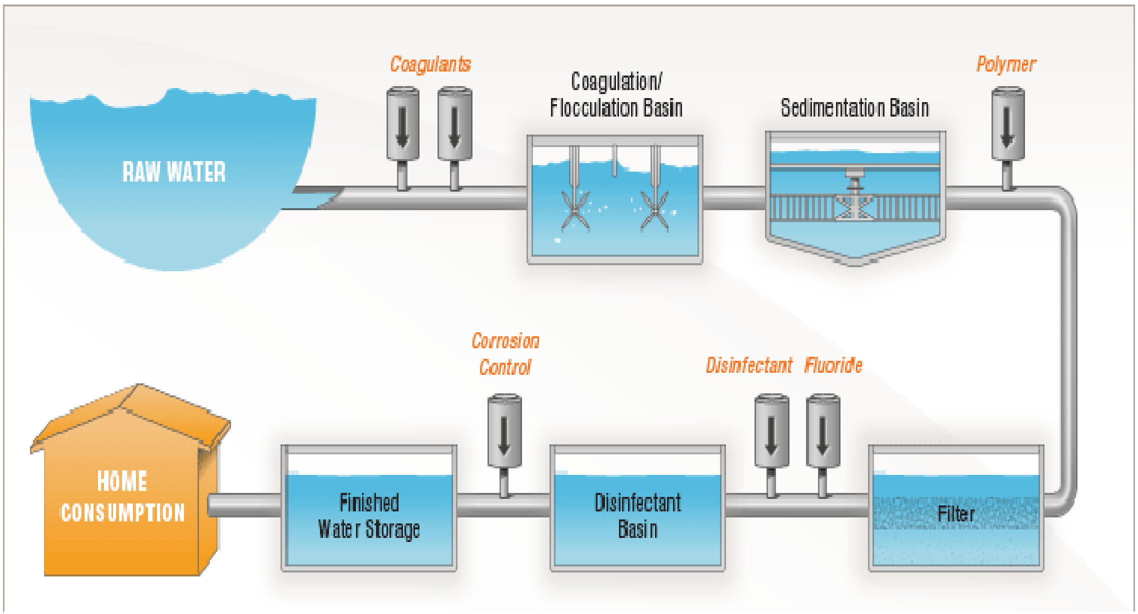


Figure 3.2 Water Supply Treatment Process

(Source: https://www.denverwater.org/your-water/treatment-process)

3.6.1 Stages in Water Supply Treatment Process

1. Water Intake

   A process of abstracting freshwater from any water resources. In this study, water intake are purely from Kuantan river basin.

2. Aeration

   This process is a process where unneeded gases such as CO_2, H_2S, NH_3 being eliminated.

3. Coagulation

   Chemicals are added in this process to make the solid particles keep apart so it can be easily remove in the next stages.

4. Flocculation

   This process involving a slow mixing process which brought all the solid particles together in contact before being removed.

5. Sedimentation

   Gravity plays an important role in this process in order to remove the suspended material from the water.

6. Filtration

   This process remove suspended solids by moving out the water through porous medium.

7. Disinfection

   Aiming to destroy pathogenic microorganisms so the water is fit for drinking.

3.7 Water Footprint Accounting

The total blue water footprint assessment is the measure of the amount of water consumption in order to provide goods and services. Blue water footprint is calculated by the summation of all the water consumed includes from starting of the process until the storage process. Besides, the rainfall intensity as well as the evaporated water from the tanks in the water supply treatment process also take into account in measuring total blue water footprint. The total blue water footprint formula can be presented as the following equation:

$$WF_{blue} = Total\ Water\ Intake + (Rainfall\ Intensity\ x\ Area) + (ET_o\ x\ Area) \qquad 3.2$$

Where:

WF_{blue} = Blue water footprint

ET_o = Evaporation for every tank

Area = Area of each tank

Most of the tanks for every process in water supply treatment process are in rectangular shape, thus the area calculation will just be the length multiply by width.

Evaporation of water at every single stages also need to be added into the blue water footprint calculation because the water losses during treatment process still considered as water consumption. After undergoing detailed review from the previous study, one method been found where the most suitable method in calculating the evaporation of water. Blaney-Criddle method is chosen since the available data just only the temperature. The formula can be presented as below:

$$ET_O = \rho(0.46\, T_{mean} + 8.128) \qquad 3.3$$

Where:

ET_O = Reference evapotranspiration (m^3/day)

ρ = Mean daily percentage of annual daytime hours

T_{mean} = Mean daily temperature (℃)

Blaney-Criddle method can be used as follows:

1. Calculate the mean daily temperature, T_{mean}

$$T_{mean} = \frac{T_{max} + T_{min}}{2} \qquad 3.4$$

2. Determine the value of ρ in the table Mean Daily Percentage of Annual Daytime Hours for Different Months.

3. Calculate ET_O using,

$$ET_O = \rho(0.46\, T_{mean} + 8.128) \qquad 3.5$$

Table 1.15 Mean Daily Percentage of Annual Daytime Hours, P, by Month for Different Northern and Southern Latitudes

Latitude North South[a]	Jan. July	Feb. Aug.	Mar. Sept.	Apr. Oct.	May Nov.	June Dec.	July Jan.	Aug. Feb.	Sept. Mar.	Oct. Apr.	Nov. May	Dec. June
60°	0.15	0.20	0.26	0.32	0.38	0.41	0.40	0.34	0.28	0.22	0.17	0.13
58°	0.16	0.21	0.26	0.32	0.37	0.40	0.39	0.34	0.28	0.23	0.18	0.15
56°	0.17	0.21	0.26	0.32	0.36	0.39	0.38	0.33	0.28	0.23	0.18	0.16
54°	0.18	0.22	0.26	0.31	0.36	0.38	0.37	0.33	0.28	0.23	0.19	0.17
52°	0.19	0.22	0.27	0.31	0.35	0.37	0.36	0.33	0.28	0.24	0.20	0.17
50°	0.19	0.23	0.27	0.31	0.34	0.36	0.35	0.32	0.28	0.24	0.20	0.18
48°	0.20	0.23	0.27	0.31	0.34	0.36	0.35	0.32	0.28	0.24	0.21	0.19
46°	0.20	0.23	0.27	0.30	0.34	0.35	0.34	0.32	0.28	0.24	0.21	0.20
44°	0.21	0.24	0.27	0.30	0.33	0.35	0.34	0.31	0.28	0.25	0.22	0.20
42°	0.21	0.24	0.27	0.30	0.33	0.34	0.33	0.31	0.28	0.25	0.22	0.21
40°	0.22	0.24	0.27	0.30	0.32	0.34	0.33	0.31	0.28	0.25	0.22	0.21
35°	0.23	0.25	0.27	0.29	0.31	0.32	0.32	0.30	0.28	0.25	0.23	0.22
30°	0.24	0.25	0.27	0.29	0.31	0.32	0.31	0.30	0.28	0.26	0.24	0.23
25°	0.24	0.26	0.27	0.29	0.30	0.31	0.31	0.29	0.28	0.26	0.25	0.24
20°	0.25	0.26	0.27	0.28	0.29	0.30	0.30	0.29	0.28	0.26	0.25	0.25
15°	0.26	0.26	0.27	0.28	0.29	0.29	0.29	0.28	0.28	0.27	0.26	0.25
10°	0.26	0.27	0.27	0.28	0.28	0.29	0.29	0.28	0.28	0.27	0.26	0.26
5°	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.28	0.27	0.27	0.27
0°	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27

Note: Appendix B details methodology and equations for computing daytime hours and P to facilitate computer application of this table.
[a] Southern latitudes: apply 6 month difference as shown.

Figure 3.3 Mean Daily Percentage of Annual Daytime Hours

After getting and calculating all the data, total blue water footprint for Semambu and Panching Water Treatment Plants from 2015-2017 can be calculated.

According to the previous study, blue water footprint assessment are important due to it is going under threat that cause by unstable climate changing, limited water supply and the increasing water demand in the world (Veettil & Mishra, 2016). This assessment little bit will help in evaluating the variability of blue water footprint and also quantify the water stress in Kuantan river basin.

3.8 Pre-processing

Firstly, pre-processing process started with the treatment of missing data. Missing data need to be treated before performing the study in order to get the better result. In this study, an average method has been used to recover the missing values. This method is the simplest and easiest method in order to get the missing values as the missing value is not affect much of the overall data and this method is suitable to be used.

Secondly, data normalization. Normalization of data is a must because the existing data range are different to each other. For blue water footprint prediction, values for water intake, rainfall and evaporation will give different range sets of data. Thus, normalization will help to reduce the data in the range of 0 to 1.

Finally, the data has been cleaned by removing the outlier value. Outlier is a data point that significantly differs from other observations. Besides, outlier might be due to the variability in the measurement or sometimes may indicate experimental error or in this study the failure of remote data collection. Removing outliers been carried out by replacing the value that are out from minimum and maximum boundary range with the average value.

3.9 The Best Algorithm Prediction

In order to achieve the last objective where to choose the best algorithm between Random Forest and Artificial Neural Network, these two algorithm need to undergo the training and produce the blue water footprint trend. The result prodcuced at the end of the training will visualize the value of Root Mean Square Error (RMSE). The lowest RMSE will be chosen because it indicates the least error that been made by the algorithm.

Meanwhile, the predicted value produced by the algorithm training will be compared with the actual value of total blue water footprint that been calculated. The least value of RMSE will produce the precise trend between the actual and predicted value of blue water footprint.

3.10 Prediction of Blue Water Footprint Accounting

After total blue water footprint been determined, blue water footprint trend will be produced by using two different algorithms that mostly used in the previous researches which is Artificial Neural Network and Random Forest. These two algorithms will undergo training in the WEKA software.

WEKA is a group of learning algorithms for data mining activities. It contains appliances for data clustering, regression, classification, visualization and preparation. WEKA got its name from a flightless bird that only found on the island of New Zealand with and curiosity behaviour. WEKA is been chose due to its friendly used and the software is easy to be used by less expert user. All types of algorithms are available in the software. Besides, there is also another software that can be used as a predicting tools which is Math Lab. This software are not recommended because it just have the algorithms for back propagation only. Moreover, in order to use the software, users need to spend some money to get the software.

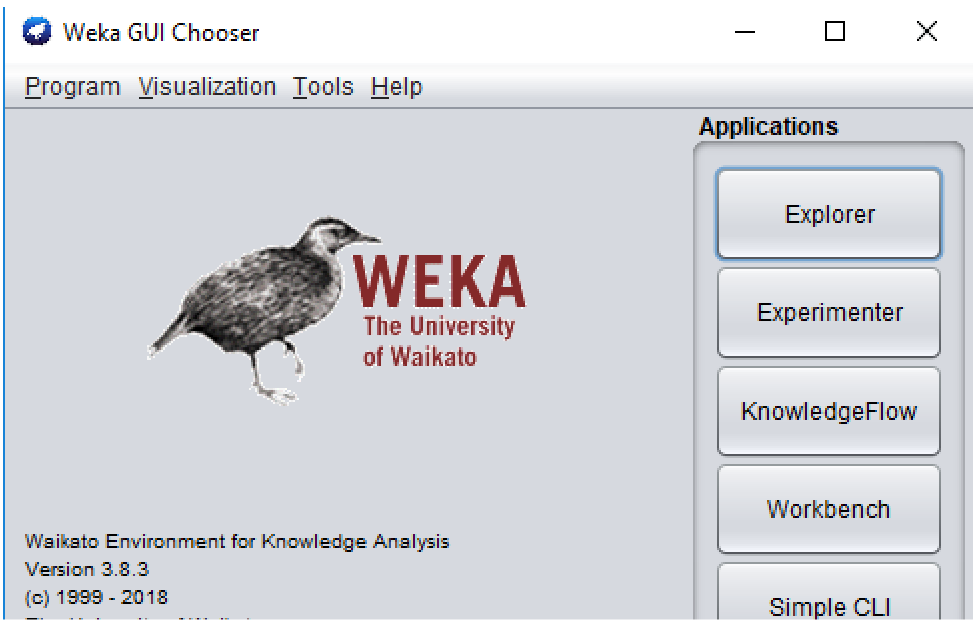


Figure 3.4 WEKA software

The use of WEKA software is as follows:

Step 1: Total blue water footprint calculation will be tabulated in Microsoft Excel by sorting it based on year and water treatment plants.

Step 2: The data in the Microsoft Excel will be converted into Common Separated Value (CSV) format. This action is needed to enable being inserted into WEKA software.

Step 3: Then, the CSV format file which contained the data sets will be trained in the WEKA software by using ANN and RF algorithm.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

This chapter will basically analyse and discuss the result acquired from the study. This study was aiming to calculate the total blue water footprint in Semambu and Panching WTPs as well as to predict blue water footprint trend and to choose the best algorithm between Artificial Neural Network and Random Forest. The study that been conducted on both water treatment plants is just from the year 2015 until 2017.

4.2 Blue Water Footprint Accounting

Blue water footprint calculation is the measure of the amount of total blue water used to produce goods or services. In this study, blue water footprint has been calculated starting from the water intake until the water storage before distributed to the consumer. The total amount of water calculated includes; water intake from Kuantan river basin, total amount of rainfall that enters the open water tank and evaporated water due to temperature.

For both water treatment plants, the conventional treatment will undergo in-line processes which are screening, aeration, mixing chamber, flocculation, sedimentation and filtration. Those processes will affect the calculation of blue water footprint due to the tank condition which exposed to the environment and will be affected by rainfall and evaporation. Total amount of rainfall and water evaporated were multiplied with the total area of all the open tanks at the water treatment plants in order to measure the total blue water footprint.

Table 4.1 Area of each tank at Semambu and Panching WTPs

Tank	Panching WTP (m^2)	Semambu WTP (m^2)
Screening	220	15.6
Aeration	167.825	400
Mixing Chamber	77.2	1000
Sedimentation	1613.7	2000
Filtration	1455.25	2000
Flocculation	1225.5	579.04
Mixing Chamber (Phase 2)	0	241.38
Clarifier (Phase 2)	0	1135.2
Filtration (Phase 2)	0	1260
Total area	**4759.475 m^2**	**8631.22 m^2**

From table 4.1, Semambu total area is higher than Panching, total area will affect the amount of water evaporation and rainfall added into the process. Semambu happens to be bigger due to this WTP is the biggest in Kuantan river basin and supplying to most of Kuantan area.

4.2.1 Total WFblue at Semambu WTP

Table 4.2 Total WFblue in 2015 at Semambu WTP

MONTH	WATER INTAKE (m^3)	TOTAL RAINFALL (m^3)	TOTAL EVAPORATION (m^3)	TOTAL BWF (m^3)
JANUARY	8933775	1394.4165	1047.240295	8936216.657
FEBRUARY	7842250	758.0385	946.8877892	7843954.926
MARCH	8405600	252.6795	1094.969768	8406947.649
APRIL	8238188	455.447	1108.956252	8239752.403
MAY	7433250	761.158	1150.08282	7435161.241
JUNE	7829900	368.101	1115.605621	7831383.707
JULY	8388500	361.862	1149.07881	8390010.941
AUGUST	8469250	2607.902	1130.583877	8472988.486
SEPTEMBER	7626000	689.4095	1096.670332	7627786.08
OCTOBER	8355275	1057.5105	1097.835247	8357430.346
NOVEMBER	8081800	1862.3415	1053.991704	8084716.333
DECEMBER	8446600	2217.9645	1083.935535	8449901.9
GRAND TOTAL	**98050388**	**12786.8305**	**13075.83805**	**98076250.67**

Table 4.2 showed, the total amount of water footprint and parameters calculated for 2015. Water intake is the major contribution to sum up the total blue water footprint.

In 2015, 98050388 m^3 of water abstracted from Kuantan river basin, while 12786.8305 m^3 amount of total rainfall utilised in the process of treatment water for Semambu WTP. 13075.83805 m^3 of water evaporated and utilised in this process for 2015.

