PREDICTION OF GREY WATER FOOTPRINT OF SUNGAI LEMBING, BUKIT SAGU AND BUKIT UBI WATER TREATMENT PLANTS

SITI FAZLINA BINTI MOHD SUHAIMI

B. ENG (HONS.) CIVIL ENGINEERING

UNIVERSITI MALAYSIA PAHANG

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	: SITI FAZLINA BINTI MOHD SUHAIMI
Date of Birth	: 19TH FEBRUARY 1996
Title	: PREDICTION OF GREY WATER FOOTPRINT OF SUNGAI LEMBING, BUKIT SAGU AND BUKIT UBI 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)
960219115450	Dr. Edriyana binti A.Aziz
New IC/Passport Number	Name of Supervisor
Date: 31st May 2019	Date: 31st May 2019

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

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:

Note: This letter should be written by the supervisor, addressed to the Librarian, *Perpustakaan Universiti Malaysia Pahang* with its copy attached to the thesis.

SUPERVISOR'S DECLARATION

I/We* hereby declare that I/We* have checked this thesis/project* and in my/our* opinion, this thesis/project* is adequate in terms of scope and quality for the award of the degree of *Doctor of Philosophy/ Master of Engineering/ Master of Science in …………………………..

____

(Supervisor's Signature)

Full Name :

Position :

Date :

____

(Co-supervisor's Signature)

Full Name :

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 : SITI FAZLINA BINTI MOHD SUHAIMI

ID Number : AA15168

Date : 31st MAY 2019

PREDICTION OF GREY WATER FOOTPRINT OF SUNGAI LEMBING, BUKIT SAGU AND BUKIT WATER TREATMENT PLANTS

SITI FAZLINA BINTI MOHD SUHAIMI

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

Thanks to Almighty Allah for giving me strength and ability to understand learn and complete this report.

I would like to express my deep gratitude to Dr Edriyana binti A.Aziz, my research supervisors, for her patient guidance, enthusiastic encouragement and useful critiques of this research work. I would also like to thank Mohd Syazwan Nizam bin Mohd Moni for his advice and assistance in keeping my progress on schedule.

My grateful thanks are also extended to all my friends especially Siti Ainifatihah binti Noor Hazlim and Ruziana binti Kamarzaman for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four years.

Finally, I wish to thank my parents for their support and encouragement throughout my study.

ABSTRAK

Faktor terpenting yang mempengaruhi kekurangan air di dalam dan luar negara dan ketersediaan sumber air tawar bukan sahaja penduduk dunia yang semakin berkembang tetapi juga peningkatan permintaan air. Dari kajian ini, tahap pencemaran air di lembah sungai Kuantan direkodkan mengikut setiap loji rawatan air (WTP) dan penilaian jejak air kelabu digunakan sebagai pendekatan untuk mengira jumlah air tawar yang digunakan untuk mengasimilasi kepekatan pencemar. Oleh itu, kajian ini bertujuan untuk mengira jejak air kelabu keseluruhan, untuk meramalkan trend keseluruhan jejak air kelabu dan membandingkan algoritma terbaik antara Rangkaian Neural Buatan (ANN) dan Bayesian Networks (BN) dalam ramalan jejak air kelabu di Sungai Lembing WTP, Bukit Sagu WTP dan Bukit Ubi WTP pada tahun 2015 hingga 2017. Sebagai hasil akhir kajian ini, jumlah air kelabu di Sungai Lembing, Bukit Sagu dan loji rawatan air Bukit Ubi di lembah sungai Kuantan dikira. Trend ramalan keseluruhan jejak air kelabu dalam tiga loji rawatan air telah dapat dihasilkan sebagai hasil akhir kajian. Algoritma Rangkaian Neural Buatan (ANN) juga dipilih sebagai algoritma terbaik.

ABSTRACT

The most important factors affecting water scarcity in local and global and the availability of fresh water resources are not only a growing world population but also an increasing water demand. From this study, the level pollution of water in Kuantan river basin is recorded according to each water treatment plant (WTP) and grey water footprint assessment was used as an approach to account the total amount of freshwater used to assimilate the pollutant's concentration. Hence, this study is aimed to calculate the total grey water footprint, to predict the trend of total grey water footprint and to compare the best algorithm between Artificial Neural Network (ANN) and Bayesian Networks (BN) in grey water footprint prediction at Sungai Lembing WTP, Bukit Sagu WTP and Bukit Ubi WTP in 2015 until 2017. As the end result of this study, the total grey water footprint in Sungai Lembing, Bukit Sagu and Bukit Ubi water treatment plant in Kuantan river basin is calculated. Prediction trend of total grey water footprint in three water treatment plants has able to be produced. Artificial Neural Network (ANN) algorithm is also be chosen as the best algorithm.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Fe^2	Iron
m^3/s	Meter cubic per second
L	Load of pollutant
NO3-N	Nitrate-nitrogen

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
BN	Bayesian Networks
BOD	Biochemical Oxygen Demand
COD	Chemical Oxygen Demand
LOCF	Last Observation Carried Forward
JPS	Jabatan Pengairan dan Saliran Pahang
MMD	Malaysian Meteorological Department
PAIP	Pengurusan Air Pahang Berhad
RMSE	Root Mean Square Error
WF	Water Footprint
WFgrey	Grey Water Footprint
WTP	Water Treatment Plants
WEKA	Wakaito Environment for Knowledge Analysis

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Water is most important substance for all living things include plants and animals to survive on the earth. People depend on water not only for drinking but also cooking, washing, carrying away wastes and other domestic needs. Water is an important factor of production contributing both directly and indirectly to economic activity across all sectors and regions of the global economy (Distefano & Kelly, 2017). Potable water or water that is safe for drinking must be free of germs and chemicals and be clear because water is a good carrier of disease germs. When the water becomes non-potable or contaminated, people can get serious illnesses if they keep use this polluted water. Diseases that produce bacteria, toxic substances and excessive amounts of minerals and organic matter should be avoided and overcome because the water used by the public must be clean and safe. Therefore, water purification works are very important to ensure that all impurities and bacteria from the air are removed and make it healthy. Water supply systems must also meet requirements for public, commercial, and industrial activities. In all cases, the water must fulfil both quality and quantity requirements.

Water supply system gets water from an assortment of areas after proper treatment; including groundwater (aquifers), surface water, for example, lakes and streams, and the ocean through desalination. Water treatment plant is the process of converting raw water which the water is taken from the river to clean water supply to the residence areas. Raw water is full of contaminants including bacteria, chemicals and other toxins. Its treatment aims to reduce the contaminants to acceptable levels to make the water safe for people to use. Therefore, water treatment plant is needed to treat raw water. The overview of water treatment plant is first take the raw water from river and it is treated with alum, lime, fluoride, chlorine. Then it has to undergo the process of filtration, send to the master station and remove the sludge in the water before distributed to the consumers.

The most important factors affecting water scarcity in local and global availability of fresh water resources not only pollution and climate change but also a growing world population and an increasing water demand. The increasing demand for fresh water is the main challenge to sustain water utilization all over the world. Freshwater represents to 2.5% of Earth's water and is progressively threatened by human activity and climate change (Distefano & Kelly, 2017). Water has been largely studied by engineers in Jordan, while little research has adopted a discourse analysis procedure to water scarcity in the country (Hussein, 2018). While the concept of water scarcity is generally current topic, it is the difficulty of getting sources of clean water for utilize amid a period of time and may result in encourages reduction and disintegration of accessible water resources. The total energy and water use in China has been clearly increasing in the last decades as China has been experiencing a dramatic economic development (Xu, Li, & Lu, 2017).

An increasing number of studies have been carried out since in 1990s, to quantify and to investigate the existing differences between water demand, water supply and the geographical distance between them, the concepts of water footprint and virtual water trade (Arto, Andreoni, & Rueda-Cantuche, 2016). The water footprint (WF) of a product or process was introduced for the first time in 2003 and is defined as the volume of freshwater consumed and polluted to produce a product. The water footprint is further analysed in three parts: the blue, green and grey water footprints. The blue WF is an indicator of the surface water or groundwater consumption, which includes the evaporated water, incorporated into the product, and lost return flow. The green WF is defined as the consumption of water from precipitation that is stored in the soil and does not run off or recharge the ground-water and thus, is available for evapotranspiration of plants. Finally, the grey WF of a process step indicates the degree of freshwater pollution that can be associated with the process step. The grey WF is defined as the volume of freshwater that is required to assimilate the load of pollutants based on natural background concentrations and existing ambient water quality standards (A Y Hoekstra & Mekonnen, 2011).

1.2 Problem Statement

It is necessary to have a knowledge of water resource consumption and pollution during the life cycle of energy production to reduce water consumption (Ding, Liu, Yang, & Lu, 2018). Water treatment plants at Sungai Lembing, Bukit Sagu and Bukit Ubi in Kuantan River Basin have enough water quantity but using ineffective long-term the management. Inadequate infrastructure and resources will bring problems in handling wastewater management efficiently and sustainably for the majority of cities in developing countries (Ding et al., 2018). According (Cha, Son, Hong, An, & Part, 2017), one of measures against water depletion and degradation by human activities by applied the efficient water management. In addition, the data for overall water which includes rain and evaporation also is not recorded. The data are important for any researches as this data can give a solution to any problems that comes from the water.

An increasing in demand for water and a decrease in availability and quality will affect the freshwater scarcity and pollution will be aggravated problems in the future (Ercin & Hoekstra, 2014). The Intergovernmental Panel on Climate Change (IPCC) reported that unabated climate change has the potential to strongly impact freshwater resources with wide ranging consequences for societies and ecosystems(Murray, Foster, & Prentice, 2012).

Water Footprints can be analysed as blue, green or grey water footprints. According to (Zhi, Yang, Yin, Hamilton, & Zhang, 2015) this sustainability analysis should be conducted for a river basin which is the common spatial unit in water planning process. Blue water footprints relate to the consumptive use of surface and groundwater, whereas green water footprints refer to the use of rainwater that ends up as runoff and does not replenish underground water supplies. Grey water footprint is an indicator of the level of freshwater pollution associated to a stage of a particular process. It is defined as the freshwater volume required assimilating the pollutant load, given the natural background concentration and existing ambient water quality standards. For this study, we will use grey water footprints as the indicator considering the places that use to study which are Sungai Lembing, Bukit Sagu and Bukit Ubi water treatment plants in Kuantan River Basin.

1.3 Objectives of Study

The objectives of this study are;

i) To calculate total grey water footprint in Sungai Lembing, Bukit Sagu and Bukit Ubi water treatment plants in Kuantan River Basin for 2015 – 2017.

ii) To compare the best algorithm between Artificial Neural Network and Bayesian Networks in grey water footprint prediction.

iii) To predict the trend of total grey water footprint for Sungai Lembing, Bukit Sagu and Bukit Ubi water treatment plants in Kuantan River Basin.

1.4 Scope of Study

In this study grey water footprint assessment will be used to asses fully water utilization in Sungai Lembing, Bukit Sagu and Bukit Ubi water supply treatment process. Grey water footprint calculation involves only water treatment processes from water abstraction to final step filtration before being distributed to all consumers. The purpose of this calculation is to determine the actual number of water used in the process of delivering to the users.

The National Physical Plan 2005 recognized Kuantan as one of the nation's future development focuses and a centre point for exchange, trade, transportation, and the travel industry, attributable to its vital area on the east coast (Kozaki, Daisuke & Idayu Binti Harun, Norhasmira & Ab. Rahim, Mohd Hasbi & Mori, Masanobu & Nakatani, Nobutake & Tanaka, 2017). Sungai Lembing, Bukit Sagu and Bukit Ubi water treatment plants are the areas that will used for this further study. This water treatment plants are connected from Kuantan River Basin before the process of distributing to all consumers. The water at this three water treatment plants will be collected to calculate total grey water footprints.

Water Footprint Assessment Manual will be used to calculate the total of grey water footprint. For grey water footprint prediction trend, the algorithms that will be used are Artificial Neural Network (ANN) and Bayesian Networks. We choose these two algorithms to predict the trend grey water footprint at Sungai Lembing, Bukit Sagu and Bukit Ubi water treatment plants as many researchers use these two algorithms in their investigations. Within this two algorithms will be compared to choose the best algorithm for grey water footprint prediction.

1.5 Significance of Study

Waste water treatment is an addition to the natural method of water purification. To maximize the utilization of natural resources, wastewater treatment plants are organized and enforced. Industrial wastewater treatment plants and sewage treatment plants are used to purify water and make it helpful again. It employed the basic concepts and working principles of essential unit operations such as coagulation, sedimentation, filtration and adsorption to ensure potable water for human consumption based on WHO standard specifications (Agudosi et al., 2018). The study of the grey water footprint for energy industry sectors is an important for the regional sustainable water utilization.

From this study, the level of water pollution in Kuantan river basin will be recorded as this study was using the grey water footprint as an indicator. The water demand for the residential areas of Kuantan river basin will be known based on data that collected during the study.

In this study, the total grey water footprint in Sungai Lembing, Bukit Sagu and Bukit Ubi water treatment plant in Kuantan river basin will be calculated. Then, prediction trend of total grey water footprint in three water treatment plants will be able to be obtained. Finally, the best algorithm also will be determined between Artificial Neural Network and Bayesian Networks.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

‘Consumption' refers to water losses in the catchment area from the available ground-surface water body. Losses occur when water evaporates, returns to another catchment area or the sea or becomes part of a product (Aldaya, Chapagain, Hoekstra, & Mekonnen, 2011). Water consumption does not mean that water disappears because water stays within the cycle and always returns somewhere. Water is a renewable resource, but there is a limit to its availability.

Water consumption and pollution can be related with specific activities, such as irrigation, bathing, washing, cleaning, cooling and treatment. Total water consumption and pollution is generally considered to be the sum of demanding and polluting independent water activities. In the future, the facts are total water consumption and pollution is related to what and how many communities consume and the structure of the global economy supplying the various consumer products and services.

Almost 75% of the world freshwater is used annually for agriculture, and most of this water returns to the atmosphere through evapotranspiration. Accordingly, an accurate estimate of evapotranspiration in agricultural fields provides critical information about water consumption on different scales and the productivity of crop water. Given the current shortage of water, a better understanding of water consumption and productivity is crucial to global food security. It helps to identify when and where to intervene and provides information that decision - makers need to implement more sustainable water policies.

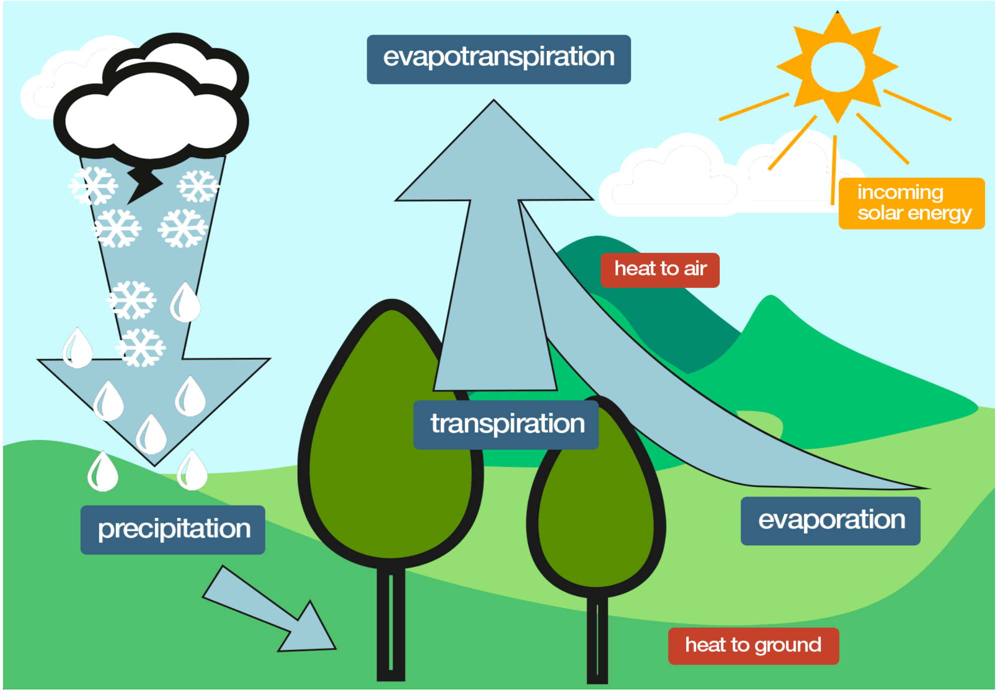

Figure 2.1 The relationship between evapotranspiration and available soil moisture provides critical information about water consumption and the productivity of crop water.

(https://www.icarda.org/dryWire/estimating-water-consumption-rates-more-effectively)

The national consumption water footprint is defined as the total quantity of fresh water used to produce the goods and services consumed by the people of the country. The footprint of national water consumption can be evaluated in two ways. The bottom - up approach is to consider the sum of all products consumed multiplied by their water footprint. The top - down approach calculates the water footprint of national consumption as the total utilization of domestic water resources plus the gross virtual import of water minus the gross virtual export of water (Aldaya et al., 2011).

Water footprint assessment is an analytical tool that can help understand how activities and products relate to water scarcity, pollution and related effects and what can be done to ensure that activities and products do not contribute to unsustainable freshwater use. The idea of the grey water footprint was introduced to express water pollution as a polluted volume, so that it can be compared to the volume of water consumption.

The detailed water footprint data provided by Hoekstra (2012) will help national governments to understand the extent to which national consumption's water footprint is. It’s related to the inefficient use of water in production and the extent to which it is inherent in the current national consumption pattern. Hence, it helps governments striving for more sustainable water use to prioritize production policies aimed at improving the efficiency of water use.

One of the most important characteristics of the revised water footprint calculation method is that it allows for a meaningful comparison of the different products and the different stages of the life cycle of a particular product (Ridoutt & Pfister, 2010). The relationship between water consumption and the effects of human health can also be analysed by quantifying freshwater availability for human needs, assessing vulnerability and estimating water scarcity-related health damages (Brown, Matlock, & Ph, 2011).

2.2 Grey Water

2.2.1 What is Grey Water Footprint?

Water footprint (WF) is an integrated measurement indicator for total water consumption, including green, blue and grey water (Arjen Y Hoekstra & Mekonnen, 2012). These categories include rainfall water resources, surface and groundwater fresh water, pollution assimilation and reflect the water requirements of different human activities. The grey element refers to the pollution of water resources and is defined as the volume of fresh water required to dilute the pollutant load generated by a given process so that the naturally occurring concentrations and water quality standards from its original source remain unchanged.

Grey water footprint is the amount of fresh water needed to assimilate pollutants in order to meet specific water quality standards (Arjen Y Hoekstra & Mekonnen, 2012). The grey water footprint considers point-source pollution discharged directly through a pipe or indirectly through runoff or leaching from the soil, impermeable surfaces or other diffuse sources to a freshwater resource. As an indicator of the appropriation of water resources by pollution, it provides a tool for assessing the sustainable, efficient and fair use of water resources (Guidelines, 2013). The application of the GWF by companies to environmental NGOs and government institutions has demonstrated its diverse usability as an indicator for the management of water resources.

Although the grey water footprint can be understood as a' dilution water requirement,' (Aldaya et al., 2011) prefer not to use this term because it seemed to cause confusion with some people who thought the term implies that should dilute pollutants rather than reduce their emissions. Of course, that's not the meaning of the concept. The grey water footprint is a pollution indicator and the better the less pollution. The treatment of wastewater before disposal will obviously lead to a reduction in the footprint of grey water, possibly to zero.

The Grey Water Footprint Index is an aggregate and weighted measure of the environmental impact at catchment level of a grey water footprint. It is based on two inputs. First, the grey water footprint of a product, consumer or producer specified by catch and per month. The other input is the level of water pollution by catch and per month. The index is obtained by multiplying both matrices and summing up the resulting matrix elements. The result can be interpreted as a grey water footprint weighed according to the level of water pollution in the areas and periods during which the different grey water footprint components occur.

2.2.2 Previous study that have grey water footprint calculation

Refers to the pollutant resources as grey water elements, recently, the use of grey water in the study of water related problems has been widely used by researchers around the world. According a study by (Zhi et al., 2015), the concept of grey water footprint (GWF) was introduced to access and to verify economic sectors’ consumption of assimilative capacity in the Haihe River Basin, China. To evaluate the ability of a water body to cleanse itself, the concept of assimilative capacity (AC) was developed. This indicates its capacity to acquire wastewater or toxic materials without harmful effects to human beings and other life forms (Zhi et al., 2015).

To explain both direct and indirect wastewater consumption of AC for all economic sectors, a concept of grey water footprint (GWF) is introduced. An accounting method was also provided for GWF based on the input - output analysis (IOA) to deal with the indirect GWF connections between economic sectors. If the grey water footprint (GWF) is smaller than the existing surface flow or groundwater flow, there is still enough water to dilute the wastewater to an acceptable concentration under the water quality standards (Herath et al., 2013). When the grey water footprint (GWF) is equal to or greater than the ambient water flow, water quality will fall below the standards (Zhi et al., 2015). The IOA model divides GWF into internal GWF (IGWF) and external GWF (EGWF). IGWF represents the GWF of products or services produced and consumed within a region while EGWF is the GWF contained in the products and services exported to other regions. The sum of IGWF and EGWF, which define as the GWF of producers, reflects the total consumption of domestic AC by local producers (Zhi et al., 2015). The equations for GWF, IGWF and EGWF can be written as follows:

$$\mathrm{GWF} = \mathrm{IGWF} \; + \mathrm{EGW} \qquad [1]$$

To show the contribution of wastewater treatment rate to the GWF, with calculation using the same method the GWF in 2007 is compared with that in an assumed scenario (without wastewater treatment). GWF values are presented for the 17 economic sectors and households in the Haihe River Basin.

In addition, the effect of time step on the calculation of annual Grey Water Footprints by utilizing 30 years of daily average nitrate-nitrogen (NO3-N) concentrations in drainage water (both leachate and runoff water derived from a process-based model) from corn and soybean production systems has investigated in this study (Vergé, VanderZaag, Smith, Grant, & Gordon, 2017). As agricultural crop production represents a major point and non-point source, nitrate-nitrogen (NO3-N) is one of the key pollutants (Fuller et al., 2010). Focus on N losses is needed to study the temporary aspects of the GW calculation. Calculations were applied to the model outputs for daily volumes of water and NO3-N concentrations in leaching and runoff from corn and soybean over 30 years. The grey water for leaching and runoff is calculated separately before summed it to determine the total grey water. The total grey water can be written as follow:

The GW footprint varied significantly when calculated for different time steps. The greatest annual footprint occurred when calculated daily as shortest time step. The GW footprint for corn ranged from 2.7 × 103 m3 ha−1, or 2700 mm of water, when estimated daily to zero for the yearly time step. For soybean it ranged from 0.5 × 103 m3 ha−1, or 500 mm of water, to zero (Vergé et al., 2017). The GW footprint results are therefore highly dependent on the time step of calculation.

2.3 Algorithm: What is algorithm?

Algorithms play an increasingly important role in selecting what information is considered to be the most relevant in our public life. Algorithms are essentially set of instructions for performing a task that generates an output from a given input. The availability of increasing computing power and data sets enables algorithms to perform tasks of an extent and complexity that human standards cannot bear. Their results hardly can be expected or even explained by their designers (Bogomolny, 2015).

An order is required for algorithms to work. Algorithms are designed to be functionally automatic and to act without regular human treatment or monitoring when triggered. This means that the information included in the database must be formalized into data so that algorithms can automatically act on it (Gillespie, 2012). In computer systems, an algorithm is essentially an instance of software logic written by software developers, so that computer data can produce output from a given input. Even with old hardware, an optimal algorithm would yield faster results than a non - optimal algorithm for the same purpose.

Not only do these algorithms help us to find information, they also provide a way to know what to know and how to know it, to participate in social and political discourse and to become familiar with the public in which we participate (Gillespie, 2012). In fact, algorithms are now essential components of self-driving cars, tests for many dieses and crime prediction frameworks, together with an increasing list of other important applications (Bogomolny, 2015).

