Non-Revenue Water (NRW): Strategy on Reducing and Managing Water Losses in Johor

by

Intan Safwanah binti Suliaman
18000797

Dissertation submitted in partial fulfilment of
the requirement for the
Bachelor of Engineering (Hons)
(Civil Engineering)

SEPTEMBER 2022

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Non-Revenue Water (NRW): Strategy on Reducing and Managing Water Losses in Johor

by

Intan Safwanah binti Suliaman
18000797

A project dissertation submitted to the
Civil Engineering Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF ENGINEERING (Hons)
(CIVIL ENGINEERING)

Approved by,

____
(Dr Husna binti Takaijudin)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
SEPTEMBER 2022

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

INTAN SAFWANAH BINTI SULIAMAN

ABSTRACT

One of the biggest issues with a distribution network's ability to supply water is water loss. The country of Malaysia has routinely reported higher than 35% during the past ten years. Water supply system water loss was caused by non-revenue water (NRW). As a result, the issue has received significant attention from the water delivery companies to boost supplies. Each state's water supply business has taken, and still needs to take, a variety of steps to reduce and regulate water loss, notably the NRW issues. As a result, this research project's objective focus is on calculating the proportion of NRW in the chosen areas of Johor and developing an appropriate model to oversee the NRW. To fulfil the objectives, data such as product intake, billing, and leakage will be analysed. Results have demonstrated that both tests have advantages and disadvantages in terms of eradicating the prominent level of NRW in Johor. The project has been conducted in two Johor districts which are Batu Pahat and Johor Bahru. In terms of NRW percentage, the difference between urban and rural areas is a part of project outcome. From the data that has been collected for this research it shows that Johor Bahru recorded about 20% to 25% of NRW while Batu Pahat recorded 35% to 40% of NRW. Therefore, Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) was conducted as the model of NRW cases in Johor.

ACKNOWLEDGMENT

I am delighted and grateful to everyone who has contributed to this project from the beginning to the end. The success and final result of this project required a lot of direction and help from many people.

I would like to use this time to offer my sincere gratitude to Dr. Husna Binti Takaijudin, my supervisor from Universiti Teknologi Petronas, for her recommendations, counsel, and the chance for me to complete the project successfully. Additionally, I would like to express my gratitude to Mr. Zainudin bin Ngadiran, Head of the NRW Control Department at Ranhill SAJ Sdn Bhd, and Miss Nisa Batrisyia binti Fadzil, Executive of the NRW Control Department at Ranhill SAJ Sdn Bhd, for all of their assistance, advice, suggestions, and comments that helped me complete the project in line with the current NRW case situation in Johor. Not only that, but I would also like to say thank you for all the data and information provided, as it is applicable for me to do all the testing and modelling designation for this project. In addition, I also want to say thank you to all the lecturers for their guidance carried me through all the stages of writing my project.

Last but not least, I want to express my gratitude to everyone who has helped and supported me throughout the course of this project. I have received a wealth of information that is really helpful for both my theoretical and practical studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Every living thing depends on water for survival on a regular basis. According to Murugan and Chandran (2019), water is the most important natural resource. The demand for water has increased globally as a result of factors including fast population expansion and economic development. Natural water sources have lost some of their dependability in terms of availability. As a result, it has emerged as one of the fundamental resources for all life on Earth. The water distribution firm has therefore offered a variety of services to please the customer in order to guarantee that the water supply is distributed fairly to all clients. However, the cost of the water supply appears to rise periodically, which has resulted in a number of incidents like water theft. The water distribution firm is now concerned about a number of incidents, like leaks and pipe bursts. These numerous instances have contributed to the production value and billed consumption amount against the water supply corporation. This water loss is referred to as non-revenue water (NRW).

Non – Revenue Water (NRW) is shortly defined as the water supplied from the water treatment plant but cannot be traced from the consumers' bills. It can also be defined as the difference between the water distributed to consumers and billed water. This has become a genuine issue because, in terms of increasing the water supply capacity, the NRW needs to be reduced, and the cost is quite expensive. Malaysia itself has recorded more than 30% of NRW cases in the past few years. NRW is recommended to be at least less than 25% by the water supply companies in the world.

Non – Revenue Water (NRW) is caused by three main components: physical losses, commercial losses, and unbilled authorized consumption. Physical losses usually happen because of a pipe burst, leaking, services, and tank overflow. While, commercial losses are due to illegal connections, meter issues billing errors and water theft. Last but not least, unbilled authorized consumption is used that cannot be billed for firefighting purposes. Nevertheless, NRW still has difficulty tracing because of leaking that might happen from an underground pipe.

Henceforth, NRW cases has becoming one of the crucial issues among the water supply companies around the world. This is because this issue is giving huge effect towards the water quality, water demand and the production of the water itself. Not only that, but water also known as one of the biggest resources that give a massive impact towards the country's economy. So, with the issues of NRW it can threaten the world economic. For that reason, a lot of method and system is still be researched and some of them has been used by the water supply companies in furtherance of controlling, managing, and reducing the NRW issues in their respective country.

1.2 Problem Statement

Non – Revenue Water (NRW) is the main issue for every water supply company in Malaysia. Significant factors lead to these issues, such as the increasing population, generating the high demand for water supply and water losses. Furthermore, the leakage cases that happen infrequently can affect the quality and quantity rate of the consumers. When the NRW cases rose, the company responsible for supplying the water had to spend a higher amount on operation costs for the water treatment process. Leakage, pipe burst, water theft is one of the reasons that can increases the cases and rate of NRW, not only in Malaysia but worldwide. Other country that has high population is the most unfortunate country when it comes to solve the NRW cases at their place. No one can deny that water losses are the biggest issues that need to be face by everyone, especially the government.

This is because the government is the one that responsible to ensure that all the people in their respective country can get enough supply for daily use. Water itself is one of the crucial resources that must have by every living thing in the entire world. According to Sururhanjaya Perkhidmatan Air Negara (SPAN), NRW average for Malaysia since 2017 to 2021 is exceeding 30%, whereas in year of 2018 has recorded the highest percentage of 33.9 % while in year of 2019 was recorded the lowest percentage of 33.2 %. This has shown that the NRW percentage is higher than the benchmark that has been recommended by the World Bank which must be less than 25%. Hence, the water losses cases in Malaysia are considered as severe conditions and need to be reduced so the target percentage is achieved.

In addition, the water supply distribution has become an essential problem towards the society and environment. According to Chee, Ngai and Ranjan (2017), Malaysia is a country that rich in water resources. However, the country still faces seasonal water supply problems and high rate of NRW that contributes the problems. Meanwhile, Şişman and Kizilöz (2020) say that Metropolitan cities such as Istanbul, Ankara and İzmir in Turkey have difficulties meeting water demand growth because of limited freshwater resources and depletion from the climate change impact, migration, heavy losses due to inconvenient infrastructure and water quality deterioration. Meanwhile, majority of Middle East countries are water stressed. Globally, seven out of ten countries in the Middle East are expected to face the highest water stress by year 2040 and there are thirteen regional nations are now referred to as ‘water scarce’.

According to Chawira, Hoko and Mhizha (2022), based on the study period of Zimbabwe in 2020, the NRW range is increases from 13% to 42% and averaged 36%. Whereby, real losses were estimated at 83% and apparent losses were 17% of NRW cases for that country. Meanwhile, Nigeria has indicated that the level of NRW ranges from 36% to 50% higher than the global average of 35% (Nasara, Zubairu, Jagaba, Azare, Yerima & Yerima, 2021). This has shown that the NRW cases around the world is severe and need to be taken into consideration as it is affecting the water demand

and the country’s economy. Therefore, a lot of research, method and system is being develop in each country in order to resolve the issues.

Last but not least, to achieve this project goal, many measurements need to be done so it will go smoothly and successfully. All the approaches that have been completed and are still ongoing by the company involved minimizing the NRW rate in problematic areas in the matter of avoiding the loss that the company will receive along with the water supply issues to the consumer. All the system, such as District Meter Area (DMA) and Geographical Information System (GIS) is used to oversee the NRW rate in Malaysia. Focusing on the Johor Bahru and Batu Pahat, this project report mainly discusses the system being introduced to achieve the goal and satisfy the objective of the project.

1.3 Objectives

The overall aim of this project is to achieve a better understanding of the extent of non-revenue water management and ways to oversee it. This project will focus mainly: -

1) To compare NRW percentage within two selected areas
2) To identify the possible approach that can reduce and manage NRW
3) To design a suitable model or method for handling NRW

1.4 Scope of Study

From this project, the research is being done by investigating the research based on the topic. After getting the understanding about the project, the methodology and objectives is planned. Two district area in Johor also been chosen to conduct the research. Later, the data is being collected and analysis accordingly. The analysis that will be used for this project is water losses percentage calculation, step test method and leakage percentage. The suitable model needs to be construct after the analysis has been done. The model that will be in the project is Multiple Linear Regression (MLR) and Artificial Neural Network (ANN). The model designation will be done by using Microsoft Excel and MatLab, respectively. Finally, is to identify the most suitable approach that can be used to fulfil the objectives.

CHAPTER 2

LITERATURE REVIEW

2.1 Concept of Non – Revenue Water

Water treated and produced by a water treatment plant that does not reach consumers and generates no income for the water distribution company is referred to as "water losses," or "Non-Revenue Water" (NRW) (Abdul, Zakaria, Zulkifli & Roslan, 2021). Rajasekhar, Ramana, and Viswanadh (2018) further emphasise that NRW is the fraction of water that the water company distributes into the distribution system but is unable to recoup its costs. According to Rajeskhar et al. (2018), important levels of NRW are causing the water utilities to lose revenue, which has an impact on their ability to remain financially viable. This has resulted in significant waste for the business, forcing it to incur losses and raise the water bill, burdening customers. The high NRW volume has revealed inadequate water system management by the water supply providers. However, the problems have gotten worse as a result of an increase in the demand for water supplies brought on by population development.

As a result, the majority of water distribution companies have conducted numerous studies to determine the NRW reduction programme in order to address the NRW difficulties. Unbilled approved usage, physical losses, and commercial losses are the other three key components of NRW. The use of "Vicious Circle" has been used to address two factors—physical losses and commercial losses—as the causes of subpar firm performance. Therefore, they also established the "Virtuous Circle" in order to break up this "Vicious Circle."

Numerous obstacles must be overcome in order for the "Vicious Circle" to become the "Virtuous Circle." This may limit NRW releases of fresh water and financial sources (Farley, Wyeth, Zainuddin, Arie, and Sher, 2011). The Vicious Circle and Virtuous Circle are seen in Figures 2.1 (a) and (b), respectively.

FIGURE 2.1: (a) Vicious Circle (b) Virtuous Circle (Farley, M., Wyeth, G., Zainuddin, M.G., Istandar, A. and Singh, S., 2008)

According to Table 2.1, which displays statistics on NRW volume for Peninsular Malaysia and the Federal Territory of Labuan from 2017 to 2021, the majority of the Malaysian state under study would experience an increase in NRW volume measured in millions of litres per day (MLD) during the year 2020. Additionally, while NRW volume in Selangor state reached elevated levels that may be considered serious, NRW volume in the Federal Territory of Labuan was under control.

TABLE 2.1: Non – Revenue Water (NRW) (Suruhanjaya Perkhidmatan Air Negara, SPAN., 2021)

Unit (MLD)	2017	2018	2019	2020	2021
Johor	433	452	470	507	463
Kedah	652	701	715	719	768
Kelantan	234	242	255	262	265
F.T Labuan	23	23	23	28	32
Melaka	101	112	114	173	188
N. Sembilan	251	249	251	253	262
Pulau Pinang	231	233	251	261	266
Pahang	526	569	566	638	632
Perak	406	407	400	419	410
Perlis	152	159	160	166	158
Selangor	1526	1540	1461	1417	1373
Terengganu	186	225	233	241	225
Pen. Malaysia & F.T. Labuan	4721	4912	4899	5084	5042

2.2 Components of Non – Revenue Water

Based on the system that has been standardised by the International Water Association, the system input value of Non-Revenue Water (NRW) can be separated into two main components (IWA). This major will be broken down into two further elements. Water losses and authorised consumption are key factors. Water losses are classified as commercial and physical losses while minor components under authorised consumption are characterised as billed and unbilled. With the exception of billed authorised consumption, all of the smaller components are connected to the NRW concerns. Figure 2.2 displays these components' typical water balances and terminology.

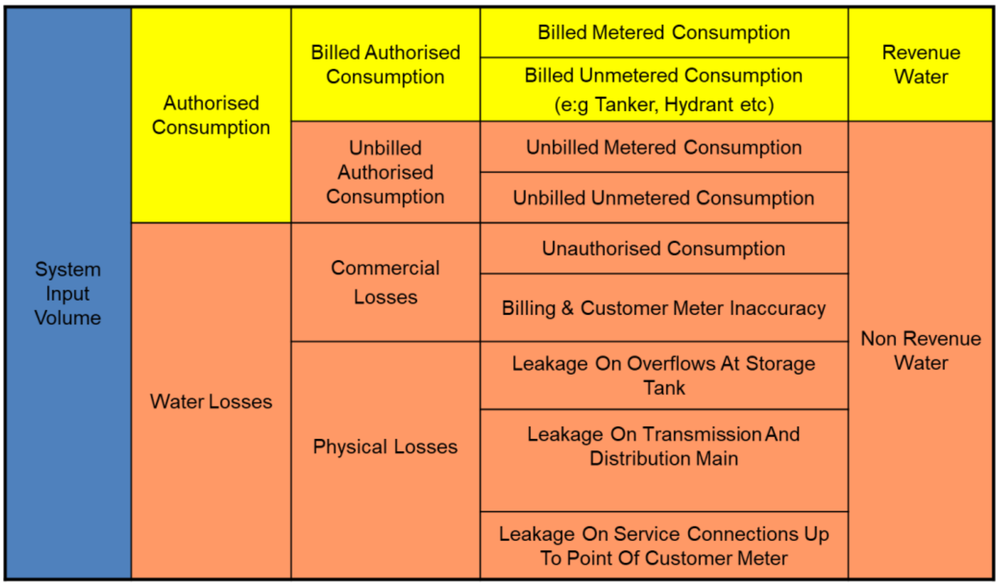

FIGURE 2.2: IWA Standard Water Balance and Terminology (Murugan and Chandran, 2019)

Water losses can be divided into two categories: commercial and physical losses. According to Rajasekhar, Ramana, and Viswanadh (2018), obtaining the records that are readily available and entering the information into the water audit worksheet is how the preliminary water loss assessment is obtained. After entering the information, the water balance compared the input volume to the distribution system with the total of customer usage and predicted losses. The mind map of typical water supply system losses is displayed in Figure 2.3.