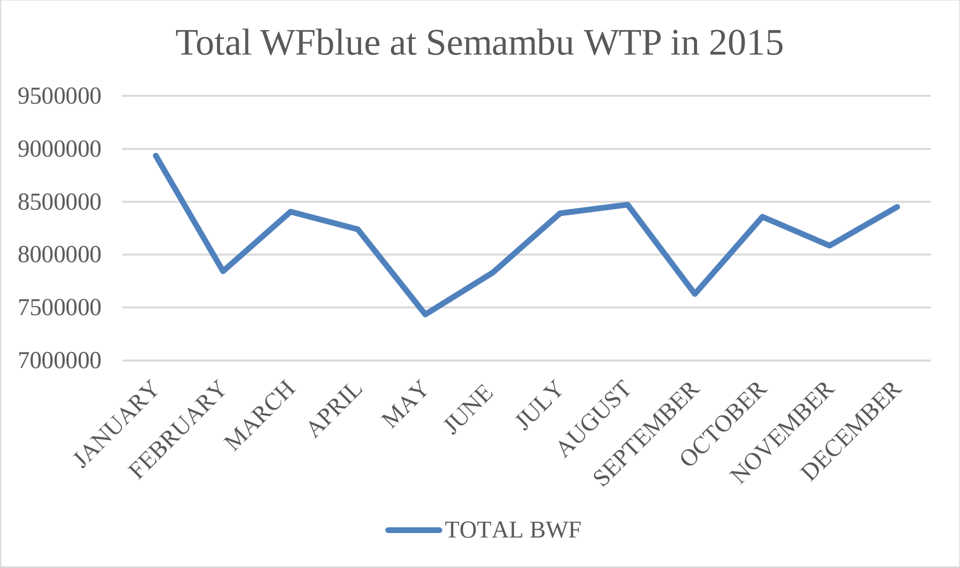


Figure 4.1 Total WFblue in 2015 at Semambu WTP

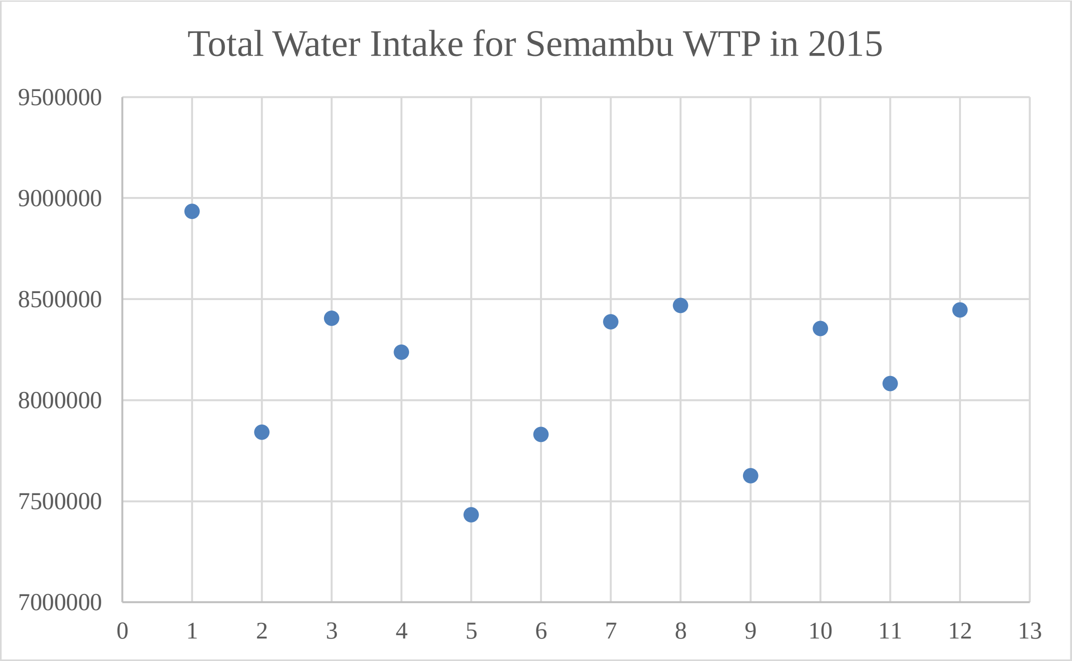


Figure 4.2 Total Water Intake for Semambu WTP in 2015

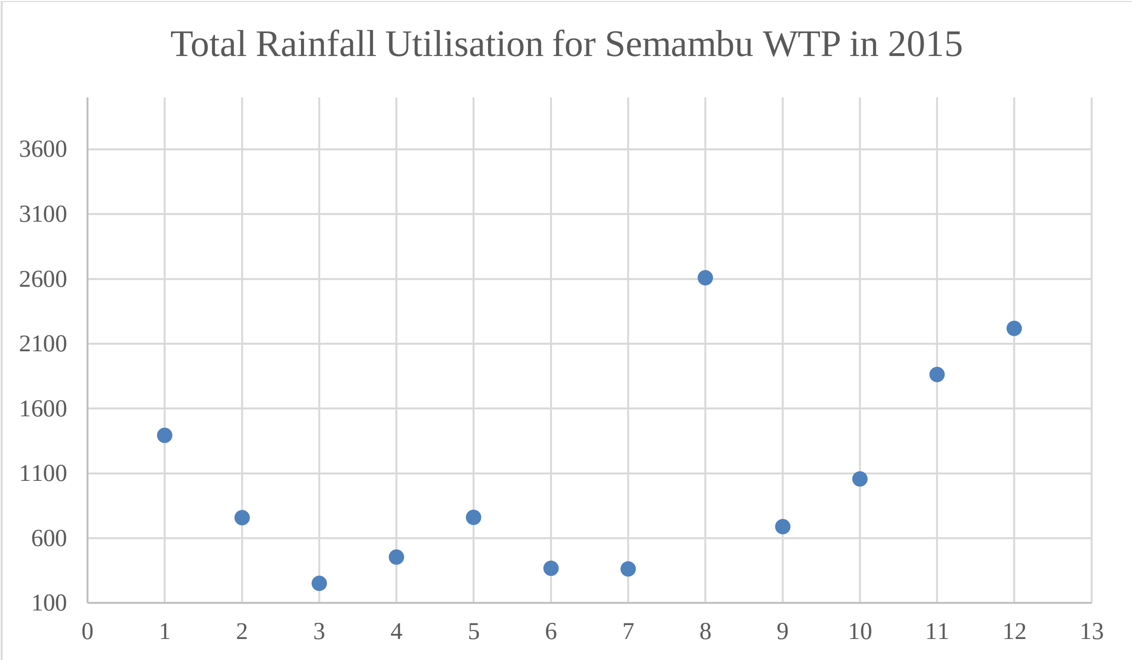


Figure 4.3 Total Rainfall Utilisation for Semambu WTP in 2015

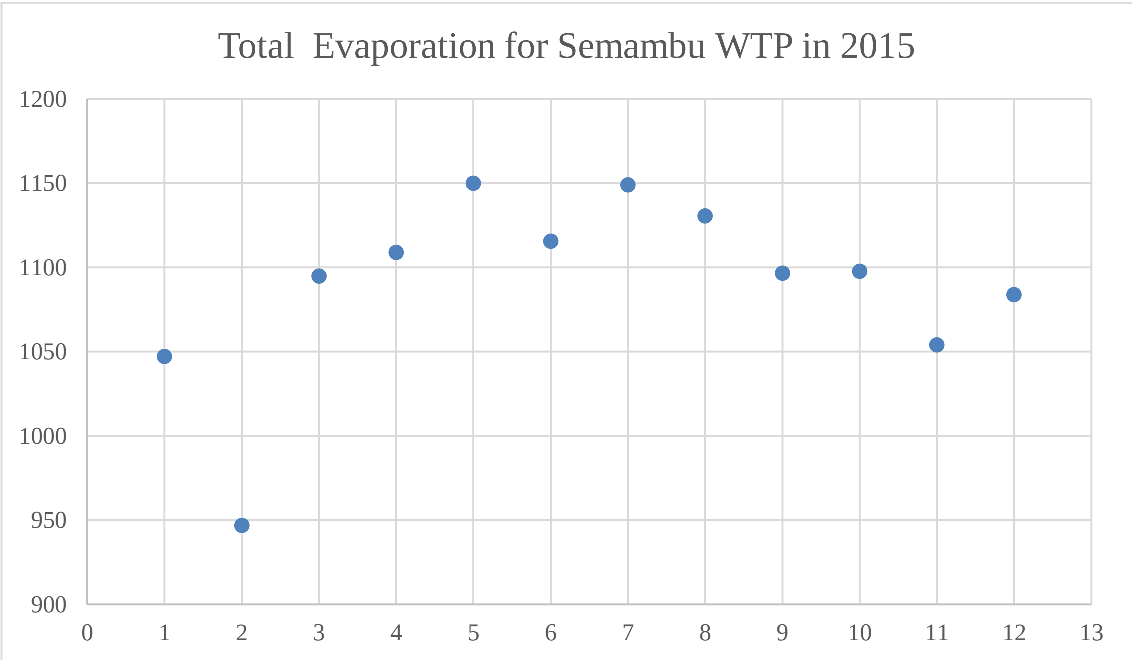


Figure 4.4 Total Evaporation for Semambu WTP in 2015

Based on the graph in Figure 4.1, in January, the WFblue amount is the highest 8936216.657 m3/month. The main contribution to this amount was total water intake, which was 8933775 m3. As the water intake remain higher among other parameters due to the WTP capacity, the amount of rainfall utilised in this process also higher in January, which was 1394.4165 m3 as shown in Figure 4.3. However, the highest amount of rainfall utilised was on December, which was; 2217.9645 m3, this is due to north east monsoon season and East Coast area of Peninsular Malaysia will be receiving abundance of rainfall during this season specifically from October - March. As seen in the Figure 4.3, the amount of rainfall utilisation were gradually increased from October to December. However, due to the global climate change, August shown the greatest amount of rainfall utilisation among all other months. As seen in Figure 4.4, evaporation amount will drop as the rainfall increased. This is mostly due to the temperature, lower temperature presents during the rainy season and will evaporates less water into the system.

Meanwhile the lowest value of total WFblue was in May, which was 7435161.241 m^3, in Malaysia or dry season lies on Southwest Monsoon season because most rainfall will affect West coast area of Peninsular Malaysia during this season which is on May until September. As seen in Figure 4.2, water intake amount were gradually decreased from March and dropped to the lowest amount on May. Dry season will affect the amount of intake because when the volume of river basin decreases, WTP will limit amount of water abstraction to avoid damage to the treatment plant if more volume of sediment abstracted rather than raw water.

Table 4.3 Total WFblue in 2016 at Semambu WTP

MONTH	WATER INTAKE (m3)	TOTAL RAINFALL (m3)	TOTAL EVAPORATION (m3)	TOTAL BWF
JANUARY	8419350	1113.6615	1078.803334	8421542.465
FEBRUARY	7949250	1020.0765	993.7593311	7951263.836
MARCH	8622680	177.8115	1091.890447	8623949.702
APRIL	7939773	62.39	1082.663602	7940917.554
MAY	8182134	427.3715	1114.061557	8183675.433
JUNE	10046538	1235.322	1056.489375	10048829.81
JULY	7847606	274.516	1088.195262	7848968.711
AUGUST	7783991	1968.4045	1095.123734	7787054.528
SEPTEMBER	7154021	1559.75	1038.321382	7156618.571
OCTOBER	8308180	1896.656	1072.413743	8311149.07
NOVEMBER	7240743	2308.43	1053.991704	7244105.422
DECEMBER	7567500	2782.594	1083.935535	7571366.53
GRAND TOTAL	**97061765**	**14826.98**	**12849.65**	**97089441.63**

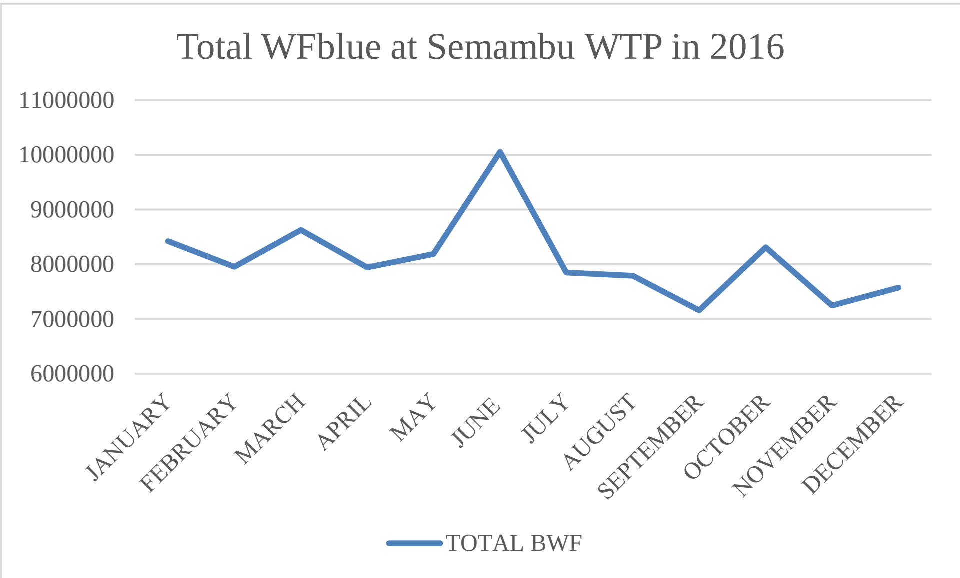


Figure 4.5 Total WFblue in 2016 at Semambu WTP

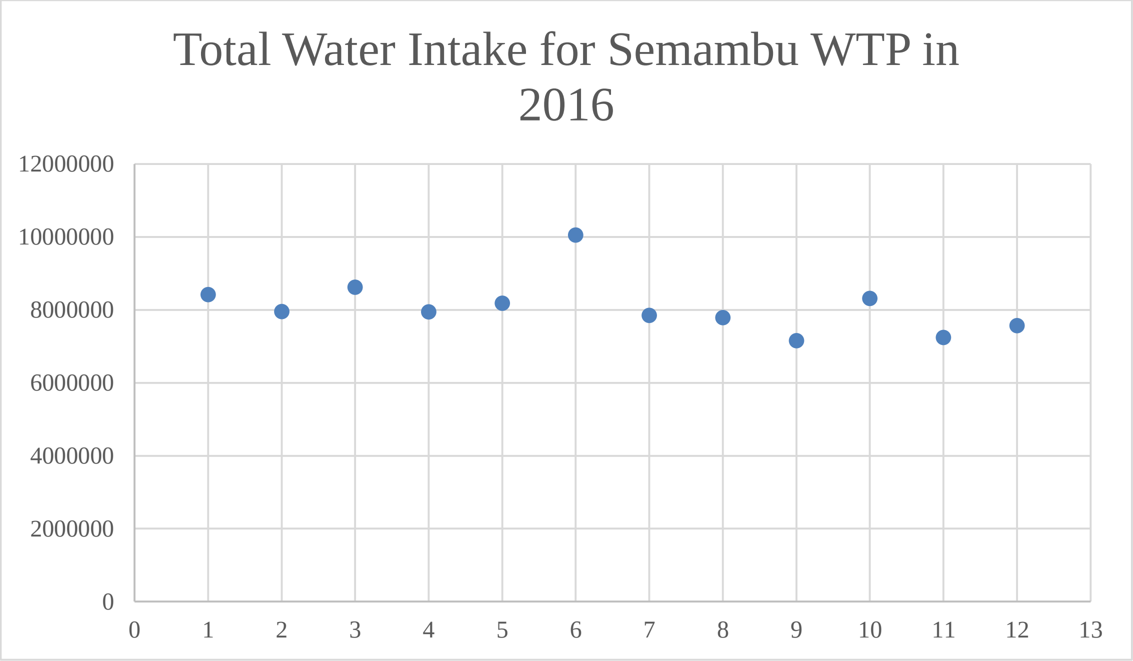


Figure 4.6 Total Water Intake in 2016 for Semambu WTP

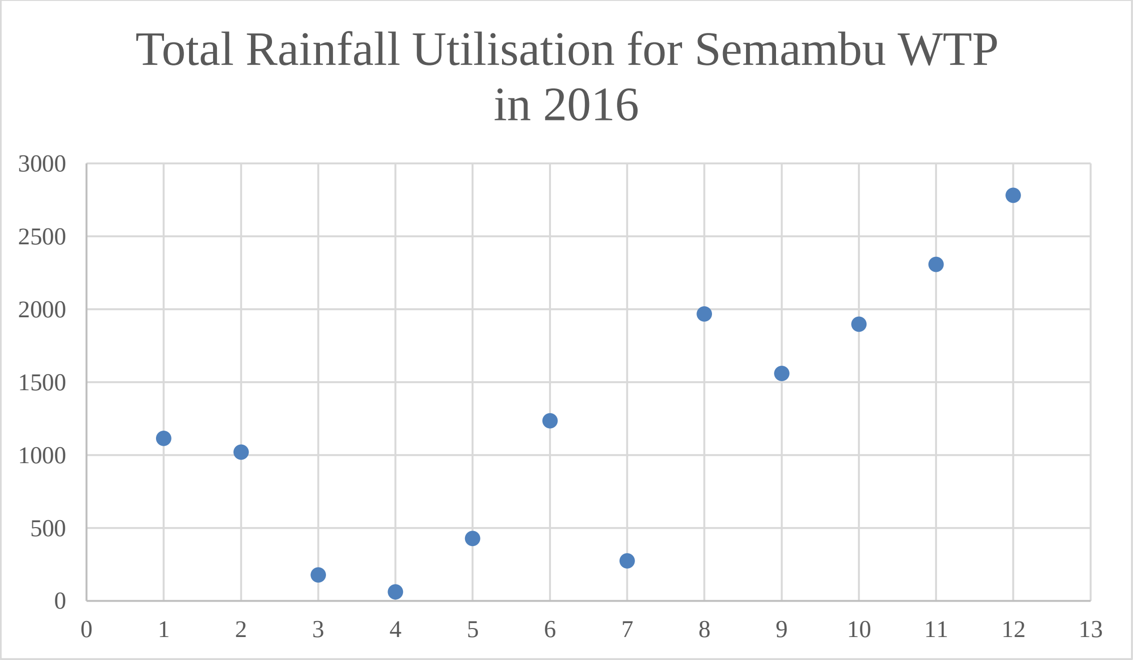


Figure 4.7 Total Rainfall Utilisation for Semambu WTP

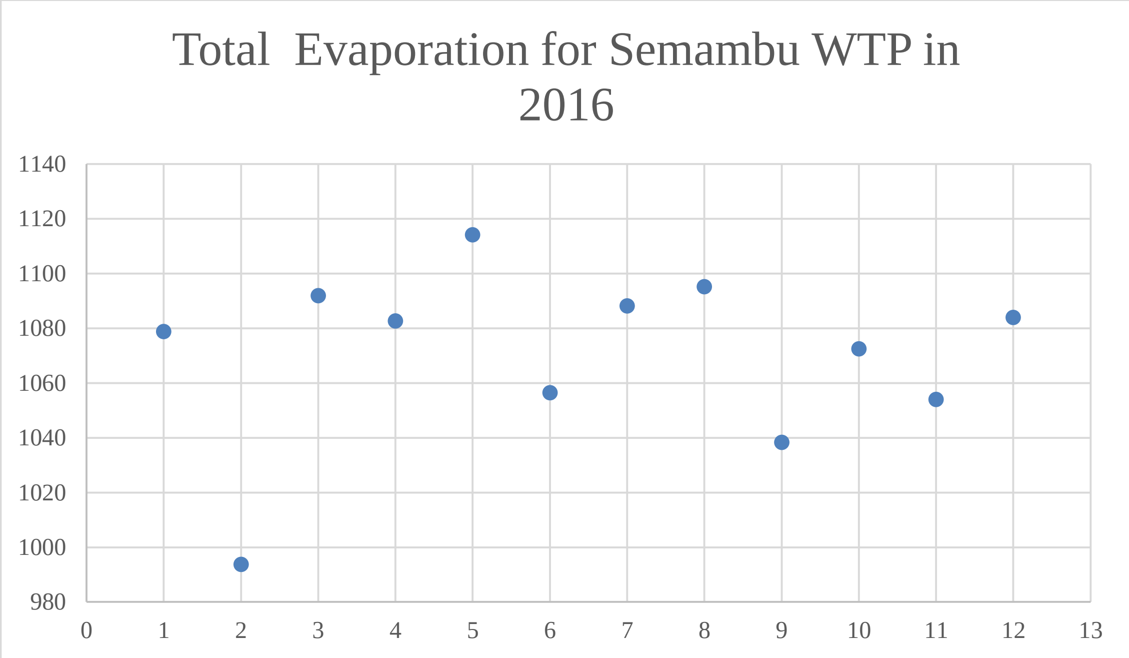


Figure 4.8 Total Evaporation for Semambu WTP in 2016

Figure 4.5 shows the highest amount of WFblue is in June and the lowest is in September with the value of 10048829.81 m^3/month and 7156618.571 m^3/month. High amount of total WFblue was due to the high amount of water intake which was 10046538 m^3. Moreover, the amount of rainfall utilised in June is also high with the amount 1235.322 $m^{3.}$ From the Figure 4.7, the amount of rainfall utilisation were gradually increased from September to December. The greatest amount of rainfall utilisation throughout the year is on December which was 2782.594 m^3. As seen in figure 4.8 , total evaporation amount will opposed the rainfall utilisation where the amount decreased right after the rainfall utilisation increased. This is because in Malaysia, there are 2 climates along the year which is hot and rainy season. During hot season, the earth temperature will affect the amount of water evaporated in the treatment process.