The advantage of increased computer infrastructure availability is that it will make our lives much easier. This development, however, carries potential risks for individuals and for society as a whole. In this regard, "algocracy" cannot be controlled by democracy. Dealing with these problems requires the installation of a system of social devices to protect the individual from the running code (Cities, Zambonelli, Salim, Loke, & Meuter, 2018). Furthermore, against potentially malicious use by individuals of the data that is collected and produced by that code.

2.3.1 How to choose algorithm to different roles

At first, it's very important to define the problem to better solve it afterwards by categorises the problem by inputs and outputs. For categorizes by input, if it is a data labelled, it is a problem of supervised learning. If the purpose of finding structure is unlabelled data, it is an unsupervised learning issue. If the solution involves optimizing an objective function by interacting with an environment, this is a problem of reinforcing learning. The data understanding process plays a key role in selecting the right algorithm for the right issue. Some algorithms may work with smaller sample sets while others require tons of samples and tons of them. Some algorithms work with categorical data while others like numerical input.

Several potential risks of using algorithms have been identified, such as the risks of manipulation, prejudice, censorship, social discrimination, infringements of privacy and property rights, abuse of market power, effects on cognitive skills and increasing heteronomy. In general, algorithmic governance is about empowering software to decide without human supervision. In addition, according to some algorithmically defined policies, regulate certain aspects of our daily human activities or certain aspects of society (Cities et al., 2018). Some early examples of algorithmic management can be found in traffic management in the area or in urban management. The example is traffic lights, which adapt their frequency according to the traffic flow. Next, public transport must adjust the bus schedule and routes in real time to meet the demand for transport. In addition, energy management must automatically adjust the price of energy depending on the immediate balance between supply and demand.

The algorithm K-Means is a non-supervised algorithm that discovers groups (or clusters) in unlabelled data. This algorithm's principle is to select K random cluster centres in unlabelled data for the first time. The class of the nearest cluster centre becomes the belonging to a group of each unlabelled data point. After each data point has been assigned a category, a new centre within the cluster is estimated. Until convergence, this step will be repeated.

The Twitter Trends algorithm, which informs the user of the terms "trend" in their area at the moment, leaves the definition of " trend " unknown (Gillespie, 2012). The criteria used to evaluate the ' trend' are described in general terms only. The speed of a certain term upsurge, whether it has previously appeared in the Trend list, whether it circulates within or through user clusters.

The genetic algorithm is a probabilistic method of copying 'real life,' a stochastic optimization algorithm based on natural selection and genetic law with global optimization characteristics, strong adaptability and strong robustness (Du, Chen, Zhu, Liu, & Zhou, 2018). It is widely used in many areas such as function optimization, estimation of nonlinear parameters, pattern recognition and image processing. The genetic algorithm consists of a population of people with randomly selected parameters.

The algorithm Levenberg-Marquardt (LM) is the most commonly used optimization algorithm. It exceeds simple gradient descent and other combined gradient methods in a wide range of issues. The LM algorithm is convergent locally, however, and the iterative difference occurs when the initial assumption is poor (Du et al., 2018).

2.4 Artificial Neural Network (ANN)

2.4.1 What is ANN?

An artificial neural network is a mathematical construction corresponding to the human brain patterns (Safa, Arkebauer, Zhu, Suyker, & Irmak, 2018). The neuron is the smallest computation unit of an ANN. The structure of an artificial neuron is inspired by biological neurons. Furthermore, ANNs have been presented to the user as a kind of ‘black box’ who’s extremely complex work transforms inputs into predetermined outputs. Most efforts in the broad field of ANN research have focused on developing new learning rules, improving the network architecture and expanding into new fields of ANN applications (Naderpour, Hossein, & Fakharian, 2018).

An ANN begins with a training phase in which it learns to recognize patterns in data, whether visually, aurally or textually. During this supervised phase, the network compares its actual output with what it was supposed to produce, in other words, the output desired. The difference between the two results is adjusted by back-propagation. This means that the network works backwards from the output unit to the input units to adjust the weight of its connections between the units to the point where the difference between the actual and the desired outcome results in the lowest possible error.

The most popular and simple ANN architecture is the feed-forward multi-layer perceptron (MLP), which is represented by different neuron layers; one input layer, one or more hidden layers and one output layer (Ascione, Bianco, Stasio, Maria, & Peter, 2017). The input layer receives data or independent variables from the outside, while the output layer provides the outcomes or objective functions of the ANN. A network may have one or more hidden intermediate layers between these two layers. The number of such layers should be selected correctly because too many hidden layers lead to an over-fitting model and insufficient layers can impede the robustness and reliability of the learning process.

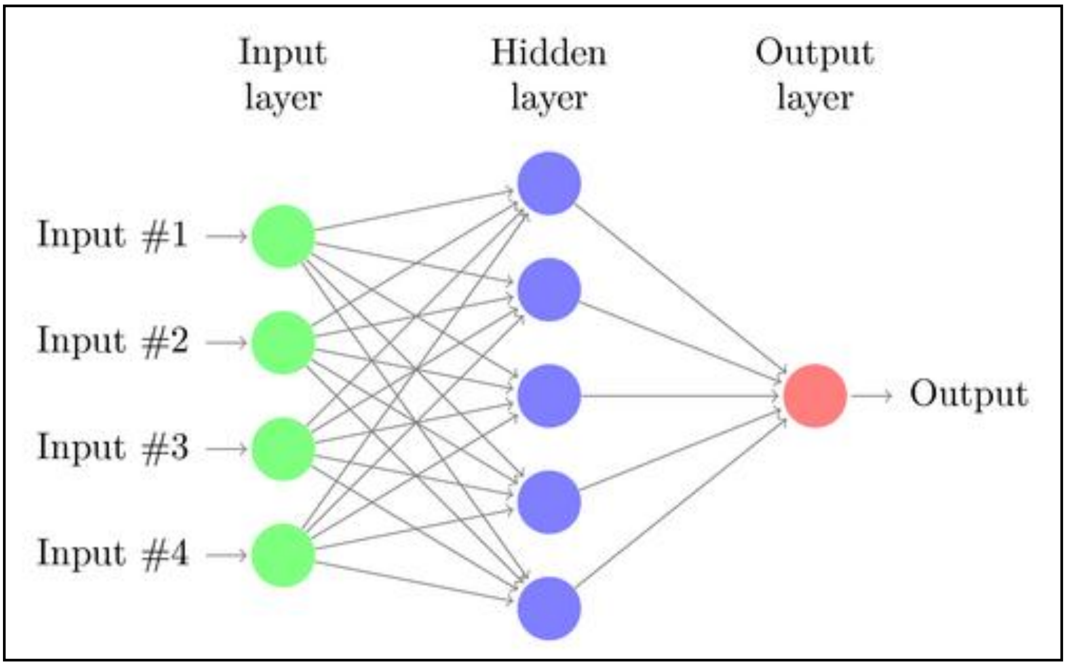

Figure 2.2 The different neuron layers

Source: http://www.texample.net/tikz/examples/neural-network/

The weights on connections between layers are adjusted during the learning process of an ANN in the same way as the processing of the human brain; where synapses are strengthened or weakened (Teixeira Júnior et al., 2015). Based on Figure above, the first layer of the ANN is the input layer, the only one that is exposed to input variables. This layer transmits the values of the input variables to the hidden layer neurons so that they can extract the relevant characteristics or patterns of the input signals and transmit the results to the output layer. Empirically, the definition of the number of neurons in each layer is carried out.

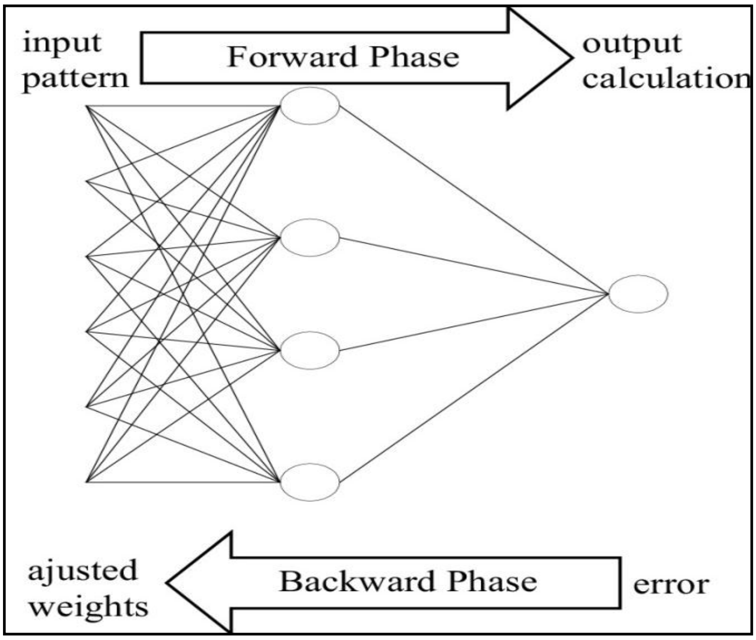

Figure 2.3 The phase of back propagation algorithm.
http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010174382015000100073

The main training algorithm is called back propagation, the fit of which weights occurs through a two-phase optimization process that is forward and backward. In the forward phase, a response is calculated for a given input pattern provided by the network. In the backward phase, to adjust the weights of the connections, the deviation (error) between the desired response (target) and the response provided by the ANN is used.

An ANN has several advantages, but the fact that it can learn from observing data sets is one of the most recognized. ANN is thus used as a random functional approach tool. These tool types help to estimate the most cost - effective and ideal ways to reach solutions while defining computer functions or distributions. ANN takes data samples instead of complete data sets to come up with solutions that save time and money. ANNs are considered relatively simple mathematical models to improve existing technologies for data analysis. The advantage of this method is its multidimensional nonlinear mapping capability of the outputs compared to the classical techniques of polynomial regression (Sezer, 2011). It will help scientists understand better the productivity of crops, the hydrological cycle and its effects on the climate (Safa et al., 2018).

2.4.2 Previous study that used ANN Application

A study by Naderpour (2018) aims to predict recycled aggregate concrete (RAC) compressive strength by using Artificial Neural Network (ANN). The experimental results on the compressive strength indicated that recycled aggregate (RA) with good quality can be used as an alternative for natural aggregate (NA) to produce concrete with mechanical properties comparable to those made with NA (Duan & Poon, 2014).

The existing tests documented in RAC literature that compiled has used as database in the study to investigate the relationship between different variables on the resulting compressive strength. A new model based on ANN therefore is developed and presented herein. The selected database consists of 139 test results containing results from important test programs which have been carried out in recent years. Back-propagation is a method used in artificial neural networks in which a gradient is calculated that is needed in the calculation of the weights to be used in the network. Generally neural network consist of three layers, input, hidden and output layer. Each layer consists of neurons and layers are interconnected by sets of correlation weights, which enable the network to process the data (Naderpour et al., 2018).

For the study, the Milne method was only applied to the connection weights in the network, as training set data was used to calculate the sensitivity analysis and the weight importance. The results of sensitivity analyses show that NN6-7 - 3 and NN6-18-3 are the best performing networks. For this study, the NN6-18-1 is selected to preserve the main objective of a single output node that predicts the concrete compressive strength. It reveals good results in the case of R-values and has the smallest MSE among all networks investigated. It is concluded that the ANN method is capable of high accuracy predictions for RAC compressive strength.

Ascione (2017) has studied and proposed a new approach using artificial neural networks (ANNs) to predict the consumption of primary energy and the thermal comfort of the occupants for any member of a building class. The results show a very satisfactory reliability of ANNs, as the values of relative errors and regression coefficients obtained are comparable to those obtained in previous studies on the use of ANNs to forecast energy performance in the building industry. It is stressed that the proposed methodology can give substantial support to rigorous approaches to building energy retrofit planning, such as cost-optimal analysis or optimization of building performance.

Based on Giwa (2016) study, the integration of electrochemical treatment and MBR for medium strength municipal wastewater treatment in Abu Dhabi (UAE) has been demonstrated. The integrated setup consisting of a pair of aluminium anode and another pair of stainless steel cathode inserted with MF membrane in a submerged MBR was able to enhance the reduction of concentrations of wastewater contaminants. Due to the initial mixed liquor compositions, an ensemble model based on artificial neural networks (ANNs) was used to model the experimental findings of chemical oxygen demand (COD), orthophosphates and ammonium removal. The components investigated in this study were chemical oxygen demand (COD), orthophosphates and ammonium.

2.4.3 What is Bayesian Networks?

Introduced by Pearl in the 1980s, Bayesian networks (BNs) are powerful tools for representing, manipulating and reasoning beliefs about the real world ((İçen & Ersel, 2019). In order to obtain a comprehensible representation of the joint probability distribution, Bayesian networks (BNs) combine graph and probability theories (Wang & Liu, 2018). In particular, a BN consists of a directed acyclic graph representing the dependent relationship between variables and a numerical section specifying the distribution of conditional probability for each variable (Wang & Liu, 2018).

In Bayesian networks, the learning task can be grouped into two subtasks: structural learning and estimation of parameters. The first subtask is to identify a network's best topology, and the second subtask is to learn the parameters that define a given network topology's conditional probability distribution (Wang & Liu, 2018).

Let $G = (X, E)$ be a directed acyclic graph (DAG), where $X = (x_1, x_2, \dots, x_n)$ is the set of nodes representing the system random variables, $E = \{e_{ij}\}$ is the set of edges representing the direct dependence relationships between variables. If there is a directed edge from node x_j to node x_i, we say x_j is a parent of x_i. $Pa(x_i)$ is defined as the set containing the parents of x_i in the graph. Let *P* be a joint probability distribution of random variables in set *V*. If (*G, P*) satisfies the Markov condition, then (*G, P*) is called a Bayesian network (BN). Together with the graph structure, the joint probability distribution of the domain can be decomposed into a product of local conditional probability distributions according to Equation, and each conditional probability distribution involves a node and its parents only.

$$P(x_1, x_2, \dots x_n) = \prod_{i=1}^{\pi} P(x_i | Pa(x_i)) \qquad [2]$$

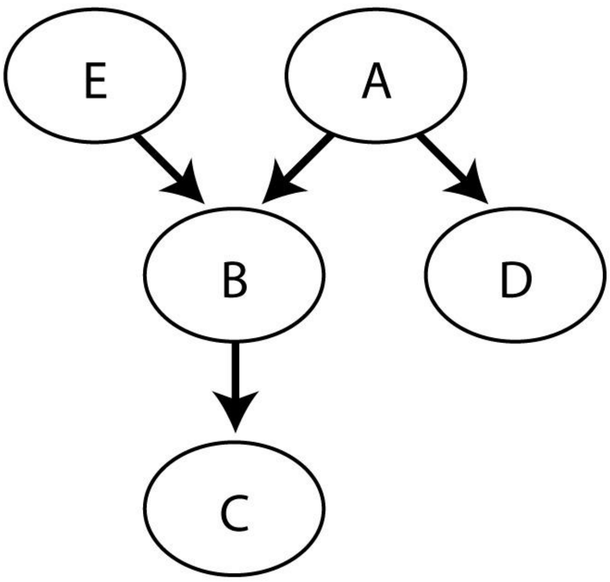

Figure 2.4 Bayesian network structure
https://stke.sciencemag.org/content/2005/281/pl4.full

Figure above is an example of a simple Bayesian network structure. This network structure implies several conditional independence statements: $(A \perp E)$, $(B \perp D|A,E)$, $(C \perp A,D,E|B)$, $(D \perp B,C,E|A)$, and $(E \perp A,D)$. The joint distribution has the product form $P(A,B,C,D,E) = P(A)P(E)P(B|A,E)P(C|B)P(D|A)$.

2.4.4 Previous study that used Bayesian Networks

A Bayesian networks is a strong probabilistic inference model and it consists of two main parts. First part is a graphical structure specifying a set of relationships of dependence and independence between its variables. Another part is a set of distributions of conditional probability quantifying the strengths of the relationships (Jackson & Mosleh, 2016).

The problem of the trader is the simpler scenario, which initially led (Bendtsen & Peña, 2016) to define GBNs. The fact that the restrictions on buying and selling places must be captured by the model and those individual BNs cannot encode these restrictions. Assume that a trader wants to buy shares of a company if it is believed that the share price will increase if this company has a positive economic climate. If the trader owns the company's shares, the trader wants to sell the shares if he believes that the share price will decrease if this company has a negative economic climate. The problem of the trader is to decide when to move between the two phases of the purchase and sale of shares so that it benefits the trader. The general problem solved by portfolio building, such as the universal portfolio or the portfolio of Markowitz, is the allocation of resources to several assets. Therefore, the problem for the study is a little different, because it only considers one asset. The need to switch between different BNs was the basis for the probabilistic graphic model that called gated Bayesian Networks (GBNs). The scenario can be modelled using the gated Bayesian Network (GBN).

A gated Bayesian Network is a probabilistic graphic model that combines several Bayesian Networks with objects called gates to model processes with different phases (Bendtsen & Peña, 2016). These gates allow the different BNs to be activated and deactivated in the model. Inference is conducted in the currently active BNs and therefore they participate in the current phase.

The results show that GBNs have consistently reduced the risk with similar or better rewards than the benchmark buy-and-hold, while remaining out of the market for considerable time. The study also discussed how GBNs differ from other existing frameworks, in particular how GBNs do not solve the expected problem of utility maximization, but rather encode a strategy for dealing with a series of evidence, where the strategy is optimized in relation to a certain score.

Next, based on a study by (Wang & Liu, 2018) a novel binary encoding water cycle algorithm is proposed for the first time to address the Bayesian network structures learning problem. In this study, the sea, rivers and streams correspond to the candidate Bayesian network structures. Three heuristic algorithms (BEWCA- BN, BNC-PSO and ABC-B) are capable of finding near-optimal structures. The results show that the BEWCA-BN algorithm is able to identify optimal or near-optimal networks with high k2 scores and small differences in structure.

In addition, (İçen & Ersel, 2019) propose an approach that incorporates the advantages of fuzzy events and fuzzy probabilities, as they have not been used together before in the literature for Bayesian Networks. İçen (2019) intend to demonstrate that this approach can be used to achieve more interpretable outcomes for real-life issues and better represent uncertainty. In addition, more effective use can be made of detailed information in the data. Due to the widespread use of FBNs in various fields such as fault detection, performance testing and safety risk analysis, various fuzzy approaches for BNs are still substantial.

Consequently, the combination of data information and expert opinion is inherent in the BNs. The approach proposed allows us to gain better knowledge from these two sources of information. With Buckley's method-based calculation of fuzzy conditional probabilities with confidence intervals, more detailed information that is buried within the data is revealed instead of a point estimate or just a single interval estimate. In addition, expert opinions are integrated with the interval arithmetic into the fuzzy probabilities by using fuzzy events to better express real-life situations. In a broader perspective, this leads to consideration of BNs.

CHAPTER 3

METHODOLOGY

3.1 Introduction

Water footprint (WF) is an integrated measurement indicator for total water consumption, including green, blue and grey water (Hoekstra & Mekonnen, 2012). The grey element refers to the pollution of water resources and grey water footprint is defined as the volume of fresh water required to dilute the pollutant load. Grey water footprint calculation involved in three Water Treatment Plants which are Sungai Lembing WTP, Bukit Sagu WTP and Bukit Ubi WTP. Grey water footprint scope of calculations only involved water treatment processes from water abstraction to final step filtration before being distributed to all consumers. The purpose of this calculation is to determine the actual number of freshwater used in the process of delivering water supply to consumer. This water treatment plants are abstracting water from Kuantan River Basin before undergo a series of in-line processes. The water quality data at this three water treatment plants was collected to calculate total grey water footprints.

Algorithms are essentially set of instructions for performing a task which produces an output from a given input. As in this case, two algorithms will be used; Artificial Neural Network and Bayesian Networks. ANNs are widely used in non-linear modelling applications because, given sufficient examples, they can primarily model any non-linear function as long as the optimal number of hidden neurons is used along with appropriate training algorithms (Giwa et al., 2016). Bayesian networks combine graph and probability theories to obtain a comprehensible representation of the joint probability distribution. Bayesian networks have been seen as one of the best way to represent causal knowledge and used in reasoning and decision making tasks in uncertain domains.

In this study, the calculation of total grey water footprint and prediction of the trend will used WEKA software and two algorithms which are Artificial Neural Network and Bayesian Networks. The Water Assessment Manual that introduced by Aldaya (2011) will be used to calculate grey water footprint and all formula will be extracted from here.

For prediction, all accounted grey water footprint data will be inserted in WEKA software system. The result from WEKA software will be analysed by graph and the calculation will be using Microsoft Excel to help as an aid of calculating various and many data. The departments that involved in data collection process are Ministry of Health (MOH), Jabatan Pengairan dan Saliran (JPS) Pahang, Pengurusan Air Pahang Berhad (PAIP) and Water Treatment Plant of Sungai Lembing, Bukit Sagu and Bukit Ubi.

3.2 Flow Chart

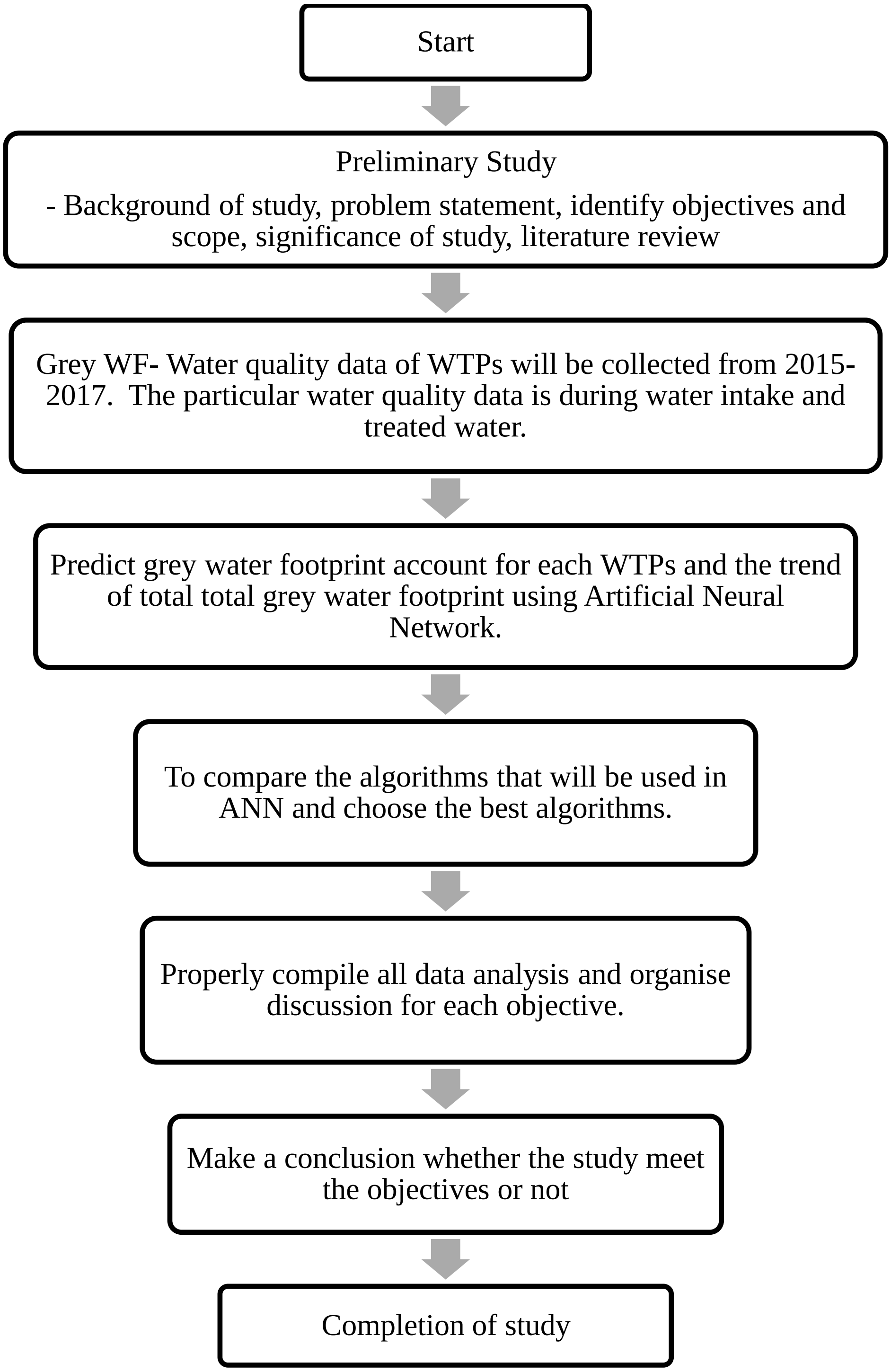

Figure 3.1 Flow chart of Methodology

3.3 Study Area

Kuantan River Basin is in the district of Kuantan at the north eastern end of Pahang State in Peninsular Malaysia. It is one of the most important river basins in Pahang and covers a catchment area of 1630 km2 that started from the forest reserved area of Mukim Ulu Kuantan through agricultural areas, the city of Kuantan (Pahang state capital) to the South China Sea.