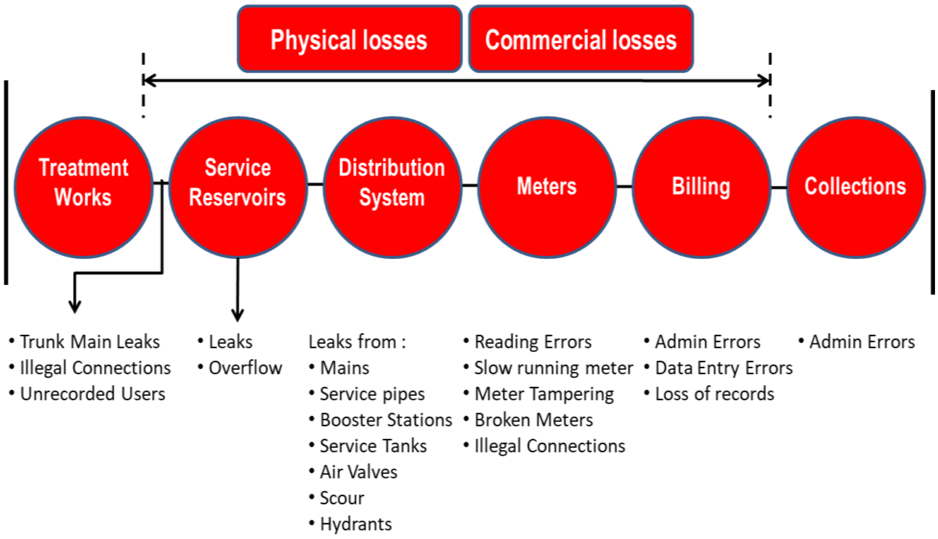

FIGURE 2.3: Typical losses from a Water Supply System (Selek, Adigüzel, Iritaş, Karaaslan, Kinaci, Muhammetoglu, & Muhammtoglu, 2018)

Real loss, also known as physical loss, is the physical loss of water from the water supply system, which includes all kinds of leaks, bursts, and overflows in service reservoirs, mains, and service connection pipes up to customer metering (Fanner, 2004). The leakage on overflows at storage tanks, transmission and distribution hubs, and service connections up to the customer metre are examples of physical losses, as shown by Figure 2.2. However, Figure 2.3 makes it clear that the two usual principal losses associated with physical losses are service reservoirs and distribution networks. The leakage instances from 2017 to 2021 in Johor are shown in Table 2.2. All major and minor cases that occurred in Johor are included in the number of cases.

TABLE 2.2: Total Leakage Cases in Johor

Year	**Number of Cases**
2017	105, 162
2018	121, 047
2019	118, 253
2020	117, 524
2021	133, 347

Unauthorized consumption, inaccurate metres, such as those used in illegal water connections and data handling problems, are what constitute apparent loss or commercial loss (Lambert & Hirner, 2000). Figures 2.2 and 2.3 illustrate how typical and standard losses relate to business losses, respectively. Both numbers demonstrate that metering and billing problems are the root cause of economic losses.

Billed and unbilled consumption make up the two halves of authorised consumption. Water uses that are identifiable and reflected in the users' bills are referred to as approved consumption that is billed. Unbilled approved consumption, on the other hand, refers to water usage that is not recognised by the metre reader and cannot be billed to the consumers. It is more concentrated on the unbilled approved consumption instances for Non-Revenue Water (NRW) cases. According to Figure 2.2, there are two categories of unbilled authorised usage: unbilled metered consumption and unbilled unmetered use.

According to Lambert and Hirner (2000), unbilled allowed usage includes water utilised for firefighting, water provided gratis to particular customer groups, and water used by water utilities for operational purposes like flushing and cleaning. This goal is essential for providing clean water and saving lives. Therefore, none of these purposes can be listed under billing consumption. Additionally, it has significantly contributed to the losses suffered by NRW.

2.3 Issues and Challenges in Reducing Non – Revenue Water

When tackling inadequate service coverage levels and rising demand for piped water delivery, Frauendorfer and Liemberger (2010) suggested that lowering NRW should be the first goal. Without managing water losses, expanding water networks will only result in a vicious circle of waste and inefficiency. The NRW challenge, according to Farley, Wyeth, Zainuddin, Istandar and Singh (2008), can only be properly comprehended if NRW and its components are quantified, the right performance measures are determined, and the lost water volume is converted into its equivalent economic value. Each NRW component's size is shown when the water balance develops. According to Farley et al. (2008), lowering NRW also opens up new financial and water sources. Reducing the excessive physical losses decreases operational expenses, produces a large amount of water that may be consumed, and delays the need to invest in news sources. More money will be made while commercial losses are reduced. Frauendorfer and Liemberger (2010) concur that a single project cannot address the problem of lowering NRW. Since institutional reforms take time to implement, pipe replacements alone will not be sufficient. There needs to be ongoing engagement and communication with the governments and water supply companies.

A study on NRW management was conducted by Farley et al. (2008) to comprehend the water losses. Not every nation or region in the globe, especially Asia, has the operational framework and infrastructure needed to administer NRW. There are numerous battles fought to guarantee that customers receive a good water supply to support health and life. Asia will provide more difficult issues for water utility firms, such as fast urbanisation, dwindling water supplies, environmental contamination, inadequate operations and maintenance procedures, a higher frequency of commercial losses, and many others. As a result, numerous strategies must be used to minimise and control the problems and challenges associated with cutting NRW in order to satisfy all parties affected by the issue.

2.4 Strategies to Manage Non – Revenue Water

According to Farley, Wyeth, Zainuddin, Istandar, and Singh (2008), the team must make sure that all NRW components are covered and that the intended plan is workable in terms of physical implementation and financial requirements. Setting proper NRW reduction targets is the first step in managing NRW. The NRW target is frequently chosen arbitrary without taking into account the financial consequences or if it is even feasible, claim Farley et al. (2008). As a result, in order to implement the strategies, the team must first determine the economic level of NRW in order to set the beginning aim and then compare the price of water lost to the price of conducting NRW reduction activities. The components of NRW reduction should be prioritised as the next tactic. By employing this method, reduction will be attained in the most efficient manner possible. This tactic is primarily concerned with the associated monetary values.

Additionally, by comparing the NRW baseline with the target level, the utility managers must determine the intended volume that can be conserved. The strategy then needs to take into account the implementation budget. The utilities manager must therefore list a few costs while creating the budget, including those for labour, supplies, equipment, and vehicles. This is so that all that must be done to lower the NRW may be conducted under budgetary constraints. Last but not least, all deeds must adhere to the laws that have been established by the government so that no one can avoid facing punishment for breaking the rules.

2.5 Existing Solutions to Manage Non – Revenue Water

There are only a handful remedies that have been implemented by the majority of nations worldwide. Inferring the fundamental principle of NRW reduction is one of the current solutions. The three steps of Awareness, Location, and Repair (ALR) time make up the fundamental tenet of NRW reduction. Any loss resulting from leaks, overflows, defective customer metres, or other causes will, according to this theory, go through these three steps (Farley, Wyeth, Zainuddin, Istandar, & Singh, 2008).

Three things are necessary for this stage: the time it takes to become aware of the leak, the time it takes to find the leak, and the time it takes to fix the leak. The District Meter Area (DMA) will then be implemented.

District Meter Zone (DMZ) is another name for DMA. This can assist in tracking the degree of water network leakage that can be identified in a particular DMA. By utilising technology like GPRS modules, transmitters, and automatic systems made to record the necessary data, smart metres can provide a lot of efficiency (Murugan & Chandran, 2019). Therefore, with the implementation of DMA, it will be simple to assess the extent of leakage and prioritise where leaks should be located, and these details will be stored in the GIS system under pipe burst situations. To construct a preliminary DMA design, certain criteria must be met. The DMA establishment flow is depicted in Figure 2.4.

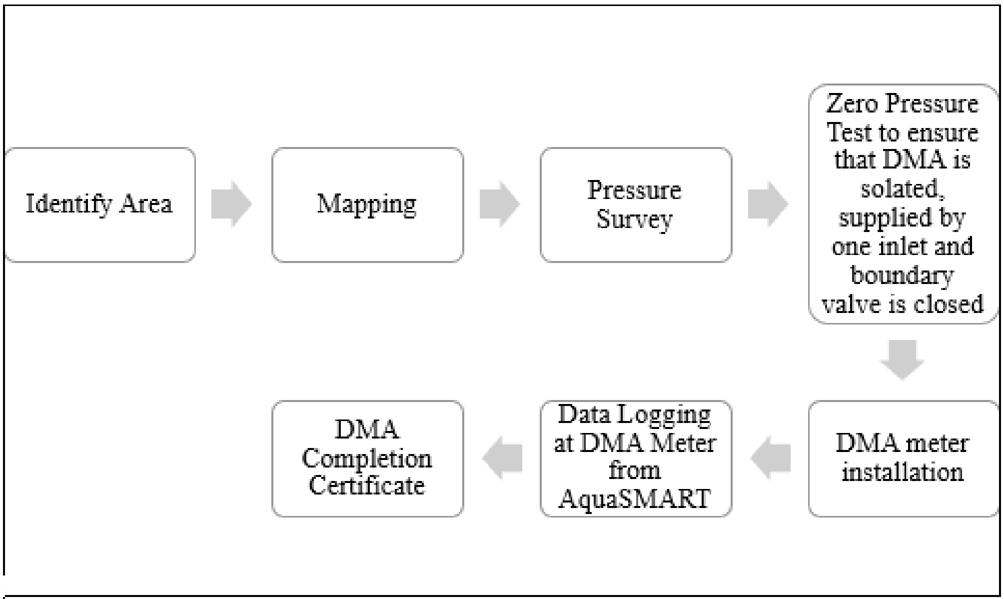

FIGURE 2.4: Flow for District Meter Area (DMA) Establishment

2.6 New Approach of Managing Non – Revenue Water

In order to identify NRW situations, a number of nations, including Egypt and Korea, have examined a number of diverse ways. Artificial Neural Networks (ANNs) and Multiple Linear Regression (MLR) are the methods used (ANN). According to Jang and Choi (2017), MLR is the most generally used approach for predicting the NRW ratio, and formulas vary depending on the employed factors and area characteristics. While this is happening, the ANN is a model that is anticipated to guide the repair and enhancement of water distribution networks and is also appropriate for DMA construction (Jang & Choi, 2017). Abdul, Zakaria, Zulkifli and Roslan (2021) added that factors including the quantity of connections, production, and consumption have a substantial impact on non-revenue water.

The MLR model can therefore be used for NRW scenarios. Elkharbotly, Mohamamed, and Abdelkawi (2021) claim that in the meantime, in order to facilitate communication, ANN obtains a complete set of neurons that are coupled in parallel patterns. Additionally, they said that the success of ANN is based on its ability to define, comprehend, and forecast distinct problems across a range of fields. Consequently, it is clear enough why nations like Egypt and Korea are researching ANN as one of the models to detect the NRW. But because MLR and ANN are not fully understood, this strategy is not extensively employed and is currently being studied in some countries. Malaysia is one of the nations that might be eligible to apply the model to determine the NRW based on various parameters required to develop the MLR and ANN model. In order to further analyse Malaysia's water distribution systems and investigate the factors that contribute to NRW, Abdul et al. (2021) performed study on MLR.

CHAPTER 3

METHODOLOGY

3.1 Research Methodology

Overall, the methodology offered is crucial for conducting NRW's study project. Every technique employed, from site identification through model design, is described in the appropriate manner.

3.1.1 Site Identifications

In order to better comprehend Non-Revenue Water (NRW) issues, Batu Pahat and Johor Bahru districts in Johor have been selected. Only a few data points, such inflow (m3) and billing, were obtained from the site (m3). The volume and proportion of NRW must be determined using this information. Other information was also required, such as population data and Minimum Night Flow (MNF), in order to choose the best method and model for the project. Based on information about water inflow (m3) and billing (m3), a district is selected. As a result, the statistics and graph demonstrate the considerable gains in value throughout the entire Johor state region.

Batu Pahat

The address of Batu Pahat, Johor, is 1°51′00′′North 102°56′00′′East. Approximately 150 DMA have been allotted to 11 administrative areas in Batu Pahat, according to the District Meter Area (DMA). Although Batu Pahat is classified as a mixed rural and urban region, the majority of the area is rural. More than 300,000 people call Batu Pahat home, and there are more than 100,000 houses there. One of the Johor districts still in the development stage is Batu Pahat. The map of the Batu Pahat, Johor district area is depicted in Figure 3.1 (a).

FIGURE 3.1 (a): Batu Pahat Map

Johor Bahru

The location of Johor Bahru, Johor, is in latitude 01°27′20′′ and longitude 103°45′40′′. Johor Bahru has allotted around 551 DMA (as of the year of the 2022 database) to 15 administrative areas, according to the District Meter Area (DMA). An urban area is classified as Johor Bahru. More than 500,000 people call Johor Bahru home, and there are more than 300,000 dwellings there. The capital of Johor is also known as Johor Bahru. The map of the Johor Bahru, Johor district is shown in Figure 3.1 (b).