Meanwhile the lowest value of total WFblue was in September, which was 7156618.571 m^3, in Malaysia or dry season lies on Southwest Monsoon season because during this season which is on May until September. As seen in Figure 4.5, water intake amount were slightly uniform throughout the year and the lowest amount was also in September which was 7152041 m^3.

Table 4.4 Total WFblue in 2017 at Semambu WTP

MONTH	WATER INTAKE (m3)	TOTAL RAINFALL (m3)	TOTAL EVAPORATION (m3)	TOTAL BWF
JANUARY	5874225	5091.024	1078.803	5880395
FEBRUARY	5926540	1213.486	959.2051	5928713
MARCH	6658578	608.3025	1091.89	6660278
APRIL	6287909	430.491	1114.743	6289454
MAY	6126220	957.6865	1147.071	6128325
JUNE	6047528	1101.184	1087.793	6049717
JULY	6060951	1378.819	1120.438	6063450
AUGUST	6333049	764.2775	1127.572	6334941
SEPTEMBER	6400301	1600.304	1069.086	6402970
OCTOBER	6089985	1297.712	1072.414	6092355
NOVEMBER	6010241	1915.373	1053.992	6013210
DECEMBER	6149515	2224.204	1080.672	6152819
GRAND TOTAL	**73965041.5**	**18528.86**	**13003.68**	**73996628**

Table 4.4 shows the total WFblue amount from January to December 2017 at Semambu WTP. From figure 11, the highest amount of WFblue is in March with the value of 6660278 m^3/month while the lowest is in January which is 5880395 m^3/month. Although the rainfall intensity in January is the highest, the amount of WFblue is still the lowest due to low water intake amount. Water intake amount will give big impact in measuring the total WFblue as shown in the Table 4.4.

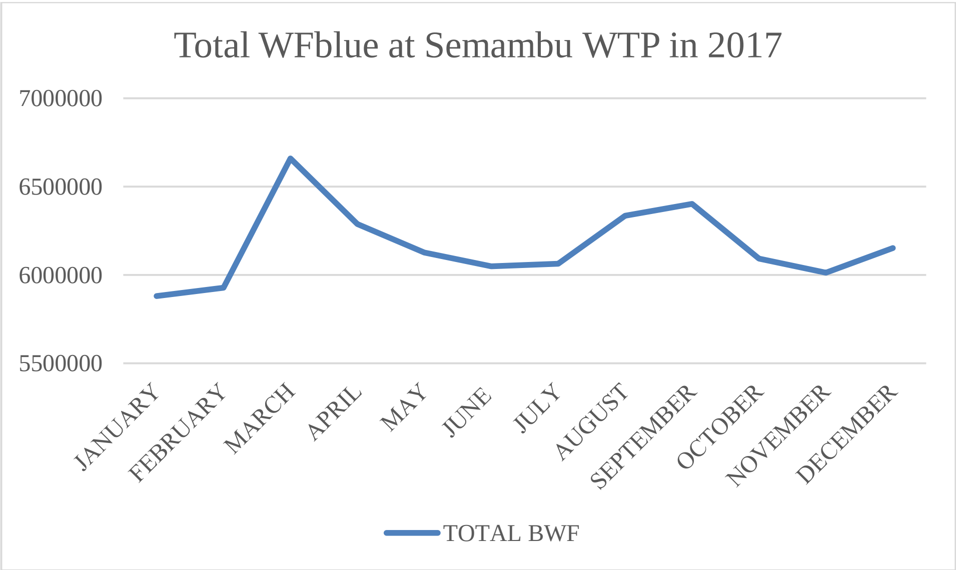


Figure 4.9 Total WFblue in 2017 at Semambu WTP

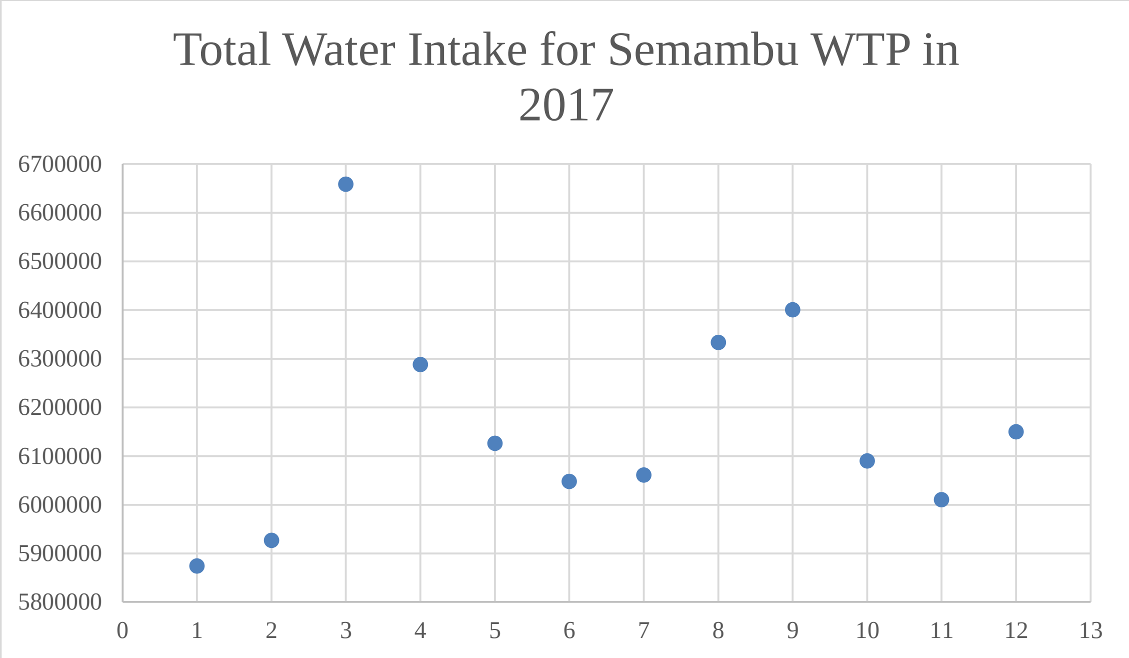


Figure 4.10 Total Water Intake for Semambu WTP in 2017

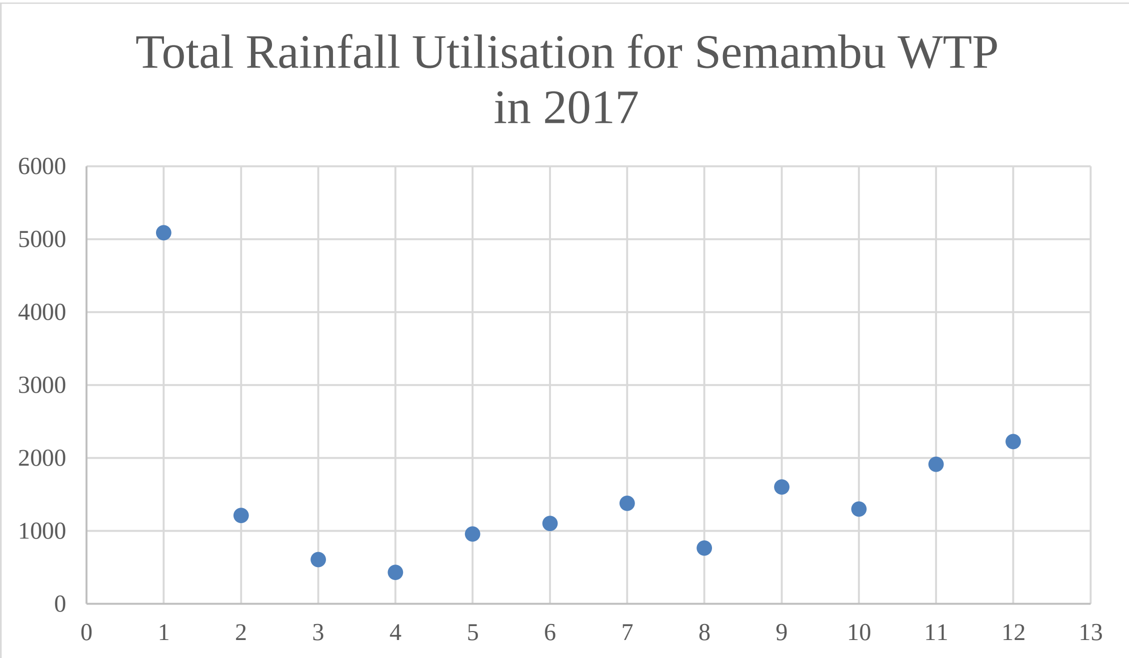


Figure 4.11 Total Rainfall Utilisation for Semambu WTP in 2017

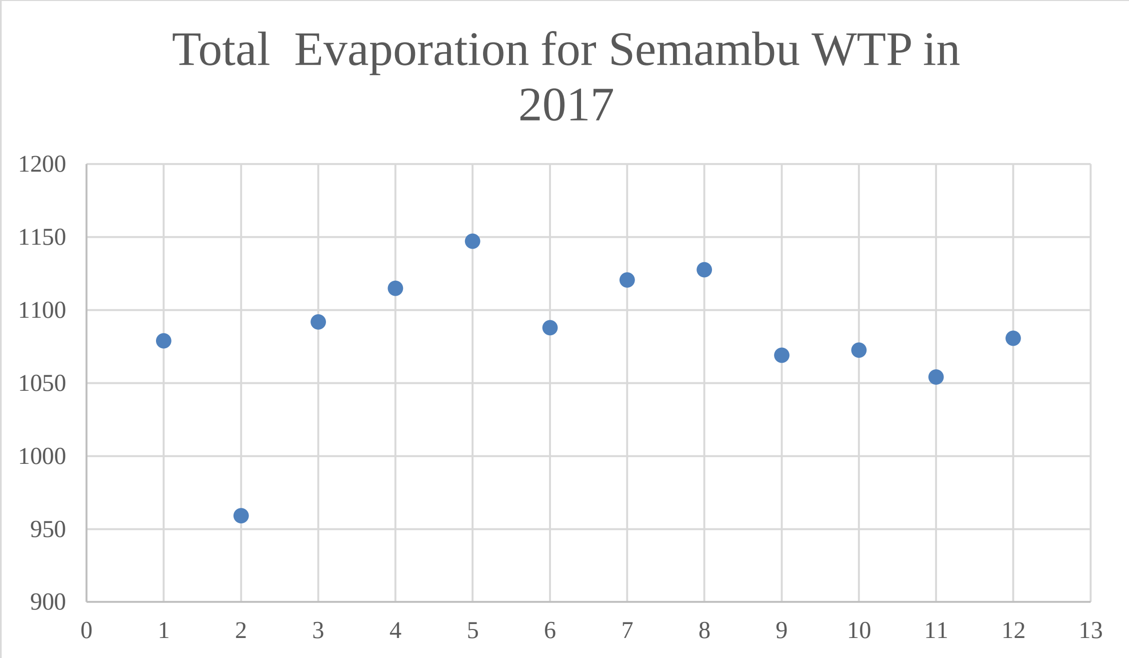


Figure 4.12 Total Evaporation for Semambu WTP in 2017

Based on the graph in Figure 4.9, in March, the WFblue amount is the highest 6660278 m^3/month. The total amount of WFblue is highly affected by the amount of water intake. As the water intake remain higher among other parameters due to the WTP capacity, the amount of rainfall utilised in this process was high in January, which was 5091.024 m^3 as shown in Figure 4.11. As the amount of rainfall utilisation is high, the evaporation amount will drop. As seen in figure 4.12 the amount of evaporation is the highest in May which was 1147.071 m^3 . This is mostly due to the temperature, higher temperature presents during the hot season and will evaporates more water from the system.

Meanwhile the lowest value of total WFblue was in January, which was 5883095 m^3. Although the amount of rainfall utilisation is the highest in that month, the total WFblue was still not being the highest. This is because total WFblue is highly affected by the amount of water intake. As seen in Figure 4.10, water intake amount were uniform throughout the year. Hot season in Malaysia will affect the amount of water in the river basin to be taken for treatment as the water will evaporated while rainy season will automatically increase the amount of water in the river basin.

Table 4.5 Total WFblue from 2015-2017 at Semambu WTP

	TOTAL BLUE WATER FOOTPRINT		
MONTH/YEAR	2015	2016	2017
JANUARY	8936216.657	8421542.465	5880394.827
FEBRUARY	7843954.926	7951263.836	5928712.691
MARCH	8406947.649	8623949.702	6660278.193
APRIL	8239752.403	7940917.554	6289454.234
MAY	7435161.241	8183675.433	6128324.757
JUNE	7831383.707	10048829.81	6049716.976
JULY	8390010.941	7848968.711	6063450.257
AUGUST	8472988.486	7787054.528	6334940.849
SEPTEMBER	7627786.08	7156618.571	6402970.39
OCTOBER	8357430.346	8311149.07	6092355.126
NOVEMBER	8084716.333	7244105.422	6013210.365
DECEMBER	8449901.9	7571366.53	6152819.376
GRAND TOTAL	**98078265.67**	**97089441.63**	**73996628.04**

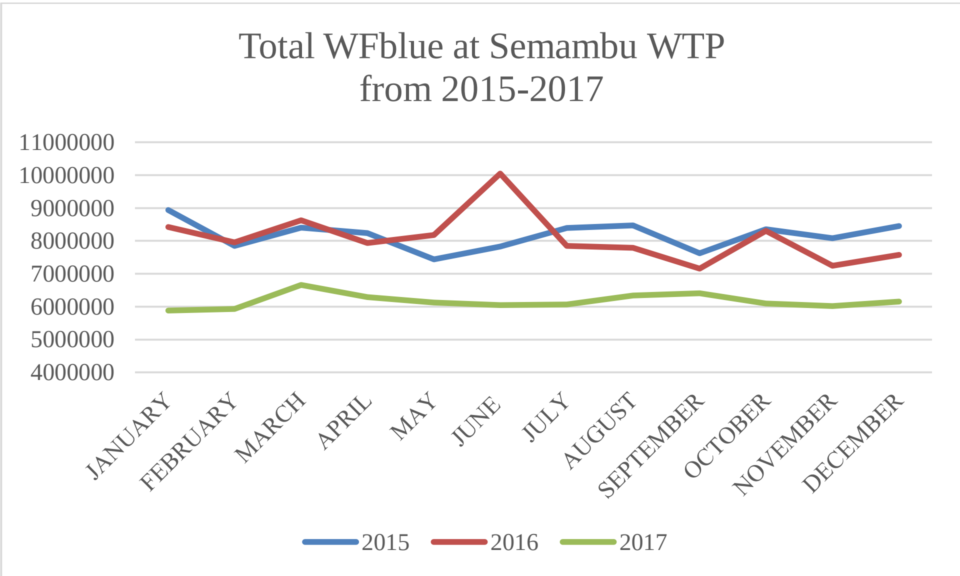


Figure 4.13 Total WFblue from 2015-2017 at Semambu WTP

From Figure 4.13, in 2015, the total blue water footprint started to increase starting from April and almost the same until the end of the year. Meanwhile, the highest value of total WFblue was in 2016 compared to other two years especially in July and decreasing towards the end of the year. While, in 2017, total WFblue is just in the range of 3,000,000 m^3 to 4,000,000 m^3 per month due to the moderate rainfall intensity and temperature rate throughout the year.

4.2.2 Total WFblue at Panching WTP

Table 4.6 Total WFblue in 2015 at Panching WTP

MONTH	WATER INTAKE (m3)	TOTAL RAINFALL (m3)	TOTAL EVAPORATION (m3)	TOTAL BWF
JANUARY	2813230	12.7166	12.00429	2813255
FEBRUARY	2620040	6.7938	10.59849	2620057
MARCH	2749360	6.4454	12.22912	2749379
APRIL	3271455	8.64032	12.02892	3271476
MAY	4399921	14.59796	11.64735	4399947
JUNE	4312251	2.99624	12.10104	4312266
JULY	4817449	4.84276	12.46413	4817466
AUGUST	4568820	18.56972	12.26351	4568851
SEPTEMBER	4390658	11.18364	11.41159	4390681
OCTOBER	4597689	9.68552	12.26113	4597711
NOVEMBER	4219390	13.47611	11.77146	4219415
DECEMBER	4288326	10.529568	12.10589	4288339
GRAND TOTAL	**47048589**	**120.47**	**142.89**	**470488423.36**

Table 4.6 shows the total WFblue amount from January to December 2015 at Panching WTP. In July, the WFblue amount is the highest with the value of 4817466 m3/month due to the high water intake as shown in Table 4.6. Figure 4.14 shows WFblue trend started to increase in the early of the year until May and being uniform to the end of the year.