Kuantan River basin which is in Kuantan District area has six administrative mukims (small district). In terms of land use, the main types of land use in this district are forest and agriculture that cover approximately 56% and 32% respectively, from the whole area of Kuantan District. The majority of forest areas are in the west or upstream of the Kuantan district. In addition, in Sungai Lembing, on the upstream or low sub-basin area there is an ex- tin mining land. The mining activity began in 1906 and came to an end in 1986 because of the economic recession.

Sungai Lembing Water Treatment Plant supplies treated water for Sungai Lembing and Panching Utara area. Sungai Lembing WTP is located at 3.9337132, 103.0501850. Sungai Lembing WTP is the smallest WTP and supply only to small area population.

Bukit Sagu Water Treatment Plant is located at 3.9111174, 103.1666351, Kampung Kuala Reman, Bukit Goh. Bukit Sagu WTP is sufficient for the bauxite mining industry in the region. In addition to the Panching and Semambu water treatment plants, the Bukit Sagu water treatment plant only provides water to a small area of Felda Bukit Sagu.

Bukit Ubi water treatment plant covers the treated water supply to the Kuala Kuantan commercial area only. Located at 3.8325003, 103.2606332 in the centre of the municipality of Permatang Temesu.

3.4 Data Collection

During this study, many departments has involved for the data collection process. Table 3.1 shows the list of department as well as the data involved in the present study.

Table 3.1 List of Department and data involved in the study

DEPARTMENT	DATA
Pengurusan Air Pahang Berhad (PAIP)	Design plan of WTP
WTP	Intake water and backwash value
Jabatan Pengairan dan Saliran (JPS) Pahang	Rainfall intensity
Malaysian Meteorological Department (MMD)	Temperature
Ministry of Heath	Water Quality Data

3.5 Site Visit

Water treatment plants in Sungai Lembing, Bukit Sagu and Bukit Ubi are the process of converting raw water from Kuantan River to clean water supply to the residence of Pahang. Through the site visit, be able to have a better understanding about the whole process of water treatment plant. The main difference will be the experience gained. During the site visit, be able to see how the water treatment plant actually works and knowing the awareness about the real working environment and the technical skills. All the processes of the water treatment plant are perfectly blend between each other in order to produce quality water to the residence of Pahang. From the site visit data water intake and backwash value, design plan of water treatment plants and information about surrounding area also can be collected.

3.6 Water Supply Treatment Process Identification for WTP

Almost all water sources in the world were contaminated by a variety of physical, chemical and biological parameters, particularly in the industry. The disturbance of clean water and the presence of biological pathogens that are harmful and cannot be removed by boiling it is the reason why water treatment processes need to provide quality water for all uses.

The first step in surface water purification is screening to remove large debris, such as sticks, leaves, trash and other large particles, which may interfere with subsequent cleaning steps. It is necessary to screening for protection of pump, valves, pipe lines, impellers.

After screening, the water is aerated by passing through a series of steps in order to obtain oxygen from the air. This helps to expel soluble gasses such as carbon dioxide and hydrogen sulphide which are acidic that making water less corrosive. Aeration also helps to remove any gaseous organic compounds that could give the water an undesirable taste. Aeration also removes iron or manganese in its insoluble form by oxidation of these substances. Pre-chlorination kills algae and bacteria for long period of contact time. This stage also reduces colour and odour problem and reduces slime formation.

Chemicals called a coagulant with a positive charge are added to the water. The positive charge of these chemicals neutralizes the negative charge of dirt and other dissolved particles in the water. The coagulation chemicals are added in a rapid mix tank, which usually has rotating paddles. When this happens, the particles bind and form larger particles called flocs. Luminium sulphate and ferrous chloride are the most common coagulants used.

Once large flocs are formed, they must be settled, and this takes place in a sedimentation process. During sedimentation, the floc settles down due to its weight to the bottom of the water supply. The coagulation and flocculation water is stored for several hours in the tank for sedimentation. Sludge is the material accumulated in the bottom of the tank; it is removed for disposal.

Once the flocs has settled down to the bottom of the water supply, the clear water above passes through filters of varying compositions such as sand, gravel and charcoal and pore sizes to remove dissolved particles such as dust, parasites, bacteria, viruses and chemicals. To clean the filter, water is quickly passed through the filter in the opposite direction to the normal direction called back flushing or backwashing to remove embedded parts. In this process, clean water and air are pumped up the filter to dislodge the trapped impurities, and if there is one, the water carrying the dirt is pumped into the sewerage system. Alternatively, it can be discharged back into the source river to remove solids after a settlement stage in a sedimentation tank.

Water is disinfected to remove any pathogenic micro- organisms that remain. Once the water has been filtered, a disinfectant can be added to kill any remaining parasites, bacteria and viruses and to protect the water from germs when it is transported to homes and businesses. Chlorine and chloramine are common disinfectants. This stays in the water through the distribution system and protects it from any microorganisms that can enter it until the water reaches the consumers.

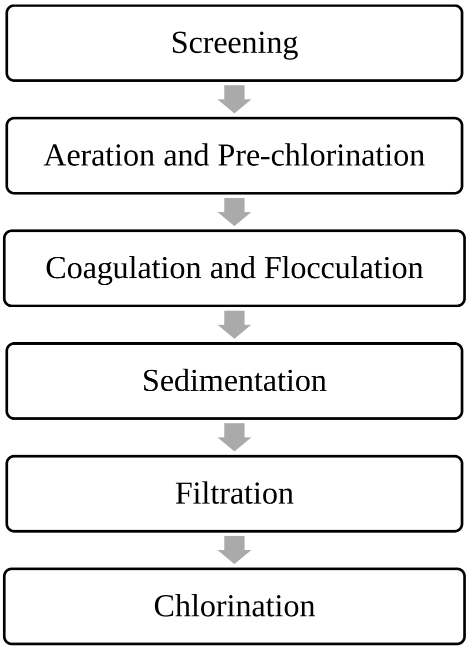

Figure 3.2 Flow chart of water supply treatment process

3.7 Water Footprint Accounting

The grey water footprint is calculated by dividing the pollutant load by the difference between the ambient water quality standard for that pollutant and its natural concentration in the receiving water body. Load of pollutant can be calculated by multiplied flow by concentration of pollutant and coefficient; 86.4. Concentration of nature can be calculated by subtracting concentration of pollutant in water intake by concentration of pollutant in treated water.

$$\mathrm{WF}_{\mathrm{grey}} = \frac{\mathrm{L}}{\mathrm{C}_{\mathrm{max}} - \mathrm{C}_{\mathrm{nat}}} \qquad [3]$$

Where;

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

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

$$C_{max} = maximum\ acceptable\ concentration\ (mass/time)$$

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

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

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

Grey water footprint calculations are performed using ambient water quality standards for the receiving body of freshwater, in other words, standards for maximum permissible concentrations. The reason for this is that the grey water footprint aims to show the required ambient water volume to absorb chemicals. Ambient water quality standards are a specific category of standards for water quality. Other types of standards include water quality standards for drinking water, irrigation quality standards and emission or effluent standards. The ambient water quality standard for a particular substance may vary from one water body to another.

According a study by (Wickramasinghe, Navaratne, & Dias, 2018), the Grey Water Footprint accounting was mainly based on equation adopted from the standard methods given in the Water Footprint Assessment Manual. In general, GWF (volume/time) is calculated through an indirect approach, namely by using an estimate of the leaching coefficient, consistently with the tier 1 method (Allocca, Marzano, Tramontano, & Celico, 2018).

3.8 Pre Processing

In water treatment plants it is required to replace the missing values of the water quality and quantity data to gain knowledge in the system and also to manage the water resources effectively. To use existing operational data as an input to a process simulation model, the missing data should be replaced. In order to provide a more accurate design proposal and system performance, a reasonable and reliable prediction of missing data is very important in determining the correct variability of water treatment plant data.

For pre-processing, the data is arranged based on input and output. For inputs data in this study are total coliform, E-coli, Ammonia, Iron (Fe^2), COD and BOD while the output data is total grey water footprint. Missing data is needed to treat before inserted in software. In this study, missing data is treated by using Mean, Median and Mode Method. In the imputation method of mean, median and mode, all missing values in a particular column are replaced with the mean, median and mode calculated using all available values in that column. It is possible to use appropriate functions in Excel to calculate the mean, median and mode by simply plugging the column range into the function input.

Mean or known as average is the sum of all column values divided by the number of column values. In Excel, the AVERAGE function can be used to compute the mean. Median is the “middle” value amongst the range of values. In Excel, MEDIAN function can be used to compute the median. Mode is the value that occurs the most often in the range of values. In Excel, to compute the mode, the MODE function can be used. In addition,, Last Observation Carried Forward (LOCF) is also used for treating missing data. LOCF is a longitudinal data analysis technique. This is a crude method where a missing value is filled in from the previous stages for a particular row with a value available.

Then, the data can be cleaned by removing the outliers. Outliers are data values that differ significantly from most data sets. These values fall outside of an overall trend in the data. Some difficulty is caused by careful examination of a set of data to look for outliers. However, there are likely to be outliers in any given sample, and it is important to avoid focusing on outliers as opposed to the trends presented overall by the data.

On the other hand, normalise the data. Data normalization is a process where data attributes are organized within a data model to increase entity type's cohesion. In other words, the objective of data normalization is to reduce data redundancy and even eliminate it. Normalization usually means to scale a variable to have a value between 0 and 1. For grey water footprint prediction, value for water intake, water discharge and load of pollutants are needed to normalise in range 0 between 1 as the data sets have different in ranges and values.

3.9 Prediction of Grey Water Footprint

WEKA is a landmark system in the history of data mining and machine learning research communities, because it is the only toolkit that has been adopted so widely and has survived for a long time (Science, 2017). The WEKA is an endemic bird from New Zealand. "WEKA" stands for the Waikato Environment for Knowledge Analysis, developed at Waikato University in New Zealand. WEKA is extensible and has become a collection of machine learning algorithms to solve data mining problems in the real world. It is free software licensed under the GNU General Public License.

WEKA is a collection of algorithms for the learning of data mining. The algorithms can either be directly applied to a dataset or be called by Java Code. WEKA includes data pre- processing, classification, regression, clustering, and rules of association and visualization tools. It is also suitable for developing new learning systems for machines. WEKA is used for research, education, and applications. The tool collects a wide range of data pre- processing tools, learning algorithms and evaluation methods, graphical user interfaces including data visualization and the learning algorithms comparison environment. WEKA is useful for a variety of methods and industries in data mining. The application enables users to identify hidden database systems, user- friendly interfaces and visualization files (Salah, Mocanu, & Florea, 2014).

There are three major implemented schemes in WEKA which are implemented schemes for classification, implemented schemes for numeric prediction and implemented "meta- schemes”. WEKA is easy to use and can be used on multiple levels. The WEKA class library can be accessed from Java's own program and new machines learning algorithms are implemented.

After the total of grey water footprint is calculated, all the data will tabulated using Microsoft Excel. The Excel converted into Common Separated Value (CSV) format will be able to be inserted in WEKA software. For prediction of grey water footprint, all accounted grey water footprint data will be inserted in WEKA software. WEKA software is chosen for prediction because it easy to handle and user-friendly. In addition, the software is free to install and has many algorithms compared to MATLAB software.

3.10 The Best Algorithm for prediction

The Artificial Neural Network (ANN) and Bayesian Networks algorithms will be used as training algorithm for prediction. The data in WEKA software will trend until get the lowest Root Mean Square Error (RMSE). The lower of value RMSE, the lower the error in RMSE is. The best algorithm between Bayesian Networks and ANN can be determined by using the graph of trend. Trend of predicted should be slightly the same with the current data. This is because there are no significant changes in five years interval especially in term of development.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

The purpose of this chapter is to analyse and discuss the result obtained from the study based on three objectives given. The result data will include the total of grey water footprint, the trend prediction of total grey water footprint and the comparison to choose the best algorithm between Bayesian Networks (BN) and Aritificial Neural Network (ANN) in grey water footprint prediction for Sungai Lembing WTP, Bukit Sagu WTP and Bukit Ubi WTP in 2015 until 2017.

4.2 Grey Water Footprint Accounting

In water treatment process at Sungai Lembing, Bukit Sagu and Bukit Ubi WTP, grey water footprint calculations involve only from water abstraction to final step filtration before being distributed to all consumers. For the input parameters that considered were total coliform, E-coli, Ammonia, Fe^2, COD and BOD because the frequency of data availability. Due to the value that can be calculated as meter cubic, among important parameters that not being calculated in this study was turbidity. This is due to the nature of turbidity is measured by using NTU and cannot be assimilated by clean water. From the result data that obtained from the study, the high value of total grey WF is due to high value of concentration Ammonia and Fe^2. The lowest value of total grey WF is also due to the lowest value of concentration in Ammonia and Fe^2. Hence, value of total grey WF is depends on value of concentration in Ammonia and Fe^2.

The high concentration in Ammonia affected total WFGrey could result from fertilization in agriculture, but also from defecation of humans or animals. Its presence in groundwater may be associated with geological factors such as the decay of fossilizing material, but if associated with microbiological compounds, it may also be associated with sewers and livestock farms (Miglietta et al.,2017).

4.2.1 Total Grey Water Footprint for Sungai Lembing Water Treatment Plant

Table 4.1 Total Grey WF in 2015 at Sungai Lembing WTP

MONTH	Total coliform	E.Coli	Ammonia	Fe²	COD	BOD	Total Grey
JAN	0.000532106	0.011155	27.88169	0.549887	0.058655	0.543098	29.04501501
FEB	0.000910897	0.005406	25.18346	0.378824	0.183672	1.020401	26.77267484
MAR	0.000739882	0.003903	0.110441	0.363183	0.183503	1.01946	1.681230058
APR	0.000863353	0.07791	3.366535	4.224281	0.183167	1.017593	8.87034922
MAY	0.002658577	1.203767	188.075	105.2276	0.181436	0.503989	295.1944184
JUN	0.000949331	0.010854	0.284492	2.473703	1.784039	0.550629	5.104666178
JUL	0.000654599	0.184775	174.5102	335.4729	0.380107	1.055852	511.6045131
AUG	0.001049413	0.003442	2.148022	4.311642	0.38504	1.069555	7.918750737
SEPT	0.002269695	0.03003	1.241958	4.908897	0.313258	0.474633	6.971045818
OCT	0.000599588	0.156527	93.41504	157.9153	0.3867	1.074168	252.9483101
NOV	0.001197617	0.007099	0.636569	2.319118	0.376709	1.046415	4.387107945
DEC	0.000931942	0.011304	0.009139	1.624413	0.131598	0.548325	2.325710795

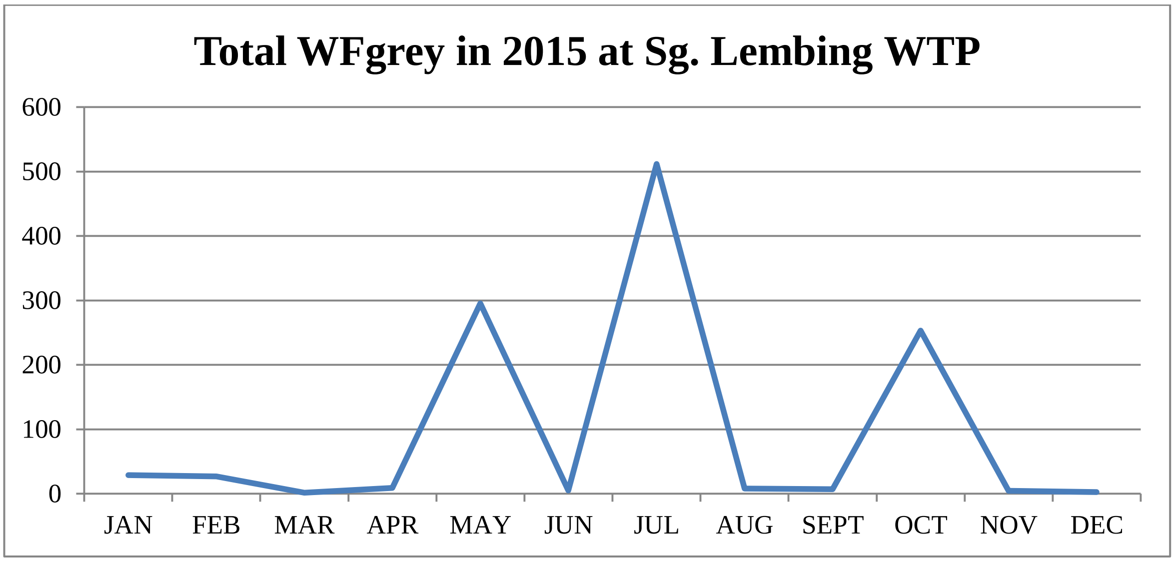

Figure 4.1 Total WFGrey in 2015 at Sg. Lembing WTP

From Table 4.1, in January and February the value of total WFgrey were 29.04501501m³/s and 26.77267484m³/s respectively. In March the value was

1.681230058m³/s and in April was 8.87034922m³/s. 295.1944184m³/s, 5.104666178m³/s, 511.6045131m³/s were the value of total WFgrey in May, June and July. For August, September and October, the values were 7.918750737 m³/s, 6.971045818m³/s and 252.9483101m³/s respectively. While for November and December, the values were 4.387107945m³/s and 2.325710795m³/s. The highest value of total WFgrey for Sg Lembing WTP was in July 2015; 511.6045131m³/s while the lowest value was in March, 1.681230058m³/s. The highest and lowest value of total WFgrey was due to the highest and lowest concentration of Ammonia and Fe^2. The high concentration of ammonia was because of using herbicides and pesticides in agriculture sector. The Figure shows that there was increasing value in May, July and October due to pollutants that happens during that month.

Table 4.2 Total Grey WF in 2016 at Sungai Lembing WTP

MONTH	Total coliform	E.Coli	Ammonia	Fe^2	COD	BOD	Total Grey
JAN	0.001270314	0.004108	0.74317	0.706501	0.057674	0.534015	2.046738007
FEB	0.001656161	0.011794	26.08287	0.577692	0.224036	0.533418	27.4314666
MAR	0.003019473	0.009715	27.88169	0.83203	0.262928	0.576597	29.56597997
APR	0.00280435	0.014607	26.98228	0.721195	0.160637	0.557769	28.43929189
MAY	0.001215537	0.01718	45.3989	62.36478	0.375279	0.539194	108.6965443
JUN	0.002226713	0.036223	26.98228	1.090658	0.377098	0.541807	29.03029256
JUL	0.002442309	0.285782	10.23686	63.98286	0.347874	0.568422	75.42424318
AUG	0.003857587	0.011305	27.88169	1.114303	0.438414	1.107107	30.55667613
SEPT	0.000978703	0.02922	26.98228	1.322981	0.156604	0.543765	29.03582837
OCT	0.004161897	3.785121	12.93061	66.4141	0.415519	0.57711	84.12661727
NOV	0.000860655	0.005464	0.827766	1.561747	0.174741	0.520062	3.090640221
DEC	0.000604963	0.00215	1.021119	2.770786	2.137602	1.619395	7.551657155

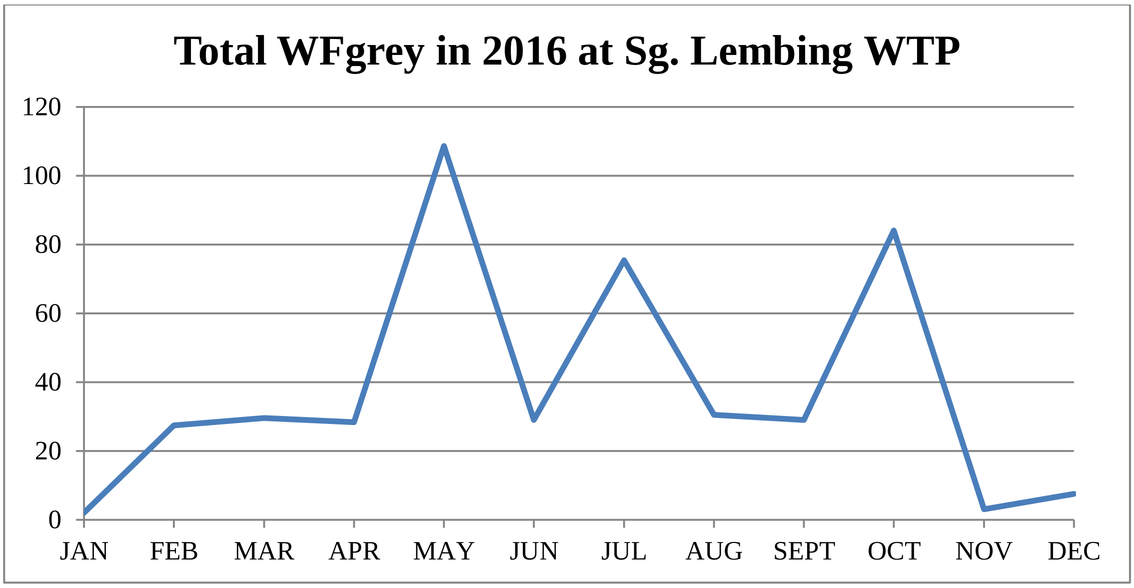

Figure 4.2 Total WFGrey in 2016 at Sg. Lembing WTP

From table above, in January and February the value of total WFgrey were 2.046738007m³/s and 27.4314666m³/s respectively. In March the value was 29.56597997m³/s and in April was 28.43929189m³/s. 108.6965443m³/s, 29.03029256m³/s, 75.42424318m³/s were the value of total WFgrey in May, June and July. For August, September and October, the values were 30.55667613m³/s, 29.03582837m³/s and 84.12661727m³/s respectively. While for November and December, the values were 3.090640221m³/s and 7.551657155m³/s. The highest value of total WFgrey for 2016 was in May with 108.6965443m³/s due to high value concentration of Ammonia and Fe^2 with 45.3989m³/s and 62.36478m³/s respectively. The lowest value of total WFgrey was in January, 2.046738007m³/s due to low concentration in Ammonia and Fe^2. The high concentration of Fe^2 was due to the bauxite mining activities that happen near the river while the concentration Ammonia was due to high agriculture activities. The graph shows that there was increasing value in May, July and October during that year. The graph was quite not uniform.