FIGURE 3.1 (b): Johor Bahru Map

3.1.2 Water Losses Calculation

Non-Revenue Water (NRW) is produced water that has not brought in any money for the utility. Water utilities are required to supply free water for a variety of functions, including firefighting. Excellent social care is provided by this service. Since the utility receives no money from this, it can also be classified as NRW. NRW is represented as a percentage of all the produced water. The equation involved to calculate the level of NRW in the system is shown below:

$$NRW\ Volume\ (m^3) = Inflow\ (m^3) - Billing\ (m^3) \qquad [1]$$

$$NRW\% = \frac{Inflow\ (m^3) - Billing\ (m^3)}{Inflow\ (m^3)} \times 100\% \qquad [2]$$

Data on non-revenue water (NRW) was gathered and examined using three different techniques to establish the NRW trend from the years 2017 to 2021, broken down by month. The calculation of the NRW %, is one of the techniques that being used by all water supply company around the world. The calculation is being done by using Microsoft Excel software. Johor Bahru and Batu Pahat, two districts in Johor, are the subjects of the comparison. Table 3.1 (a) shows the classification parameter that being used to collect the data. While Table 3.1 (b) is to describes the variables used.

TABLE 3.1 (a): Classification Parameter Data Collection

Classification	**Parameter**	**Data Collection**
Operational parameter	Billing consumption	Simulated data
	Non-revenue water volume	Measured data
	Water supply quantity	Simulated data

TABLE 3.1 (b): Variables Description

Variables	Description
Non-revenue water	The measurement of water losses (meter cube per day and percentage)
Production quantity	Total volume of treats water that are distributed from the water supply (meter cube per day)
Consumption quantity	Total volume of treats water that passes the consumer's registered (meter cube per day)

The standard and permissible percentage rates for each level are shown in Table 3.1 (c) below is based on the requirements from each research country. From the table, it is possible to identify and categorise the NRW situations according to their distinct levels and the suggested handling strategies. The table is crucial since it can serve as the benchmark for calculating water losses. This table can be used to debate and categorise the NRW proportion for each district under study.

TABLE 3.1 (c): Percentage range of NRW cases

Percentage Range (%)	Level	Description
Below 0	Under Control	Only need to do regular services
0 – 30	Optimal	Do inspection from time to time
30 - 60	Alarming	Need more observation and control
More than 60	Dangerous	Need extra monitoring and inspection

3.1.3 Step Test Method

One technique is the step test, which involves covering a portion of the suspected region without impacting water distribution in order to find leaks in a big DMA. Typically, the Step Test Method is used to calculate the physical losses in NRW scenarios. All boundary valves must be closed in the DMA region, and the state of each valve within that area must be determined. When the water flow originates from a single metered source, the step test can be performed. To determine the Minimum Night Flow (MNF) value, step tests must be conducted at night. The leakage that occurs in the DMA is also computed using the step test data. This operation typically begins at 12:00 am, and it lasts for 12 hours. This test's primary objective is to pinpoint the troublesome areas just where the excessive leakage happened. Additionally, this test reduces the amount of time needed to locate leaks so that leak detection may be done more affordably. Boundary valves (BV), circulating valves (CV), step valves (SV), and redundant valves (RV) are the valves used in this step test. The stage test method's flowchart is shown below. The Step Test Method flowchart is depicted in Figure 3.1 (c).

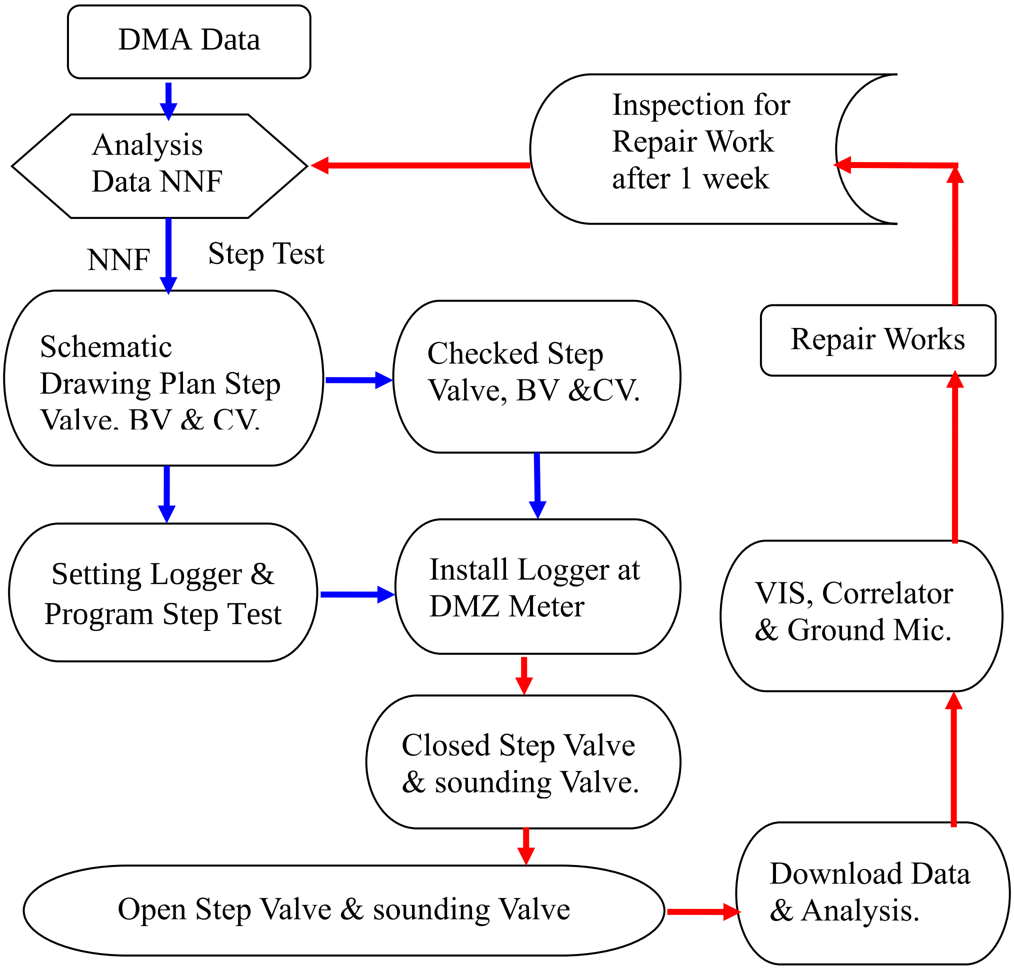

FIGURE 3.1 (c): Flow Chart of Step Test

3.1.4 Calculation of Leakage Percentage

Leakage percentage is important in terms of determined the non-revenue water reading. Leakage percentage can be calculated using the step test method that being conducted at night which is MNF. Table 3.1 (d) shows the Standard Leakage Index. Below is the step of computation the leakage percentage.

1. The Minimum Night Flow to zone, MNF was determined from base 24 hours graph of MNF (unit: 1/s).
2. Unit of MNF was converted from step 1 into 1/c/h:

   MNF = (MNF/C) x 3600

   Where, C = Number of connections at the DMA
3. Legitimate Night Flow, LNF was determined (from field reading), (unit: 1/c/h).
4. The current Net Night Flow, NNF was calculated (unit: 1/c/h).

   NNF = MNF (from step 2) – LNF
5. Total leakage in zone, L was calculated (unit: m^3/d).

   L = (NNF x T Factor x C) x 1000

   Whereby,

   T = $\frac{Total\ Leakage\ Index}{Highes\ Leakage\ Index} \times 2\ hours$

   T = time factor

   C = Number connections at the DMA

   Leakage Index can be referred from Table 4.1.
6. The Total Flow to zone, F was determined based on 24 hours graph of MNF (unit: m^3/d).
7. Current Leakage Percentage, CLP was calculated

   CLP = $\frac{L}{F} \times 1000$

   Whereby,

   L = Total Leakage in zone (from step 5)

   F = Total Flow to zone (from step 6)

TABLE 3.1 (d): Standard Leakage Index (Source: Ranhill Water Services)

Average Area Night Pressure	Leakage Index	Average Area Night Pressure	Leakage Index	Average Area Night Pressure	Leakage Index
10	6	31	20.5	52	39
11	7	32	21	53	40
12	7.5	33	22	54	41
13	8	34	23	55	42
14	8.5	35	23.5	56	43
15	9	36	24	57	44
16	10	37	25	58	45
17	10.5	38	26	59	46
18	11	39	26.5	60	47
19	11.5	40	27	61	48
20	12	41	28	62	49
21	13	42	29	63	50
22	13.5	43	30	64	51
23	14	44	31	65	52
24	15	45	32	66	53
25	16	46	33	67	54
26	16.5	47	34	68	55
27	17	48	35	69	56
28	18	49	36	70	57
29	19	50	37	71	58
30	20	51	38		

3.1.5 Design Model

Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) has been chosen as one of the models that suitable to define the determinants of Non-Revenue Water (NRW). Both models can be construct by using the MATLAB or Excel software. To ascertain the relationship between the two key parameters used for the model, two additional tests must be performed before completing the MLR model. The additional test are Pearson correlation and single regression analysis.

Pearson Correlation Analysis

According to Srivastav (2022), Pearson correlation analysis is a test used to measure the strength between the different variables and their relationship. The relationship can be defined as:

$$r = \frac{n(\Sigma\, xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\,\Sigma x^2 - (\Sigma x)^2][n\,\Sigma y^2 - (\Sigma y)^2]}} \qquad [3]$$

Whereby,	r	= Pearson Coefficient
	N	= number of the pairs of the stock
	∑xy	= sum of products of the paired stocks
	∑x	= sum of the x scores
	∑y	= sum of the y scores
	∑x2	= sum of the squared x scores
	∑y2	= sum of the squared y scores

The value of Pearson's correlation coefficient, r, typically ranged from -1 to 1. A strong and perfect negative correlation between the variables is indicated if the correlation coefficient is -1. If the correlation coefficient is 1, it means that there is a strong and perfect positive association between the variables, while a correlation coefficient of 0 shows that there is no relationship.

In addition, Pearson correlation analysis can also be used to evaluate either the relationship between two variables is significant or insignificant (Turney, 2022). In order to determine the significant, test statistic (t statistic) need to be conducted. The t statistic value can be determined by using formula:

$$t_{statistic} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \qquad [4]$$

Where,	r	= Pearson's correlation coefficient
	n	= Number of Observation

Next, the test critical (t critical) is calculate in order to compare with the t statistic that has been calculated before. The t critical can be determined by using the degrees of freedom, df, and significance level, α. The common tails for t critical are two – tailed correlation. The t critical is calculate by using the Microsoft Excel software. Hence, t statistic and t critical value is compared to accept or reject the null hypothesis. If t statistic greater than t critical, thus the relationship is statistically significant, and the null hypothesis can be rejected. While, if the t statistic is less than t critical, therefore the relationship is not statistically significantly, and it does not allow to reject the null hypothesis.

Simple Linear Regression (Single Regression Analysis)

According to Bevans, simple linear regression is used to estimate the relationship within two quantitative variables in terms of determine the strength relationship between variables as well as value of dependent variable at a certain value of independent value. Simple linear regression is basically a parametric test that able to makes certain assumptions regarding of the data.

First assumption is based on the homogeneity of variance (homoscedasticity), where the size of error in prediction does not significantly across the independent variable's values. Second assumption is based on independence of observations that been collected using statistically valid sampling methods and no hidden relationship among the observations. Third assumption is the normality that follow the normal distribution. Last assumption, the relationship between the dependent and independent variable is linear whereas line of best fit is a straight line.

Therefore, if the data conducted do not meet the assumptions of homoscedasticity or normality, the relationship need to use nonparametric test, such as Spearman rank test. Simple linear regression is calculate using Microsoft Excel software.

Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR) is an analytical technique that be used in terms of estimate the relationship between two or more independent variables and one dependent variable (Bevans, 2020). The formula used for the multiple linear regression model can be expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon_i \qquad [5]$$

Whereby,	Y	= Non – Revenue Water
	β_0	= Regression intercept
	$\beta_{1,2,3,4}$	= Regression Coefficient
	X_1	= Production Quantity
	X_2	= Customer Billing Quantity
	X_3	= Others Quantity
	X_4	= Operational Use Quantity
	ε_i	= Error

There are several conditions that needed to be identify in order to construct the best model of MLR one of it by finding the best line for each independent variable. MLR model is determine by using Microsoft Excel.

Artificial Neural Network (ANN)

Arash and Nastaran (2021) says that Artificial Neural Network (ANN) is a model that involves computations and mathematics that stimulate the human-brain process. They also stated that static ANN model is known as a multilayer perceptron neural network model, dynamic neural network models and statistical neural network models. According to Jang and Choi (2017), the ANN can be expressed as:

$$y_j = f\left(\sum_{i=1}^{n} w_{ji} x_i\right) = \frac{1}{\left(1 + e^{-net_j}\right)} \quad [6]$$

Whereby, the input for each neuron j in the hidden layer is the sum of the weighted input signal x_i ($\sum_{i=1}^{n} w_{ji} x_i = net_j$, in which w_{ji} is the interconnecting weight between neuron j in the hidden layer and neuron i in the input layer). The output y_j from the neuron based on the equation above.

There are several conditions that needed to be identify in order to construct the best model of ANN one of it by determine the neurons scatter from the output and input value. The ANN model is constructed using MatLab software.

MLR and ANN Variables

Non – Revenue model is models that used as the machine learning techniques to assists the NRW cases. Models used for this study are Multiple Linear Regression (MLR) model, and the Artificial Neural Network (ANN) model are the three techniques now in use. Microsoft Excel and MatLab are the two pieces of software that are being used for this investigation for both chosen district in Johor. Table 3.1 (e) shows the classification parameter that being used for model designation while Table 3.1 (f) used to describe the variables used for the model.