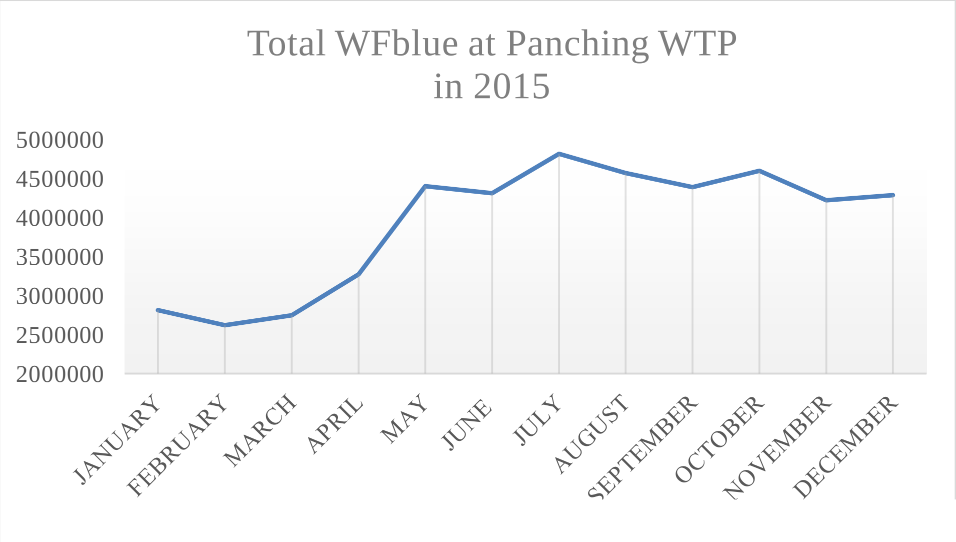


Figure 4.14 Total WFblue at Panching WTP in 2015

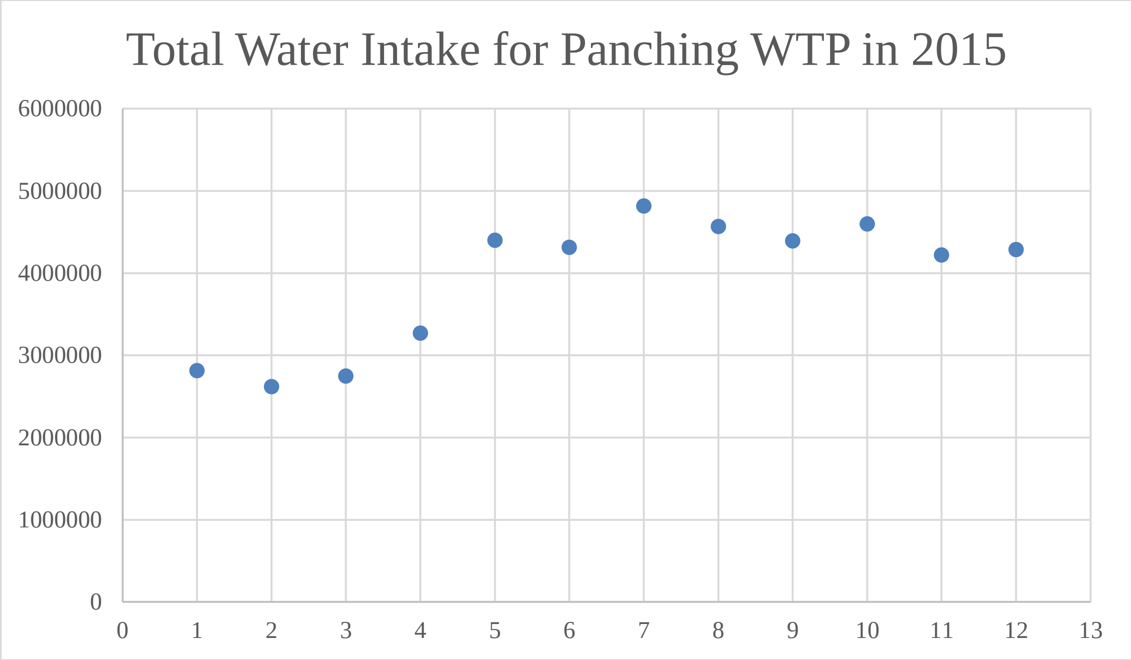


Figure 4.15 Total Water Intake for Panching WTP in 2015

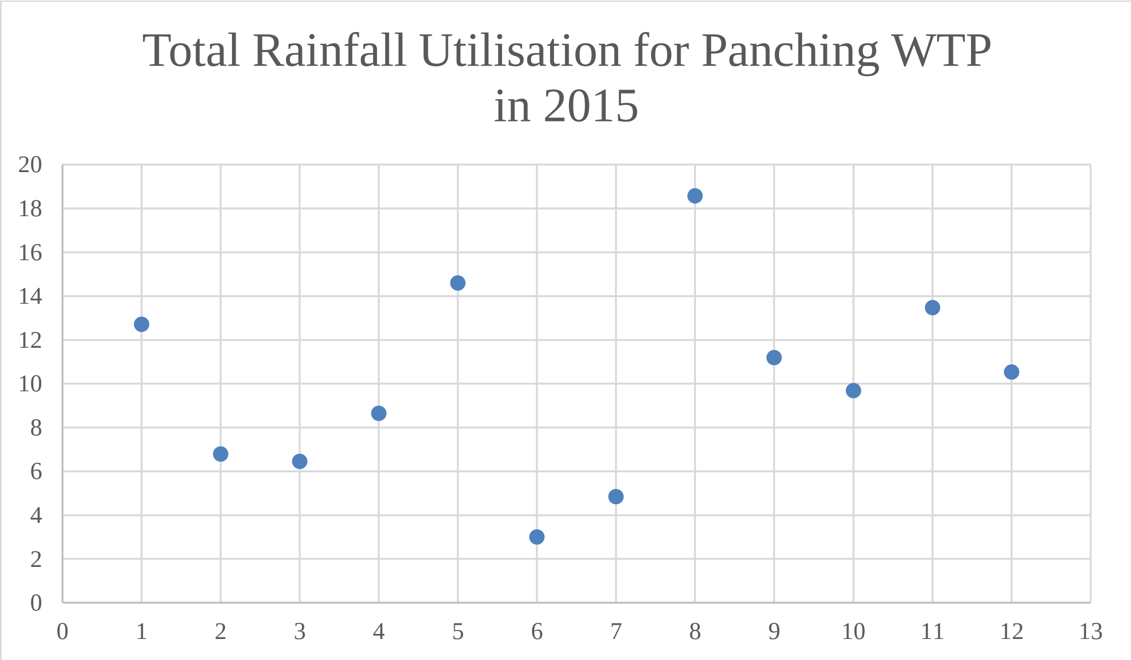


Figure 4.16 Total Rainfall Utilisation for Panching WTP in 2015

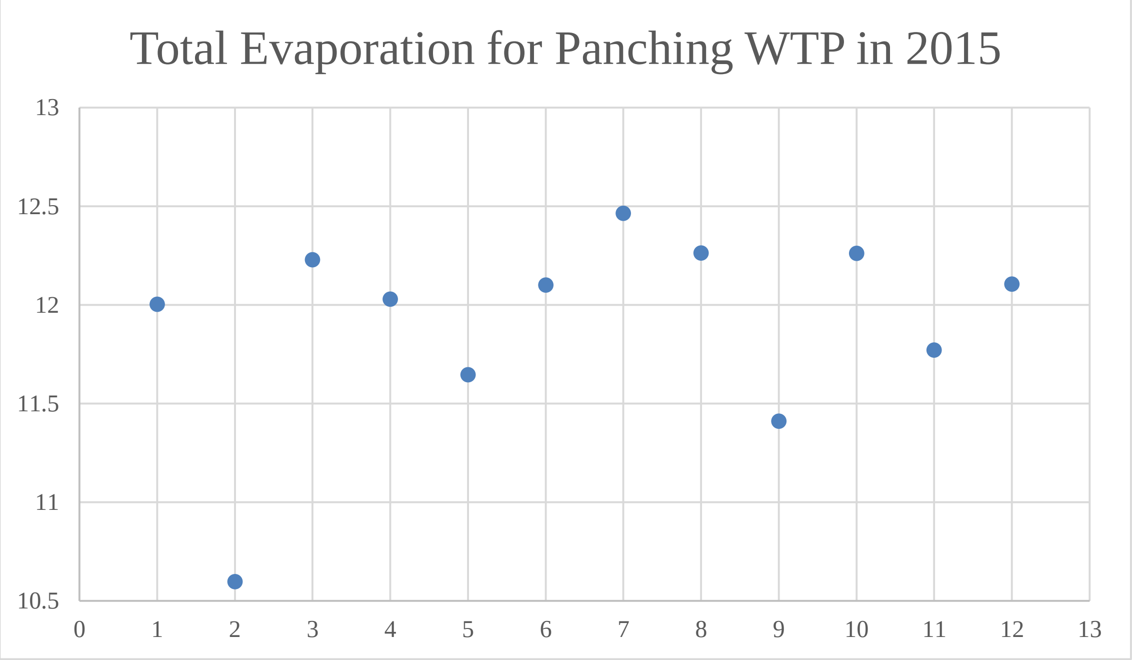


Figure 4.17 Total Evaporation for Panching WTP in 2015

Based on the graph in Figure 4.14, in July, the WFblue amount is the highest 4817466 m^3/month. The main contribution to this amount was total water intake, which was 4817449 m^3. The amount of rainfall utilised in this process is the highest in August which was 18.56972 m^3 as shown in Figure 4.16.This is due to north east monsoon season and East Coast area of Peninsular Malaysia will be receiving abundance of rainfall during this season specifically from October - March. As seen in the Figure 4.16, the amount of rainfall utilisation were not uniform along the year. As seen in figure 4.17, evaporation amount will drop as the rainfall increased. This is mostly due to the temperature, lower temperature presents during the rainy season and will evaporates less water into the system.

Meanwhile the lowest value of total WFblue was in February, which was 2620057 m^3.. As seen in Figure 4.15, water intake amount dropped to the lowest amount on February and gradually increased up to May before been un-uniform till the end of the year. Dry season will affect the amount of intake because when the volume of river basin decreases, WTP will limit amount of water abstraction to avoid damage to the treatment plant if more volume of sediment abstracted rather than raw water. Total evaporation amount was the highest in July with 12.46413 m^3 of water been evaporated.

Table 4.7 Total WFblue in 2016 at Panching WTP

MONTH	WATER INTAKE (m3)	TOTAL RAINFALL (m3)	TOTAL EVAPORATION (m3)	TOTAL BWF
JANUARY	4176069	3.484	12.04856808	4176084.53
FEBRUARY	3928827	9.02356	11.09875784	3928847.12
MARCH	4434521	4.49436	12.19473094	4434537.69
APRIL	4287649	0.20904	12.09168133	4287661.3
MAY	4758630	9.35728	11.63356108	4758650.99
JUNE	4664012	9.02356	11.79935561	4664032.82
JULY	4995475	8.60548	12.15346143	4995495.76
AUGUST	4688972	8.57064	12.23084177	4688992.8
SEPTEMBER	4836755	16.16576	11.59644718	4836782.76
OCTOBER	3481698	17.97744	11.97720622	3481727.95
NOVEMBER	3861220	26.72228	11.37118442	3861258.09
DECEMBER	3804719	6.58476	11.71325329	3804737.3
GRAND TOTAL	**51918547**	**120.22**	**141.91**	**51918809**

Table 4.7 shows the total WFblue amount from January to December 2016 at Panching WTP. The highest amount of WFblue is in July while the lowest is in October with the value of 4664032.82 m^3/month and 3481727.95 m^3/month. Meanwhile, in figure 4.18, there is slightly decrease of WFblue between September and October due to the sudden drop amount of water intake from 4836755 m^3 to 3481698 m^3. The temperature rate along the year is almost the same due to the uniform value of total evaporation that been calculated as in table 4.7.