Table 4.3 Total Grey WF in 2017 at Sungai Lembing WTP

MONTH	Total coliform	E.Coli	Ammonia	Fe²	COD	BOD	Total Grey
JAN	0.000488753	0.005421	0.732508	1.71145	0.05755	0.532868	3.0402861
FEB	0.001212014	0.005639	0.754820	1.510823	0.287614	0.479357	3.0394653
MAR	0.000838439	0.013522	0.888269	0.55788	0.318252	0.53042	2.3091830
APR	0.001711494	0.071922	26.98228	1.48694	0.116556	0.922731	29.582144
MAY	0.002002282	0.553989	126.5118	46.1482	0.393646	1.038789	174.64844
JUN	0.000381454	0.073887	26.98228	1.254786	0.39594	1.099835	29.582144
JUL	0.00062076	0.454749	10.33426	72.0961	0.061697	0.976862	83.924290
AUG	0.001698895	0.003704	0.208557	0.765315	0.300542	0.439389	1.7192058
SEPT	0.00226353	0.003826	0.037407	2.300821	0.246529	1.002149	3.5929956
OCT	0.001510858	0.002206	40.09113	138.5815	0.357033	0.770136	179.80348
NOV	0.001185699	0.002911	0.826998	1.615708	0.345516	0.745292	3.5376117
DEC	0.000443047	0.00403	0.588275	0.723552	0.065689	0.54741	1.9294006

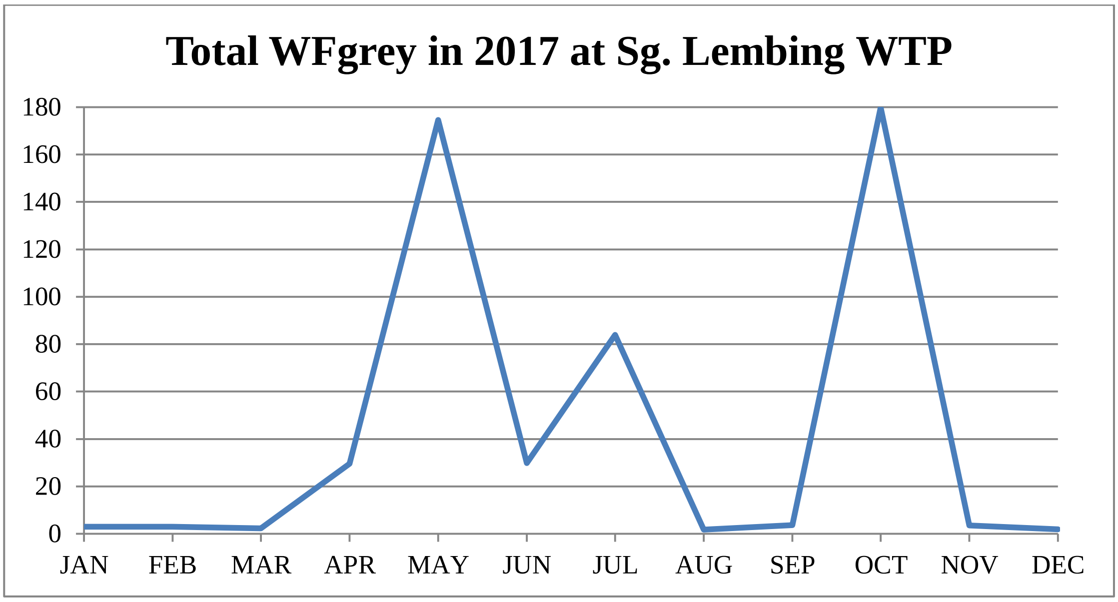

Figure 4.3 Total WFGrey in 2017 at Sg. Lembing WTP

From the table, in January and February the value of total WFgrey were 3.0402861m³/s and 3.0394653m³/s respectively. In March the value was 2.3091830m³/s and in April was 29.582144m³/s. The values of total WFgrey were 174.64844m³/s, 29.582144m³/s, 83.924290m³/s in May, June and July. For August, September and October, the values were 1.7192058m³/s, 3.5929956m³/s and 179.80348m³/s respectively. While for November and December, the values were 3.5376117m³/s and 1.9294006m³/s. The value, 179.8034797 m³/s of total WFgrey in October was the highest value while the value, 1.71920586m³/s was the lowest value of total WFgrey in 2016. This high and low value of total WFgrey was affected by the value concentration of Ammonia and Fe^2. The concentration of Ammonia may cause by the high concentration of herbicides and pesticides that used in agriculture area nearby the river. The graph in Figure 4.3 was not uniform as there was a sharp rise and drop in the value of total WFgrey in 2017 year.

Table 4.4 Total Grey WF in three years at Sungai Lembing WTP

MONTH	2015	2016	2017
JAN	29.04501501	2.04674	3.040286095
FEB	26.77267484	27.4315	3.039465254
MAR	1.681230058	29.566	2.309182988
APR	8.87034922	28.4393	29.58214402
MAY	295.1944184	108.697	174.6484402
JUN	5.104666178	29.0303	29.80710991
JUL	511.6045131	75.4242	83.92428991
AUG	7.918750737	30.5567	1.719205786
SEPT	6.971045818	29.0358	3.592995619
OCT	252.9483101	84.1266	179.8034797
NOV	4.387107945	3.09064	3.537611699
DEC	2.325710795	7.55166	1.929400579

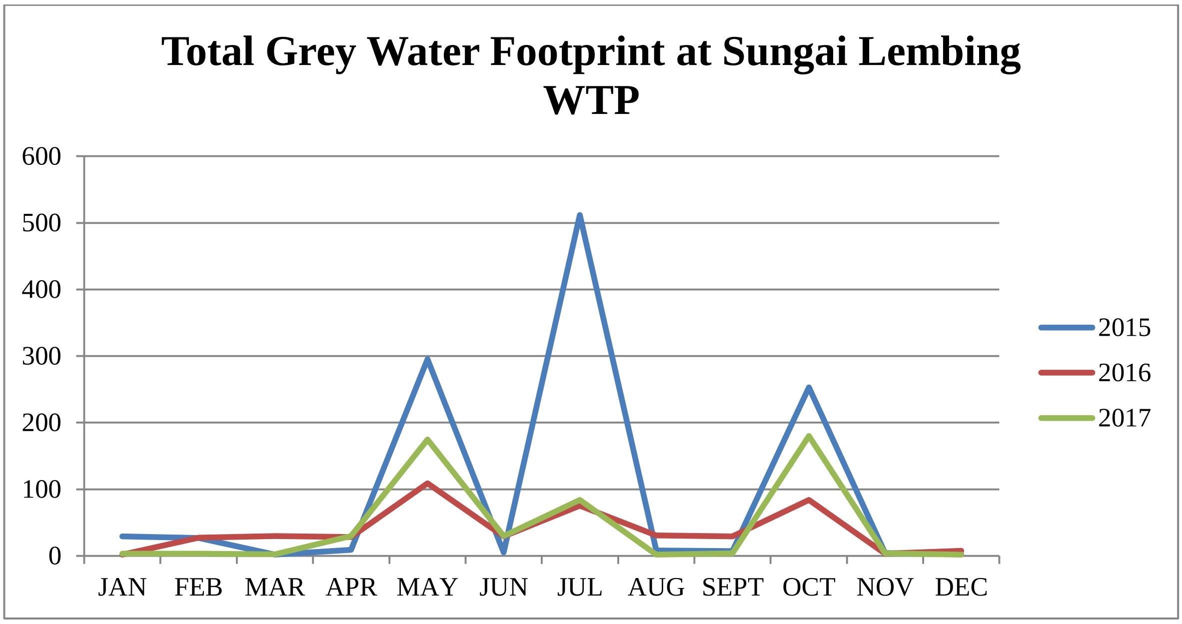

Figure 4.4 Total WFGrey in 2015 - 2017 at Sg. Lembing WTP

From Figure 4.4, in 2015, the highest value of total grey water footprint was in July, 511.6045131 m³/s and also the highest value among three years due to high of concentration in ammonia and Fe^2. The lowest value of total grey WF was in March 2015 which 1.681230058m³/s due to low concentration in ammonia and Fe². The graphs for all years were not uniform as there were sharply increase and decrease in the total of WFgrey.

4.2.2 Total Grey Water Footprint for Bukit Sagu Water Treatment Plant

Table 4.5 Total Grey WF in 2015 at Bukit Sagu WTP

MONTH	Total coliform	E.Coli	Ammonia	Fe²	COD	BOD	Total Grey
JAN	0.000908052	0.005978	21.67376828	1.754529	0.745091	0.611867	24.79214228
FEB	0.000528653	0.003994	19.57630683	2.175979	0.2408	0.501667	22.49927512
MAR	0.000597878	0.001876	0.754118056	0.723744	0.201098	0.837909	2.519343136
APR	0.000364607	0.014797	0.152800926	1.084341	0.130972	0.218287	1.601562628
MAY	0.002065605	0.404199	10.25875496	92.30301	0.463063	0.551265	103.9823604
JUN	0.002193742	0.159375	20.97461447	1.872798	1.133372	0.497093	24.63944592
JUL	0.003641473	0.483953	42.60885417	343.6182	0.454482	1.26245	388.4315648
AUG	0.003101186	0.015973	21.67376828	5.105693	0.352539	0.979275	28.13034969
SEP	0.004008399	0.05257	20.97461447	2.639894	0.479111	0.415895	24.56609284
OCT	0.005786499	9.582182	12.77714616	187.9307	0.352982	1.470757	212.1195771
NOV	0.005586878	0.121443	1.204461017	2.700627	0.542007	1.505576	6.079702306
DEC	0.004066122	0.020865	21.67376828	2.431338	0.98325	0.630289	25.74357677

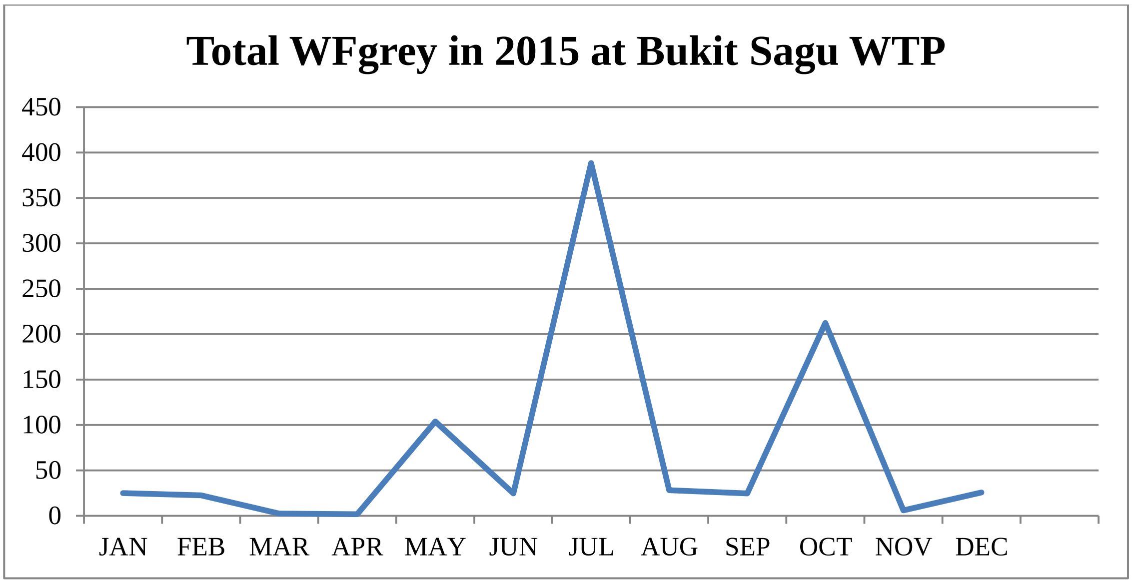

Figure 4.5 Total WFGrey in 2015 at Bukit Sagu WTP

From Table 4.5, in January and February the value of total WFgrey were 24.79214228m³/s and 22.49927512m³/s respectively. In March the value was 2.519343136m³/s and in April was 1.601562628m³/s. The values of total WFgrey were 103.9823604m³/s, 24.63944592m³/s, 388.4315648m³/s in May, June and July. For August, September and October, the values were 28.13034969m³/s, 24.56609284m³/s and 212.1195771m³/s respectively. While for November and December, the values were 6.079702306m³/s and 25.74357677m³/s. The value 388.4315648m³/s was the highest value of total WFgrey in July 2015 while 1.601562628m³/s was the lowest value of total WFgrey in April 2015. The different value of total WFrey was depends on concentration of Ammonia and Fe^2. The graph in Figure 4.5 was not uniform as there was a dramatic rise and drop in July and August. The high value of total WFgrey was depends on high concentration of Ammonia and Fe^2 that high in pollutants in that month.

Table 4.6 Total Grey WF in 2016 at Bukit Sagu WTP

Month	**Total coliform**	**E.Coli**	**Ammonia**	**Fe^2**	**COD**	**BOD**	**Total Grey**
JAN	0.001907987	0.006341	2.150257137	2.723824	0.704937	0.743605	6.330872557
FEB	0.001565113	0.008055	1.062992012	2.396542	0.222338	0.71262	4.404111733
MAR	0.001085524	0.006521	21.67376828	1.219759	0.066432	0.615108	23.5826737
APR	0.002005459	0.044042	20.97461447	2.395139	0.368129	0.766935	24.55086494
MAY	0.004668472	3.822526	10.04422259	82.60779	0.42834	0.759468	97.6670198
JUN	0.002064106	0.024743	20.97461447	1.674785	0.48736	0.688362	23.8519288
JUL	0.00247221	2.412126	15.55277005	111.8069	1.497473	0.656786	131.9284894
AUG	0.0024346	0.009853	0.709782373	4.236656	1.143593	1.191243	7.293562593
SEPT	0.001460807	0.036999	20.97461447	5.424616	2.20714	0.634235	29.27906567
OCT	0.002218957	0.097415	42.53228877	167.7434	0.505398	0.701941	211.5826921
NOV	0.003892604	0.013389	20.97461447	41.97825	1.534577	0.639407	65.1441311
DEC	0.00445011	0.015832	2.814318988	2.455684	1.438299	1.410097	8.138681596

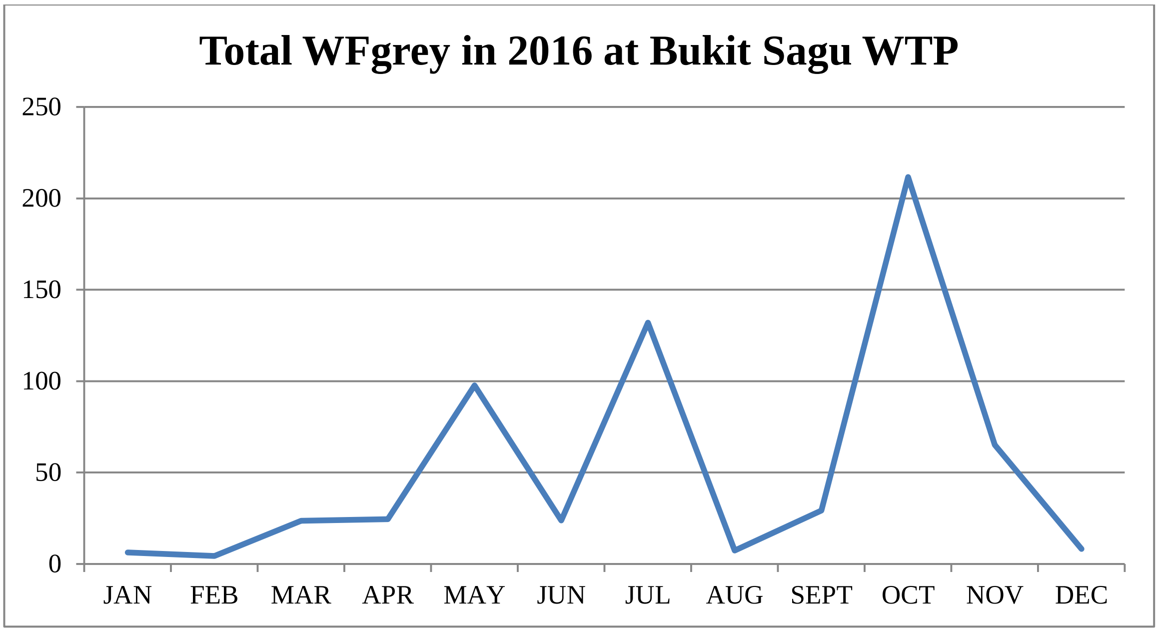

Figure 4.6 Total WFGrey in 2016 at Bukit Sagu WTP

From the table above, in January and February the value of total WFgrey were 6.330872557m³/s and 4.404111733m³/s respectively. In March the value was 23.5826737m³/s and in April was 24.55086494m³/s. The values of total WFgrey were 97.6670198m³/s, 23.8519288m³/s, 131.9284894m³/s in May, June and July. For August, September and October, the values were 7.293562593m³/s, 29.27906567m³/s and 211.5826921m³/s respectively. While for November and December, the values were 65.1441311m³/s and 8.138681596m³/s. The highest value of total WFgrey in 2016 at Bukit Sagu WTP is 211.582692m³/s in October due to high value of concentration in Ammonia and Fe^2, 42.53228877m³/s and 167.7434m³/s respectively. For the lowest value of total WFgrey is in February which is 4.404111733m³/s. The high concentration of Ammonia was produced by bacteria in water and soil as an end product of plant and animal waste decomposition. The graph is not uniform as there was an increasing value of total WFgrey in May, July and October in the year.

Table 4.7 Total Grey WF in 2016 at Bukit Sagu WTP

Month	Total coliform	E.Coli	Ammonia	Fe^2	COD	BOD	Total Grey
JAN	0.003212033	0.00415	0.670344237	5.40715	1.155466	0.687777	7.928100536
FEB	0.001294139	0.001786	0.10328588	1.86389	0.141931	1.123619	3.235805871
MAR	0.002630299	0.020356	0.245047538	4.88703	0.753484	0.654066	6.562614051
APR	0.002927732	0.037305	4.168369698	2.711598	0.159437	1.262206	8.341842777
MAY	0.002193677	0.035416	210.1236138	116.6624	0.958279	1.37934	329.1612278
JUN	0.003662547	0.029061	20.97461447	3.612091	0.403452	0.672421	25.69530173
JUL	0.004228956	0.023608	61.48767881	201.7915	2.258826	1.375567	266.9414345
AUG	0.002687827	0.013995	21.67376828	3.17324	1.391601	0.682157	26.93744906
SEPT	0.005146865	0.118211	20.97461447	3.813533	0.499686	0.71794	26.12913112
OCT	0.007479607	0.023585	45.26924989	111.7977	0.745091	0.844999	158.6880814
NOV	0.004653155	0.056792	4.088696166	3.265229	0.721056	0.817741	8.954167485
DEC	0.004706832	0.016919	1.751464568	3.615867	0.471483	0.654838	6.51527948

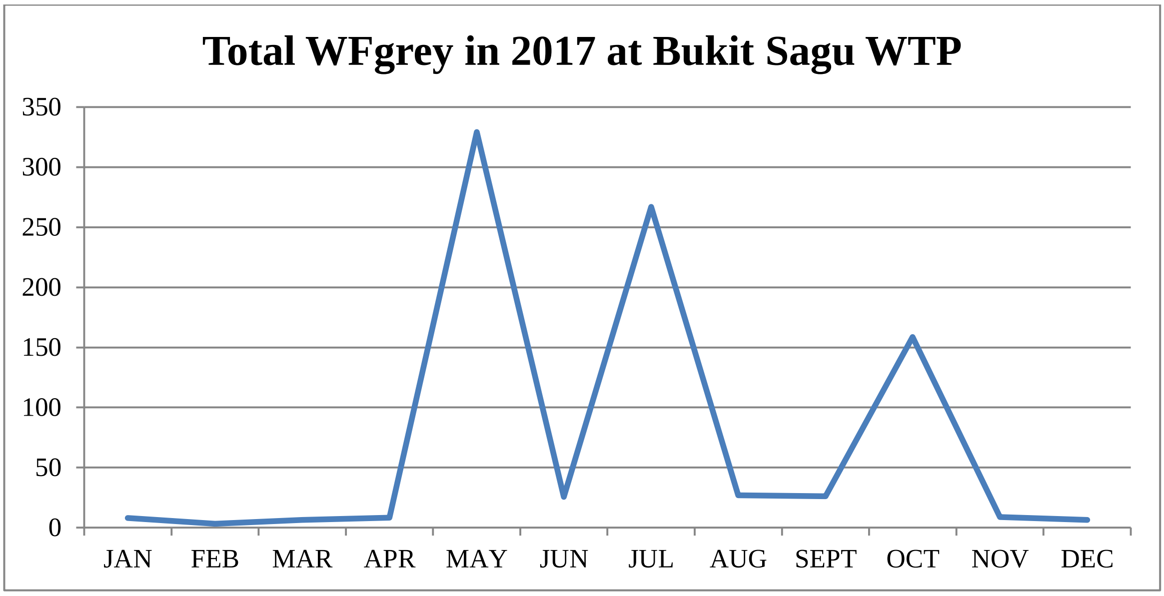

Figure 4.7 Total WFGrey in 2017 at Bukit Sagu WTP

From Table 4.7, for 2017, the value of total WFgrey in January and February were 7.928100536m³/s and 3.235805871m³/s respectively. In March the value was 6.562614051m³/s and in April was 8.341842777m³/s. The values of total WFgrey were 329.1612278m³/s, 25.69530173m³/s, 266.9414345m³/s in May, June and July. For August, September and October, the values were 26.93744906m³/s, 26.12913112m³/s and 158.6880814m³/s respectively. While for November and December, the values were 8.954167485m³/s and 6.51527948m³/s. The value 329.1612278m³/s was the highest value of total WFgrey in May 2017 while 3.235805871m³/s was the lowest value of total WFgrey in February 2017. The high and low value of total WFrey was depends on concentration of Ammonia and Fe^2. The high concentration of Fe^2 was due to bauxite mining activities that occur at the nearby river. The graph in Figure 4.7 was not uniform as there was a dramatic rise and drop in May and June.

Table 4.8 Total Grey WF in three years at Bukit Sagu WTP

MONTH	2015	2016	2017
JAN	24.79214228	6.330872557	7.928100536
FEB	22.49927512	4.404111733	3.235805871
MAR	2.519343136	23.5826737	6.562614051
APR	1.601562628	24.55086494	8.341842777
MAY	103.9823604	97.6670198	329.1612278
JUN	24.63944592	23.8519288	25.69530173
JUL	388.4315648	131.9284894	266.9414345
AUG	28.13034969	7.293562593	26.93744906
SEP	24.56609284	29.27906567	26.12913112
OCT	212.1195771	211.5826921	158.6880814
NOV	6.079702306	65.1441311	8.954167485
DEC	25.74357677	8.138681596	6.51527948

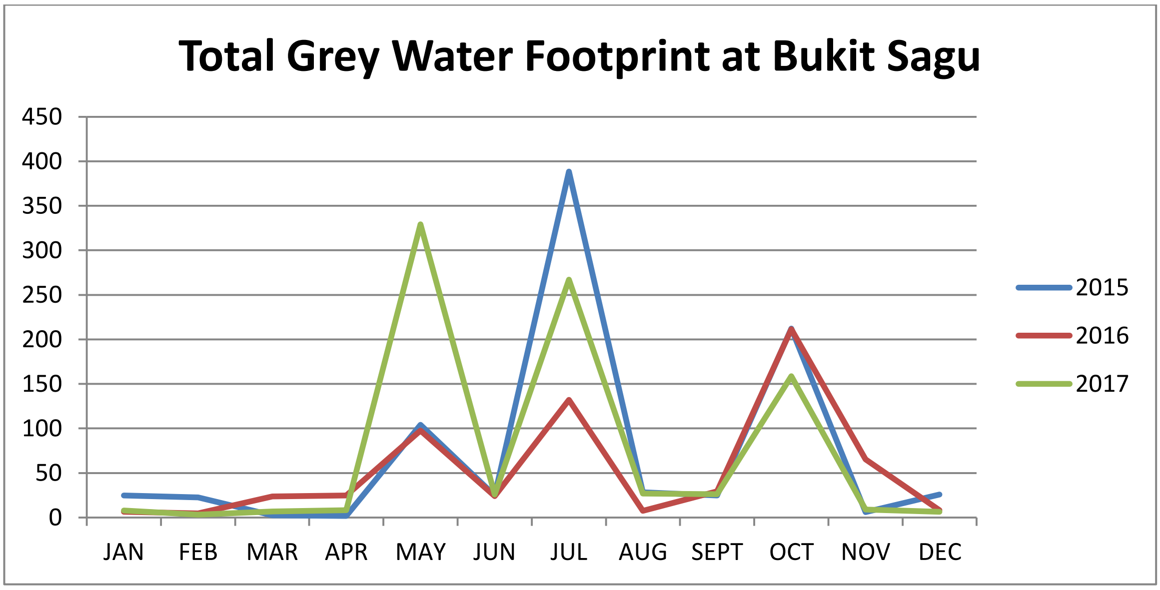

Figure 4.8 Total WFGrey in 2015 - 2017 at Bukit Sagu WTP

From Figure 4.8, the graph shows that there has been a slight increase and the highest value of total grey WF which 388.4315648m³/s in July 2015. The highest value of total grey WF was affected by high of concentration in Ammonia and Fe^2. In July 2015 was the highest value of total WFgrey among the three years. In 2015, the lowest value of total grey WF was 1.601562628m³/s on April due to low concentration of Ammonia and Fe^2. In April 2015 also was the lowest value of total WFgrey among three years. The graphs for all years were not uniform as there were dramatically declined in the value of total WFgrey.