TABLE 3.1 (e): Classification Parameter for Model Designation

Classification	Parameter	Data collection
Operational parameter	Billing consumption	Simulated data
	Non-revenue water volume	Measured data
	Water supply quantity	Simulated data
	Operational Use quantity	Simulated data
Physical parameter	Leak Quantity	Simulated data
	Unbilled Consumption	Simulated data

TABLE 3.1 (f): Variables Description for Model Designation

Variables	Description
Non-revenue water	The measurement of water losses (%)
Production quantity	Total volume of treated water that are distributed from the water supply (m^3/month)
Consumption quantity	Total volume of treated water that passes the consumer's registered (m^3/month)
Other Bill Quantity	Total volume of treated water that do not passes the customer’s registered and operational use (m3/month)
Operational Use Quantity	Total volume of treated water that being use for water valve services and fire hydrant (m3/month)

3.2 Project Activities

3.2.1 Data Collection in Study Area

There are several methods can be used in order to collect the data used for NRW. The data can be collected by using the DMA method, water system and step test method. Data that need to be collect is water inflow (m3) and billing (m3), as well as data that necessary for the model designation.

3.2.2 Observations in Study Area and Document Analysis

Observation in Study Area

Observation is an important method for this project as in identifying factors of non-revenue water (NRW) that occur in the study area by using DMA method. This method used to detect a reasonable and acceptable solutions for reducing and controlling NRW. From this method, any changes that effected the NRW rate can be discovered and identified. Thus, the result and data from the observations is collected.

Document Analysis

Several document that related with the project has been identified and analysed. Document related is leakage data, number of consumer and connection, and NRW data. All data has been collected and provided by the company. The data is collected from 2017 to 2021.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Non - Revenue Water

4.1.1 Non – Revenue Water Percentage

Non – revenue water was analysed by using the water audit method. The data used for NRW calculation were inflow and billing consumption. NRW percentage is important in order to identify the level of NRW in terms of controlling and overseeing the issues. Formula used to calculate the NRW is by using equation [1] and [2] that has been defined in the previous chapter. Appendix VII shows the sample calculation for both Johor Bahru and Batu Pahat. The determination of the production, consumption, and NRW volume trend in the Johor Bahru and Batu Pahat areas, respectively, is shown in Figures 4.1 (a) and (b) below. The NRW% 12-month average from July 2017 to February 2020 showed the lowest gradient on the Johor Bahru trend graph, as opposed to the high value range from February 2020 to April 2021. In contrast, the production component in Johor Bahru shows a considerable increase between August 2019 and January 2020, followed by a decrease, and increase in both 2020 and 2021. As for Johor Bahru, client billing is drastically declining from the first quarter of 2020 to the last month of 2021. The fact that the production exceeds the consumer billing is really concerning. This demonstrates how risky leaking or unauthorised cases were at the time because they fuelled an increase in NRW cases in the Johor Bahru area.

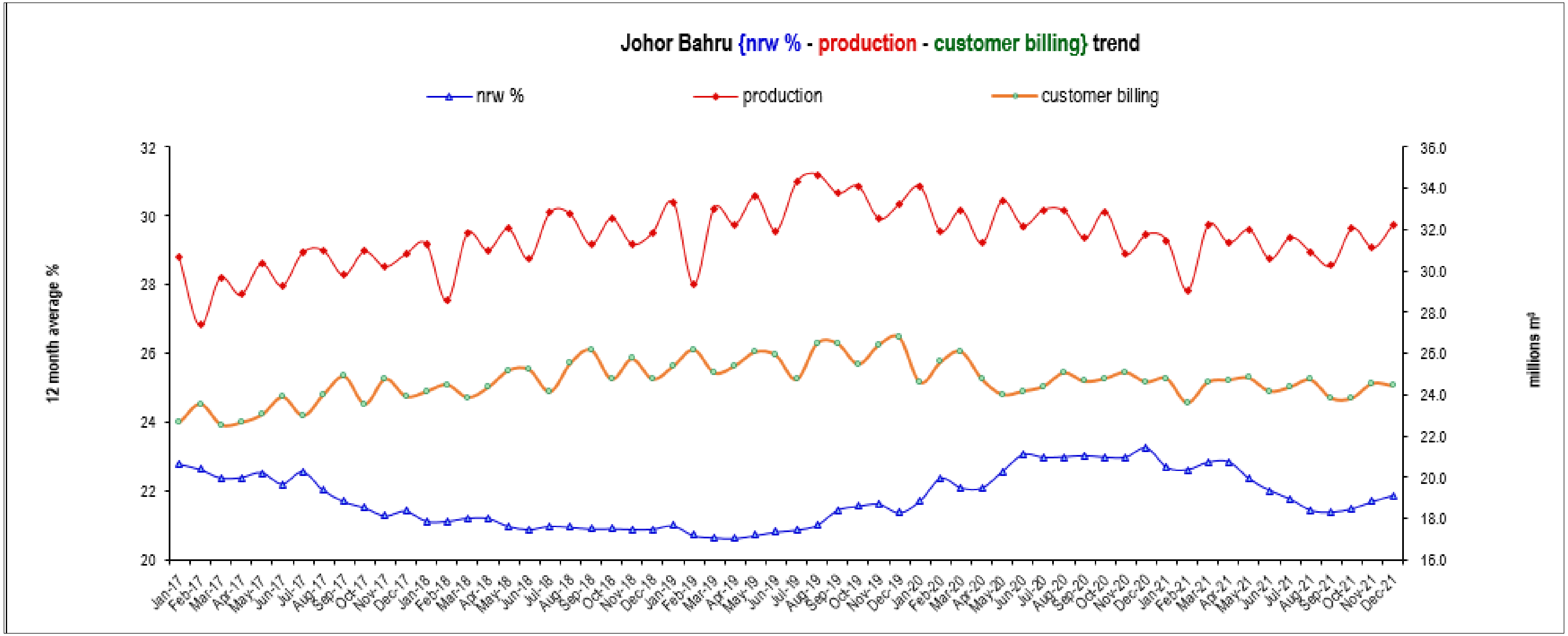

FIGURE 4.1 (a): Johor Bahru Trend

From Batu Pahat trend graph, the production and customer billing within the 12-month average from year 2017 to 2021, shows that the graph is significant increasing and decreasing towards all the year. There is no sudden increase or decrease value as the graph line is slightly smooth. Differently for NRW % line, it shows that during the year of 2021 the percentage value is decreasing significantly. This means that there is changing drastically towards the management of NRW cases even though it still in exceeded the 30% of allowable NRW percentage that has been set by water supply companies.

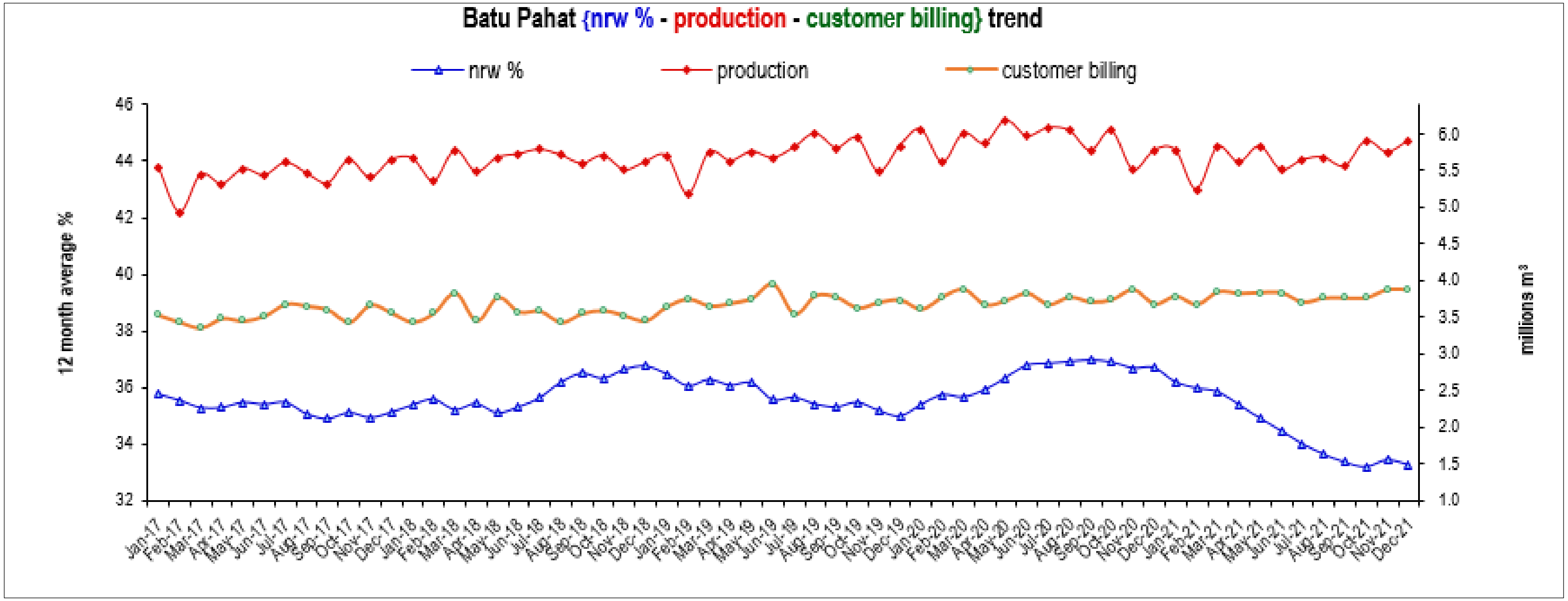

FIGURE 4.1 (b): Batu Pahat Trend

The NRW (%) for Johor Bahru and Batu Pahat from 2017 to 2021 is shown in Figure 4.1 (c) and (d). According to the statistics below, Johor Bahru, a district with a higher population than Batu Pahat, has an NRW (percentage) that is primarily within the optimal level of 0 to 30% but still requires continual monitoring. In contrast, the data for Batu Pahat indicates that it has reached the alarming level, which is worrying given that the majority of cases are greater than 30%. Johor Bahru has stated an average of 21.67% within the optimal level and Batu Pahat is having an average of 35.29% within the alarming level of NRW percentage.

The figure for Johor Bahru NRW % shows that the value is decreasing drastically in February as on Year 2019 has recorded the lowest percentage value among the year of 10.72%, while significantly increasing on July and October from 20% to 26% for 5 years trend. Year 2020 has stated the high value of NRW percentage in January and May with value of 27.66% and 28.03%, respectively compared to other month and year. In addition, for the same year, from February to April shows that the percentage value almost constant or slightly increase compared to others and still under the average NRW % for 5 years trend.

Furthermore, for Batu Pahat NRW % figure, it shows that the value is decreasing significantly in February as on Year 2019 has recorded the lowest percentage value among the year of 27.69 %, while significantly increasing on October within 36% to 40% for 5 years trend. Year 2020 has stated the high value of NRW percentage in January with value of 40.48% compared to other month and year of the studied period. In addition, March to September of Year 2021 shows that the percentage value slightly increases compared to May on the same year is slight decreased from the average value of Batu Pahat NRW % for 5 years trend. Batu Pahat also shows that majority of its value is in alarming level and need more observing and inspection to reducing the NRW percentage.

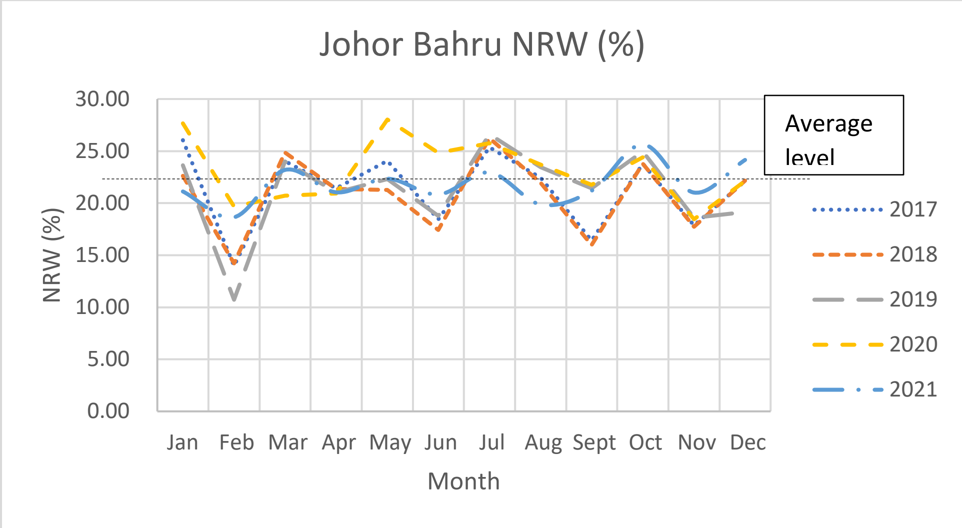

FIGURE 4.1 (c): NRW (%) for Johor Bahru district from the year of 2017 to 2021

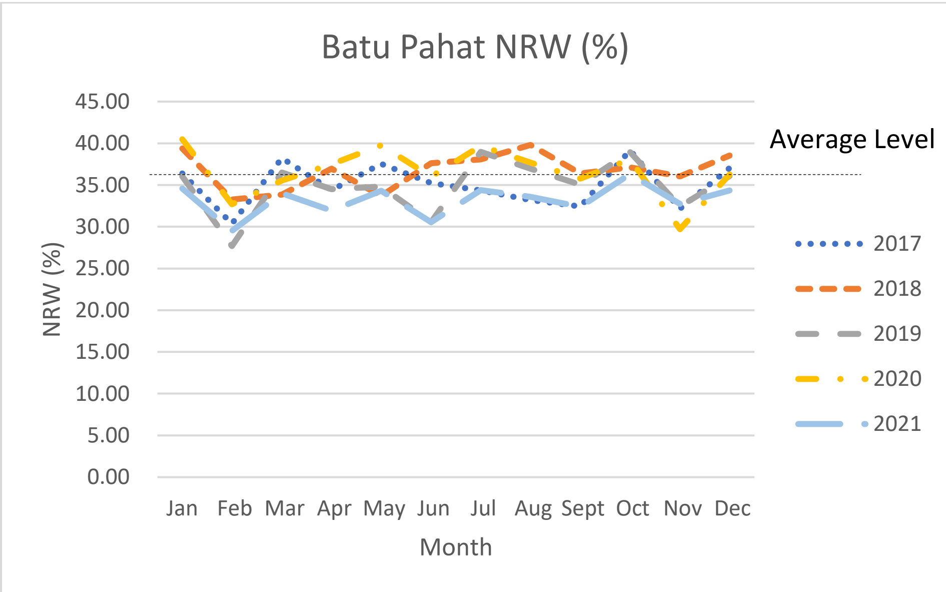

FIGURE 4.1 (d): NRW (%) for Batu Pahat district from the year of 2017 to 2021

4.2 Non – Revenue Water Model

4.2.1 Multiple Linear Regression Model

Pearson Correlation Analysis

Appendix VIII shows the sample calculation for the Pearson Correlation Analysis for both districts based on the parameters used. The Pearson Correlation Analysis is analysed based on the hypothesis that suitable with the parameters that being used for the analysis. The parameters are customers billing, other bills and operational use that corresponding with the production value. Thus, the hypothesis is shown as below:

Null Hypothesis, H0: The production does not affect the NRW
Rejected H1: The production does affect the NRW

The Pearson Correlation Analysis applied to the data from Johor Bahru is shown in Appendix I. Appendix I demonstrates that there is a substantial correlation between production and customer billing because the coefficient, r, is 0.52331. As a result, the customer billing increases along with the growth in production. When the r value is within 0 to 1, the variables are changing in the same directions and the direction is positive. The critical value of t critical is 2.0 for a two-tailed test of significance with $\alpha = 0.05$ and degrees of freedom of 58. So, the t statistic is greater than t critical, $4.68 > 2.0$, then the result is statistically significant. Hence, the result of p value, 0.00002 is less than α, 0.05 is accepted and null hypothesis is rejected.

Meanwhile, Appendix I also demonstrates that there is a substantial correlation between production and other bills because the coefficient, r, is 0.32422. As a result, the other bills increase along with the growth in production. When the r value is within 0 to 1, the variables are changing in the same directions and the direction is positive. The critical value of t critical is 2.0 for a two-tailed test of significance with $\alpha = 0.05$ and degrees of freedom of 58. So, the t statistic is greater than t critical, $2.61 > 2.0$, then the result is statistically significant. Hence, the result of p value, 0.01150 is less than α, 0.05 is accepted and null hypothesis is rejected.

Last but not least, Appendix I also demonstrates that there is a substantial correlation between production and operational use because the coefficient, r, is 0.17066. As a result, the operational use increases along with the growth in production. When the r value is within 0 to 1, the variables are changing in the same directions and the direction is positive. The critical value of t critical is 2.0 for a two-tailed test of significance with $\alpha = 0.05$ and degrees of freedom of 58. So, the t statistic is less than t critical, $1.32 < 2.0$, then the result is insignificant. While, the result of p value, 0.1923 is less than α, 0.05 is accepted and null hypothesis is rejected.

Meantime, the Pearson Correlation Analysis applied to the data from Batu Pahat is shown in Appendix II. Appendix II demonstrates that there is a substantial correlation between production and customer billing because the coefficient, r, is 0.40337. As a result, the customer billing increases along with the growth in production. When the r value is within 0 to 1, the variables are changing in the same directions and the direction is positive. The critical value of t critical is 2.0 for a two-tailed test of significance with $\alpha = 0.05$ and degrees of freedom of 58. So, the t statistic is greater than t critical, $3.36 > 2.0$, then the result is statistically significant. Hence, the result of p value, 0.0014 is less than α, 0.05 is accepted and null hypothesis is rejected.

In addition, Appendix II also demonstrates that there is a substantial correlation between production and other bills because the coefficient, r, is -0.14745. As a result, the other bills increase along with the growth in production. When the r value is -1, the variable of the direction is negative. The critical value of t critical is 2.0 for a two-tailed test of significance with α = 0.05 and degrees of freedom of 58. So, the t statistic is greater than t critical, -1.14 < 2.0, then the result is insignificant. While, the result of p value, 0.2609 is higher than α, 0.05, therefore the null hypothesis is accepted.

Lastly, Appendix II also demonstrates that there is a substantial correlation between production and operational use because the coefficient, r, is -0.04420. As a result, the operational use increases along with the growth in production. When the r value is within -1 to 0, the variables direction is negative. The critical value of t critical is 2.00 for a two-tailed test of significance with α = 0.05 and degrees of freedom of 58. So, the t statistic is less than t critical, -0.34 < 2.0, then the result is insignificant. While, the result of p value, 0.7373617 is more than α, 0.05, therefore the null hypothesis is accepted.

Simple Linear Regression (Single Regression Analysis)

Appendix III shows the Simple Linear Regression of Johor Bahru. The Simple Linear Regression is discuss based on its various of variables. First analysis is the result of summary output for the production that responded with the customer's billing. From the summary, the regression statistic shows that Multiple R value is 0.52331 bigger than 0, hence the linear relationship of the correlation coefficient is far from perfect positive relationship which is 1. While the R2 for 60 observations data is 0.279853 and the adjusted R2 value is 0.261334. If the variable is more than one, value of adjusted R2 is used. The R2 value is greater than 0 and less than 1, so, the value is accepted but considered as low. Next, the Analysis of Variance (ANOVA) shows that the total sum of squares, SS is 1.33E+14. Regression MS is calculated from regression sum of squares, SS divided with the regression degrees of freedom, df, thus the value of 3.63E+13 is obtained. Meanwhile, in order to calculate Residual MS, the value of residual sum of squares, SS divided with the residual degrees of freedom, df, later multiply with the mean squared error, therefore

1.66E+12 is obtained. When the total amount of water billed is increased by 1, the coefficient value for this data is 11998305. Consequently, the direction is in the favour. The standard error serves as a rough estimation of the coefficient's standard deviation. According to the table, the standard error is 4197357. Then, the coefficient value is higher than the standard error. So, the coefficients likely not equal to 0. The coefficient is divided by its standard error to produce the t statistic. The t statistic obtained from the table is 2.858538. P –Value obtained for this data is 0.005903 lesser than significant $\alpha = 0.05$.

Second analysis is the result of summary output for the production that responded with the other bills. From the summary, the regression statistic shows that Multiple R value is 0.324216 bigger than 0, hence the linear relationship of the correlation coefficient is far from perfect positive relationship which is 1. While the R2 for 60 observations data is 0.105116 and the adjusted R2 value is 0.089687. If the variable is more than one, value of adjusted R2 is used. The R2 value is greater than 0 and less than 1, so, the value is accepted but considered as low. Next, the Analysis of Variance (ANOVA) shows that the total sum of squares, SS is 1.33E+14. Regression MS is calculated from regression sum of squares, SS divided with the regression degrees of freedom, df, thus the value of 1.39E+13 is obtained. Meanwhile, in order to calculate Residual MS, the value of residual sum of squares, SS divided with the residual degrees of freedom, df, later multiply with the mean squared error, therefore 2.05E+12 is obtained. When the total amount of water billed is increased by 1, the coefficient value for this data is 31240966. Consequently, the direction is in the favour. The standard error serves as a rough estimation of the coefficient's standard deviation. According to the table, the standard error is 233439.6. Then, the coefficient value is higher than the standard error, so, the coefficients likely not equal to 0. The coefficient is divided by its standard error to produce the t statistic. The t statistic obtained from the table is 133.8289. P –Value obtained for this data is 6E-74 lesser than significant $\alpha = 0.05$.

Last analysis for Simple Linear Regression of Johor Bahru is the result of summary output for the production that responded with the operational use. From the summary, the regression statistic shows that Multiple R value is 0.170656426 bigger than 0, hence the linear relationship of the correlation coefficient is far from perfect positive relationship which is 1. While the R2 for 60 observations data is 0.029123616 and the adjusted R2 value is 0.012384368. If the variable is more than one, value of adjusted R2 is used. The R2 value is greater than 0 and less than 1, so, the value is accepted but considered as low. Next, the Analysis of Variance (ANOVA) shows that the total sum of squares, SS is 1.33E+14. Regression MS is calculated from regression sum of squares, SS divided with the regression degrees of freedom, df, thus the value of 3.86E+12 is obtained. Meanwhile, in order to calculate Residual MS, the value of residual sum of squares, SS divided with the residual degrees of freedom, df, later multiply with the mean squared error, therefore 2.22E+12 is obtained. When the total amount of water billed is increased by 1, the coefficient value for this data is 31215442.42. Consequently, the direction is in the favour. The standard error serves as a rough estimation of the coefficient's standard deviation. According to the table, the standard error is 357961.4. Then, the coefficient value is higher than the standard error, so, the coefficients likely not equal to 0. The coefficient is divided by its standard error to produce the t statistic. The t statistic obtained from the table is 87.20338. P –Value obtained for this data is 3.26E-63 lesser than significant $\alpha = 0.05$.

In the meantime, Appendix IV shows the Simple Linear Regression of Batu Pahat. The Simple Linear Regression is discuss based on its various of variables. First analysis is the result of summary output for the production that responded with the customer's billing. From the summary, the regression statistic shows that Multiple R value is 0.403373 bigger than 0, hence the linear relationship of the correlation coefficient is far from perfect positive relationship which is 1. While the R2 for 60 observations data is 0.16271 and the adjusted R2 value is 0.148274. If the variable is more than one, value of adjusted R2 is used. The R2 value is greater than 0 and less than 1, so, the value is accepted but considered as low. Next, the Analysis of Variance (ANOVA) shows that the total sum of squares, SS is 3.39E+12. Regression MS is calculated from regression sum of squares, SS divided with the regression degrees of freedom, df, thus the value of 5.52E+11 is obtained. Meanwhile, in order to calculate Residual MS, the value of residual sum of squares, SS divided with the residual degrees of freedom, df, later multiply with the mean squared error, therefore 4.89E+10 is obtained. When the total amount of water billed is increased by 1, the coefficient value for this data is 3129653. Consequently, the direction is in the favour. The standard error serves as a rough estimation of the coefficient's standard deviation. According to the table, the standard error is 758192.2. Then, the coefficient value is higher than the standard error. So, the coefficients likely not equal to 0. The coefficient is divided by its standard error to produce the t statistic. The t statistic obtained from the table is 4.127783. P –Value obtained for this data is 0.000119 lesser than significant $\alpha = 0.05$.

Second analysis is the result of summary output for the production that responded with the other bills. From the summary, the regression statistic shows that Multiple R value is 0.147449607 bigger than 0, hence the linear relationship of the correlation coefficient is far from perfect positive relationship which is 1. While the R2 for 60 observations data is 0.021741387 and the adjusted R2 value is 0.004874859. If the variable is more than one, value of adjusted R2 is used. The R2 value is greater than 0 and less than 1, so, the value is accepted but considered as low. Next, the Analysis of Variance (ANOVA) shows that the total sum of squares, SS is 3.39E+12. Regression MS is calculated from regression sum of squares, SS divided with the regression degrees of freedom, df, thus the value of 7.37E+10 is obtained. Meanwhile, in order to calculate Residual MS, the value of residual sum of squares, SS divided with the residual degrees of freedom, df, later multiply with the mean squared error, therefore 5.72E+10 is obtained. When the total amount of water billed is increased by 1, the coefficient value for this data is 5698993.625. Consequently, the direction is in the favour. The standard error serves as a rough estimation of the coefficient's standard deviation. According to the table, the standard error is 38286.404. Then, the coefficient value is higher than the standard error, so, the coefficients likely not equal to 0. The coefficient is divided by its standard error to produce the t statistic. The t statistic obtained from the table is 148.8516. P –Value obtained for this data is 1.28E-76 lesser than significant $\alpha = 0.05$.

Final analysis for Simple Linear Regression of Batu Pahat is the result of summary output for the production that responded with the operational use. From the summary, the regression statistic shows that Multiple R value is 0.04420206 bigger than 0, hence the linear relationship of the correlation coefficient is far from perfect positive relationship which is 1. While the R2 for 60 observations data is 0.001953822 and the adjusted R2 value is -0.015253871. If the variable is more than one, value of adjusted R2 is used. The R2 value is greater than 0 and less than 1, so, the value is accepted but considered as low. Next, the Analysis of Variance (ANOVA) shows that the total sum of squares, SS is 3.39E+12. Regression MS is calculated from regression sum of squares, SS divided with the regression degrees of freedom, df, thus the value of 6.62E+09 is obtained. Meanwhile, in order to calculate Residual MS, the value of residual sum of squares, SS divided with the residual degrees of freedom, df, later multiply with the mean squared error, therefore 5.83E+10 is obtained. When the total amount of water billed is increased by 1, the coefficient value for this data is 5683477.628. Consequently, the direction is in the favour. The standard error serves as a rough estimation of the coefficient's standard deviation. According to the table, the standard error is 43443.402. Then, the coefficient value is higher than the standard error, so, the coefficients likely not equal to 0. The coefficient is divided by its standard error to produce the t statistic. The t statistic obtained from the table is 130.8249. P –Value obtained for this data is 2.23E-73 lesser than significant $\alpha = 0.05$.