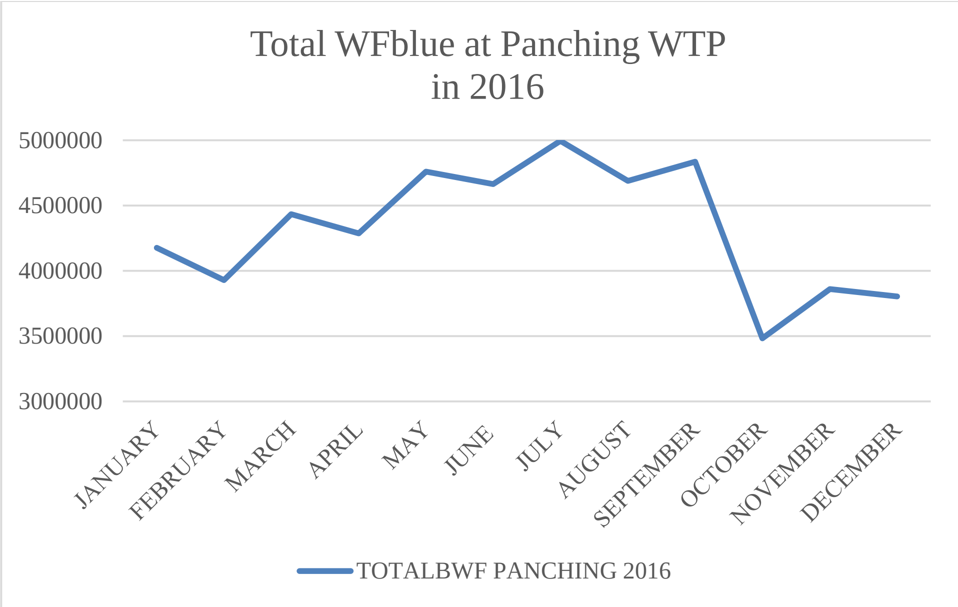


Figure 4.18 Total WFblue at Panching WTP in 2016

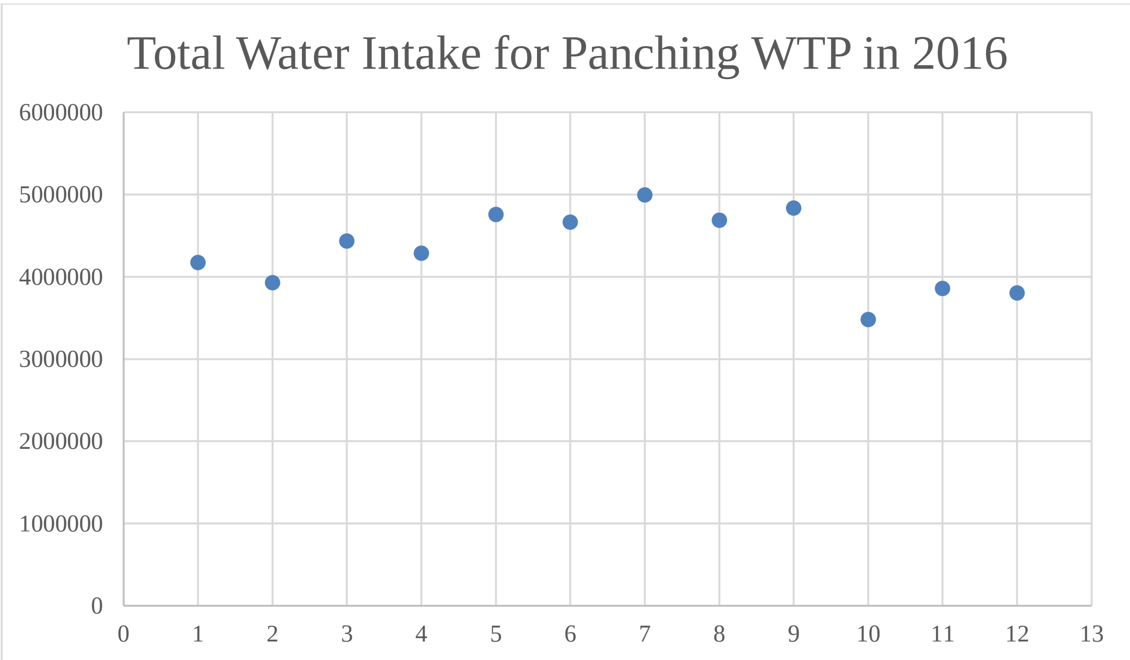


Figure 4.19 Total Water Intake for Panching WTP in 2016

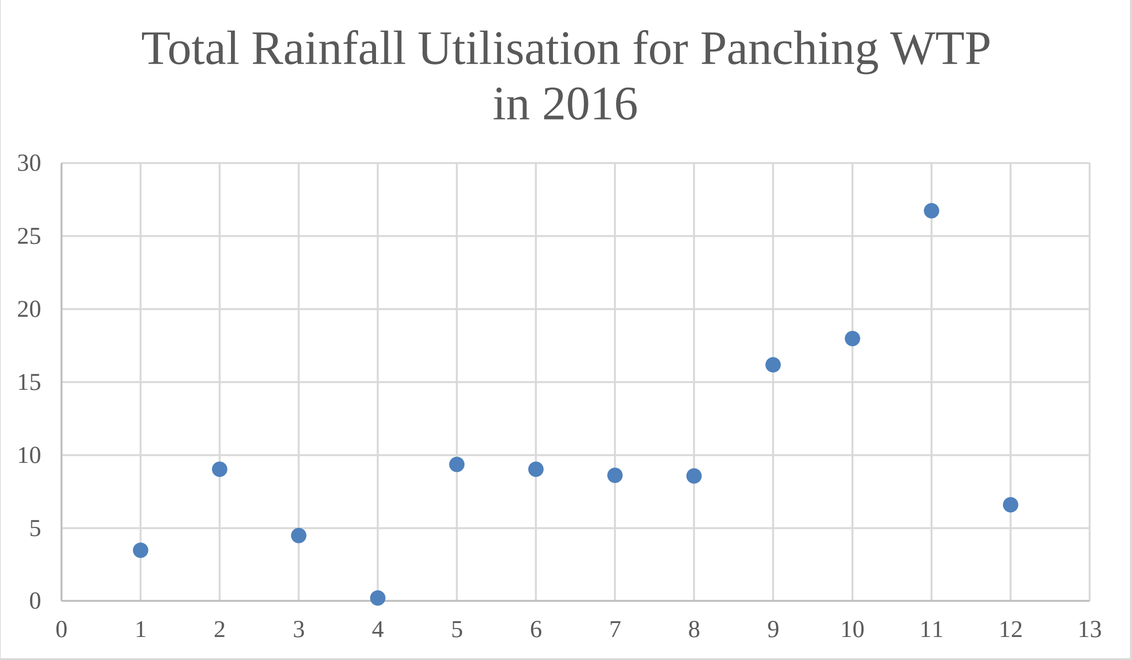


Figure 4.20 Total Rainfall Utilisation for Panching WTP in 2016

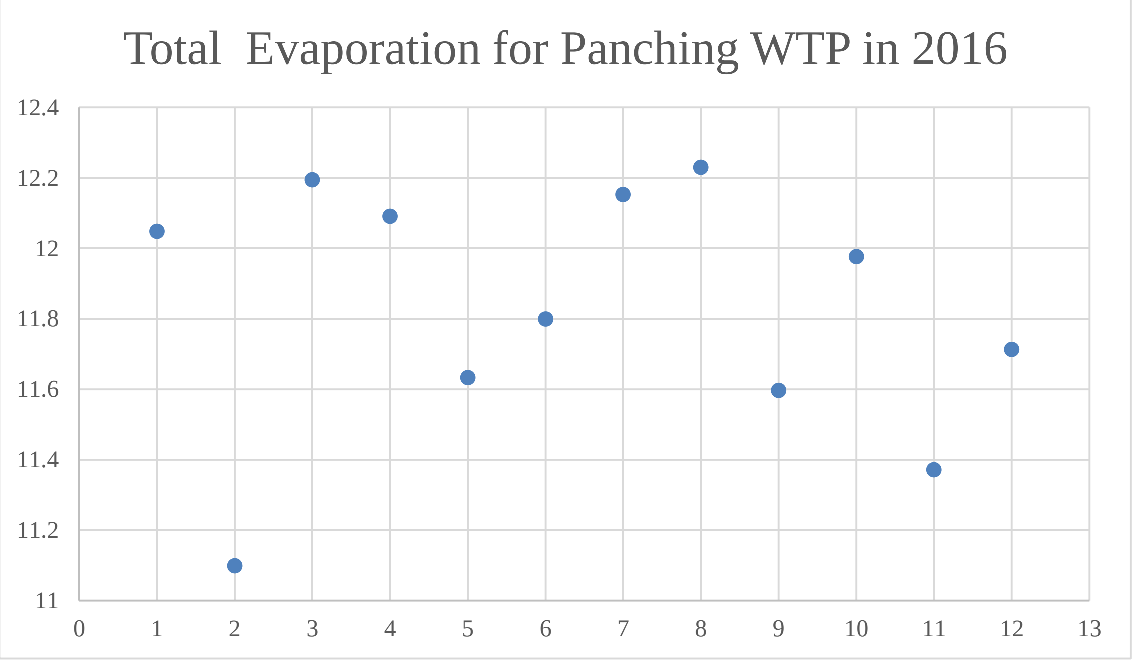


Figure 4.21 Total Evaporation for Panching WTP in 2016

Figure 4.18 shows the highest amount of WFblue is in July and the lowest is in October with the value of 4995495.76 m^3/month and 3481727.95 m^3/month. High amount of total WFblue was due to the high amount of water intake which was 4995475 m^3. Moreover, the amount of rainfall utilised is the highest in November with the amount 26.72228 $m^{3.}$ From the Figure 4.20, the amount of rainfall utilisation were gradually increased from August to November. The amount of rainfall utilisation in July was 8.60548 m^3 and this showed that the amount of rainfall utilisation also affect the WFblue amount but just a little because water intake give higher impact towards amount of total WFblue. As seen in figure 4.21, total evaporation amount will opposed the rainfall utilisation where the amount decreased right after the rainfall utilisation increased. This is because in Malaysia, there are 2 climates along the year which is hot and rainy season.

Meanwhile the lowest value of total WFblue was in October, which was 3481727.95 m^3, in Malaysia or dry season lies on Southwest Monsoon season because during this season which is on May until September. As seen in Figure 4.19, water intake amount were slightly uniform throughout the year and the lowest amount was also in October which was 3481698 m^3. However, amount of water evaporated also affects the amount of total WFblue. The highest amount of water evaporated was in August with the value of 12.23084177 m^3.

Table 4.8 Total WFblue in 2017 at Panching WTP

MONTH	WATER INTAKE (m3)	TOTAL RAINFALL (m3)	TOTAL EVAPORATION (m3)	TOTAL BWF
JANUARY	4109859	16.7302	11.70294	4109921
FEBRUARY	3796846	16.31906	10.56496	3796873
MARCH	4007553.5	6.86348	11.94711	4007572
APRIL	3748308	6.16668	12.09168	3748326
MAY	3407634	8.750136	11.63356	3407654
JUNE	3322259	7.643896	11.79936	3322278
JULY	3430447	15.25992	12.15346	3430474
AUGUST	3626760	8.91904	12.23084	3626781
SEPTEMBER	3431649	14.49344	11.59645	3431675
OCTOBER	3570427	22.50664	11.97721	3570461
NOVEMBER	3497273	28.84752	11.37118	3497313
DECEMBER	3591922	17.14128	11.6005	3591951
GRAND TOTAL	**43540397.5**	**169.6412**	**140.67**	**43541281.12**

Table 4.8 shows the total WFblue amount from January to December 2017 at Panching WTP. In June, the WFblue amount is the lowest with the value of 3322278 m^3/month due to the low amount of water intake. Moreover, the rainfall intensity in that month is also one of the lowest throughout the year that will affect the amount of WFblue. Figure 4.22 shows the amount of WFblue is the highest in January with the value of 4109921 m^3/month. The WFblue trend in 2017 started to decrease until June and become more stable towards the end of the year.

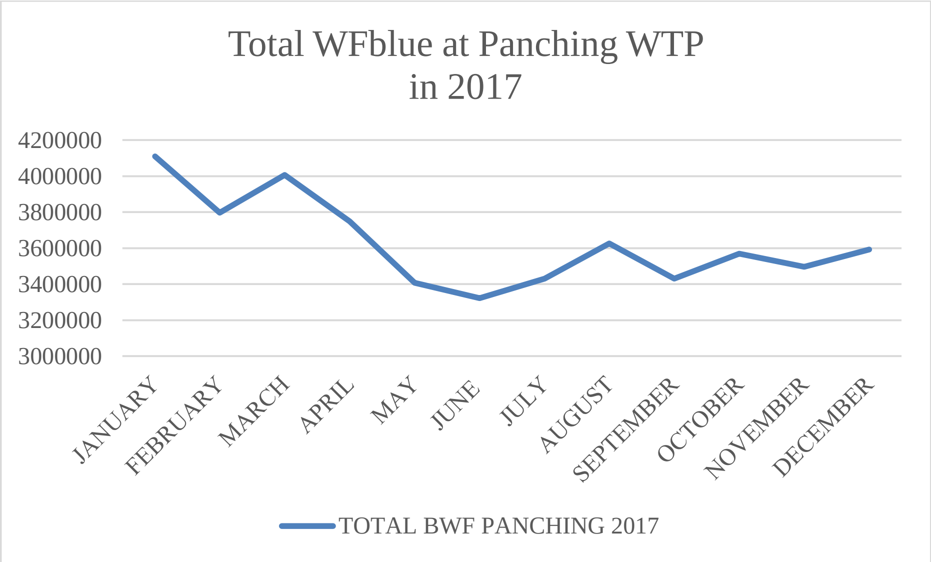


Figure 4.22 Total WFblue in 2017 at Semambu WTP

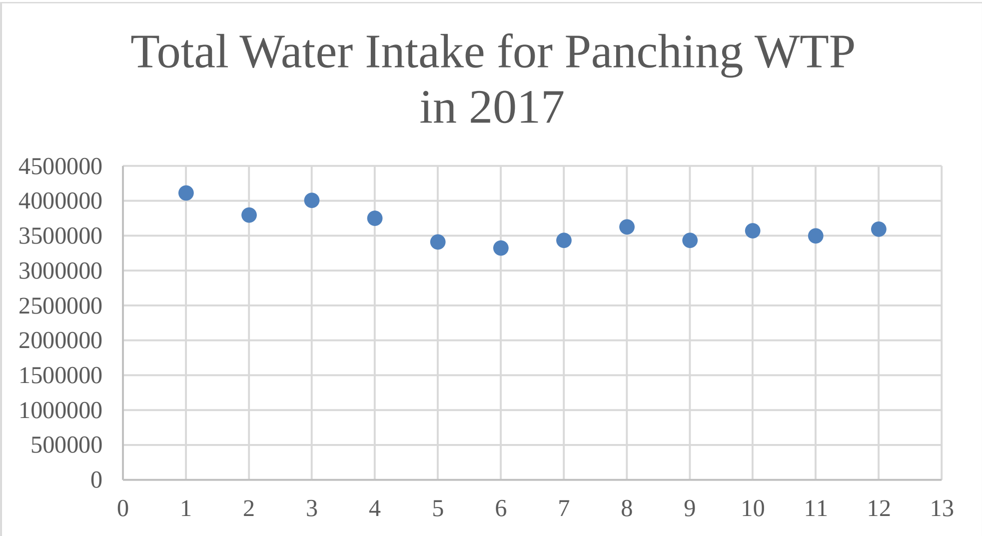


Figure 4.23 Total Water Intake for Panching WTP in 2017

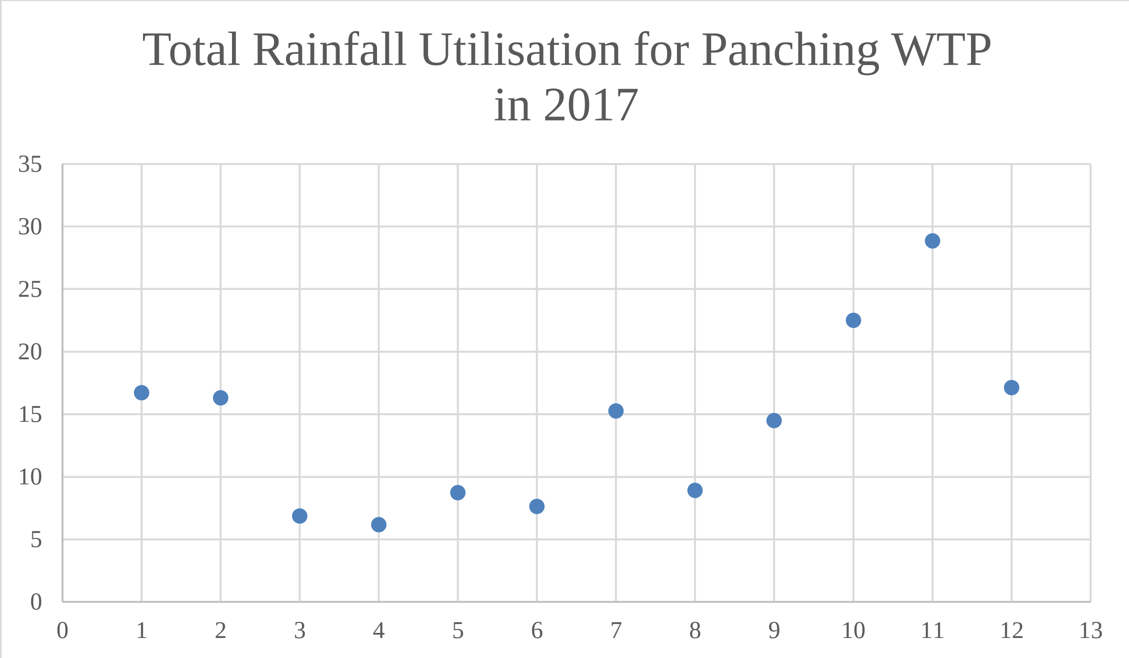


Figure 4.24 Total Rainfall Utilisation for Panching WTP in 2017

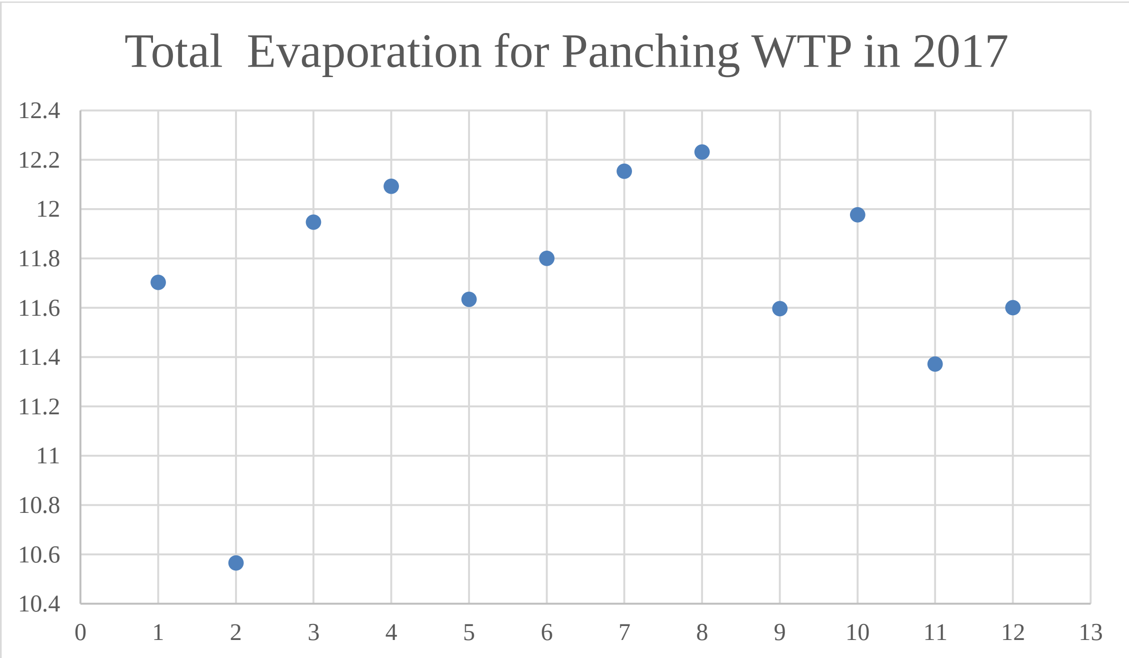


Figure 4.25 Total Evaporation for Panching WTP in 2017

Based on the graph in Figure 4.22, in January, the WFblue amount is the highest 4109921 m^3/month. The total amount of WFblue is highly affected by the amount of water intake. As the water intake remain higher among other parameters due to the WTP capacity, the amount of rainfall utilised in this process also high in January, which was 16.7302 m^3 as shown in Figure 4.24. Apart from that, the highest amount of rainfall was in November with the amount 28.84752 m^3 as seen in Figure 4.24. As the amount of rainfall utilisation is high, the evaporation amount will drop. As seen in figure 4.25, the amount of evaporation is the highest in August which was 12.23084 m^3 . This is mostly due to the temperature, higher temperature presents during the hot season and will evaporates more water from the system.

Meanwhile the lowest value of total WFblue was in June, which was 3322278 m^3. Al though the amount of rainfall utilisation is the highest in that month, the total WFblue was still not being the highest. This is because total WFblue is highly affected by the amount of water intake. As seen in Figure 4.22, water intake amount were uniform throughout the year. Hot season in Malaysia will affect the amount of water in the river basin to be taken for treatment as the water will evaporated while rainy season will automatically increase the amount of water in the river basin.

Table 4.9 Total WFblue from 2015-2017 at Panching WTP

	TOTAL BLUE WATER FOOTPRINT		
MONTH/YEAR	2015	2016	2017
JANUARY	2813254.721	4176084.533	4109920.747
FEBRUARY	2620057.392	3928847.122	3796872.884
MARCH	2749378.675	4434537.689	4007572.311
APRIL	3271475.669	4287661.301	3748326.258
MAY	4399947.245	4758650.991	3407654.384
JUNE	4312266.097	4664032.823	3322278.443
JULY	4817466.307	4995495.759	3430474.413
AUGUST	4568850.833	4688992.801	3626781.15
SEPTEMBER	4390680.595	4836782.762	3431675.09
OCTOBER	4597710.947	3481727.955	3570461.484
NOVEMBER	4219415.248	3861258.093	3497313.219
DECEMBER	4288338.635	3804737.298	3591950.742
GRAND TOTAL	**47048842.36**	**51918809**	**43541281.12**

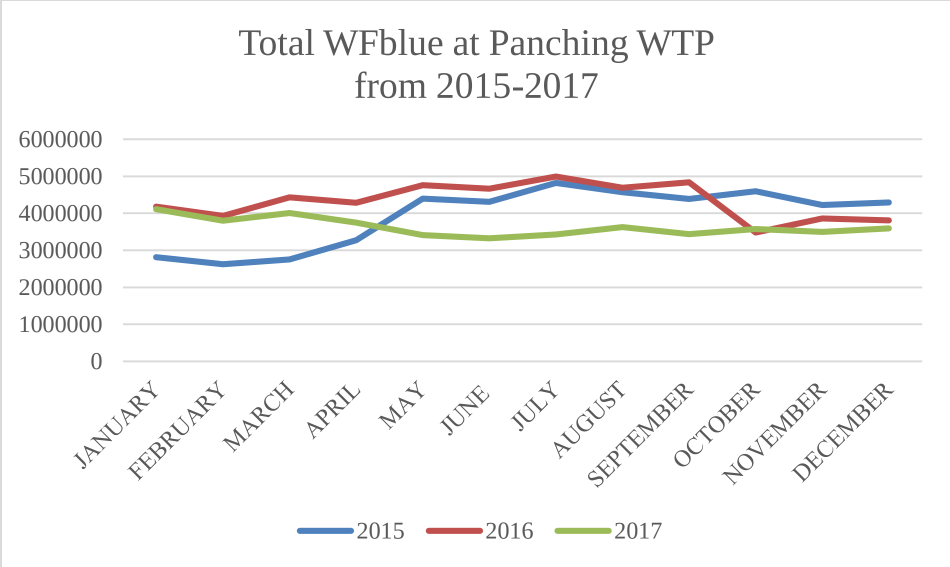


Figure 4.26 Total WFblue at Panching WTP from 2015-2017

In the figure 4.26, in 2015 and 2016, the highest value of total WFblue was on July due to rainy season. This will directly affects water footprint amount because the rainfall intensity is one of the parameters in accounting WFblue. Meanwhile in 2017, the highest value of total WFblue was on January. Comparing the total blue water footprint for the three years, the highest value was in 2016, which was 51,918,809 m^3 while, 47,048,842 m^3 in 2015 and 43,541,281 m^3 in 2017. This is due to the increasing amount of rainfall.