4.2.3 Total Grey Water Footprint for Bukit Ubi Water Treatment Plant

Table 4.9 Total Grey WF in 2015 at Bukit Ubi WTP

Month	Total coliform	E.Coli	Ammonia	Fe^2	COD	BOD	Total Grey
JAN	0.0016704	0.03924	5.538181	5.64557	2.66045	6.33442	20.2195248
FEB	0.000982	0.00247	5.002228	6.79129	1.12538	6.25204	19.1743804
MAR	0.0001491	0.00198	5.538181	8.07773	1.64818	6.86737	22.1335911
APR	0.009747	0.00614	5.35953	12.2574	1.60206	3.33762	22.5637221
MAY	0.0045078	0.15908	5.53817	10.4709	2.4787	6.88529	22.5366688
JUN	0.002124	0.00759	0.83694	3.96405	3.65208	3.04341	11.5061989
JUL	0.0010076	0.00695	3.434831	29.0843	2.30054	6.3904	41.2180596
AUG	0.0030142	0.0164	5.538181	19.653	2.40138	6.67049	34.2824402
SEPT	0.007002	0.05433	1.40541	9.85902	0.62778	2.75343	14.706969
OCT	0.0208305	0.01312	5.538181	14.0572	0.71467	5.65787	25.9918435
NOV	0.0048616	0.01439	3.91263	9.03507	1.84131	4.10625	18.914507
DEC	0.002855	0.00801	3.511835	9.69231	1.19691	2.90135	17.3132629

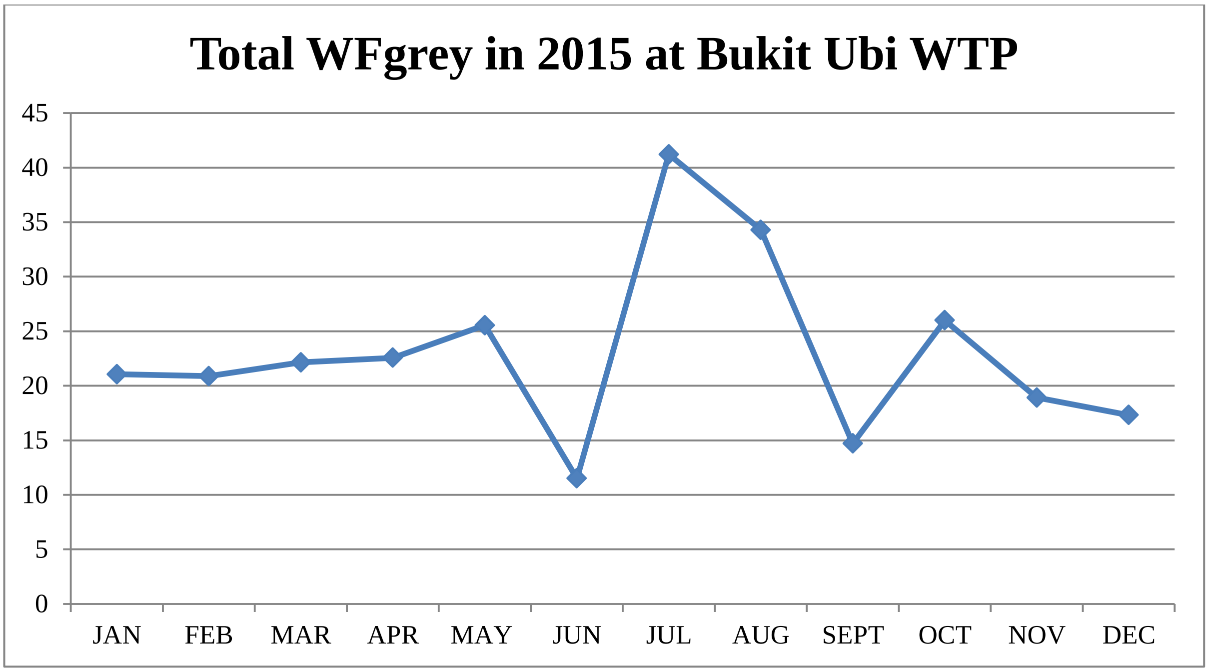

Figure 4.9 Total WFGrey in 2015 at Bukit Ubi WTP

From Table 4.9, the value of total WFgrey in January and February were 20.2195248m³/s and 19.1743804m³/s respectively. In March the value was 22.1335911m³/s and in April was 22.5637221m³/s. The values of total WFgrey were 22.5366688m³/s, 11.5061989m³/s, 41.2180596m³/s in May, June and July. For August, September and October, the values were 34.2824402m³/s, 14.706969m³/s and 25.9918435m³/s respectively. While for November and December, the values were 18.914507m³/s and 17.3132629m³/s. The highest value of total WFgrey for Bukit Ubi WTP was in July 2015 which is 41.2180596m³/s while the lowest value was in June, 11.5061989 m³/s. The highest and lowest value of total WFgrey was due to the highest and lowest concentration of Ammonia and Fe^{2}.^{2}. The high concentration of ammonia was because of using herbicides and pesticides in agriculture sector. The high concentration of Fe^{2} was due to the bauxite mining activities that happen near the river. The figure shows that there was sharp rise value in July from the drop value in June due to different concentration of pollutant in that month.

Table 4.10 Total Grey WF in 2016 at Bukit Ubi WTP

Month	Total coliform	E.Coli	Ammonia	Fe²	COD	BOD	Total Grey
JAN	0.0029427	0.0151447	10.519385	4.786431	0.692664	2.886286	18.9028534
FEB	0.0023189	0.0067826	12.396543	5.456379	1.680579	5.386692	24.9292945
MAR	0.0005337	0.0024677	5.538181	7.533031	1.515497	6.314824	20.9045344
APR	0.0006491	0.0021708	5.35953	5.85558	0.98232	2.72883	14.9290799
MAY	0.0020789	0.0086314	5.538181	9.811686	1.214983	2.736525	19.3120853
JUN	0.0012602	0.0118152	5.35953	9.11292	2.26845	2.5206	19.2745754
JUL	0.0019302	0.0074942	5.538181	12.92173	1.539522	2.618384	22.6272414
AUG	0.0071567	0.012936	2.235069	12.5537	1.901664	5.560625	22.2711487
SEPT	0.00309	0.0065866	5.35953	6.82011	8.36973	2.40522	22.9642666
OCT	0.013833	0.38513	5.538181	9.286701	2.880365	2.526748	20.630958
NOV	0.0191648	0.046829	5.35953	9.23883	3.11469	2.35971	20.1387538
DEC	0.0023796	0.013337	5.538181	7.529745	4.225827	8.804155	26.1136246

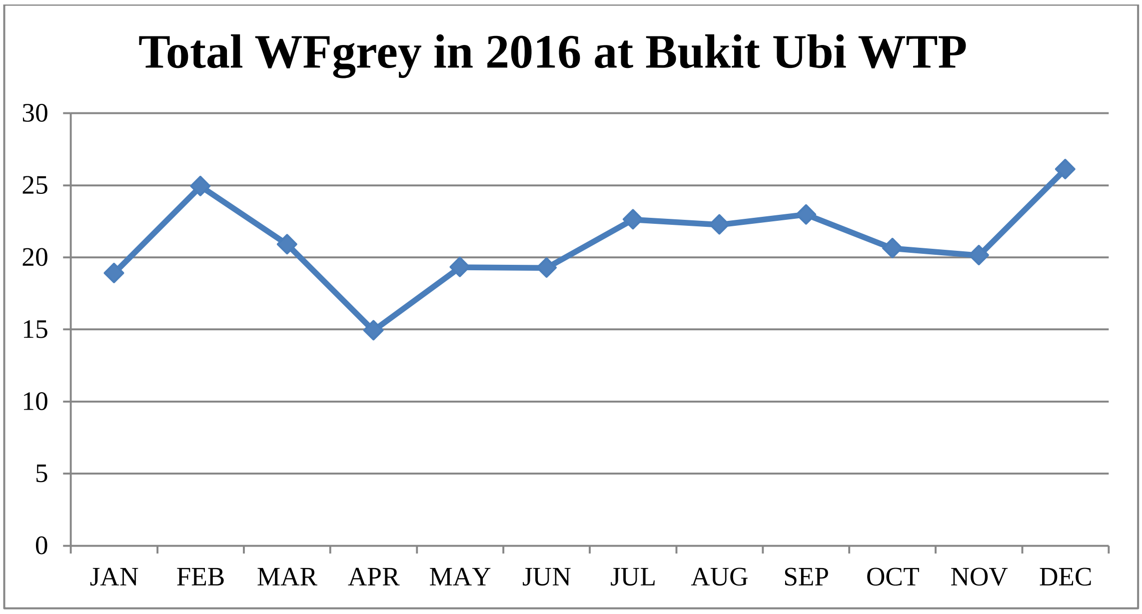

Figure 4.10 Total WFGrey in 2016 at Bukit Ubi WTP

From the table above, the value of total WFgrey in January and February were 18.9028534m³/s and 24.9292945m³/s respectively. In March the value was 20.9045344m³/s and in April was 14.9290799m³/s. The values of total WFgrey were 19.3120853m³/s, 19.2745754m³/s, 22.6272414m³/s in May, June and July. For August, September and October, the values were 22.2711487m³/s, 22.9642666m³/s and 20.630958m³/s respectively. While for November and December, the values were 20.1387538m³/s and 26.1136246m³/s. The highest value of total WFgrey in 2016 was in

December with 26.1136246m³/s due to high value concentration of COD and BOD which 4.225827m³/s and 8.804155m³/s respectively. The lowest value of total WFgrey was in April, 14.9290799m³/s due to low concentration in COD. The graph shows that there was steady decrease in value of May and the graph was seems like uniform.

Table 4.11 Total Grey WF in 2017 at Bukit Ubi WTP

Month	Total coliform	E.Coli	Ammonia	Fe^2	COD	BOD	Total Grey
JAN	0.0050618	0.01336	5.538181	9.80518	3.10403	2.5867	21.0525148
FEB	0.0083177	0.0508	5.002228	9.36824	2.09121	4.35669	20.8774857
MAR	0.0037224	0.01656	2.051549	8.38082	2.64287	2.53149	15.6270164
APR	0.004373	0.0299	11.42229	9.30681	1.95327	5.42574	28.142386
MAY	0.0025568	0.01684	2.929097	16.6464	3.79849	5.01189	28.4052898
JUN	0.0061655	0.06722	5.35953	10.4874	1.44213	3.00444	20.3668805
JUL	0.005718	0.04882	5.538181	11.1226	0.70423	5.57504	22.994566
AUG	0.005289	0.02477	5.538181	10.4775	1.6435	2.40275	20.091992
SEP	0.007002	0.05433	1.40541	9.85902	0.62778	2.75343	14.706969
OCT	0.0108305	0.01312	5.538181	14.0572	0.71467	5.65787	25.9918435
NOV	0.0048616	0.01439	3.91263	9.03507	1.84131	4.10625	18.914507
DEC	0.002855	0.00801	3.511835	9.69231	1.19691	2.90135	17.3132629

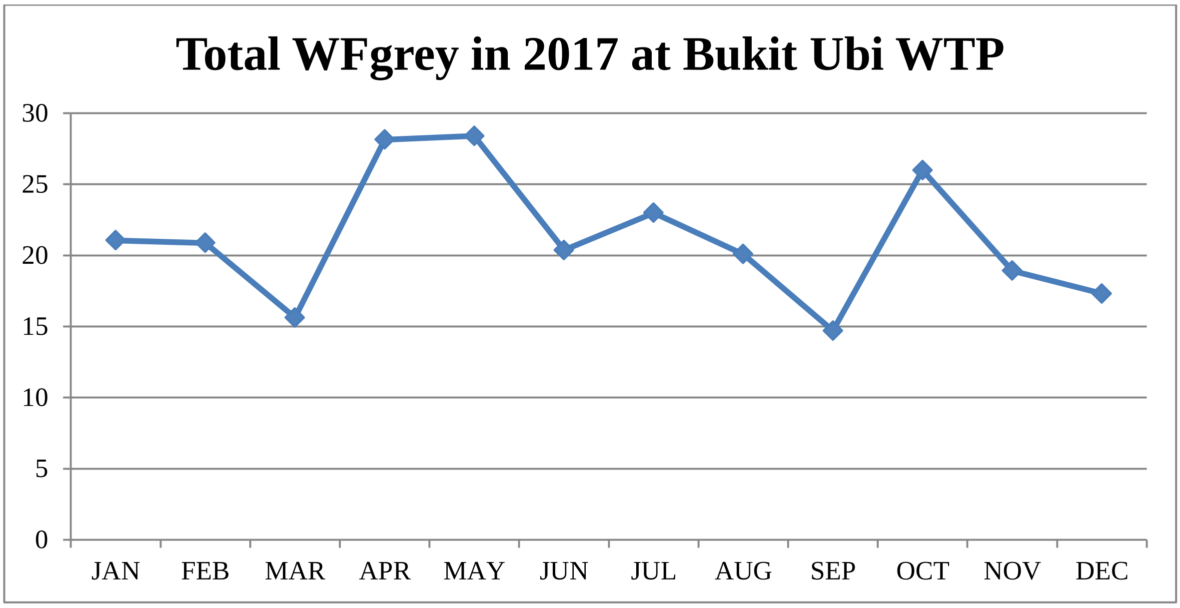

Figure 4.11 Total WFGrey in 2017 at Bukit Ubi WTP

From Table 4.11, the value of total WFgrey in January and February were 21.0525148m³/s and 20.8774857m³/s respectively. In March the value was

15.6270164m³/s and in April was 28.142386m³/s. The values of total WFgrey were 28.4052898m³/s, 20.3668805m³/s, 22.994566m³/s in May, June and July. For August, September and October, the values were 20.091992m³/s, 14.706969m³/s and 25.9918435m³/s respectively. While for November and December, the values were 18.914507m³/s and 17.3132629m³/s. The value 28.4052898m³/s was the highest value of total WFgrey in May 2017 while 14.706969m³/s was the lowest value of total WFgrey in September 2017. The high and low value of total WFrey was depends on concentration of Fe^{2}. The highest and lowest value of total WFgrey was due to the highest and lowest concentration of Ammonia and Fe^{2}. The high concentration of ammonia was because of using herbicides and pesticides in agriculture sector. The graph in Figure 4.11 was not uniform as there was a dramatic drop in March and September and rise in April and October.

Table 4.12 Total Grey WF in three years at Bukit Ubi WTP

MONTH	2015	2016	2017
JAN	21.05251	18.90285	21.05251
FEB	20.87749	24.92929	20.87749
MAR	22.13359	20.90453	15.62702
APR	22.56372	14.92908	28.14239
MAY	25.53667	19.31209	28.40529
JUN	11.5062	19.27458	20.36688
JUL	41.21806	22.62724	22.99457
AUG	34.28244	22.27115	20.09199
SEP	14.70697	22.96427	14.70697
OCT	25.99184	20.63096	25.99184
NOV	18.91451	20.13875	18.91451
DEC	17.31326	26.11362	17.31326

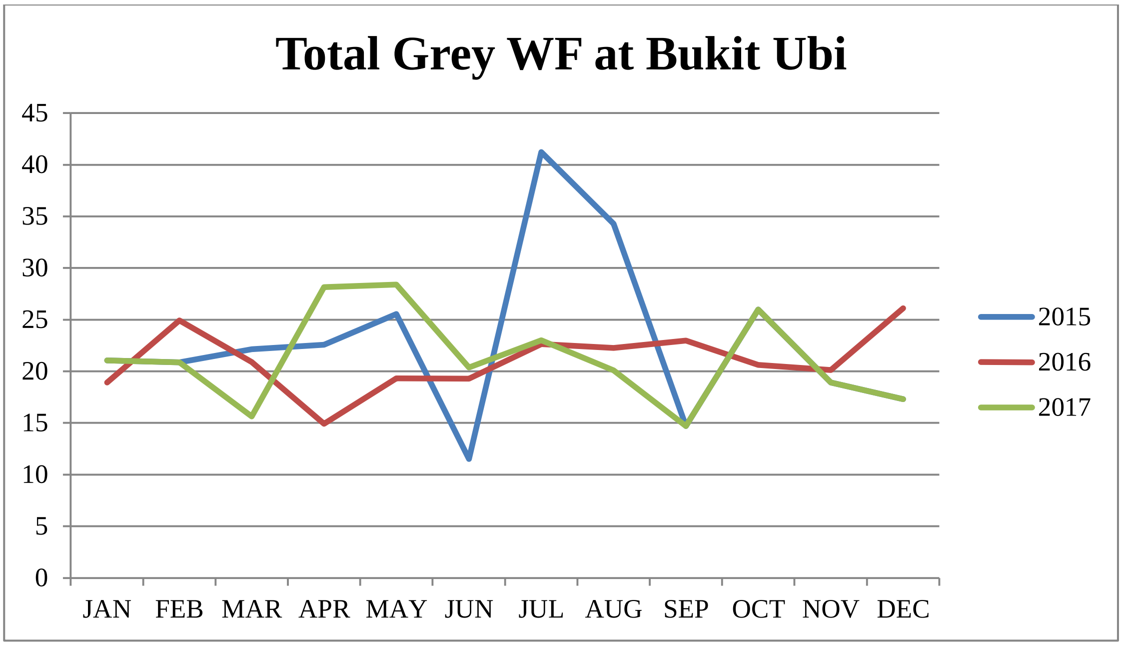

Figure 4.12 Total WFGrey in 2015 - 2017 at Bukit Ubi WTP

From Figure 4.12, for 2015, the graph shows that there has a sharp increase in value of total grey WF in July, 41.21806 m^3/s, from the lowest value of total grey WF in June, 11.5062 m^3/s. The value of total WFgrey in July 2015 was the highest value between among three years. The lowest value of total WFgrey was 11.5062m^3/s which in June 2015 and the lowest value among three years.

4.3 The Best Algorithm in Grey Water Footprint Prediction

4.3.1 Sungai Lembing Water Treatment Plant

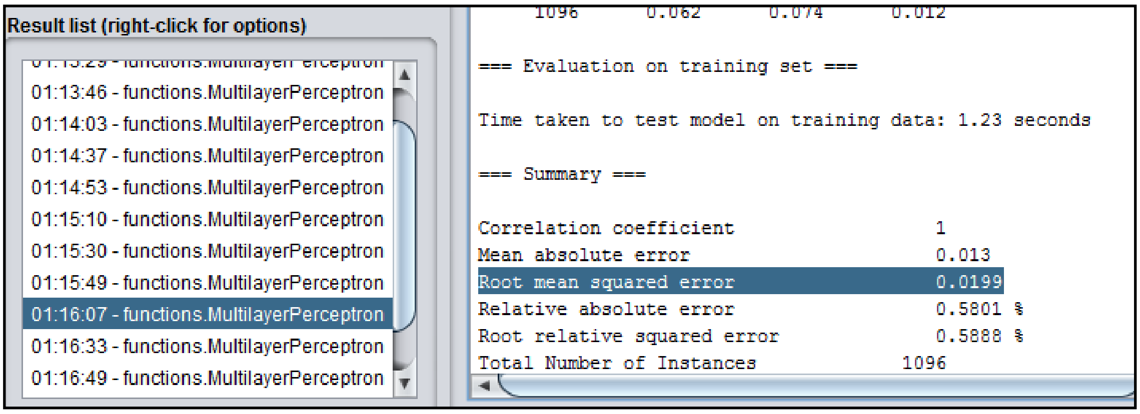

Figure 4.13 RMSE value after training process using ANN

Table 4.13 Values of RMSE with the hidden neurons by using ANN Algorithm

Hidden Neuron	RMSE	Hidden Neuron	RMSE
1	0.0225	**11**	0.0023
2	0.0029	**12**	0.0032
3	0.0027	**13**	0.0017
4	0.0028	**14**	0.0047
5	0.0021	**15**	0.0068
6	0.0022	**16**	0.0064
7	0.0020	**17**	0.0062
8	0.0016	**18**	0.0057
9	0.0021	**19**	0.0052
10	0.0018	**20**	0.0130

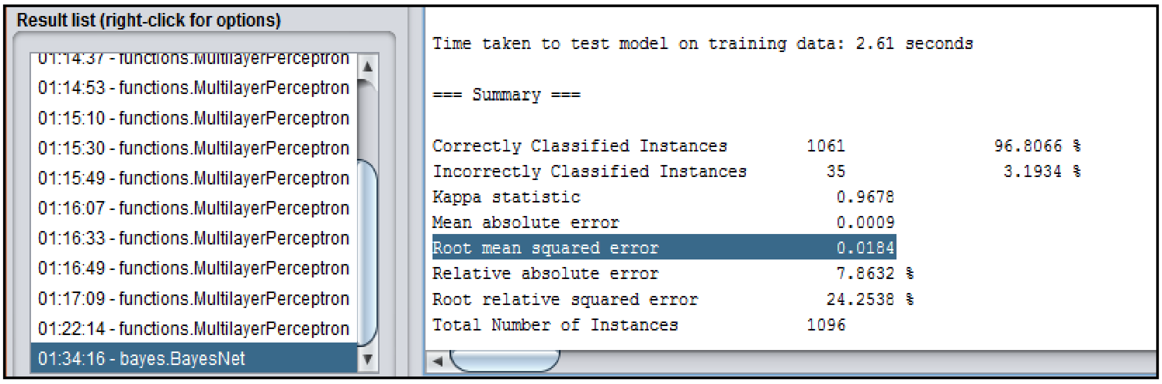

Figure 4.14 RMSE value after training process using BN

Figure 4.13 and 4.14 show the result after training process using Artificial Neural Network (ANN) and Bayesian Networks algorithm. From the figure above, ANN produced the RMSE value was 0.0199 and BN produced RMSE value which 0.0184. The lowest RMSE value indicates the least error that algorithm made.