Multiple Linear Regression (MLR)

Appendix V shows the statistic of multiple linear regression analysis of Johor Bahru. From the regression statistic table, Multiple R value is 0.998394983 bigger than 0, hence the linear relationship of the correlation coefficient has perfect positive relationship which is near to 1. Next, value of the coefficient of determination, R2 for 60 observations data is 0.996792542 while the adjusted R2 value is 0. .996559272. The R2 value is greater than 0 and less than 1, thus, the value is accepted. The standard error of the regression is the precision that the regression coefficient is measured. The value for standard error is 0.201387729; therefore, the coefficient is larger than the standard error, then coefficient is probably different from 0.

The analysis of variance (ANOVA) results for Johor Bahru shows, the total of sum of squares, SS is 695.4529. Regression MS is calculated from regression sum of squares, SS divided with the regression degrees of freedom, df, thus the value of 173.3056 is obtained. Meanwhile, in order to calculate Residual MS, the value of residual sum of squares, SS divided with the residual degrees of freedom, df, later multiply with the mean squared error, therefore 0.040557 is obtained. F value is used for the F test for null hypothesis, while Significance F is associated with p-value.

The regression coefficient for Johor Bahru's multiple linear regression analysis shows that, the independent variables is increased by 1, the dependent variables is increase based on the coefficient value, which is 20.0285291. Consequently, the direction is in the favour. The standard error serves as a rough estimation of the coefficient's standard deviation. According to the table, the standard error is 0.717488. When the coefficient's value is high in relation to its standard error, the coefficient is likely not equal to 0. The coefficient is divided by its standard error to produce the t statistic. The t statistic obtained from the table is 27.91481. P – Value obtained for this data is 3.7E-34 lesser than significant $\alpha = 0.05$.

Meanwhile, Appendix VI shows the statistic for the multiple linear regression analysis of Batu Pahat. From the table, Multiple R value is 0.998467089 bigger than 0, hence the linear relationship of the correlation coefficient has perfect positive relationship which is near to 1. Meanwhile, value of the coefficient of determination, R2 for 60 observations data is 0.996936528 while the adjusted R2 is 0.99671373. The R2 value is greater than 0 and less than 1, thus, the value is accepted. The standard error of the regression is the precision that the regression coefficient is measured. The value for standard error is 0.16353659; therefore, the coefficient is larger than the standard error, then coefficient is probably different from 0.

The analysis of variance (ANOVA) results for Batu Pahat shown the total of sum of squares, SS is 480.1518938. Regression MS is calculated from regression sum of squares, SS divided with the regression degrees of freedom, df, thus the value of 119.6702405 is obtained. Meanwhile, in order to calculate Residual MS, the value of residual sum of squares, SS divided with the residual degrees of freedom, df, later multiply with the mean squared error, therefore 0.026744216 is obtained. F value is used for the F test for null hypothesis, while Significance F is associated with p-value.

The regression coefficient for the multiple linear regression analysis for Batu Pahat, shows that the coefficient value for this data is 34.38069 when the independent variables is raised by 1. As a result, things are moving in the right way. The standard error serves as a rough estimation of the coefficient's standard deviation. According to the table, the standard error is 0.6615597. When the coefficient's value is high in relation to its standard error, the p-value is likely not equal to 0. The coefficient is divided by its standard error to produce the t statistic. The t statistic obtained from the table is 51.969145. P – Value obtained for this data is 1.937E-48 lesser than significant $\alpha = 0$. Sample calculation of Multiple Linear regression (MLR) can be referred in Appendix IX. The calculation is done by using equation [5].

The model adequacy for the regression model found needs to be checked to satisfy the assumption of multiple linear regression model. The scatter matrix was plotted in Figure 4.2 (a) and (b). As shown in the figure, the distribution is scattered nicely in Johor Bahru, while the Batu Pahat figure shows that the distribution is gather closely with one another. The value of R square for Johor Bahru is 0.9968 lesser than Batu Pahat which is 0.9969. Johor Bahru statically shows that within the range of 15 to 27 percent the scattered is more than other NRW percentage ratio. Meanwhile, Batu Pahat statically shows that within the range of 29 to 41 percent the scattered is more than other NRW percentage ratio.

Both figures also shown that the measured NRW percentage ratio is significantly increases with the predicted NRW percentage ratio. Hence, the predicted value is comparable with the existing value. The figures also shown a positive linear relationship between the measured and predicted NRW (%) ratio. The scatterplot in figures reveals that the variance is constant and that there is no clear pattern. It was discovered that the error variances are homogeneous because this p-value is greater than the level of significance ($\alpha=0.05$). The regression model discovered is valid and capable of making predictions because all the requirements for multiple linear regression were met.

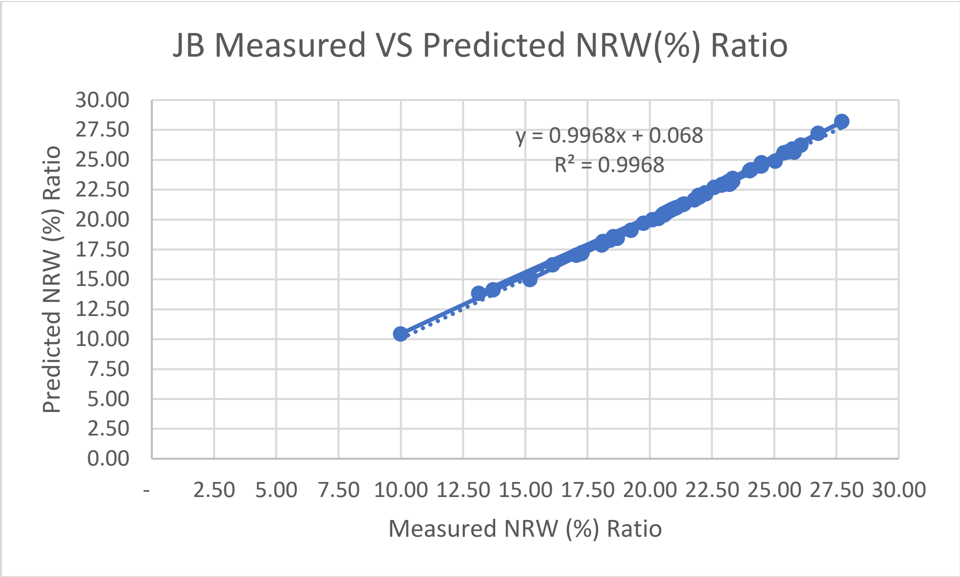

FIGURE 4.2 (a): MLR for Measured VS Predicted NRW Ratio (%) for Johor Bahru

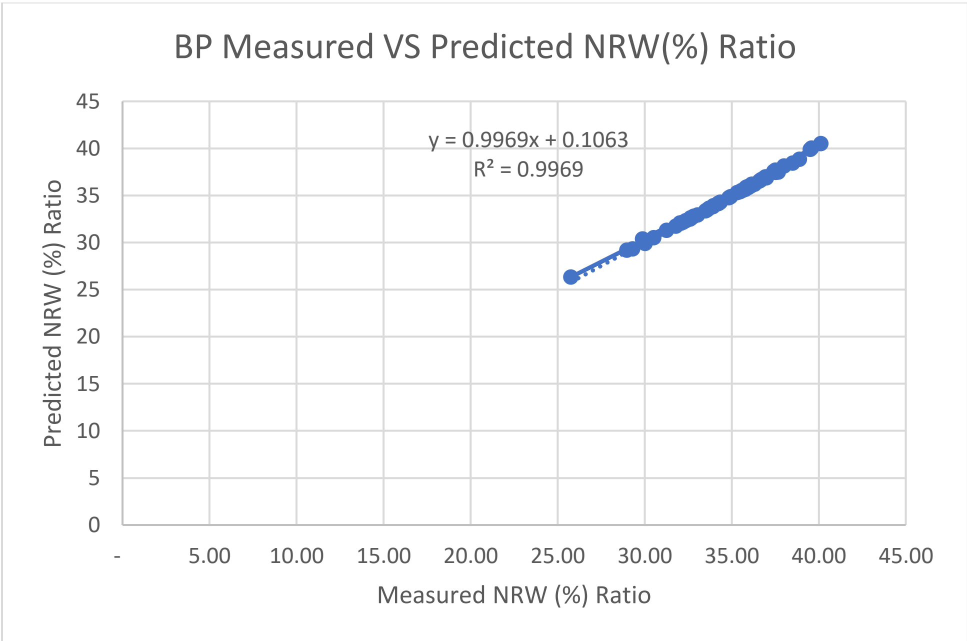

FIGURE 4.2 (b): MLR for Measured VS Predicted NRW Ratio (%) for Batu Pahat

4.2.2 Artificial Neural Network Model

Artificial Neural Networks can also be used to identify the key factors influencing non-revenue water (ANN). This analysis is being done with MatLab. In the study of the ANN model, Non-Revenue Water is employed as the dependent variable, and Production Quantity, Consumption Quantity, Leakage Quantity, and Number of Connections are all used as independent variables. Figure 4.2. (c), displays the various layers of ANN data based on the model's constituent parts.

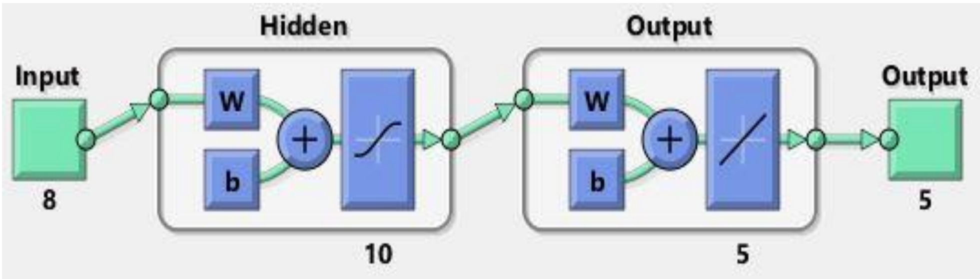

FIGURE 4.2 (c): Layers of ANN Model

Every single Figure 4.2 (d) and (e) illustrates how NRW data can be predicted using two components: production in m3/day and customer billing in m3/day in response to NRW volume in m3. The neurons are scatter by using classification plot in the MatLab, hence, the scatter plot is limited. According to Figure 4.2 (d), if production and customer billing grow, Johor Bahru neurons will cluster together more tightly. While Figure 4.2 (e) demonstrates that when production and customer billing rose, Batu Pahat neurons were nicely dispersed. The data indicates that Johor Bahru has more output and billing than Batu Pahat when comparing the results between those two districts. All This demonstrates how population, connections, and consumer numbers have a significant impact on both production and the demand for water supplies. As the contributing components rise, so does the demand for water. This occurred as a result of the production and customer billing being crucial factors in determining the amount and proportion of NRW. The data's reaction to the

NRW volume demonstrates that the NRW% is falling as the NRW volume falls. Coloured neuron was defined as the value of respective NRW volume that correspondence with the parameters.

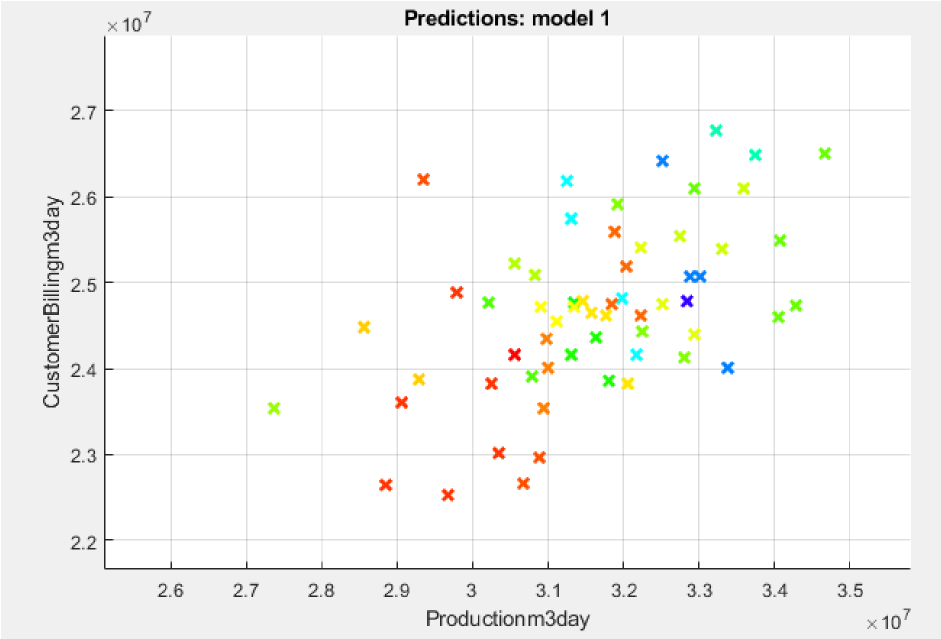

FIGURE 4.2 (d): ANN Prediction of NRW cases in Johor Bahru

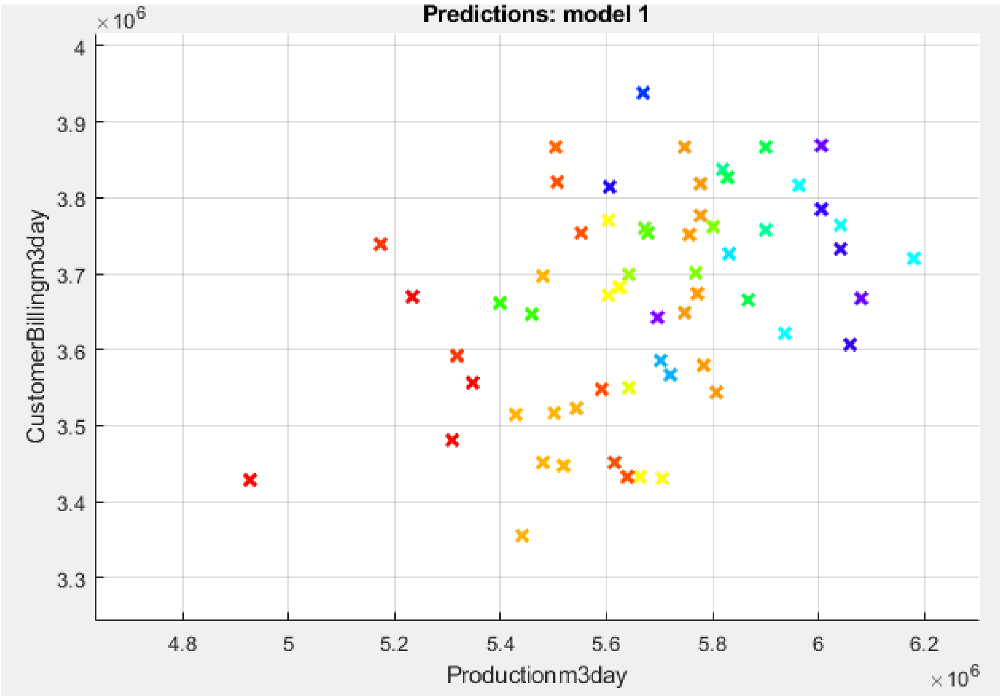

FIGURE 4.2 (e): ANN Prediction of NRW cases in Batu Pahat

4.3 Discussion

There are various things that may be done to manage and cut down on non-revenue water (NRW). One of them involves doing the studies utilising the DMA system. Based on NRW elements, the system was utilised to investigate the area. This is so that it can capture a specific spot that is difficult to detect on a map when a customer makes a report, or the water level drops. As a result, the DMA facility has been divided into sections based on its water service tank, highest point, and water level critical point. This has expedited inspections and the provision of problem solutions for the reported incidents. DMA can also determine the area's length, size, and asset unit.