4.3 The Best Algorithm Selection

4.3.1 Semambu WTP

```
=== Evaluation on training set ===

Time taken to test model on training data: 0.56 seconds

=== Summary ===

Correlation coefficient                  1
Mean absolute error                      0.0002
Root mean squared error                  0.0002
Relative absolute error                  1.3353 %
Root relative squared error              0.7734 %
Total Number of Instances             1096
```

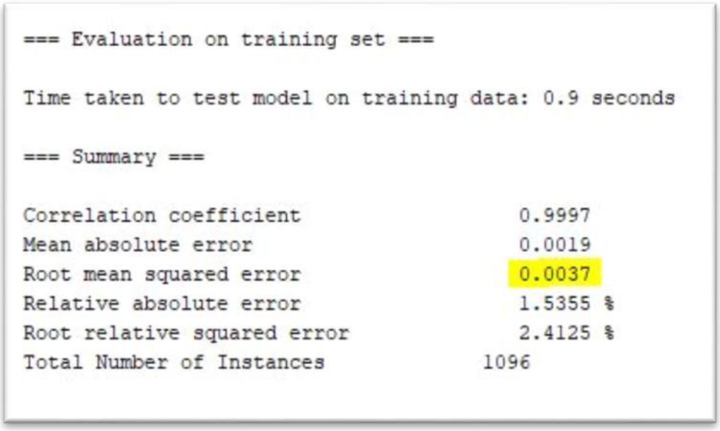
Figure 4.27 Result after the training using ANN

```
=== Evaluation on training set ===

Time taken to test model on training data: 0.69 seconds

=== Summary ===

Correlation coefficient                  0.9746
Mean absolute error                      0.0007
Root mean squared error                  0.0113
Relative absolute error                  4.9077 %
Root relative squared error             35.2562 %
Total Number of Instances             1096
```

Figure 4.28 Result after the training using RF

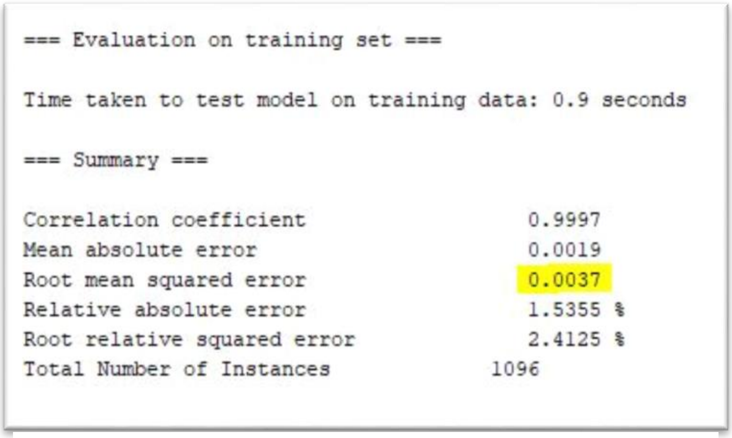
Refer to figure 4.27 and 4.28, ANN and RF algorithms produced low error in training the WFblue data sets. ANN produced the lowest RMSE value which is 0.0002 while RF produce RF came out with the RMSE value of 0.0113. Based on the result, ANN been chosen as the best algorithm in predicting WFblue trend at Semambu WTP.

4.3.2 Panching WTP



```
=== Evaluation on training set ===

Time taken to test model on training data: 0.9 seconds

=== Summary ===

Correlation coefficient                  0.9997
Mean absolute error                      0.0019
Root mean squared error                  0.0037
Relative absolute error                  1.5355 %
Root relative squared error              2.4125 %
Total Number of Instances             1096
```

Figure 4.29 Result after the training using ANN



```
=== Evaluation on training set ===

Time taken to test model on training data: 0.9 seconds

=== Summary ===

Correlation coefficient                  0.9997
Mean absolute error                      0.0019
Root mean squared error                  0.0037
Relative absolute error                  1.5355 %
Root relative squared error              2.4125 %
Total Number of Instances             1096
```

Figure 4.30 Result after the training using RF

Refer to figure 4.29 and 4.30, ANN and RF algorithms produced low error in training the WFblue data sets with the RMSE value of 0.0008 and 0.0037. Based on the result, ANN been chosen as the best algorithm in predicting WFblue trend at Panching WTP.

4.3.3 Overall Best Algorithm

For Semambu WTP, ANN algorithm came out with the least value of RMSE which is 0.0002 compared to RF algorithm which is 0.0113.For Panching WTP, RMSE value from ANN algorithm is 0.0008 while from RF algorithm is 0.0037. The best algorithm is ANN due to the least error which led to more precise trend of blue water footprint between actual and predicted value.

This result concluded that ANN is the best algorithm in predicting WFblue trend. The adjustable hidden neurons of this algorithm will allow the user to produce the least error in prediction. The least error produced will lead to the precise trend between the actual and predicted value.

4.4 Prediction of Blue Water Footprint Accounting

4.4.1 Prediction WFblue at Semambu WTP

4.4.1.1 Artificial Neural Network Algorithm

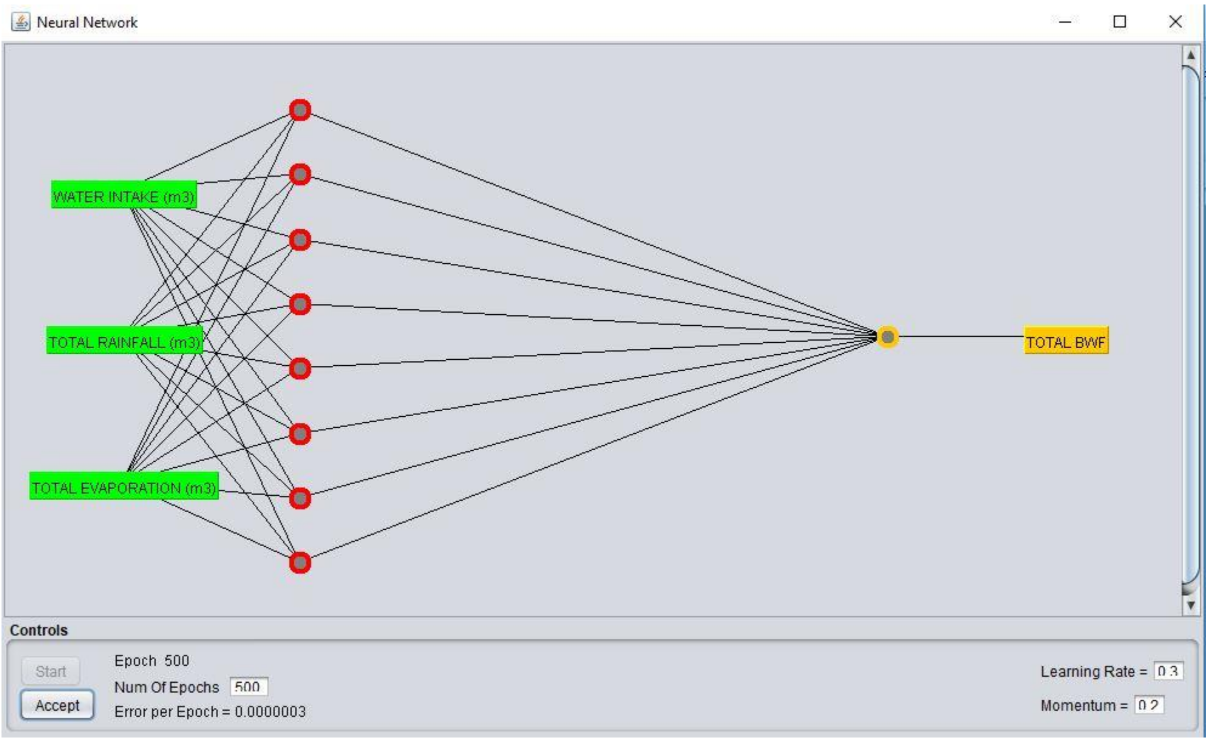


Figure 4.31 Number of hidden layers for ANN in WEKA software

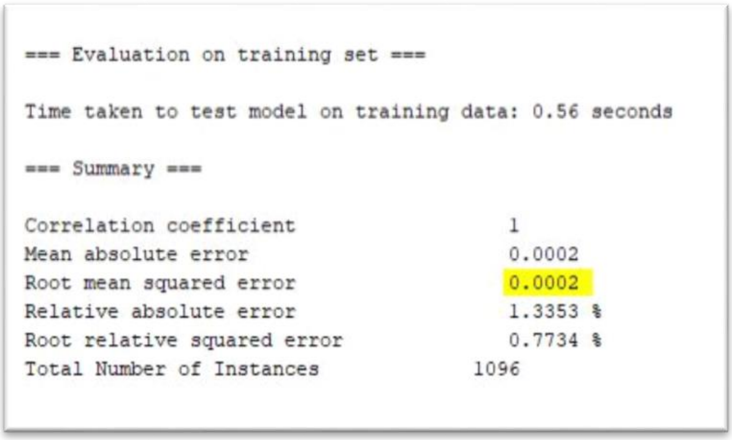
Figure 4.31 shows 8 of hidden layers when performing a training to a data sets using ANN algorithm. ANN algorithms will trained the WFblue data sets in order to produce the predicted value. The number of hidden layers in ANN defined the number of neurons that been created by ANN to produce the output. Meanwhile, the number of

Epochs means the number of the data sets will undergo the training. In this study, 20 hidden neurons been tested and 500 Epochs been set. The lowest value of RMSE produce by the training been chose to construct the predicted trend.

Table 4.10 Analysis of RMSE and hidden neurons

Hidden Neuron	**RMSE-value**	**Hidden Neuron**	**RMSE-value**
1	0.0006	11	0.0004
2	0.0004	12	0.0004
3	0.0003	13	0.0004
4	0.0003	14	0.0005
5	0.0003	15	0.0005
6	0.0003	16	0.0005
7	0.0003	17	0.0004
8	0.0002	18	0.0005
9	0.0003	19	0.0006
10	0.0004	20	0.0007

Refer to Table 4.10, the training sets with 8 neurons have been chosen to predict the total WFblue because the lowest RMSE value was obtained from the training process. The value of RMSE is different depends on the number of hidden neurons that used to train the data sets.



```
=== Evaluation on training set ===

Time taken to test model on training data: 0.56 seconds

=== Summary ===

Correlation coefficient                  1
Mean absolute error                      0.0002
Root mean squared error                  0.0002
Relative absolute error                  1.3353 %
Root relative squared error              0.7734 %
Total Number of Instances             1096
```

Figure 4.32 Result after the training using ANN

Figure 4.32 shows the result after the WFblue data sets undergoes the training. The RMSE value which is 0.0002 produced after the hidden layer been set to 8.

Table 4.11 Analysis of actual and predicted value of WFblue by using ANN

MONTH	ACTUAL (m^3)	PREDICTED (m^3)
JANUARY 2015	8936216.657	8926427.72
FEBRUARY 2015	7843954.926	7833589.377
MARCH 2015	8406947.649	8396601.907
APRIL 2015	8239752.403	8226048.011
MAY 2015	7435161.241	7409858.719
JUNE 2015	16221394.65	16188218.09
JULY 2015	8390010.941	8374898.626
AUGUST 2015	8472988.486	8455520.779
SEPTEMBER 2015	7627786.08	7606254.991
OCTOBER 2015	8357430.346	8339029.391
NOVEMBER2015	8084716.333	8062373.116
DECEMBER 2015	8449901.9	8429526.715
JANUARY 2016	8421542.465	8406828.187
FEBRUARY 2016	7951263.836	7933335.445
MARCH 2016	8340814.763	8328153.25
APRIL 2016	7940917.554	7928625.905
MAY 2016	8183675.433	8163238.507
JUNE 2016	10048829.81	10029324.81
JULY 2016	7848968.711	7840065.225
AUGUST 2016	7787054.528	7760781.529
SEPTEMBER 2016	7156618.571	7134166.171
OCTOBER 2016	8311149.07	8297870.232
NOVEMBER2016	7244105.422	7214545.494
DECEMBER 2016	7571366.53	7546661.875
JANUARY 2017	5880394.827	5814563.303
FEBRUARY 2017	5928712.691	5918601.573
MARCH 2017	6660278.193	6640437.455
APRIL 2017	6289454.234	6272230.489
MAY 2017	6128324.757	6114831.101
JUNE 2017	6049716.976	6026465.093
JULY 2017	6063450.257	6055428.883
AUGUST 2017	6334940.849	6322092.74
SEPTEMBER 2017	6402970.39	6381878.313
OCTOBER 2017	6092355.126	6085225.287
NOVEMBER2017	6013210.365	5996215.535
DECEMBER 2017	6152819.376	6137830.297

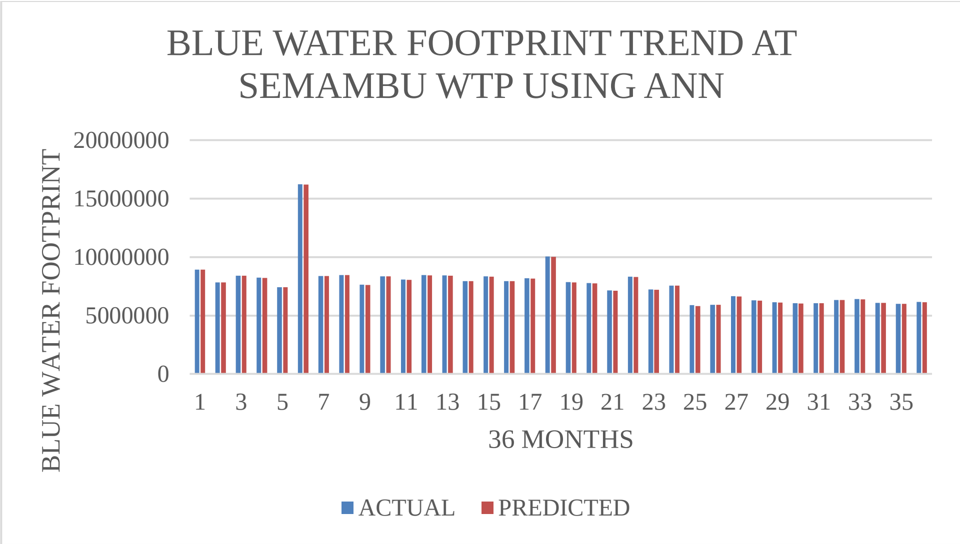
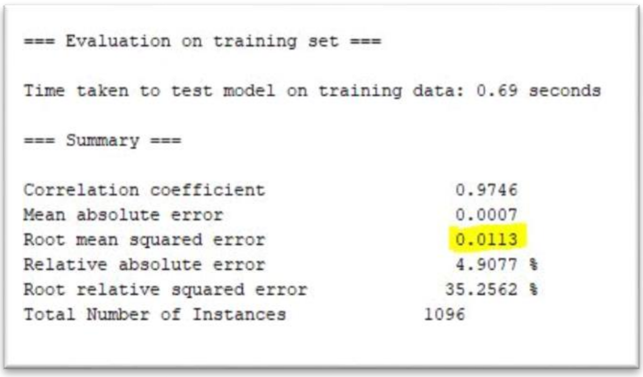


Figure 4.33 WFblue trend at Semambu WTP using ANN

Figure 4.33 illustrates the actual and predicted value of WFblue at Semambu WTP after undergoes training using ANN. The highest value of WFblue was in June 2015 which for actual value is 16221394.65 m³/month and for predicted value is 16188218.09 m³/month. There is not much difference in both values. Besides that, the lowest of WFblue amount of the actual and predicted was in January 2017 with the amount of 5880394.827 m³/month and 5814563.303 m³/month.

4.4.1.2 Random Forest Algorithm



```
=== Evaluation on training set ===

Time taken to test model on training data: 0.69 seconds

=== Summary ===

Correlation coefficient                  0.9746
Mean absolute error                      0.0007
Root mean squared error                  0.0113
Relative absolute error                  4.9077 %
Root relative squared error             35.2562 %
Total Number of Instances             1096
```

Figure 4.34 Result after the training using RF

Figure 4.34 shows the result after the WFblue data sets undergoes the training with a RMSE value of 0.0113.

Table 4.12 Analysis of actual and predicted value of WFblue by using RF

MONTH	ACTUAL (m^3)	PREDICTED (m^3)
JANUARY 2015	8936216.657	8923446.689
FEBRUARY 2015	7843954.926	7837919.803
MARCH 2015	8406947.649	8395973.005
APRIL 2015	8239752.403	8235517.138
MAY 2015	7435161.241	7438154.117
JUNE 2015	16221394.65	16230212.69
JULY 2015	8390010.941	8393099.252
AUGUST 2015	8472988.486	8458754.636
SEPTEMBER 2015	7627786.08	7629568.924
OCTOBER 2015	8357430.346	8350559.174
NOVEMBER2015	8084716.333	8078261.001
DECEMBER 2015	8449901.9	8447643.628
JANUARY 2016	8421542.465	8417286.011
FEBRUARY 2016	7951263.836	7945579.932
MARCH 2016	8340814.763	8335590.735
APRIL 2016	7940917.554	7955736.526
MAY 2016	8183675.433	8174806.01
JUNE 2016	10048829.81	9101490.549
JULY 2016	7848968.711	7750415.336
AUGUST 2016	7787054.528	7786973.341
SEPTEMBER 2016	7156618.571	7163097.524
OCTOBER 2016	8311149.07	8255850.564
NOVEMBER2016	7244105.422	7243102.284
DECEMBER 2016	7571366.53	7593525.567
JANUARY 2017	5880394.827	5851320.041
FEBRUARY 2017	5928712.691	5960806.983
MARCH 2017	6660278.193	6648353.289
APRIL 2017	6289454.234	6294147.388
MAY 2017	6128324.757	6138656.659
JUNE 2017	6049716.976	6056167.482
JULY 2017	6063450.257	6084467.447
AUGUST 2017	6334940.849	6357722.817
SEPTEMBER 2017	6402970.39	6390828.413
OCTOBER 2017	6092355.126	6125064.426
NOVEMBER2017	6013210.365	6024796.575
DECEMBER 2017	6152819.376	6174946.681

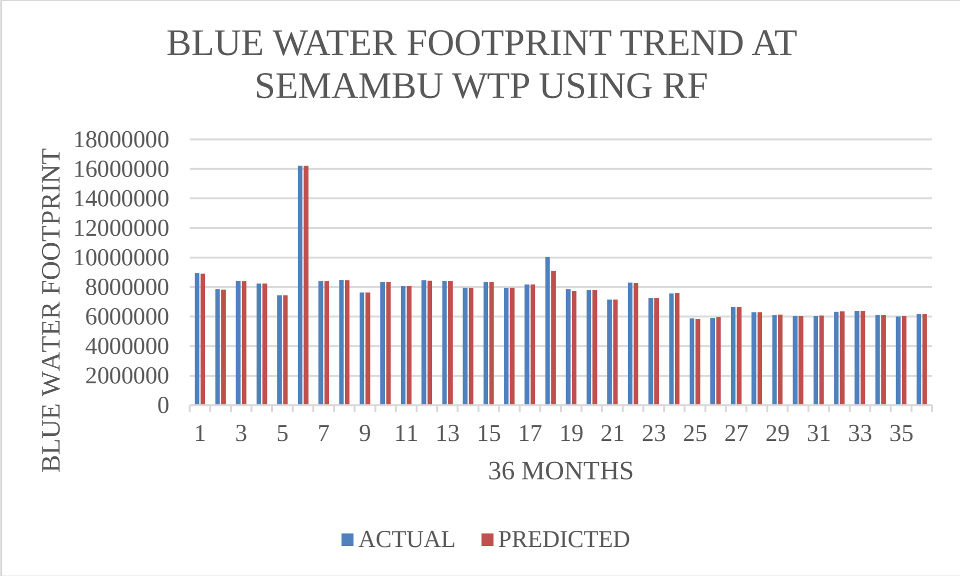


Figure 4.35 WFblue trend at Semambu WTP using RF

Figure 4.35 illustrates the actual and predicted value of WFblue at Semambu WTP after undergoes training using RF. The highest value of WFblue was in June 2015 which for actual value is 16221394.65 m³/month and for predicted value is 16230212.69 m³/month. There is not much difference in both values. Besides that, the lowest of WFblue amount of the actual and predicted was in January 2017 with the amount of 5880394.827 m³/month and 5851320.041 m³/month.