4.3.2 Bukit Sagu Water Treatment Plant

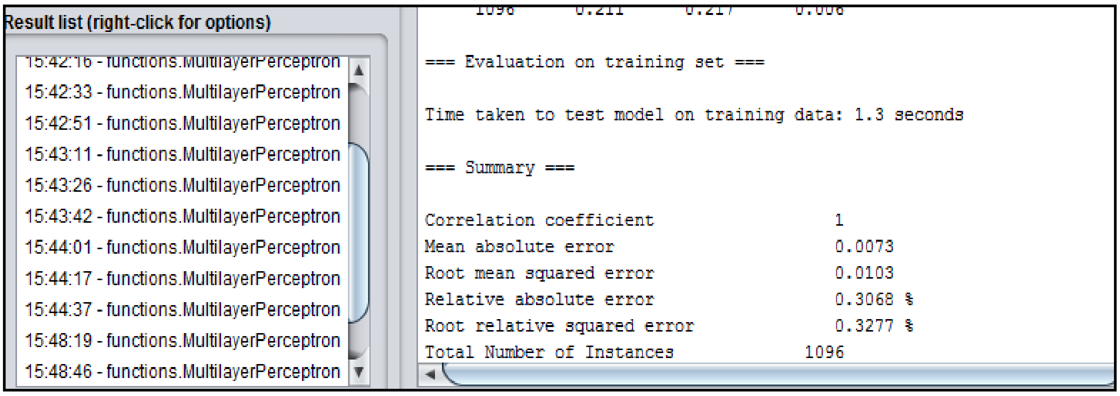

Figure 4.15 RMSE value after training process using ANN

Table 4.14 Values of RMSE with the hidden neurons by using ANN Algorithm

Hidden Neuron	RMSE	Hidden Neuron	RMSE
1	0.0961	**11**	0.0151
2	0.0425	**12**	0.0150
3	0.0195	**13**	0.0164
4	0.0145	**14**	0.0195
5	0.0169	**15**	0.0173
6	0.0151	**16**	0.0172
7	0.0137	**17**	0.0196
8	0.0175	**18**	0.0189
9	0.0172	**19**	0.0123
10	0.0165	**20**	0.0103

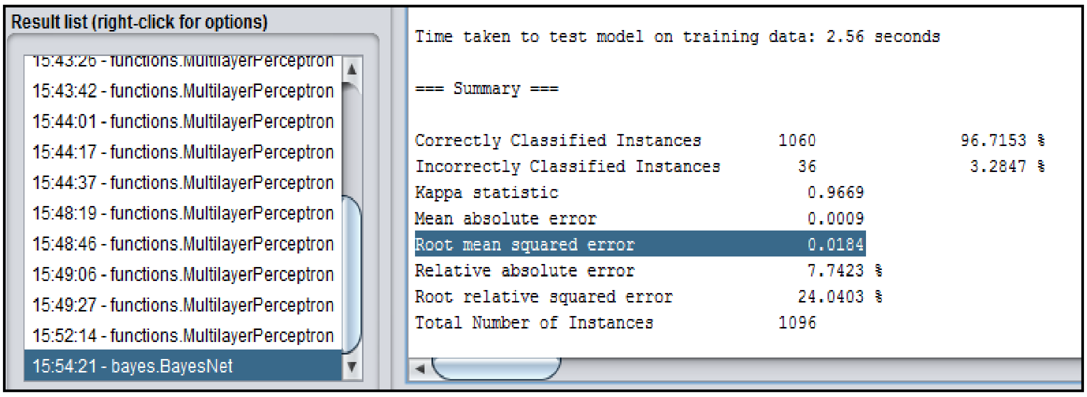

Figure 4.16 RMSE value after training process using BN

Figure 4.15 and 4.16 show the result after training process using Artificial Neural Network (ANN) and Bayesian Networks algorithm. From the figure above, ANN produced the RMSE value was 0.0103 while BN produced RMSE value was 0.0184. Based on the result, due to the lowest RMSE value, ANN has been chosen as the best algorithm compared to Bayesian Networks algorithm.

4.3.3 Bukit Ubi Water Treatment Plant

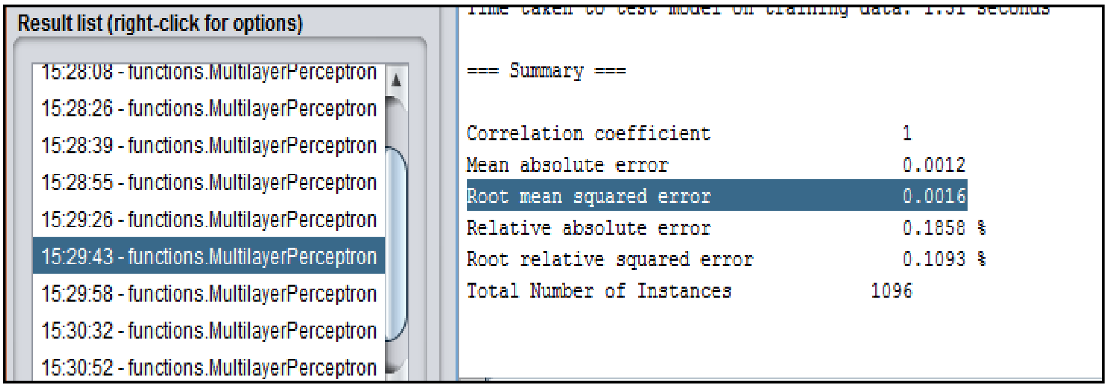

Figure 4.17 RMSE value after training process using ANN

Table 4.15 Values of RMSE with the hidden neurons by using ANN Algorithm

Hidden Neuron	RMSE	Hidden Neuron	RMSE
1	0.0225	**11**	0.0023
2	0.0029	**12**	0.0032
3	0.0027	**13**	0.0017
4	0.0028	**14**	0.0047
5	0.0021	**15**	0.0068
6	0.0022	**16**	0.0064
7	0.0020	**17**	0.0062
8	0.0016	**18**	0.0057
9	0.0021	**19**	0.0052
10	0.0018	**20**	0.0130

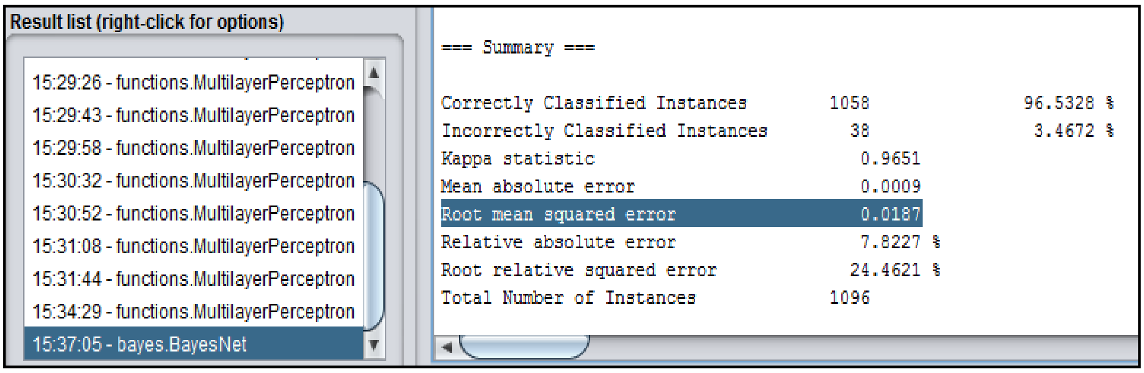

Figure 4.18 RMSE value after training process using BN

Figure 4.17 and 4.18 show the result after training process using Artificial Neural Network (ANN) and Bayesian Networks algorithm. From the figure above, ANN produced the RMSE value which 0.0016 and BN produced RMSE value was 0.0187. Based on the result, due to the lowest RMSE value, ANN has been chosen as the best algorithm compared to Bayesian Networks algorithm.

In order to determine the best algorithm between Artificial Neural Network (ANN) and Bayesian Networks (BN), the results from total of grey WF was trained accordingly to produce the grey water footprint trend. The algorithm with the lowest value of Root Mean Square Error (RMSE) was indicated as the least error and was chosen as best algorithm.

For Sungai Lembing WTP, ANN algorithm RMSE value was 0.0199 while for Bayesian Networks algorithm RMSE value was 0.0184. For Bukit Sagu WTP, value RMSE of ANN algorithm, 0.0103 was the least value compared to Bayesian Networks which 0.0184. For Bukit Ubi WTP, the value RMSE for ANN algorithm was 0.0016 while for Bayesian Networks algorithm was 0.0187. The best algorithm was Artificial Neural Network as the lowest value of RMSE indicates as the least error.

4.4 Prediction of Grey Water Footprint

4.4.1 Sungai Lembing Water Treatment Plants

4.4.1.1 Artificial Neural Network (ANN) Algorithm

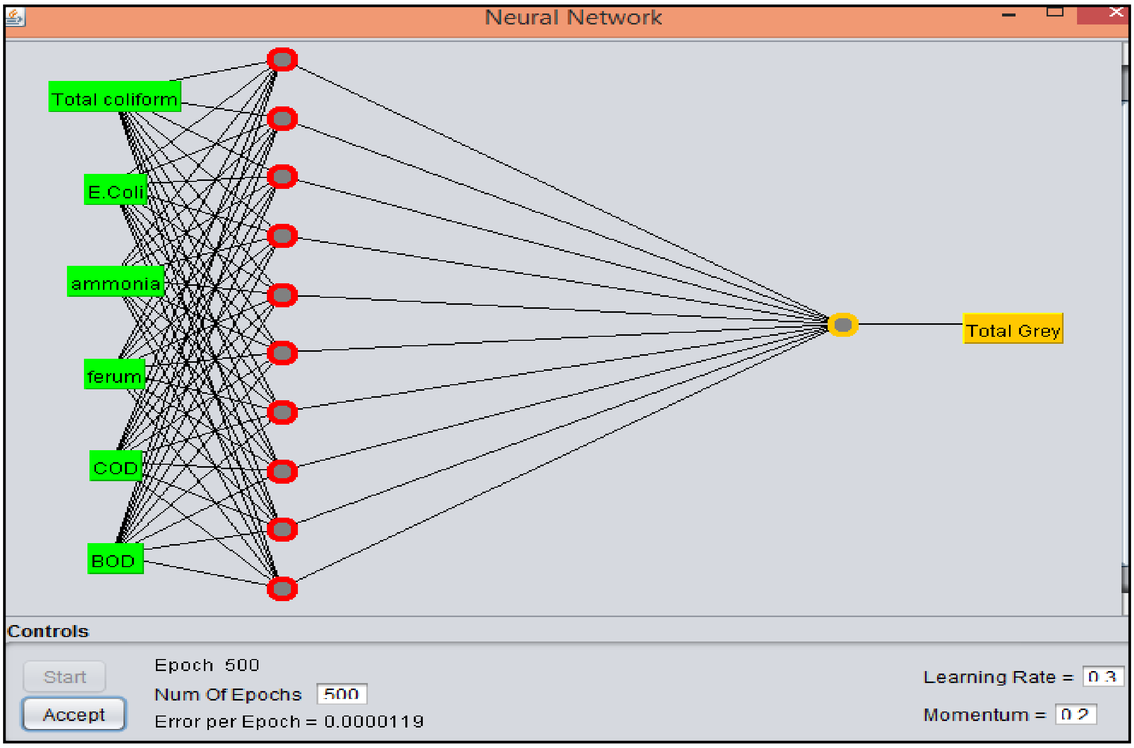

Figure 4.19 Number of hidden layers for ANN in WEKA software

Figure above illustrates ten numbers of hidden layers when performing training to the data sets using ANN algorithm. The inputs to the network were total coliform, E-coli, Ammonia, Iron (Fe^2), COD and BOD parameters. The total grey water footprint was set as the outputs of the network. The number of neurons in the hidden layer was varied from one to twenty, based on trial and error during the training process and the optimum number obtained is ten. To construct the predicted trend, the lowest value of RMSE produced by the training has been chosen.

Table 4.16 Values of RMSE with the hidden neurons by using ANN Algorithm

Hidden Neuron	RMSE	Hidden Neuron	RMSE
1	0.1258	**11**	0.0202
2	0.0389	**12**	0.0212
3	0.0317	**13**	0.0212
4	0.0273	**14**	0.0227
5	0.0225	**15**	0.0212
6	0.0212	**16**	0.0220
7	0.0204	**17**	0.0209
8	0.0214	**18**	0.0249
9	0.0207	**19**	0.0238
10	0.0199	**20**	0.0222

Referring to Table 4.13, the lowest RMSE value was obtained from training process and the training sets with 10 neurons have been chosen to predict the total WFgrey. Because of the adjustable hidden neurons, the ANN algorithm can be trained several times.

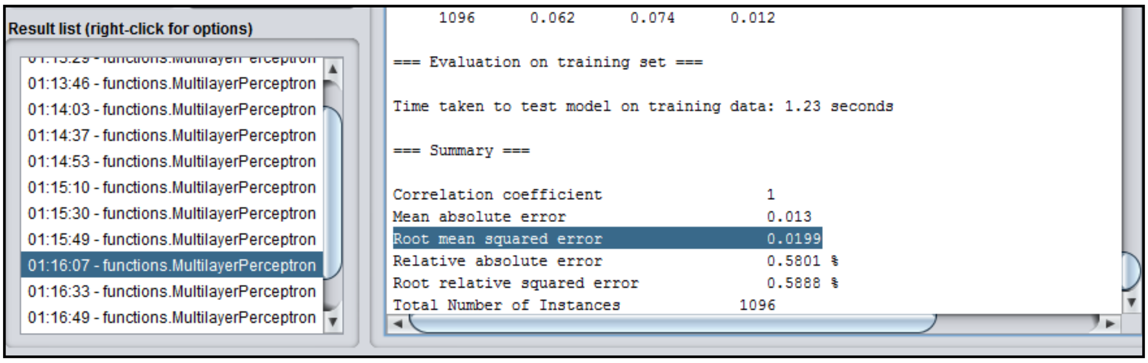

Figure 4.20 RMSE value after training using ANN

Figure above shows the result after the WFgrey data sets undergoes the training. The hidden layer that been set to 10 gives the smallest value of RMSE which is 0.0199.

Table 4.17 Analysis of actual and predicted value by using ANN

Month	Actual (m^3/s)	Predicted (m^3/s)
Jan-15	29.053	29.141
Feb-15	26.768	26.346
Mar-15	1.674	5.991
Apr-15	8.864	11.841
May-15	295.203	296.544
Jun-15	5.106	9.363
Jul-15	511.608	499.180
Aug-15	7.909	11.205
Sep-15	6.966	10.186
Oct-15	252.951	257.783
Nov-15	4.385	7.997
Dec-15	2.333	6.760
Jan-16	2.046	6.360
Feb-16	27.434	27.102
Mar-16	29.566	28.746
Apr-16	28.440	27.677
May-16	108.694	105.444
Jun-16	29.032	28.475
Jul-16	75.430	73.448
Aug-16	30.551	29.036
Sep-16	29.040	28.888
Oct-16	84.128	80.087
Nov-16	3.091	7.265
Dec-16	7.564	11.168
Jan-17	3.044	7.496
Feb-17	3.036	6.800
Mar-17	2.308	6.786
Apr-17	29.578	28.847
May-17	174.645	169.167
Jun-17	29.800	29.488
Jul-17	83.922	82.170
Aug-17	1.714	6.060
Sep-17	3.600	6.968
Oct-17	179.800	181.043
Nov-17	3.540	7.440
Dec-17	1.922	6.546

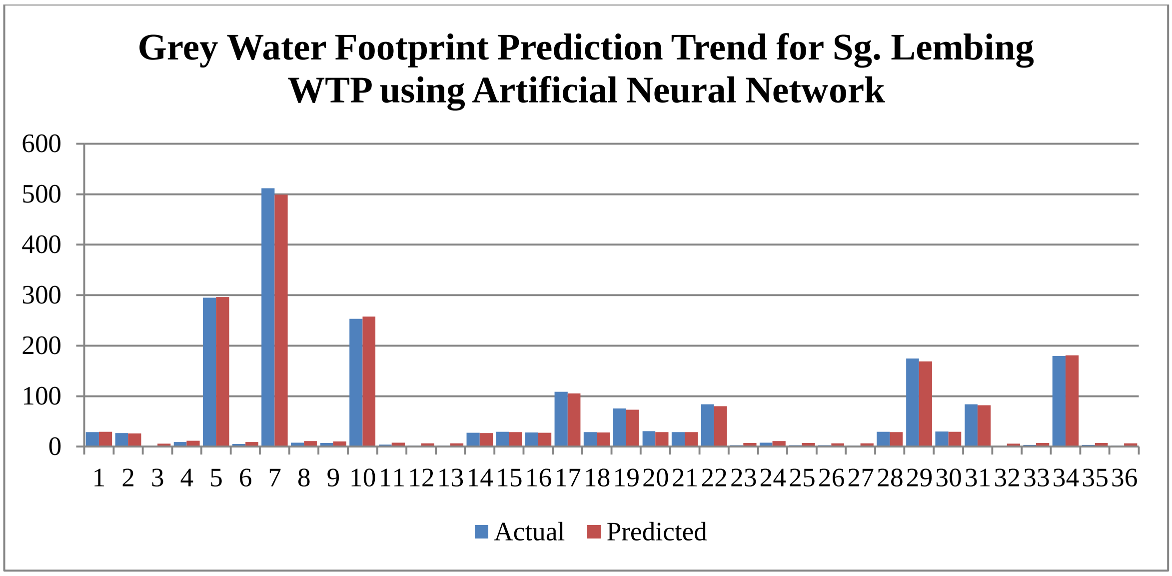

Figure 4.21 WFgrey trend at Sungai Lembing WTP using ANN

Figure above reveals the actual and predicted values of WFgrey trend at Sg. Lembing WTP after go through training using ANN algorithm. The actual values of WFgrey in 2015 until 2017 were 1152.820m³/s, 455.016m³/s and 516.909m³/s respectively. The predicted value for next three years; 2018, 2019 and 2020 were 1172.337m³/s, 453.696m³/s and 538.811m³/s respectively. The value for 2015 and 2017 were increase in 1.69% and 4.24% while for 2016 was decrease in 0.29%. The value of actual and predicted has differences due to pollution. This is because the pollution will need more water to assimilate the pollutants.

4.4.1.2 Bayesian Networks (BN) Algorithm

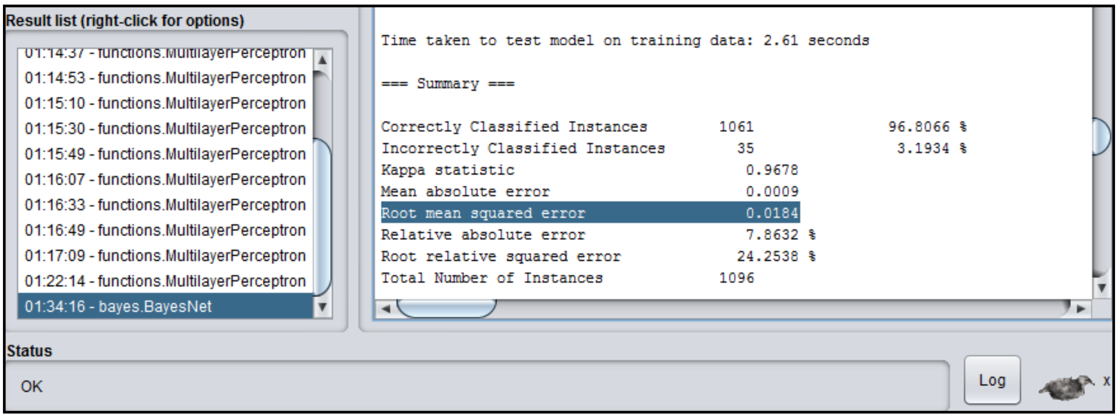

Figure 4.22 Result after the training using Bayesian Networks

Figure above shows the result after training process of WFgrey data with a RMSE value is 0.0184.

Table 4.18 Analysis of actual and predicted value by using BN

Month	**Actual (m^3/s)**	**Predicted (m^3/s)**
Jan-15	29.045006	29.045006
Feb-15	26.77252	26.77252
Mar-15	1.681227	1.681227
Apr-15	8.870346	8.802776
May-15	295.19428	295.19428
Jun-15	5.104662	5.101308
Jul-15	511.525	511.525
Aug-15	7.918753	7.918753
Sep-15	6.971043	6.971037
Oct-15	252.94821	252.94821
Nov-15	4.387112	4.387112
Dec-15	2.325709	2.327518
Jan-16	2.046734	2.046734
Feb-16	27.431466	27.432483
Mar-16	29.56583	29.56583
Apr-16	28.439291	28.439291
May-16	108.69641	108.69641
Jun-16	29.03008	29.03008
Jul-16	75.42415	75.42415
Aug-16	30.55556	30.55556
Sep-16	29.03566	29.03566
Oct-16	84.12651	84.12651
Nov-16	3.090636	3.089646
Dec-16	7.551658	7.551634
Jan-17	3.04029	3.039901
Feb-17	3.039456	3.039434
Mar-17	2.309188	2.309188
Apr-17	29.58202	29.58202
May-17	174.64835	174.64835
Jun-17	29.807	29.807
Jul-17	83.92414	83.92414
Aug-17	1.719208	1.719208
Sep-17	3.593	3.593252
Oct-17	179.80329	179.80329
Nov-17	3.537616	3.537352
Dec-17	1.929405	1.929405

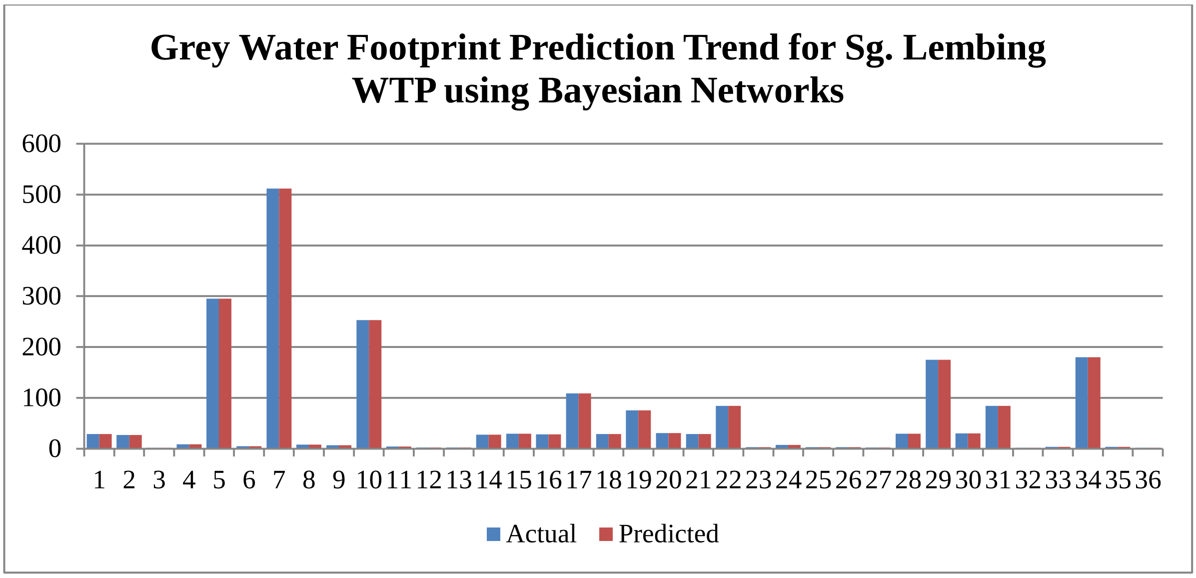

Figure 4.23 WFgrey trend at Sungai Lembing WTP using Bayesian Networks

Figure above reveals the actual and predicted values of WFgrey trend at Sg. Lembing WTP after go through training using Bayesian Networks algorithm. The actual values of WFgrey in 2015 until 2017 were 1152.744m³/s, 454.994m³/s and 516.933m³/s respectively. The predicted value for next three years; 2018, 2019 and 2020 were 1152.675m³/s, 454.994m³/s and 516.932m³/s respectively. For 2015, the value was decrease in 0.006% and for 2017 was decrease in 0.00008%. The value of actual and predicted has differences due to pollution. This is because the pollution will need more water to assimilate the pollutants.