Because it ignores crucial elements like the network's length, the quantity and length of connections, as well as the pressure and hydraulic load of the water supply network, percentage water losses, which are still commonly employed in Malaysia and around the world, are unreliable. The variability of water losses in a multi-year distribution system can be evaluated using this method, though. In order to acquire more full information on the quantity of wasted water lost from the network, it is advised that the IWA standard approach be used in addition to the traditional technique of calculating the percentage of losses.

It is possible to draw the conclusion that water losses can be determined simply on the increasing and decreasing data based on the graph figure of NRW percentage that has been produced for the examined area. Therefore, it will be much simpler to identify which area needs to focus more on controlling and reducing the percentage by applying the water losses computation to a smaller component each month to each DMA on the district. It can also ascertain whether the water demand was met and balance with customer billing. The figures also demonstrate that while the NRW is under control for some months, it is exceedingly high for the remaining months. In order to identify the issue and find a workable solution, it is necessary to thoroughly study the month that saw more than 30% of NRW incidents. This will allow for the reduction and control of NRW.

Based on two designation models that has been conducted by using different software shows that even though the same parameters used for different method, the distribution is difference. Multiple Linear Regression (MLR) shows that the standardize residuals and residuals value is equivalent with the predicted residuals value. In order to conducted MLR model, two test has been done to determine the relationship between two main parameters which is production (m^3/month) and total billing (m^3/month). The two-test used are Pearson correlation analysis and simple linear regression analysis. Based on these two tests, the production relationship is comparable with the total billing. This has defined by the result from the tests whereas when the production increases, the total billing also increase. Hence, MLR model is can only obtained by certain parameters that limited for each result.

Meanwhile, the Artificial Neural Network (ANN) data shows the different layers that applicable to understand the model. All the suitable parameters can also be used at once. Thus, the data accumulated for ANN is more accurate than MLR. Thence, ANN model is more suitable to be used in order to detect NRW percentage and distribution compared to MLR model. This is because ANN is more accurate than MLR. By using ANN, not only the getting the data of distribution scatter but also can determine the validation, error as well as the terrain of the data.

In the context of plans and strategies to reduce and control losses for each risk level determined based on the data analysis and experience, it is simpler to estimate and monitor NRW values using the provided model in order to meet the minimal target. By using the provided methodology and the assistance of specialists in accordance with the realisations and risk levels for each year, NRW monitoring may be conducted relatively quickly. The NRW forecast is crucial for planning water utility projects. The unpredictability of NRW variables makes it difficult for deterministic algorithms to anticipate NRW loss.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

To determine the NRW instances in Malaysia, this research is crucial. Numerous studies have been conducted to upgrade the state of Malaysia's water supply system. A benchmark to help reduce water losses can be a few techniques that were employed in the past study from many different countries. The study also introduces some techniques to help readers comprehend the NRW examples better. As a result, the method presented used for data collecting and analysis. The discrepancy between average flow and water consumption is known as non-revenue water (NRW). In this project, a thorough evaluation of NRW was conducted.

Two districts in Johor were selected based on a set of criteria because the Johor State in Malaysia was the major focus of this research. Johor Bahru and Batu Pahat are the two districts that are being researched for the project. Based on the study performed in Chapter 4, it is possible to draw the conclusion that both districts exhibit a considerable difference that is acceptable when referring to each district's type from the point of view of its population, urban or rural area, and kind of land.

In order to identify the percentage (%) level either it is to be observed or to alter any of the water supply components, such as pipes, valves, or tanks, it is crucial to compute the water losses percentage, also known as NRW percentage. Additionally, the District Meter Area (DMA) system has been extremely beneficial in many ways in overcoming NRW instances. The area's highest and most critical points of water pressure, the number of customers, the placement of the valve, the water level, and the use of water tanks may all be determined using DMA. By using the DMA method, it shows that Johor Bahru has an average of 21.67%, less than 30% of optimal level compared to Batu Pahat that has an average of 35.29% higher than 30%. Thus, the first objective of the project was achieved.

To obtain a better outcome in reducing and regulating the NRW, all systems and approaches are cooperating. Regarding minimising and managing NRW in Johor, each of them has advantages. While managing the cases, other techniques also need to be emphasised and taken into account. To achieve the goals of having zero percent instances in Malaysia, all methods and systems must be periodically upgraded. Therefore, lot of approaches had been researched and undergo in other to overcome the NRW cases. The approaches used for most of the world had been discussed in Chapter 2. Johor itself has do their absolute best in order to determine the NRW by using method such as DMA, Step Test, Leakage Calculation and Geographic Information Systems (GIS). So, the second objective had achieved.

Finally, to demonstrates the accuracy assessment of the MLR and ANN models' ideal models by employing two models as the project's novel technique. The physical and operational factors in water distribution networks were essentially used to calculate MLR. A comparison of prediction accuracy between the ANN and MRL models revealed that the ANN model was more accurate. The accuracy of NRW ratio estimation increased when an ANN was applied compared to the prior MLR technique. The selection of parameters for MLR limited to be done at once compared to ANN. Hence, ANN is more flexible compared to MLR. This conclude that the third objective achieved.

5.2 Recommendation

This experiment has a number of consequences for how decisions are made in relation to the decrease of NRW rations in Johor. The first implication is that it makes it possible to use various parameters via the water distribution network. The next conclusion is that all of the outcomes can be used to inform both case-reduction efforts and NRW management strategies. Furthermore, using other machine learning system for Artificial Neural Analysis to get a better result compared to the MatLab. The study's conclusion is that the machine-learning system is more adept at seeing intricate connections between several characteristics. Results show that the machine-learning model is appropriate for water supply firms looking to upgrade their water distribution network in a sustainable and economical manner. In order to improve water distribution systems and anticipate NRW more accurately, future research on machine learning approaches must be conducted. Findings based on machine learning techniques demand more precise outcomes by considering a variety of system elements.

REFERENCES

Sisman, E. and Kiziloz, B. (2020). Trend-risk model for predicting non-revenue water: An application in Turkey. Utilities Policy, 67(2020), 101 – 137. doi: 10.1016/j.jup.2020.101137

Farok, GG. M. G. (2017). Non-Revenue Water (NRW) is a challenge for Global Water Supply System Management: A case study of Dhaka Water Supply System Management. Journal of Mechanical Engineering, ME (46), 28 – 35. doi: 10.3329/jme.v46i1.32520

Ayad, A., Khalifa, A., Fawy, M. E. L. and Moawad, A. (2021). An integrated approach for non-revenue water reduction in water distribution networks based on field activities, optimisation, and GIS applications. Ain Shams Engineering Journal, 12(2021) 3509 – 3520. doi: 10.1016/j.asej.2021.04.007

Marques, L. O.D. A., Carvalho, R. S., Sa, M. O. M. D. and Malheiros, T. F. (2021). Benchmarking as a management tool to reduce non-revenue water. Special Issue: Sustentare and International Workshop on Sustainability Indicators (WIPIS), 24. doi: 10.1590/1809-4422asoc20200025vu2021L4DE.

See, F. K. and Ma, Z. (2018). Does non-revenue water affect Malaysia's water services industry productivity? Utilities Policy, 54(2018), 125 – 131. doi: 10.1016/j.jup.2018.04.006

Jang, D. and Choi, G. (2017). Estimation of Non-Revenue Water Ratio Using MRA and ANN in Water Distribution Networks. Water 2018, 10, 2. doi: 10.3390/w10010002

Caroline van der Berg. (2015). Drivers of non-revenue water: A cross-national analysis. Utilities Policy, 36 (2015), 71 - 78. doi: 10.1016/j.jup.2015.07.005

Kanakoudis, V. and Gonelas, K. (2014). Forecasting the Residential Water Demand, Balancing Full Water Cost Pricing and Non-revenue Water Reduction Policie. Procedia Engineering, 89(2014), 958 -966. doi: 10.1016/j.proeng.2014.11.530

Ociepa-Kubieka, A. and Wilezak, K. (2017). Water Loss Reduction as the Basis of Good Water Supply Companies' Management. doi: 10.1051/e3sconf/20171902015

Chee, H. L., Ngai, W. C., & Ranjan, R. (2017). Understanding Public Perception of and Participation in Non-Revenue Water Management in Malaysia to Support Urban Water Policy. doi:10.3390/w9010026.

Cassidy, J., Barbosa, B., Damião, M., Ramalho, P., Ganhão, A., Santos, A., and Feliciano, J. (2021). Taking water efficiency to the next level: digital tools to reduce non-revenue water. Journal of Hydroinformatics, 453 – 465. doi: 10.2166/hydro.2020.072

Elkharbotly, M. R., Mohamed, S. and Abdelkawi, K. (2022). Toward Sustainable Water: Prediction of non-revenue water via Artificial Neural Network and Multiple Linear Regression modelling approach in Egypt. Ain Shams Engineering Journal, 13 (2022), 101673. doi: 10.1016/j.asej.2021.101673

Frauendorfer, R., & Liemberger, R. (2010). The Issues and Challenges of Reducing Non – Revenue Water. Maidenhead: Asian Development Bank.

Lai, C. H., Tan, D. T., Roy, R., Chan, N. W. and Nor, A. Z. (2020). Systems thinking approach for analysing non-revenue water management reform in Malaysia. Water Policy, (2020) 22 (2), 237–251. doi: 10.2166/wp.2020.165

Liemberger, R. and Wyatt, A. (2019). Quantifying the global non-revenue water problem. Water Supply, (2019) 19 (3), 831–837. doi: 10.2166/ws.2018.129

Rajasekhar, G., Ramana, G. V., and Viswanadh, G. K. (2018). Non – Revenue Water Assessment in Urban Water Supply. International Journal of Scientific Engineering and Science, 2(5), 37-41.

Selek, B., Adigüzel, A., Iritaş, Ö., Karaaslan, Y., Kinaci, C., Muhammetoglu, A., and Muhammtoglu, H. (2018). Management of Water Losses in Water Supply and Distribution Networks in Turkey. doi: 10.31807/TJWSM.354298

Farley, M., Wyeth, G., Zainuddin, M.G., Istandar, A. and Singh, S. (2008). The Manager's Non-Revenue Water Handbook: A Guide to Understanding Water Losses.

Suruhanjaya Perkhidmatan Air Negara, SPAN. (2021). Water And Sewerage Fact Book 2021 Peninsular Malaysia and F.T. Labuan.

Winarni, W. (2009). Infrastructure Leakage Index (ILI) as Water Losses Indicator. Report from Volume 11, Number 2, Environmental Engineering Department, Trisakti University, Jakarta, Indonesia

Murugan, S.S. and Chandran, S. (2019). Assessment of Non-Revenue Water in a Water Distribution System and Strategies to Manage the Water Supply. International Research Journal of Engineering and Technology (IRJET), 06(04), 3488-3492.

Abdul, M. H., Zakaria, M. H., Zulkifli, M. L., and Roslan, N. F. (2021). Determination of Non – Revenue Water. Malaysian Journal of Computing, 6 (1), 642-649.

Nasara, M. A., Zubairu. I., Jagaba, A. H., Azare, A. A., Yerima, Y. M., and Yerima, B. (2021). Assessment of Non-Revenue Water Management Practices in Nigeria (A Case Study of Bauchi State Water and Sewerage Cooperation). American Journal of Engineering Research (AJER), 10(5), 390-401. Retrieved from https://www.researchgate.net/profile/A-H-Jagaba/publication/352291911_Assessment_of_Non-Revenue_Water_Management_Practices_in_Nigeria_A_Case_Study_of_Bauchi_State_Water_and_Sewerage_Cooperation/links/60c2237992851ca6f8d95e37/Assessment-of-Non-Revenue-Water-Management-Practices-in-Nigeria-A-Case-Study-of-Bauchi-State-Water-and-Sewerage-Cooperation.pdf

Chawira, M., Hoko, Z., and Mhizha, A. (2022). Partitioning non-revenue water for Juru Rural Service Centre, Goromonzi District, Zimbabwe. Physics and Chemistry of the Earth. doi.org/10.1016/j.pce.2022.103113

Arash, M. and Nastaran, C. (2021). Concepts, procedures, and applications of artificial neural network models in streamflow forecasting. Advances in Streamflow Forecasting. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780128206737000032

Bevans, R. (2020). Multiple Linear Regression | A Quick Guide (Examples). Retrieved from https://scribbr.com/statistics/multiple-linear-regression/#:~:text=Multiple%20linear%20regression%20is%20a,variables%20using%20a%20straight%20line

Turney, S. (2022). Pearson Correlation Coefficient (r) | Guide & Examples. Retrieved from https://www.scribbr.com/statistics/pearson-correlation-coefficient/#:~:text=The%20Pearson%20correlation%20coefficient%20(r,the%20relationship%20between%20two%20variables.&text=When%20one%20variable%20changes%2C%20the,changes%20in%20the%20same%20direction.