4.4.2 Prediction WFblue at Panching WTP

4.4.2.1 Artificial Neural Network Algorithm

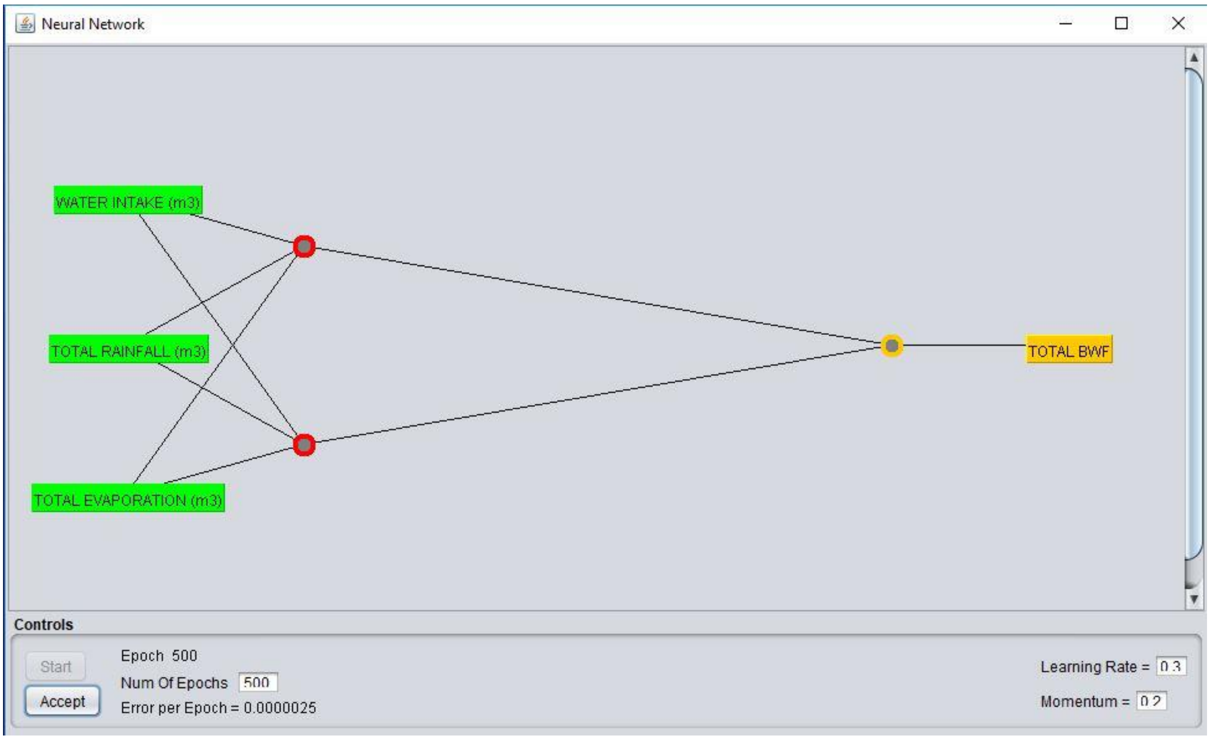


Figure 4.36 Number of hidden layers for ANN in WEKA software

Figure 4.36 shows 2 hidden layers when performing a training to a data sets using ANN algorithm. ANN algorithms will trained the WFblue data sets in order to produce the predicted value. The number of hidden layers in ANN defined the number of neurons that been created by ANN to produce the output. Meanwhile, the number of Epochs means the number of the data sets will undergo the training. In this study, 20 hidden neurons been tested and 500 Epochs been set. The lowest value of RMSE produce by the training been chose to construct the predicted trend.

```
=== Evaluation on training set ===

Time taken to test model on training data: 0.55 seconds

=== Summary ===

Correlation coefficient                  1
Mean absolute error                      0.0004
Root mean squared error                  0.0008
Relative absolute error                  0.3459 %
Root relative squared error              0.5491 %
Total Number of Instances             1096
```

Figure 4.37 Result after the training using ANN

Figure 23 shows the result after the WFblue data sets undergoes the training. The RMSE value which is 0.0008 produced after the hidden layer been set to 2.

Table 4.13 Analysis of RMSE and hidden neurons

Hidden Neuron	**RMSE-value**	**Hidden Neuron**	**RMSE-value**
1	0.0047	**11**	0.0011
2	0.0008	**12**	0.0011
3	0.001	**13**	0.001
4	0.0011	**14**	0.0009
5	0.0012	**15**	0.0011
6	0.0011	**16**	0.0011
7	0.0011	**17**	0.0009
8	0.001	**18**	0.001
9	0.0011	**19**	0.0011
10	0.0011	**20**	0.001

Refer to table 4.13, the training sets with 2 neurons have been chosen to predict the total WFblue because the lowest RMSE value was obtained from the training process. The value of RMSE is different depends on the number of hidden neurons that used to train the data sets.

Table 4.14 Analysis of actual and predicted value of WFblue by using ANN

MONTH	ACTUAL (m^3)	PREDICTED (m^3)
JANUARY 2015	2813254.71	2807313.56
FEBRUARY 2015	2620057.48	2616490.336
MARCH 2015	2749378.61	2748404.569
APRIL 2015	3271475.67	3273135.245
MAY 2015	4399947.24	4400150.454
JUNE 2015	9129732.38	9131639.986
JULY 2015	4817466.29	4818163.847
AUGUST 2015	4568850.82	4571381.423
SEPTEMBER 2015	4390680.65	4393599.429
OCTOBER 2015	4597710.95	4599067.743
NOVEMBER2015	4219415.25	4221915.525
DECEMBER 2015	4288338.61	4289659.746
JANUARY 2016	4176084.52	4177461.938
FEBRUARY 2016	3928847.13	3930381.369
MARCH 2016	4287665.31	4288688.651
APRIL 2016	4287661.31	4287938.508
MAY 2016	4758650.98	4760496.276
JUNE 2016	4664032.83	4665809.976
JULY 2016	4995495.78	4996529.581
AUGUST 2016	4688992.81	4691355.566
SEPTEMBER 2016	4836782.77	4837145.954
OCTOBER 2016	3481727.96	3484131.369
NOVEMBER2016	3861258.08	3862327.668
DECEMBER 2016	3804737.29	3806635.181
JANUARY 2017	4109920.76	4111366.767
FEBRUARY 2017	3796872.9	3798684.474
MARCH 2017	4007572.28	4007926.624
APRIL 2017	3748326.27	3749736.688
MAY 2017	3407654.38	3407932.151
JUNE 2017	3322278.45	3320853.632
JULY 2017	3430474.44	3428514.648
AUGUST 2017	3626781.15	3627025.52
SEPTEMBER 2017	3431675.07	3429813.623
OCTOBER 2017	3570461.51	3568449.317
NOVEMBER2017	3497313.22	3494227.488
DECEMBER 2017	3591950.77	3590203.544

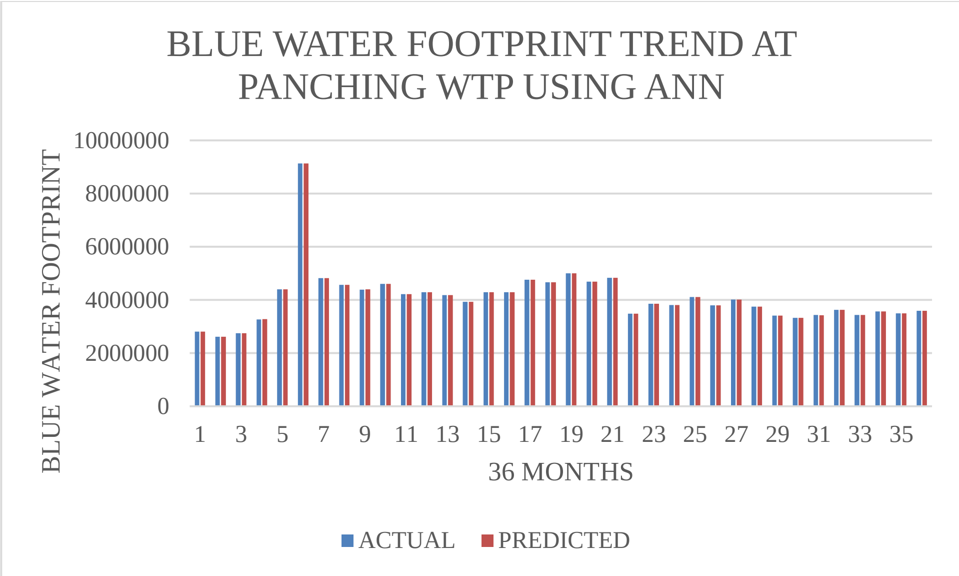
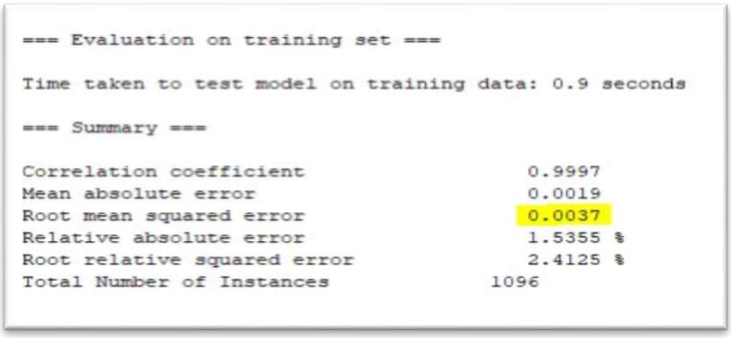


Figure 4.38 WFblue trend at Panching WTP using ANN

Figure 4.38 illustrates the bar graph comparison between actual and predicted value of WFblue at Semambu WTP after undergoes training using ANN. The highest value of WFblue was in June 2015 which for actual value is 9129732.38 m³/month and for predicted value is 9131639.986 m³/month. There is not much difference in both values. Besides that, the lowest of WFblue amount of the actual and predicted was in February 2015 with the amount of 2620057.48 m³/month and 2616490.336 m³/month.

4.4.2.2 Random Forest Algorithm



```
=== Evaluation on training set ===

Time taken to test model on training data: 0.9 seconds

=== Summary ===

Correlation coefficient                  0.9997
Mean absolute error                      0.0019
Root mean squared error                  0.0037
Relative absolute error                  1.5355 %
Root relative squared error              2.4125 %
Total Number of Instances             1096
```

Figure 4.39 Result after the training using RF

Figure 4.39 shows the result after the WFblue data sets undergoes the training with a RMSE value of 0.0037.

Table 4.15 Analysis of actual and predicted value of WFblue by using RF

MONTH	ACTUAL (m^3)	PREDICTED (m^3)
JANUARY 2015	2813254.71	2806882
FEBRUARY 2015	2620057.48	2615836.841
MARCH 2015	2749378.61	2750765.571
APRIL 2015	3271475.67	3290103.173
MAY 2015	4399947.24	4397767.335
JUNE 2015	9129732.38	9121353.475
JULY 2015	4817466.29	4810227.652
AUGUST 2015	4568850.82	4571699.539
SEPTEMBER 2015	4390680.65	4385536.583
OCTOBER 2015	4597710.95	4589183.204
NOVEMBER2015	4219415.25	4220794.125
DECEMBER 2015	4288338.61	4287415.669
JANUARY 2016	4176084.52	4173613.175
FEBRUARY 2016	3928847.13	3926347.338
MARCH 2016	4287665.31	4287607.758
APRIL 2016	4287661.31	4289256.29
MAY 2016	4758650.98	4757480.964
JUNE 2016	4664032.83	4667161.512
JULY 2016	4995495.78	4990480.158
AUGUST 2016	4688992.81	4685022.053
SEPTEMBER 2016	4836782.77	4827884.855
OCTOBER 2016	3481727.96	3929517.625
NOVEMBER2016	3861258.08	3863518.653
DECEMBER 2016	3804737.29	3832597.986
JANUARY 2017	4109920.76	4107475.756
FEBRUARY 2017	3796872.9	3794361.91
MARCH 2017	4007572.28	4008090.447
APRIL 2017	3748326.27	3756337.686
MAY 2017	3407654.38	3408606.418
JUNE 2017	3322278.45	3325437.356
JULY 2017	3430474.44	3436299.906
AUGUST 2017	3626781.15	3626360.621
SEPTEMBER 2017	3431675.07	3429049.409
OCTOBER 2017	3570461.51	3565365.302
NOVEMBER2017	3497313.22	3498361.558
DECEMBER 2017	3591950.77	3591088.349

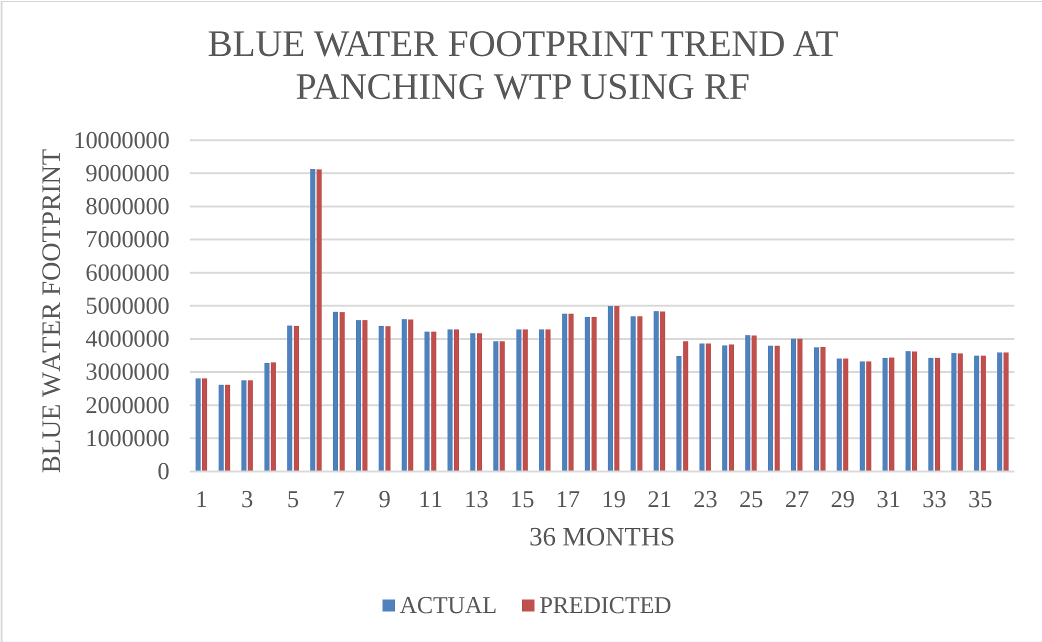


Figure 4.40 WFblue trend at Panching WTP using RF

Figure 4.40 illustrates the actual and predicted value of WFblue at Semambu WTP after undergoes training using RF algorithm. The differences between the actual and predicted value is slightly the same due to the least error that been made during the training.

The highest value of WFblue for actual value is 9129732.38 m³/month and for predicted value is 9121353.475 m³/month. There is not much difference in both values. Besides that, the lowest of WFblue value was in February 2015. The actual and predicted values are 2620057.48 m³/month and 2615836.841 ³/month.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

At the end of the study, all the objectives has been achieved and successfully stated. The total WFblue in Semambu and Panching WTPs has been calculated. All the parameters involved in measuring WFblue which is the amount of water intake, rainfall utilisation and total evaporation are clearly stated and tabulated in a table.The total blue water footprint gives clear visualization about the amount of water been used in water treatment plants which are; in Panching WTP is 47,048,842 m^3 (2017), 51,918,809 m^3 (2017) and 43,541,281 m^3 (2017) and Semambu WTP is 98,076,250 m^3 (2017), 97,089,442 m^3 (2017) and 73,996,628 m^3 (2017). This results indicate that our water resource is sustainable for the supplies. Moreover, it also can be concluded that Semambu WTP consumed high amounnt of fresh water compared to Panching WTP due to its larger capacity of WSTP tanks.

Meanwhile, after the actual results were undergo a series of training by using different algorithms, the predicted trend of blue water footprint for three years have been produced. The predicted trend that produced by ANN and RF algorithms been compared to the actual value of WFblue. This comparison showed the precision of the algorithms in predicting the value. Besides, the predicted trend of WFblue in Semambu WTP showed a decrement and this pattern shows that the amount of total WFblue in the next year are predicted to be decrease. In Panching WTP, total WFblue is also predicted to be decrease based on the predicted data produced after undergo series of training by using ANN and RF algorithms. This prediction information can be used and applied by any authorities or stakeholders in maintaining the sustainability of water treatment plant as well as the water management.

Lastly, Artificial Neural Network been declared as the best algorithm to be used in prediction. This conclusion been made after comparing the RMSE value produced by both algorithms after the training process. Least RMSE values that generated by ANN defined that this algorithm will make less error in predicting values. Furthermore, based on the capabilities of ANN where it can train the data sets or inputs by applying the previous experienced in order to produce the ouput regarding to the required objectives. High numbers of training in ANN algorithms will give the greater ouput as its has experienced lots of data with different parameters. This result is parallel with the previous study which showed the effectiveness of ANN application as a machine learning.