4.4.2 Bukit Sagu Water Treatment Plant

4.4.2.1 Artificial Neural Network (ANN) Algorithm

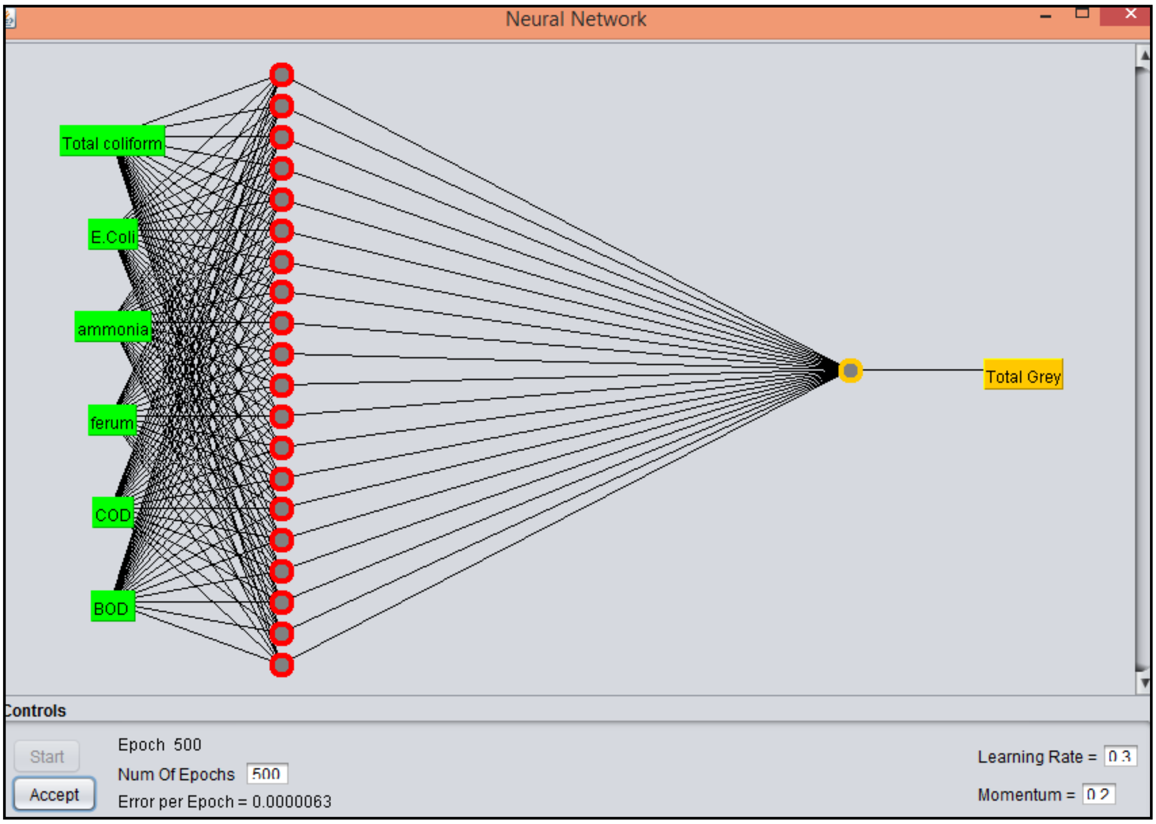

Figure 4.24 Number of hidden layers for ANN in WEKA software

Figure above illustrates 20 numbers of hidden layers when performing training to the data sets using ANN algorithm. The inputs to the network are total coliform, E-coli, Ammonia, Iron (Fe^2), COD and BOD parameters. The total grey water footprint was set as the outputs of the network. The number of neurons in the hidden layer was varied from one to twenty, based on trial and error during the training process and the optimum number obtained is 20. To construct the predicted trend, the lowest value of RMSE produced by the training has been chosen.

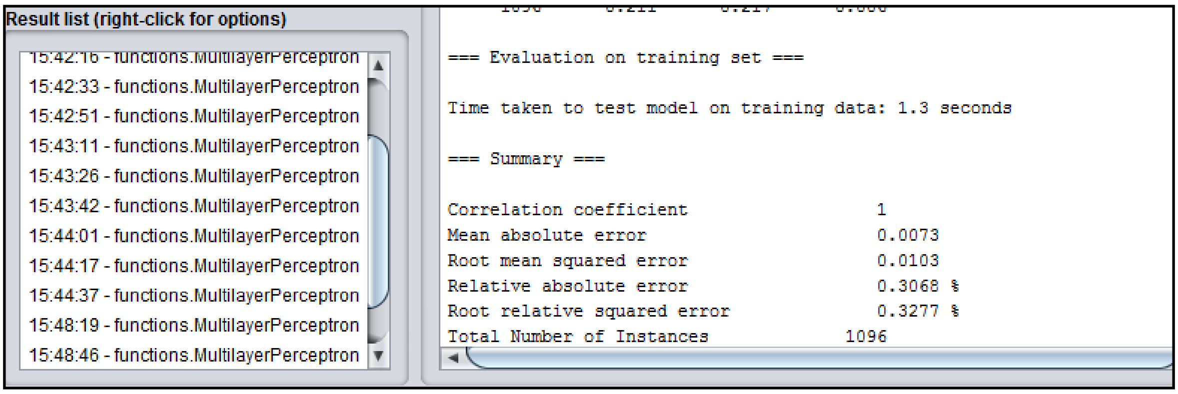

Figure 4.25 RMSE value after training using ANN

Figure above shows the result after training process of the WFgrey data sets. The hidden layer that been set to 20 gives the smallest value of RMSE was 0.0103.

Table 4.19 Values of RMSE with the hidden neurons by using ANN Algorithm

Hidden Neuron	RMSE	Hidden Neuron	RMSE
1	0.0961	**11**	0.0151
2	0.0425	**12**	0.0150
3	0.0195	**13**	0.0164
4	0.0145	**14**	0.0195
5	0.0169	**15**	0.0173
6	0.0151	**16**	0.0172
7	0.0137	**17**	0.0196
8	0.0175	**18**	0.0189
9	0.0172	**19**	0.0123
10	0.0165	**20**	0.0103

Referring to Table above, the lowest RMSE value was obtained from training process and the training sets with 20 neurons have been chosen to predict the total WFgrey. Because of the adjustable hidden neurons, the ANN algorithm can be trained several times.

Table 4.20 Analysis of actual and predicted value by using ANN

Month	Actual (m^3/s)	Predicted (m^3/s)
Jan-15	24.800	24.732
Feb-15	22.498	22.400
Mar-15	2.511	2.598
Apr-15	1.603	1.868
May-15	103.986	103.335
Jun-15	24.640	24.504
Jul-15	388.429	387.895
Aug-15	28.125	27.871
Sep-15	24.561	24.507
Oct-15	212.115	212.086
Nov-15	6.084	5.840
Dec-15	25.746	25.668
Jan-16	6.330	6.434
Feb-16	4.408	4.410
Mar-16	23.591	23.484
Apr-16	24.546	24.429
May-16	97.663	97.058
Jun-16	23.851	23.759
Jul-16	131.936	131.398
Aug-16	7.292	7.338
Sep-16	29.280	29.121
Oct-16	211.578	212.029
Nov-16	65.144	64.496
Dec-16	8.136	8.049
Jan-17	7.936	7.987
Feb-17	3.232	3.335
Mar-17	6.561	6.633
Apr-17	8.340	8.225
May-17	329.155	328.566
Jun-17	25.702	25.578
Jul-17	266.938	266.537
Aug-17	26.939	26.836
Sep-17	26.121	26.043
Oct-17	158.684	158.759
Nov-17	8.952	9.018
Dec-17	6.522	6.629

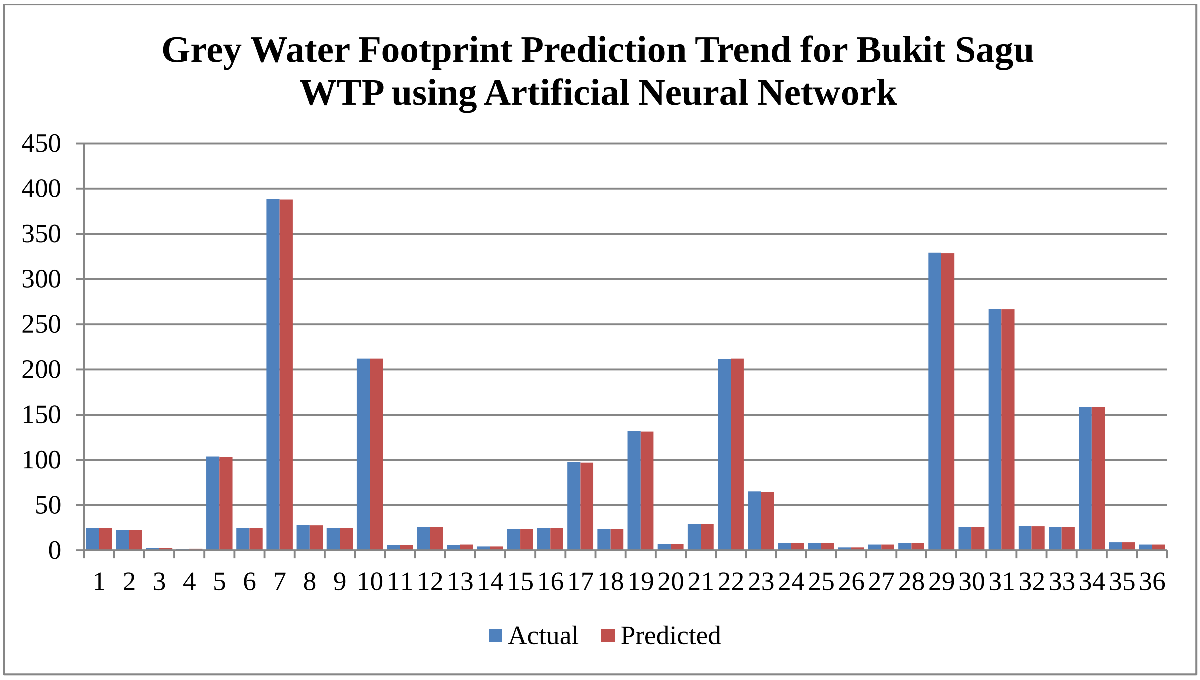

Figure 4.26 WFgrey trend at Bukit Sagu WTP using ANN

Figure above illustrates the actual and predicted values of WFgrey trend at Bukit Sagu WTP after undergoes training using ANN algorithm. The actual value of WFgrey in 2015 was 865.098m³/s while predicted value was 863.304m³/s. the value was decrease in 0.21%. For 2016, the actual value was 633.755m³/s while predicted value was 632.005m³/s and the value was decrease in 0.28%. For 2017 also the value was decrease in 0.11% as the actual value was 875.082m³/s and predicted value was 874.146m³/s. The predicted trend value has different with actual value due to pollution. This is because the pollution will need more water to assimilate the pollutants.

4.4.2.2 Bayesian Networks (BN) Algorithm

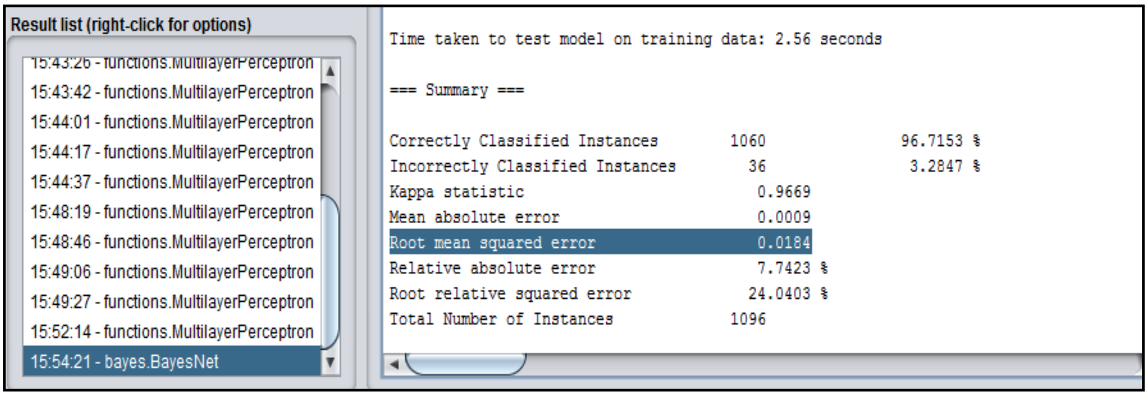

Figure 4.27 RMSE value after training using Bayesian Networks

Figure above shows the result after training process of WFgrey data with a RMSE value was 0.0184.

Table 4.21 Analysis of actual and predicted value by using BN

Month	Actual (m^3/s)	Predicted (m^3/s)
Jan-15	24.79214	24.79214
Feb-15	22.49928	22.49968
Mar-15	2.51935	2.51927
Apr-15	1.60156	1.58882
May-15	103.98213	103.98213
Jun-15	24.63945	24.63882
Jul-15	388.42990	388.42990
Aug-15	28.13188	28.13888
Sep-15	24.56609	24.55783
Oct-15	212.11941	213.93252
Nov-15	6.07970	6.07970
Dec-15	25.74347	25.74347
Jan-16	6.33087	6.33087
Feb-16	4.40412	4.40401
Mar-16	23.58268	23.58268
Apr-16	24.55087	24.55087
May-16	97.66685	79.45241
Jun-16	23.85193	23.83479
Jul-16	131.92828	131.92828
Aug-16	7.29356	7.29358
Sep-16	29.27887	29.27887
Oct-16	211.58257	211.58257
Nov-16	65.14395	65.14395
Dec-16	8.13869	8.13869
Jan-17	7.92811	7.92811
Feb-17	3.23581	3.23581
Mar-17	6.56262	6.56262
Apr-17	8.34185	8.34185
May-17	329.15980	329.15980
Jun-17	25.69512	25.69512
Jul-17	266.94128	266.94128
Aug-17	26.93731	26.93731
Sep-17	26.12895	26.12895
Oct-17	158.68794	158.68794
Nov-17	8.95417	8.95429
Dec-17	6.51529	6.51529

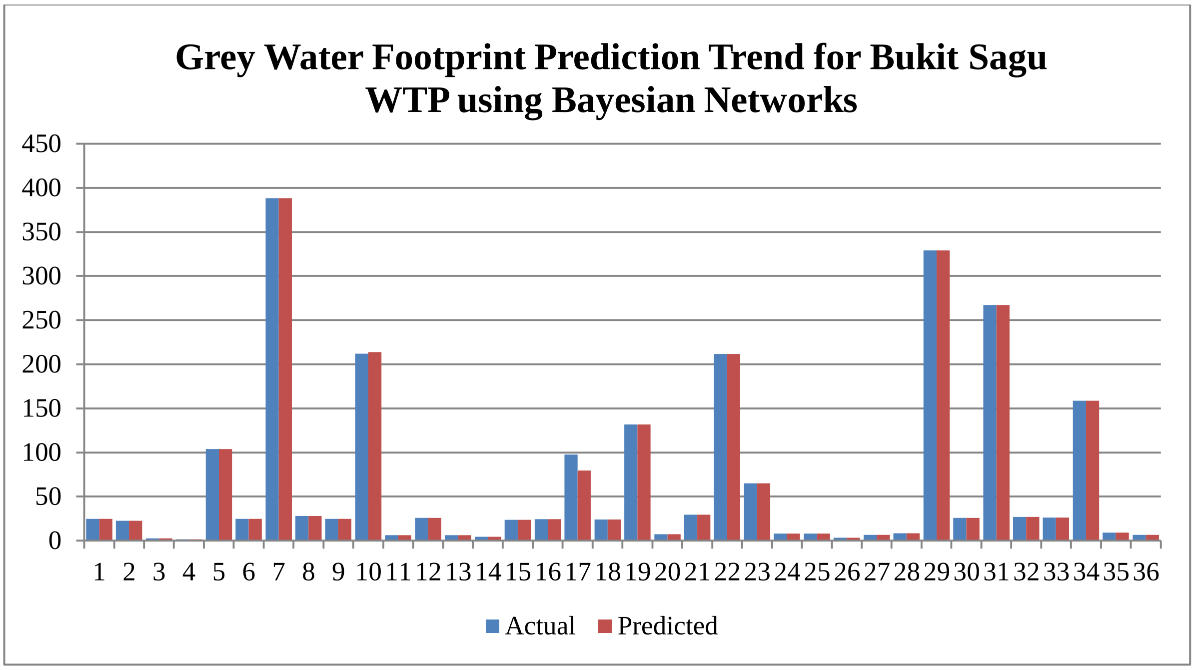

Figure 4.28 WFgrey trend at Bukit Sagu WTP using Bayesian Networks

Figure above shows the actual and predicted values of WFgrey trend at Bukit Sagu WTP after go through training using Bayesian Networks algorithm. For 2015 until 2017, the actual values of WFgrey were 865.104m³/s, 633.753m³/s, 875.088m³/s respectively. Hence, for 2018 until 2020, the predicted values were 866.903m³/s, 615.522m³/s and 875 088m³/s respectively. For 2015, the value increase was 0.21% however for 2016 was decrease in 2.88% and for 2017 was increase in 0.000014%. The predicted trend value has different with actual value due to pollution. This is because the pollution will need more water to assimilate the pollutants.

4.4.3 Bukit Ubi Water Treatment Plant

4.4.3.1 Artificial Neural Network (ANN) Algorithm

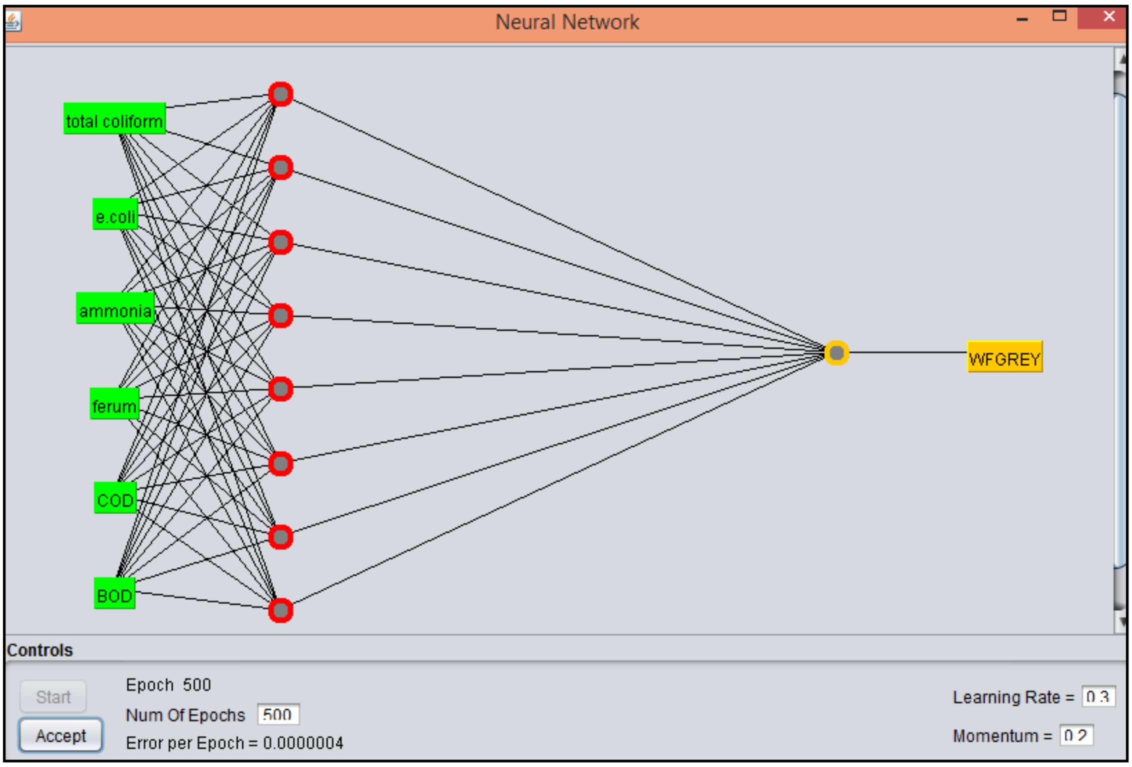

Figure 4.29 Number of hidden layers for ANN in WEKA software

Figure above illustrates eight numbers of hidden layers when performing training to the data sets using ANN algorithm. The inputs to the network are total coliform, E-coli, Ammonia, Iron (Fe^2), COD and BOD parameters. The total grey water footprint was set as the outputs of the network. The number of neurons in the hidden layer was varied from one to twenty, based on trial and error during the training process and the optimum number obtained is eight. To construct the predicted trend, the lowest value of RMSE produced by the training has been chosen.

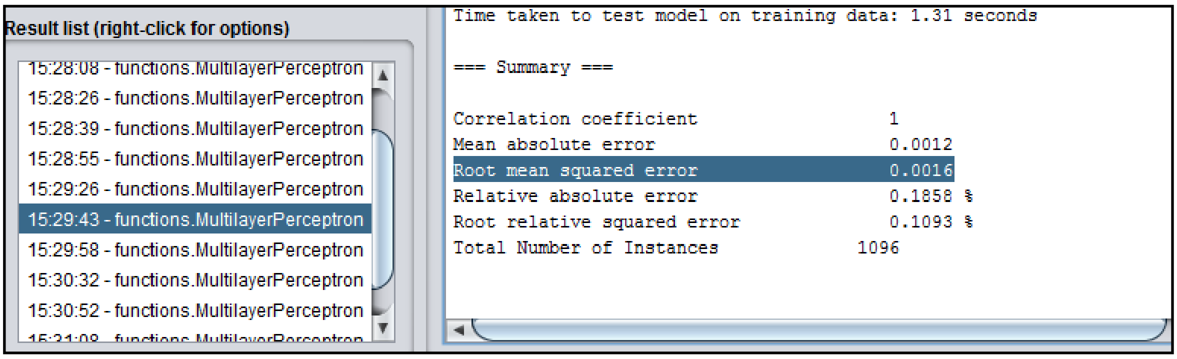

Figure 4.30 RMSE value after training using ANN

Figure above shows the result after training process of the WFgrey data sets. The hidden layer that been set to 8 gives the smallest value of RMSE was 0.0016.

Table 4.22 Values of RMSE with the hidden neurons by using ANN Algorithm

Hidden Neuron	**RMSE**	**Hidden Neuron**	**RMSE**
1	0.0225	**11**	0.0023
2	0.0029	**12**	0.0032
3	0.0027	**13**	0.0017
4	0.0028	**14**	0.0047
5	0.0021	**15**	0.0068
6	0.0022	**16**	0.0064
7	0.0020	**17**	0.0062
8	0.0016	**18**	0.0057
9	0.0021	**19**	0.0052
10	0.0018	**20**	0.0130

Based on Table 4.19, the lowest RMSE value is obtained from training process and the training sets with 8 neurons have been chosen to predict the total WFgrey. Because of the adjustable hidden neurons, the ANN algorithm can be trained several times.

Table 4.23 Analysis of actual and predicted value by using ANN

Month	Actual (m^3/s)	Predicted (m^3/s)
Jan-15	20.216	20.196
Feb-15	19.180	19.208
Mar-15	22.134	22.165
Apr-15	22.563	22.533
May-15	248.902	248.893
Jun-15	11.498	11.580
Jul-15	41.223	41.323
Aug-15	34.274	34.298
Sep-15	24.584	24.526
Oct-15	172.159	172.103
Nov-15	23.620	23.585
Dec-15	21.686	21.622
Jan-16	18.904	18.960
Feb-16	24.931	24.969
Mar-16	20.894	20.925
Apr-16	14.940	15.003
May-16	19.313	19.326
Jun-16	19.273	19.251
Jul-16	22.634	22.603
Aug-16	22.264	22.263
Sep-16	22.960	22.983
Oct-16	20.631	20.631
Nov-16	20.137	20.137
Dec-16	26.113	26.126
Jan-17	21.058	21.002
Feb-17	20.880	20.844
Mar-17	15.633	15.633
Apr-17	28.146	28.146
May-17	28.407	28.334
Jun-17	20.365	20.330
Jul-17	22.987	22.996
Aug-17	20.084	20.079
Sep-17	14.715	14.778
Oct-17	25.988	26.020
Nov-17	18.915	18.908
Dec-17	17.314	17.352

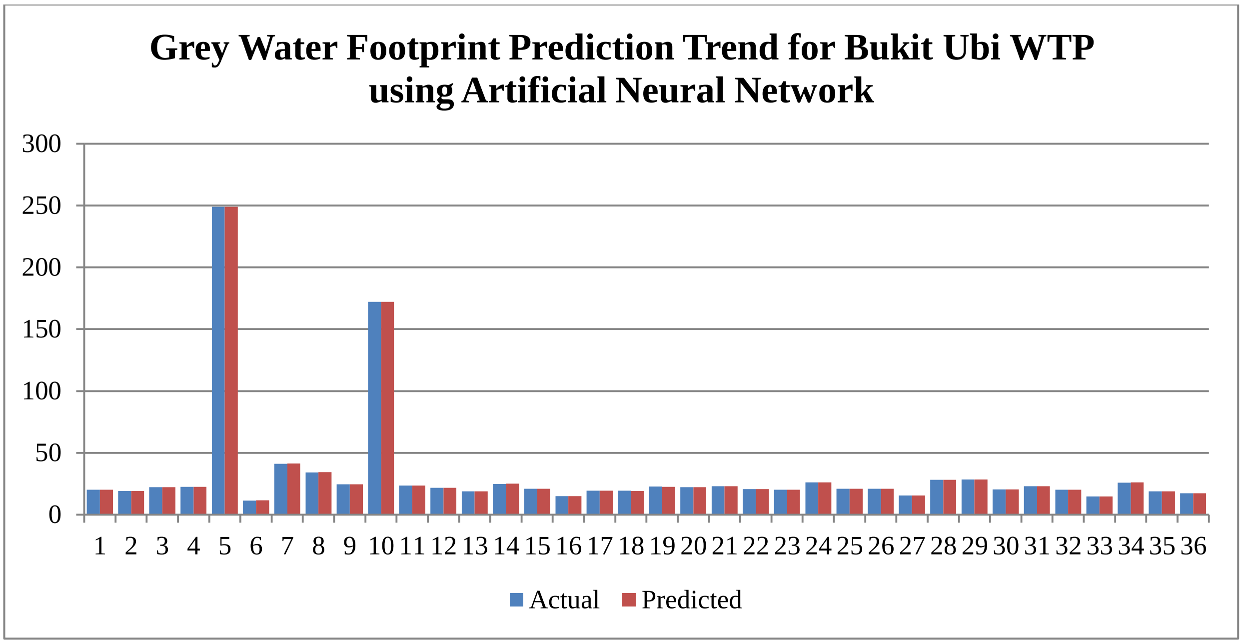

Figure 4.31 WFgrey trend at Bukit Ubi WTP using ANN

Figure above illustrates the actual and predicted values of WFgrey trend at Bukit Ubi WTP after go through training using ANN algorithm. For 2015, the actual value of WFgrey was 662.039m³/s while predicted value was 662.032m³/s and has decrease about 0.001%. For 2016, the actual value was 252.994m³/s while predicted value was 253.177m³/s and the value was increase in 0.072%. For 2017, the actual value was 254.492m³/s and predicted value was 254.422m³/s. The value was decrease in 0.28%. The predicted value was for next three years; 2018, 2019 and 2020. The predicted trend value has different with actual value due to pollution. This is because the pollution will need more water to assimilate the pollutants.