Srivastav, A. K. (2022). Pearson Correlation Coefficient. Retrieved from https://www.wallstreetmojo.com/pearson-correlation-coefficient/

Bevans, R. (2020). Simple Linear Regression | An Easy Introduction & Examples. Retrieved from https://www.scribbr.com/statistics/simple-linear-regression/

Frost, J. (2022). Null Hypothesis: Definition, Rejecting & examples. Retrieved from https://statisticsbyjim.com/hypothesis-testing/null-hypothesis/

APPENDIX

Appendix I: Pearson Correlation Table for Johor Bahru

Production VS Customer Billing

Coefficient (r):	**0.52331**
N:	60.00
T Statistic:	4.68
DF:	58.00
p value:	0.0000179
T Critical, t*	2.00
Significance α	0.05

Production VS Other Billing

Coefficient (r):	**0.32422**
N:	60.00
T Statistic:	2.61
DF:	58.00
p value:	0.0114975
T Critical, t*	2.00
Significance α	0.05

Production VS Operational Use

Coefficient (r):	**0.17066**
N:	60.00
T Statistic:	1.32
DF:	58.00
p value:	0.1923432
T Critical, t*	2.00
Significance α	0.05

Appendix II: Pearson Correlation Table for Batu Pahat

<u>Production VS Customer Billing</u>

Coefficient (r):	**0.40337**
N:	60.00
T Statistic:	3.36
DF:	58.00
p value:	0.0013947
T Critical, t*	2.00
Significance α	0.05

<u>Production VS Other Billing</u>

Coefficient (r):	**-0.14745**
N:	60.00
T Statistic:	-1.14
DF:	58.00
p value:	0.2608976
T Critical, t*	2.00
Significance α	0.05

<u>Production VS Operational Use</u>

Coefficient (r):	**-0.04420**
N:	60.00
T Statistic:	-0.34
DF:	58.00
p value:	0.7373617
T Critical, t*	2.00
Significance α	0.05

Appendix III: Simple Linear Regression Result for Johor Bahru

Summary Output (Production VS Customer Billing)

Regression Statistics	
Multiple R	0.52331
R Square	0.273853
Adjusted R Square	0.261334
Standard Error	1288668
Observations	60

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3.63E+13	3.63E+13	21.87368	1.79E-05
Residual	58	9.63E+13	1.66E+12		
Total	59	1.33E+14			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	11998305	4197357	2.858538	0.005903	3596382	20400228
Customer Billing (m3/month)	0.794504	0.169877	4.676931	1.79E-05	0.454458	1.13455

Summary Output (Production VS Other Bills)

Regression Statistics	
Multiple R	0.324216
R Square	0.105116
Adjusted R Square	0.089687
Standard Error	1430581
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1.39E+13	1.39E+13	6.812854	0.011497
Residual	58	1.19E+14	2.05E+12		
Total	59	1.33E+14			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	31240966	233439.6	133.8289	6E-74	30773685	31708246
Other bills (m3/month)	8.605464	3.29693	2.610144	0.011497	2.005941	15.20499

Summary Output (Production to Operational Use)

Regression Statistics	
Multiple R	0.170656426
R Square	0.029123616
Adjusted R Square	0.012384368
Standard Error	1490084.988
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3.86E+12	3.86E+12	1.73984	0.192343
Residual	58	1.29E+14	2.22E+12		
Total	59	1.33E+14			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	31215442.42	357961.4	87.20338	3.26E-63	30498905	31931980
Operational Use (m3/month)	2.754251991	2.088089	1.31903	0.192343	-1.42551	6.934016

Appendix IV: Simple Linear Regression Result for Batu Pahat

Summary Output (Production VS Customer Billing)

Regression Statistics	
Multiple R	0.403373
R Square	0.16271
Adjusted R Square	0.148274
Standard Error	221238.4
Observations	60

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	5.52E+11	5.52E+11	11.27111	0.001395
Residual	58	2.84E+12	4.89E+10		
Total	59	3.39E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	3129653	758192.2	4.127783	0.000119	1611966	4647339
Customer Billing (m3/month)	0.6942	0.206777	3.357248	0.001395	0.280292	1.108109

Summary Output (Production VS Other Bills)

Regression Statistics	
Multiple R	0.147449607
R Square	0.021741387
Adjusted R Square	0.004874859
Standard Error	239138.5009
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	7.37E+10	7.37E+10	1.289026	0.260898
Residual	58	3.32E+12	5.72E+10		
Total	59	3.39E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5698993.625	38286.40389	148.8516	1.28E-76	5622355	5775632
Other bills (m3/month)	-9.279955783	8.173632941	-1.13535	0.260898	-25.6413	7.081348

Summary Output (Production to Operational Use)

Regression Statistics	
Multiple R	0.04420206
R Square	0.001953822
Adjusted R Square	-0.015253871
Standard Error	241544.96
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	6.62E+09	6.62E+09	0.113544	0.737362
Residual	58	3.38E+12	5.83E+10		
Total	59	3.39E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5683477.628	43443.4017	130.8249	2.23E-73	5596516	5770439
Operational Use (m3/month)	-0.310408082	0.92119555	-0.33696	0.737362	-2.15438	1.533565

Appendix V: Multiple Linear Regression Result for Johor Bahru

Summary Output

Regression Statistics	
Multiple R	0.998394983
R Square	0.996792542
Adjusted R Square	0.996559272
Standard Error	0.201387729
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	693.2223	173.3056	4273.134	7.46E-68
Residual	55	2.230636	0.040557		
Total	59	695.4529			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	20.0285291	0.717488	27.91481	3.7E-34	18.59065	21.46641
Production (m3/day)	2.56894E-06	2.2E-08	116.8438	1.33E-67	2.52E-06	2.61E-06
Customer Billing (m3/day)	-3.21309E-06	3.17E-08	-101.38	3.16E-64	-3.3E-06	-3.1E-06
Other bills (m3/day)	-4.06651E-06	4.96E-07	-8.19511	4.2E-11	-5.1E-06	-3.1E-06
Operational Use (m3/day)	-3.62661E-06	2.89E-07	-12.5595	8.63E-18	-4.2E-06	-3E-06

Appendix VI: Multiple Linear Regression Result for Batu Pahat

Summary Output

Regression Statistics	
Multiple R	0.998467089
R Square	0.996936528
Adjusted R Square	0.99671373
Standard Error	0.16353659
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	478.6809619	119.6702405	4474.62134	2.11077E-68
Residual	55	1.47093189	0.026744216		
Total	59	480.1518938			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	34.38069	0.6615597	51.969145	1.937E-48	33.054894	35.706485
Production (m3/day)	1.169E-05	9.904E-08	118.04295	7.609E-68	1.149E-05	1.189E-05
Customer Billing (m3/day)	-1.78E-05	1.766E-07	-100.9744	3.936E-64	-1.82E-05	-1.75E-05
Other bills (m3/day)	-2.11E-05	5.686E-06	-3.717621	0.0004728	-3.25E-05	-9.74E-06
Operational Use (m3/day)	-1.81E-05	6.563E-07	-27.58952	6.764E-34	-1.94E-05	-1.68E-05

Appendix VII: Water Losses Volume and Percentage Sample Hand Calculation

<u>Johor Bahru (December 2021)</u>

Production Inflow (m3/month) = 32, 252, 323.90

Customer's Billing (m3/month) = 24,508, 430.03

$NRW\ Volume\ (m^3) = Inflow\ (m^3) - Billing\ (m^3)$

$NRW\ Volume\ (m^3) = 32\ 252\ 323\ .90 - 24\ 508\ 430.03$

$NRW\ Volume\ (m^3) = 7\ 743\ 893.87\ m^3$/Month

$NRW\% = \frac{Inflow\ (m^3) - Billing\ (m^3)}{Inflow\ (m^3)} \times 100\%$

$NRW\ (\%) = \frac{7\ 743\ 893.87}{32\ 252\ 323.90} \times 100\%$

$NRW\ (\%) = 24.01\%$

<u>Batu Pahat (December 2021)</u>

Production Inflow (m3/month) = 5, 899 289.42

Customer's Billing (m3/month) = 3, 872 698.99

$NRW\ Volume\ (m^3) = Inflow\ (m^3) - Billing\ (m^3)$

$NRW\ Volume\ (m^3) = 5\ 899\ 289.42 - 3\ 872\ 698.99$

$NRW\ Volume\ (m^3) = 2\ 026\ 590.43\ m^3$/Month

$NRW\% = \frac{Inflow\ (m^3) - Billing\ (m^3)}{Inflow\ (m^3)} \times 100\%$

$NRW\ (\%) = \frac{2\ 026\ 590.43}{5\ 899\ 289.42} \times 100\%$

$NRW\ (\%) = 34.35\%$

Appendix VIII: Pearson Correlation Analysis Sample Hand Calculation

Johor Bahru (Production responses to Customer's Billing)

N = 60

$\Sigma xy = 46\ 875\ 835\ 040\ 885\ 500.00$

$\Sigma x = 1\ 896\ 817\ 731.45$

$\Sigma y = 1\ 481\ 326\ 772.45$

$\Sigma x^2 = 60\ 097\ 935\ 322\ 001\ 600.00$

$\Sigma y^2 = 36\ 629\ 695\ 780\ 762\ 700.00$

$$r = \frac{N(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

$$r = \frac{2.74321 \times 10^{15}}{5.24204 \times 10^{15}}$$

$$r = 0.52331$$

$$t_{statistic} = \frac{r}{\sqrt{\frac{1 - r^2}{N - 2}}}$$

$$t_{statistic} = \frac{0.52331}{\sqrt{\frac{1 - 0.52331^2}{58}}}$$

$$t_{statistic} = 4.68$$

p – Value = 0.0000179

Significant α = 0.05

Degree of Freedom, DF = 58

$t_{critical}$= 2.00

p – value < Significant α

$t_{statistic} > t_{critical}$

Batu Pahat (Production responses to Operational Use)

N = 60

$\Sigma xy = 11\ 155\ 726\ 695\ 464.30$

$\Sigma x = 340\ 397\ 115.65$

$\Sigma y = 1\ 970\ 122.72$

$\Sigma x^2 = 1\ 1934\ 560\ 513\ 715\ 990.00$

$\Sigma y^2 = 133\ 442\ 828\ 131.44$

$$r = \frac{N(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

$$r = \frac{-1.28049 \times 10^{12}}{2.8969 \times 10^{12}}$$

$$r = -0.04420$$

$$t_{statistic} = \frac{r}{\sqrt{\frac{1 - r^2}{N - 2}}}$$

$$t_{statistic} = \frac{-0.04420}{\sqrt{\frac{1 - (-0.04420)^2}{58}}}$$

$$t_{statistic} = -0.345$$

p – Value = 0.7373617

Significant α = 0.05

Degree of Freedom, DF = 58

$t_{critical}$= 1.67

p – value > Significant α

$t_{statistic} < t_{critical}$

Appendix IX: Multiple Linear Regression Sample Hand Calculation

Johor Bahru (December 2021)

$\beta_0 = 20.0285$
$\beta_1 = 2.5689 \times 10^{-6}$
$\beta_2 = -3.2131 \times 10^{-6}$
$\beta_3 = -4.0665 \times 10^{-6}$
$\beta_4 = -3.6266 \times 10^{-6}$

$\varepsilon_i = 0.2014$
$X_1 = 32\ 252\ 323.90$
$X_2 = 24\ 432\ 636.45$
$X_3 = 31\ 272.47$
$X_4 = 44\ 521.11$

$\beta_1 X_1 = 2.5689 \times 10^{-6} \times 32\ 252\ 323.90 = 82.854$

$\beta_2 X_2 = -3.2131 \times 10^{-6} \times 324\ 432\ 636.45 = -78.504$

$\beta_3 X_3 = -4.0665 \times 10^{-6} \times 31\ 272.47 = -0.127$

$\beta_4 X_4 = -3.6266 \times 10^{-6} \times 44\ 521.11 = -0.161$

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon_i$

$Y = 20.0285 + 82.854 - 78.504 - 0.127 - 0.161 + 0.2014$

$Y = 24.29\%$

Batu Pahat (December 2021)

$\beta_0 = 34.3807$
$\beta_1 = 1.169 \times 10^{-5}$
$\beta_2 = -1.78 \times 10^{-5}$
$\beta_3 = -2.11 \times 10^{-5}$
$\beta_4 = -1.81 \times 10^{-5}$

$\varepsilon_i = 0.1635$
$X_1 = 5\ 899\ 289.42$
$X_2 = 3\ 867\ 213.00$
$X_3 = 5\ 485.99$
$X_4 = 5\ 539.32$

$\beta_1 X_1 = 1.169 \times 10^{-5} \times 5\ 899\ 289.42 = 68.97$

$\beta_2 X_2 = -1.78 \times 10^{-5} \times 3\ 867\ 213.00 = -68.96$

$\beta_3 X_3 = -2.11 \times 10^{-5} \times 5\ 485.99 = -0.116$

$\beta_4 X_4 = -1.81 \times 10^{-5} \times 5\ 539.32 = -0.1003$

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon_i$

$Y = 34.3807 + 68.97 - 68.96 - 0.116 - 0.1003 + 0.1635$

$Y = 34.33\%$