5.2 Recommendation

According to the result that showing the decrement of WFblue over time, some recommendation are suggested for water management sustainabilty in water treatment plants. The sustainability of WTPs is very important in order to provide sufficient amount of water to the consumers.

In order to control the amount of WFblue, the correct monitoring of the water intake amount need to be implemented. Excessive amount of water intake will burden the water treatment plants itself. The stakeholders can plan or produce a daily schedule for the amount of water that will be abstracted for treatment process. Besides, the proper management of water treatment plant need to be improve instead of using the old management system. The actual amount of water abstraction and distribution must be recorded accordingly. In Malaysia, climate changes will also affect the amount of water resources. During hot season, the amount of water might be less due to the evaporation while during rainy season, the amount of water is sufficient for the treatment and to be provided to the consumers. This uncontrolled climate changes sometimes will harm and gives bad effects to the WTP which is excessive amount of water, lot of sediments and lack of water to be distributed. Hence, proper management of WTP will ensure the sustainability of the water supply.

In addition, it is suggested that all WTPs applies water footprint assessment as an approach to ensure the efficiency of water utilisation. This approach will also help in ensuring the sustainbility of the water supply. The total amount of WFblue will gives clear visualisation of total amount of water been used during the treatment process. Furthermore, WFblue amount will be lower if people know how to utilised the rainfall. Good rainfall utilisation will little bit help to reduce the amount of water been taken from water resources and still can be used or treated in order to provide good quality of water to the consumers. Water footprint assessment also can provide awareness to the community about the importance of the knowledge about water consumption. This indication will make the users to be aware and used the water wisely without wasting it.

In terms of prediction trend, it shows that strategy can be developed in order to maintain the sustainability of water treatment plants. Nowadays, artificial intellegence could be used in various type of problems including the issue that related to water. Lot of researches has been made to solve water scarcity by using machine learning. This method will help the user or decision maker to solve the problems just in a split seconds. On the other hand, machine learning can be used for classification, regression, sorting, logging and many more. Thus, people are suggested to learn and used machine learning in order to solve any problems related to any fields.

Lastly, in order to ensure the better output been produce, data cleaning or pre-processing process need to be conducted. This process will clean the data without any uncertainties. Besides, data cleaning will also indirectly help the users to produce least error in performing the series of training using the algorithm. Data cleaning that were used in this study also can be used for other problems involving numerical types of parameters. Treatment of missing data is a compulsory before inserting the data into the software, so the software and algorithm chose will able to produce the ouput without any missing values. Next, data normalisation is also needed because some of the data prepared has different range of values. To fix this problem, data normalisation will convert the range value of the input in the range of 0 to 1, thus the data can be easily trained by the algorithm. The last step of data cleaning would be removing outliers. Some of the data might be out of the boundary limits which is upper boundary and lower boundary. By performing outliers removal, the out-range data will be replace with the average value of the parameters.

REFERENCES

Ahmad, S., & Simonovic, S. P. (2005). An artificial neural network model for generating hydrograph from hydro-meteorological parameters, *315*, 236–251. https://doi.org/10.1016/j.jhydrol.2005.03.032

Ali, U., Muhammad, W., Brahme, A., Skiba, O., & Inal, K. (2019). Application of artificial neural networks in micromechanics for polycrystalline metals. *International Journal of Plasticity*, (May), 1–15. https://doi.org/10.1016/j.ijplas.2019.05.001

Ao, Y., Li, H., Zhu, L., Ali, S., & Yang, Z. (2019). Journal of Petroleum Science and Engineering The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. *Journal of Petroleum Science and Engineering*, *174*(November 2018), 776–789. https://doi.org/10.1016/j.petrol.2018.11.067

Asare, I., Zhao, X., Asante, H., & Nyarko, C. (2018). Technology in Society Urban water supply systems improvement through water technology adoption. *Technology in Society*, *55*(January), 70–77. https://doi.org/10.1016/j.techsoc.2018.06.005

Banerjee, P., Singh, V. S., Chatttopadhyay, K., Chandra, P. C., & Singh, B. (2011). Artificial neural network model as a potential alternative for groundwater salinity forecasting. *Journal of Hydrology*, *398*(3–4), 212–220. https://doi.org/10.1016/j.jhydrol.2010.12.016

Buchtele, J., Richta, K., & Chlumecky, M. (2017). Application of random number generators in genetic algorithms to improve rainfall-runoff modelling, *553*, 350–355. https://doi.org/10.1016/j.jhydrol.2017.08.025

Cai, B., Liu, B., & Zhang, B. (2019). Evolution of Chinese urban household ' s water footprint. *Journal of Cleaner Production*, *208*, 1–10. https://doi.org/10.1016/j.jclepro.2018.10.074

Casella, P., Rosa, L. De, Salluzzo, A., & Gisi, S. De. (2019). Combining GIS and FAO ' s crop water productivity model for the estimation of water footprinting in a temporary river catchment. *Sustainable Production and Consumption*, *17*, 254–268. https://doi.org/10.1016/j.spc.2018.11.002

Chang, F., Chang, L., Huang, C., & Kao, I. (2016). Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. *Journal of Hydrology*, *541*, 965–976. https://doi.org/10.1016/j.jhydrol.2016.08.006

Chapagain, A. K., & Hoekstra, A. Y. (2011). The blue , green and grey water footprint of rice from production and consumption perspectives. *Ecological Economics*, *70*(4), 749–758. https://doi.org/10.1016/j.ecolecon.2010.11.012

Chen, Z., Han, F., Wu, L., Yu, J., Cheng, S., Lin, P., & Chen, H. (2018). Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents. *Energy Conversion and Management*, *178*(October), 250–264. https://doi.org/10.1016/j.enconman.2018.10.040

Civit, B., Piastrellini, R., Curadelli, S., & Pablo, A. (2018). The water consumed in the production of grapes for vini fi cation (Vitis vinifera) . Mapping the blue and green water footprint. *Ecological Indicators*, *85*(March 2017), 236–243. https://doi.org/10.1016/j.ecolind.2017.10.037

Ðozi, D. J., & Uro, B. D. G. (2019). Application of arti fi cial neural networks for testing long-term energy policy targets, *174*, 488–496. https://doi.org/10.1016/j.energy.2019.02.191

Dur, P. (2018). Management effectiveness assessment in wastewater treatment plants through a new water footprint indicator, *198*, 463–471. https://doi.org/10.1016/j.jclepro.2018.07.062

Gerbens-leenes, P. W., Hoekstra, A. Y., & Bosman, R. (2018). The blue and grey water footprint of construction materials : Steel , cement and glass. *Water Resources and Industry*, *19*(October 2017), 1–12. https://doi.org/10.1016/j.wri.2017.11.002

Girolamo, A. M. De, Miscioscia, P., Politi, T., & Barca, E. (2019). Improving grey water footprint assessment : Accounting for uncertainty. *Ecological Indicators*, *102*(March), 822–833. https://doi.org/10.1016/j.ecolind.2019.03.040

Gush, M., Dzikiti, S., Laan, M. Van Der, Steyn, M., Manamathela, S., & Pienaar, H. (2019). Agricultural and Forest Meteorology Field quanti fi cation of the water footprint of an apple orchard , and extrapolation to watershed scale within a winter rainfall Mediterranean climate zone. *Agricultural and Forest Meteorology*, *271*(December 2018), 135–147. https://doi.org/10.1016/j.agrformet.2019.02.042

Harding, G., Courtney, C., & Russo, V. (2017). When geography matters . A location-adjusted blue water footprint of commercial beef in South Africa. *Journal of Cleaner Production*, *151*, 494–508. https://doi.org/10.1016/j.jclepro.2017.03.076

Hess, T., Andersson, U., Mena, C., & Williams, A. (2015). The impact of healthier dietary scenarios on the global blue water scarcity footprint of food consumption in the UK. *Food Policy*, *50*, 1–10. https://doi.org/10.1016/j.foodpol.2014.10.013

Hoekstra, A. Y. (2019). Advances in Water Resources Green-blue water accounting in a soil water balance. *Advances in Water Resources*, *129*(December 2018), 112–117. https://doi.org/10.1016/j.advwatres.2019.05.012

Hogeboom, R. J., Knook, L., & Hoekstra, A. Y. (2018). Advances in Water Resources The blue water footprint of the world ’ s artificial reservoirs for hydroelectricity , irrigation , residential and industrial water supply , flood protection , fishing and recreation. *Advances in Water Resources*, *113*, 285–294. https://doi.org/10.1016/j.advwatres.2018.01.028

Hou, X., Yuan, J., Ma, C., & Sun, C. (2019). Neurocomputing Parameter estimations of uncooperative space targets using novel mixed artificial neural network. *Neurocomputing*, *339*, 232–244. https://doi.org/10.1016/j.neucom.2019.02.038

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

Laan, M. Van Der, Vahrmeijer, T., Bristow, K. L., & Annandale, J. G. (2017). Science of the Total Environment Establishing and testing a catchment water footprint framework to inform sustainable irrigation water use for an aquifer under stress. *Science of the Total Environment*, *599–600*, 1119–1129. https://doi.org/10.1016/j.scitotenv.2017.04.170

Li, C., Xu, M., Wang, X., & Tan, Q. (2018). Spatial analysis of dual-scale water stresses based on water footprint accounting in the Haihe River Basin , China. *Ecological Indicators*, *92*(February 2017), 254–267. https://doi.org/10.1016/j.ecolind.2017.02.046

Lindauer, M., Rijn, J. N. Van, & Kotthoff, L. (2019). The Algorithm Selection Competitions 2015 and 2017. *Artificial Intelligence*, (October 2018). https://doi.org/10.1016/j.artint.2018.10.004

Liu, X., Guo, P., Li, F., & Zheng, W. (2019). Optimization of planning structure in irrigated district considering water footprint under uncertainty. *Journal of Cleaner Production*, *210*, 1270–1280. https://doi.org/10.1016/j.jclepro.2018.10.339

Lucia, A., Maiello, A., & Quintslr, S. (2018). Water supply system in the Rio de Janeiro Metropolitan Region : Open issues , contradictions , and challenges for water access in an emerging megacity. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2018.02.045

Ma, X., Yang, D., Shen, X., Zhai, Y., Zhang, R., & Hong, J. (2018). Science of the Total Environment How much water is required for coal power generation : An analysis of gray and blue water footprints. *Science of the Total Environment*, *636*, 547–557. https://doi.org/10.1016/j.scitotenv.2018.04.309

Morera, S., Corominas, L., Poch, M., Aldaya, M. M., & Comas, J. (2016). Water footprint assessment in wastewater treatment plants. *Journal of Cleaner Production*, *112*, 4741–4748. https://doi.org/10.1016/j.jclepro.2015.05.102

Nguyen-ky, T., Mushtaq, S., Loch, A., Reardon-smith, K., An-vo, D., Ngo-cong, D., & Tran-cong, T. (2017). Predicting water allocation trade prices using a hybrid Artificial Neural Network-Bayesian modelling approach. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2017.11.049

Nouri, H., Stokvis, B., Galindo, A., Blatchford, M., & Hoekstra, A. Y. (2019). Science of the Total Environment Water scarcity alleviation through water footprint reduction in agriculture : The effect of soil mulching and drip irrigation. *Science of the Total Environment*, *653*, 241–252. https://doi.org/10.1016/j.scitotenv.2018.10.311

Novoa, V., Ahumada-rudolph, R., Rojas, O., Sáez, K., De, F., & Luis, J. (2019). Science of the Total Environment Understanding agricultural water footprint variability to improve water management in Chile. *Science of the Total Environment*, *670*, 188–199. https://doi.org/10.1016/j.scitotenv.2019.03.127

Owusu-sekyere, E., Jordaan, H., & Chouchane, H. (2017). Evaluation of water footprint and economic water productivities of dairy products of South Africa. *Ecological Indicators*, *83*(June), 32–40. https://doi.org/10.1016/j.ecolind.2017.07.041

Partopour, B., Paffenroth, R. C., & Dixon, A. G. (2018). Random Forests for mapping and analysis of microkinetics models. *Computers and Chemical Engineering*, *115*, 286–294. https://doi.org/10.1016/j.compchemeng.2018.04.019

Piotrowski, A. P., Napiorkowski, M. J., Napiorkowski, J. J., & Osuch, M. (2015). Comparing various artificial neural network types for water temperature prediction in rivers. *Journal of Hydrology*, *529*, 302–315. https://doi.org/10.1016/j.jhydrol.2015.07.044

Poort, J. P., Ramdin, M., Kranendonk, J. Van, & Vlugt, T. J. H. (2019). Fluid Phase Equilibria Solving vapor-liquid fl ash problems using arti fi cial neural networks. *Fluid Phase Equilibria*, *490*, 39–47. https://doi.org/10.1016/j.fluid.2019.02.023

Ramı, V., Cleofe, M., & Jesus, N. (2005). Artificial neural network technique for rainfall forecasting ˜ o Paulo region applied to the Sa, *301*, 146–162. https://doi.org/10.1016/j.jhydrol.2004.06.028

Rosecrans, C. Z., Nolan, B. T., & Gronberg, J. M. (2017). Prediction and visualization of redox conditions in the groundwater of Central Valley , California. *Journal of Hydrology*, *546*, 341–356. https://doi.org/10.1016/j.jhydrol.2017.01.014

Safari, M., Aksoy, H., & Mohammadi, M. (2016). Artificial neural network and regression models for flow velocity at sediment incipient deposition. *Journal of Hydrology*, *541*, 1420–1429. https://doi.org/10.1016/j.jhydrol.2016.08.045

Tauer, G., Date, K., Nagi, R., & Sudit, M. (2019). An incremental graph-partitioning algorithm for entity resolution. *Information Fusion*, *46*(March 2018), 171–183. https://doi.org/10.1016/j.inffus.2018.06.001

Udimal, T. B., Jincai, Z., Ayamba, E. C., & Owusu, S. M. (2017). China ' s water situation ; the supply of water and the pattern of its usage. *International Journal of Sustainable Built Environment*, *6*(2), 491–500. https://doi.org/10.1016/j.ijsbe.2017.10.001

Veettil, A. V., & Mishra, A. K. (2016). Water security assessment using blue and green water footprint concepts. *Journal of Hydrology*, *542*, 589–602. https://doi.org/10.1016/j.jhydrol.2016.09.032

Villacampa, Y., Navarro-gonzález, F. J., Compañ-rosique, P., & Satorre-cuerda, R. (2019). A guided genetic algorithm for diagonalization of symmetric and Hermitian matrices. *Applied Soft Computing Journal*, *75*, 180–189. https://doi.org/10.1016/j.asoc.2018.11.004

Wang, F., Wang, S., Li, Z., You, H., Aviso, K. B., & Tan, R. R. (2019). Resources , Conservation & Recycling Water footprint sustainability assessment for the chemical sector at the regional level. *Resources, Conservation & Recycling*, *142*(July 2018), 69–77. https://doi.org/10.1016/j.resconrec.2018.11.009

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, *527*, 1130–1141. https://doi.org/10.1016/j.jhydrol.2015.06.008

Xie, X., Zhang, T., Wang, M., & Huang, Z. (2019). Science of the Total Environment Impact of shale gas development on regional water resources in China from water footprint assessment view. *Science of the Total Environment*, *679*, 317–327. https://doi.org/10.1016/j.scitotenv.2019.05.069

Xu, M., Li, C., Wang, X., Cai, Y., & Yue, W. (2018). Optimal water utilization and allocation in industrial sectors based on water footprint accounting in Dalian City , China. *Journal of Cleaner Production*, *176*, 1283–1291. https://doi.org/10.1016/j.jclepro.2017.11.203

Zhai, Y., Tan, X., Ma, X., An, M., Zhao, Q., & Shen, X. (2019). Water footprint analysis of wheat production. *Ecological Indicators*, *102*(December 2018), 95–102. https://doi.org/10.1016/j.ecolind.2019.02.036

Zhenzhen, T., Heating, B., & Wang, P. R. (2019). ScienceDirect ScienceDirect August water analysis of blue and footprint for Beijing-Tianjin-Hebei Urban Agglomeration A three-scale analysis and grey water a of blue footprint for Beijing-Tianjin-Hebei Urban Agglomeration Assessing the feasibility using heat demand forecast. *Energy Procedia*, *158*(2018), 4049–4054. https://doi.org/10.1016/j.egypro.2019.01.833

Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). ScienceDirect Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *CJ*, *4*(3), 212–219. https://doi.org/10.1016/j.cj.2016.01.008

Zhuo, L., Hoekstra, A. Y., Wu, P., & Zhao, X. (2019). Science of the Total Environment Monthly blue water footprint caps in a river basin to achieve sustainable water consumption : The role of reservoirs. *Science of the Total Environment*, *650*, 891–899. https://doi.org/10.1016/j.scitotenv.2018.09.090

APPENDIX A
DATA ANALYSIS

Table 5.1 Total WFblue from 2015-2017

Total WFblue (m^3)	**2015**	**2016**	**2017**
Semambu WTP	98078265.67	97089441.6	72996628
Panching WTP	47048842.36	51918809	43541281.1

Table 5.2 RMSE comparison

Algorithm\RMSE	Semambu WTP	Panching WTP
ANN	0.0002	0.0008
RF	0.0113	0.0037

Table 5.3 Prediction of WFblue at Semambu WTP

Algorithm	Artificial Neural Network (ANN)			Random Forest (RF)		
Year	**2015**	**2016**	**2017**	**2015**	**2016**	**2017**
Actual	98078266	97089442	72996628	98078266	97089442	72996628
Predicted	97954663	96973241	71995834	98912454	97286633	74233512
Percentage (%)	-0.12602	-0.11968	-1.37101	0.850533	0.203103	1.69444
Overall (%)	-0.46263			0.8458437		

Table 5.4 Prediction of WFblue at Panching WTP

Algorithm	Artificial Neural Network (ANN)			Random Forest (RF)		
Year	**2015**	**2016**	**2017**	**2015**	**2016**	**2017**
Actual	47048842	51918809	43541281	47048842	51918809	43541281
Predicted	47066282	51866331	43441896	47266944	51593321	43265891
Percentage (%)	-0.04703	-0.10108	-0.22825	0.379109	-0.62692	-0.63248
Overall (%)	-0.15382			-0.58204322		