4.4.3.2 Bayesian Networks (BN) Algorithm

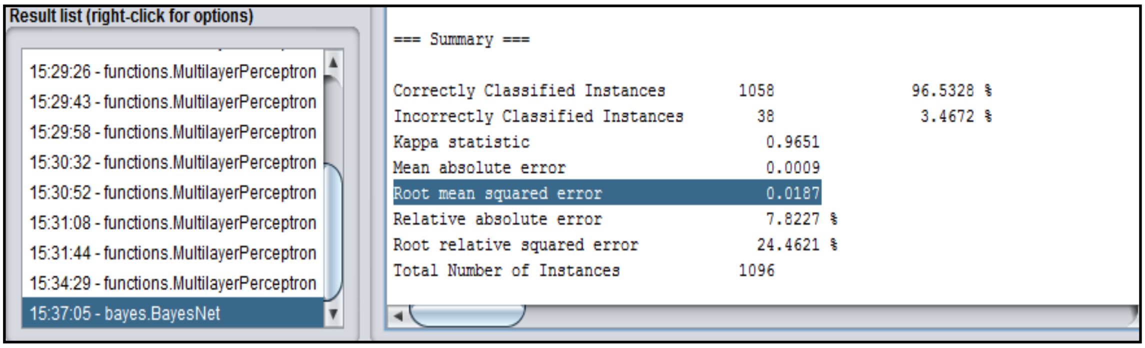

Figure 4.32 RMSE value after training using Bayesian Networks

Figure above shows Bayesian Networks algorithm trained the data sets and gives the value of RMSE, 0.0187.

Table 4.24 Analysis of actual and predicted value by using BN algorithm

Month	**Actual (m^3/s)**	**Predicted (m^3/s)**
Jan-15	20.21953	20.21953
Feb-15	19.17438	19.17438
Mar-15	22.13359	22.13339
Apr-15	22.56373	22.56373
May-15	248.90936	248.90936
Jun-15	11.50620	11.50620
Jul-15	41.21806	41.21806
Aug-15	34.28245	34.28245
Sep-15	24.54833	24.54833
Oct-15	172.15021	172.15021
Nov-15	23.62208	23.62208
Dec-15	21.68030	21.68432
Jan-16	18.90285	18.90285
Feb-16	24.92897	24.92897
Mar-16	20.90453	20.90472
Apr-16	14.92908	14.92908
May-16	19.31209	19.31209
Jun-16	19.27458	19.27501
Jul-16	22.62725	22.62725
Aug-16	22.27115	22.27256
Sep-16	22.96306	22.96306
Oct-16	20.63096	20.63096
Nov-16	20.13875	20.14155
Dec-16	26.11362	26.11362
Jan-17	21.05252	21.05252
Feb-17	20.87748	20.87748
Mar-17	15.62701	15.62701
Apr-17	28.14239	28.14239
May-17	28.40424	28.40424
Jun-17	20.36688	20.36688
Jul-17	22.99456	22.99456
Aug-17	20.09199	20.09199
Sep-17	14.70697	14.70697
Oct-17	25.99184	25.99184
Nov-17	18.91450	18.91432
Dec-17	17.31326	17.31326

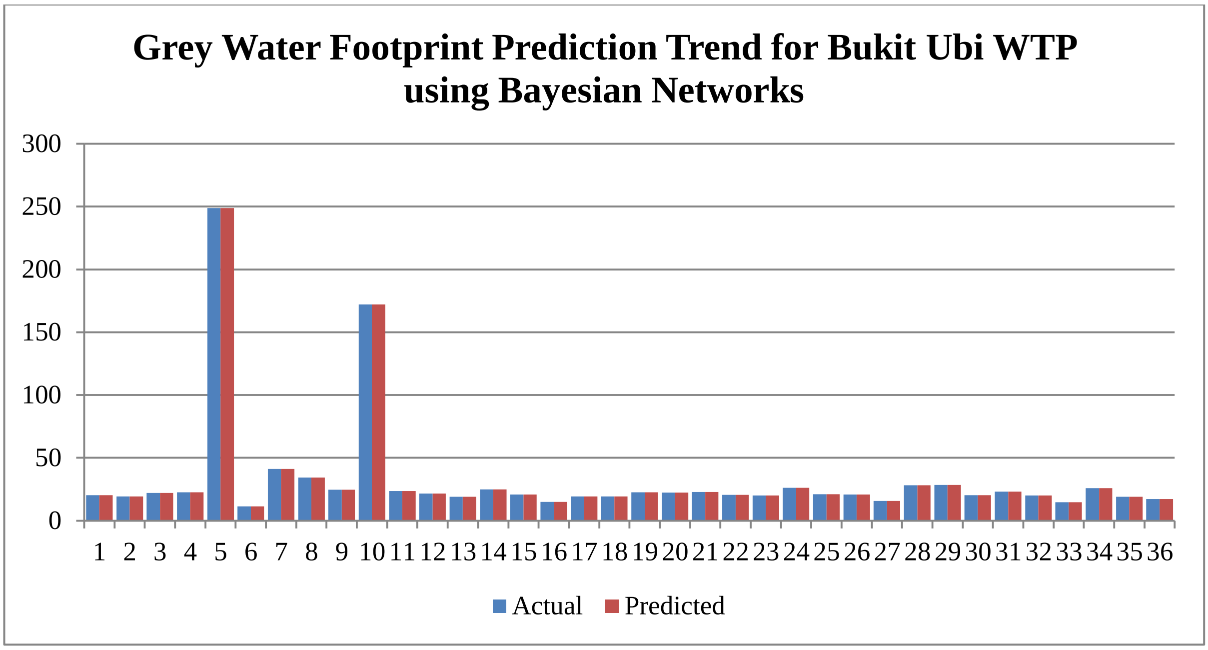

Figure 4.33 WFgrey trend at Bukit Ubi WTP using Bayesian Networks

Figure above reveals the actual and predicted values of WFgrey trend at Bukit Ubi WTP after go through training using Bayesian Networks algorithm. For 2015 until 2017, the actual values of WFgrey were 662.008m³/s, 252.997m³/s and 254.484m³/s respectively. Hence, for 2018 until 2020, the predicted values were 662.012m³/s, 253.002m³/s and 254.483m³/s respectively. The different value for 2015 was increase in 0.0006% while for 2016 was increase about 0.002%. However, for 2017 the value was decrease in 0.000073%. The predicted trend value has different with actual value due to pollution. This is because the pollution will need more water to assimilate the pollutants.

CHAPTER 5

CONCLUSION

5.1 Conclusion

In conclusion, the total grey water footprint in Sungai Lembing, Bukit Sagu and Bukit Ubi water treatment plant in Kuantan river basin is calculated. For Sungai Lembing WTP, the total of WFgrey in three years; 2015, 2016 and 2017 were 1152.824m^3/s, 454.996m^3/s and 516.934m^3/s respectively. For Bukit Sagu WTP, 865.105m^3/s, 633.754m^3/s and 875.090m^3/s were the value of total WFgrey in 2015, 2016 and 2017. In addition, for Bukit Ubi WTP the value of total WFgrey in 2015, 2016 and 2017 were 276.097m^3/s, 252.998m^3/s and 254.485m^3/s respectively. In 2015, Sg. Lembing WTP was the highest value while Bukit Ubi WTP was the lowest value of total WFgrey. For 2016 and 2017, the highest value of total WFgrey was at Bukit Sagu WTP and the lowest value was at Bukit Ubi WTP.

On the other hand, the best algorithm was determined between Bayesian Networks and Artificial Neural Network. For the best algorithm, Artificial Neural Network (ANN) has been selected as the best algorithm to be used in WFgrey prediction. This algorithm has generated the least RMSE values as indicates that it has least error in predicting values. For Sungai Lembing WTP, from ANN algorithm RMSE value was 0.0199 while for Bayesian Networks algorithm was 0.0184. For Bukit Sagu WTP, value of RMSE for ANN algorithm, 0.0103 was the least value compared to Bayesian Network was 0.0184. For Bukit Ubi WTP, the value RMSE for ANN algorithm was 0.0016 while for Bayesian Network algorithm was 0.0187. Hence, the best algorithm was Artificial Neural Network as the lowest value of RMSE indicates as the least error.

Prediction trend of total grey water footprint in three water treatment plants was be able to be produced as the end result of the study. This prediction trend of total WFgrey was for three next years which are 2018, 2019 and 2020. For Sungai Lembing WTP, by using ANN, the predicted value for next three years; 2018, 2019 and 2020 were 1172.337m³/s, 453.696m³/s and 538.811m³/s respectively. By using Bayesian Networks, the predicted value for 2018, 2019 and 2020 were 1152.675m³/s, 454.994m³/s and 516.932m³/s respectively. Furthermore, for Bukit Sagu WTP, by using ANN, the predicted value were 863.304m³/s, 632.005m³/s and 874.146m³/s respectively while by Bayesian Networks the predicted values were 866.903m³/s, 615.522m³/s and 875 088m³/s for 2018, 2019 and 2020 respectively. Last but not least, for Bukit Ubi WTP, for next three years; 2018, 2019 and 2020, the predicted values were 662.032m³/s, 253.177m³/s and 254.422m³/s respectively by using ANN. However, with Bayesian Networks the predicted values were 662.012m³/s, 253.002m³/s and 254.483m³/s respectively.

5.2 Recommendation

In the future, it is recommended that, the pollutants should be controlled and monitored to prevent the excessive pollutant in our drinking water. As agriculture is one of water resources, so it is essential to have climate-friendly crops, efficient irrigation that reduces the need for water and energy-efficient food production. Green agriculture is also crucial to limit the chemicals from entering the water. In addition, practice organic farming is suggested because organically grown crops reduce the quantity of herbicides and pesticides used in farming. It is also promoting the use of organic fertilizer which promotes natural growth and eliminates the use of toxic chemicals found in synthetic fertilizers, which can penetrate into the ground and pollute water supplies. If the concentration of pollutants is lower, the freshwater use to assimilate pollutant will also be lower. Thus, the total grey water footprint will be reduced also.

As Artificial Neural Network (ANN) has been chosen as the best algorithm due to low value of RMSE compared to the Bayesian Networks, ANN was recommended for further study in the future. ANN can predicted the data with accurate and precisely. For further study, grey water footprint can be assessed in further process beginning from the water treatment plant. There are also other algorithms that can be explored for example, linear regression; this is due to the linear trend of data.

REFERENCES

Agudosi, E. S., Abdullah, E. C., Mubarak, N. M., Khalid, M., Pudza, M. Y., Agudosi, N. P., & Abutu, E. D. (2018). SC. *Biochemical Pharmacology*. https://doi.org/10.1016/j.jece.2018.11.001

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

Allocca, V., Marzano, E., Tramontano, M., & Celico, F. (2018). Environmental impact of cattle grazing on a karst aquifer in the southern Apennines (Italy): Quantification through the grey water footprint. *Ecological Indicators*, *93*(April), 830–837. https://doi.org/10.1016/j.ecolind.2018.05.075

Arto, I., Andreoni, V., & Rueda-Cantuche, J. M. (2016). Global use of water resources: A multiregional analysis of water use, water footprint and water trade balance. *Water Resources and Economics*, *15*, 1–14. https://doi.org/10.1016/j.wre.2016.04.002

Ascione, F., Bianco, N., Stasio, C. De, Maria, G., & Peter, G. (2017). Arti fi cial neural networks to predict energy performance and retro fi t scenarios for any member of a building category : A novel approach. *Energy*, *118*, 999–1017. https://doi.org/10.1016/j.energy.2016.10.126

Bendtsen, M., & Peña, J. M. (2016). International Journal of Approximate Reasoning Gated Bayesian networks for algorithmic trading. *International Journal of Approximate Reasoning*, *69*, 58–80. https://doi.org/10.1016/j.ijar.2015.11.002

Bogomolny, A. (2015). What Is Algorithm? *Cut-the-Knot.Org*. Retrieved from http://www.cut-the-knot.org/WhatIs/WhatIsAlgorithm.shtml

Brown, A., Matlock, M. D., & Ph, D. (2011). A Review of Water Scarcity Indices and Methodologies.

Cha, K., Son, M., Hong, S., An, S., & Part, S. (2017). Method to assess water footprint, a case study for white radishes in Korea. *International Soil and Water Conservation Research*, *5*(2), 151–157. https://doi.org/10.1016/j.iswcr.2017.02.005

Chen, Q., Norcio, A. F., & Wang, J. (2000). Neural network based stereotyping for user profiles. *Neural Computing and Applications*, *9*(4), 259–265. https://doi.org/10.1007/s005210070003

Cities, S., Zambonelli, F., Salim, F., Loke, S. W., & Meuter, W. De. (2018). [E] Algorithmic Governance in Smart Cities, (June).

Ding, N., Liu, J., Yang, J., & Lu, B. (2018). Water footprints of energy sources in China : Exploring options to improve water ef fi ciency. *Journal of Cleaner Production*, *174*, 1021–1031. https://doi.org/10.1016/j.jclepro.2017.10.273

Distefano, T., & Kelly, S. (2017). Are we in deep water? Water scarcity and its limits to economic growth. *Ecological Economics*, *142*, 130–147. https://doi.org/10.1016/j.ecolecon.2017.06.019

Du, H., Chen, W., Zhu, Q., Liu, S., & Zhou, J. (2018). Identification of weak peaks in X-ray fluorescence spectrum analysis based on the hybrid algorithm combining genetic and Levenberg Marquardt algorithm. *Applied Radiation and Isotopes*, *141*(July), 149–155. https://doi.org/10.1016/j.apradiso.2018.07.009

Duan, Z. H., & Poon, C. S. (2014). Properties of recycled aggregate concrete made with recycled aggregates with different amounts of old adhered mortars. *Materials and Design*, *58*, 19–29. https://doi.org/10.1016/j.matdes.2014.01.044

Ercin, A. E., & Hoekstra, A. Y. (2014). Water footprint scenarios for 2050 : A global analysis. *Environment International*, *64*, 71–82. https://doi.org/10.1016/j.envint.2013.11.019

Fuller, K. D., Gordon, R., Grimmett, M., Fillmore, S., Madani, A., Vanroestel, J., … George, E. S. (2010). Agriculture , Ecosystems and Environment Seasonal and crop rotational effects of manure management on nitrate – nitrogen leaching in Nova Scotia. *"Agriculture, Ecosystems and Environment," 137*(3–4), 267–275. https://doi.org/10.1016/j.agee.2010.02.012

Gillespie, T. (2012). The Relevance of Algorithms, (Light 1999).

Giwa, A., Daer, S., Ahmed, I., Marpu, P. R., & Hasan, S. W. (2016). Journal of Water Process Engineering Experimental investigation and artificial neural networks ANNs modeling of electrically-enhanced membrane bioreactor for wastewater treatment. *Journal of Water Process Engineering*, *11*, 88–97. https://doi.org/10.1016/j.jwpe.2016.03.011

Guidelines, S. (2013). R ESEARCH R EPORT S ERIES N O . 65.

Herath, I., Green, S., Singh, R., Horne, D., Zijpp, S. Van Der, & Clothier, B. (2013). Water footprinting of agricultural products : a hydrological assessment for the water footprint of New Zealand ' s wines. *Journal of Cleaner Production*, *41*, 232–243. https://doi.org/10.1016/j.jclepro.2012.10.024

Hoekstra, A. Y., & Mekonnen, M. M. (2011). National Water Footprint Accounts: The Green, Blue and Grey Water Footprint of Production and Consumption. Value of Water Research Report Series No. 50. *Delft, the Netherlands: UNESCO-IHE*.

Hoekstra, A. Y., & Mekonnen, M. M. (2012). The water footprint of humanity, *109*(9).

https://doi.org/10.1073/pnas.1109936109

Hussein, H. (2018). Lifting the veil: Unpacking the discourse of water scarcity in Jordan. *Environmental Science and Policy*, *89*(March), 385–392. https://doi.org/10.1016/j.envsci.2018.09.007

İçen, D., & Ersel, D. (2019). A new approach for probability calculation of fuzzy events in Bayesian Networks. *International Journal of Approximate Reasoning*, *108*, 76–88. https://doi.org/10.1016/j.ijar.2019.03.004

Jackson, C., & Mosleh, A. (2016). Bayesian inference with overlapping data: Reliability estimation of multi-state on-demand continuous life metric systems with uncertain evidence. *Reliability Engineering and System Safety*, *145*(October 2015), 124–135. https://doi.org/10.1016/j.ress.2015.09.006

Kozaki, Daisuke & Idayu Binti Harun, Norhasmira & Ab. Rahim, Mohd Hasbi & Mori, Masanobu & Nakatani, Nobutake & Tanaka, K. (2017). Determination of Water Quality Degradation Due to Industrial and Household Wastewater in the Galing River in Kuantan, Malaysia Using Ion Chromatograph and Water Quality Data. *Environments*.

Miglietta, P. P., Toma, P., Fanizzi, F. P., Donno, A. De, Coluccia, B., Migoni, D., … Serio, F. (n.d.). A Grey Water Footprint Assessment of Groundwater Chemical Pollution : Case Study in Salento. https://doi.org/10.3390/su9050799

Murray, S. J., Foster, P. N., & Prentice, I. C. (2012). Future global water resources with respect to climate change and water withdrawals as estimated by a dynamic global vegetation model. *Journal of Hydrology*, *448–449*, 14–29. https://doi.org/10.1016/j.jhydrol.2012.02.044

Naderpour, H., Hossein, A., & Fakharian, P. (2018). Compressive strength prediction of environmentally friendly concrete using arti fi cial neural networks, *16*(October 2017), 213–219. https://doi.org/10.1016/j.jobe.2018.01.007

Ridoutt, B. G., & Pfister, S. (2010). A revised approach to water footprinting to make transparent the impacts of consumption and production on global freshwater scarcity, *20*, 113–120. https://doi.org/10.1016/j.gloenvcha.2009.08.003

Rosegrant, M. W., Cai, X., & Cline, S. A. (2002). *World water and food to 2025: dealing with scarcity*. Intl Food Policy Res Inst.

Safa, B., Arkebauer, T. J., Zhu, Q., Suyker, A., & Irmak, S. (2018). Latent heat and sensible heat fl ux simulation in maize using arti fi cial neural networks. *Computers and Electronics in Agriculture*, *154*(September), 155–164. https://doi.org/10.1016/j.compag.2018.08.038

Salah, H. A., Mocanu, M., & Florea, A. (2014). Analysis of Data Mining Tools Used for Water Resources Management inTigris River Analysis of Data Mining Tools Used for Water Resources Management inTigris River, (October 2018). https://doi.org/10.7508/AMS-V3-N2-1-10

Science, C. (2017). a Comparative Analysis for Smart Water Resource Using, *5*, 24–30. https://doi.org/10.5281/zenodo.840201

Teixeira Júnior, L. A., Souza, R. M. de, Menezes, M. L. de, Cassiano, K. M., Pessanha, J. F. M., & Souza, R. C. (2015). Artificial Neural Network and Wavelet Decomposition in the Forecast of Global Horizontal Solar Radiation. *Pesquisa Operacional*, *35*(1), 73–90. https://doi.org/10.1590/0101-7438.2015.035.01.0073

Vergé, X., VanderZaag, A., Smith, W., Grant, B., & Gordon, R. (2017). The consideration of time step in calculating grey water footprints of agricultural cropping systems. *Ecological Indicators*, *78*, 31–36. https://doi.org/10.1016/j.ecolind.2017.03.006

Wang, J., & Liu, S. (2018). Knowle dge-Base d Systems Novel binary encoding water cycle algorithm for solving Bayesian network structures learning problem. *Knowledge-Based Systems*, *150*, 95–110. https://doi.org/10.1016/j.knosys.2018.03.007

Wickramasinghe, W. M. S., Navaratne, C. M., & Dias, S. V. (2018). Building resilience on water quality management through grey water footprint approach: A case study from Sri Lanka. *Procedia Engineering*, *212*, 752–759. https://doi.org/10.1016/j.proeng.2018.01.097

Xu, M., Li, C., & Lu, S. (2017). Sustainable Water Resources Utilization on Energy Industry Based on the Gray Water Footprints Assessment in Hunan Province. *Energy Procedia*, *105*, 3758–3764. https://doi.org/10.1016/j.egypro.2017.03.872

Zhi, Y., Yang, Z., Yin, X., Hamilton, P. B., & Zhang, L. (2015). Using gray water footprint to verify economic sectors ' consumption of assimilative capacity in a river basin : model and a case study in the Haihe River Basin , China. *Journal of Cleaner Production*, *92*, 267–273. https://doi.org/10.1016/j.jclepro.2014.12.058

APPENDIX A
SAMPLE APPENDIX 1

Table 5.1 Prediction of grey water footprint at Sungai Lembing WTP

	Artificial Neural Network		
Years	2015	2016	2017
Actual	1152.820	455.016	516.909
Predicted	1172.337	453.696	538.811
Percentage (%)	1.69298	-0.29010	4.23711
	1.88724		

Table 5.2 Prediction of grey water footprint at Sungai Lembing WTP

	Bayesian Networks		
Years	2015	2016	2017
Actual	1152.744	454.994	516.933
Predicted	1152.675	454.994	516.932
Percentage (%)	-0.005996	6.593x10^-7	-0.0000818
	-0.00327		

Table 5.3 Prediction of grey water footprint at Bukit Sagu WTP

	Artificial Neural Network		
Years	2015	2016	2017
Actual	865.098	6333.755	875.082
Predicted	863.304	632.005	874.146
Percentage (%)	-0.20738	-0.27613	-0.10696
	-0.18872		

Table 5.4 Prediction of grey water footprint at Bukit Sagu WTP

	Bayesian Networks		
Years	2015	2016	2017
Actual	865.104	633.753	875.082
Predicted	866.903	615.522	875.088
Percentage (%)	0.20793	-2.87678	0.000014
	-0.692221		

Table 5.5 Prediction of grey water footprint at Bukit Ubi WTP

	Artificial Neural Network		
Years	2015	2016	2017
Actual	662.039	252.994	254.492
Predicted	662.032	253.177	254.422
Percentage (%)	-0.00106	0.007233	-0.02751
	0.00906		

Table 5.6 Prediction of grey water footprint at Bukit Ubi WTP

	Bayesian Networks		
Years	2015	2016	2017
Actual	662.008	252.997	254.484
Predicted	662.012	253.002	254.483
Percentage (%)	0.0005766	0.0005766	-0.000073
	0.000723		