APPLICATION OF ARTIFICIAL NEURAL NETWORKS FOR PREDICTING ACID MINE DRAINAGE CHARACTERISTICS IN THE WITWATERSRAND WESTERN BASIN

MSc RESEARCH

Prepared by

Mbhoni Shiviti

457847

Submitted to

School of Chemical and Metallurgical Engineering, Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, South Africa

Supervisor(s): Prof. Jean Mulopo

11 October 2021

DECLARTION

I declare that this dissertation is my own, unaided work. It is being submitted for the Master of Science Degree at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University

Signature of Candidate

11th day of October 2021 in Braamfontein

ABSTRACT

South Africa is amongst the countries experiencing water supply challenges, a problem partly exacerbated by pollution of both ground and surface water with Acid Mine Drainage (AMD). The now-defunct mines in the Witwatersrand Western Basin amongst other places have been identified as sources of AMD water. As a way of adding to efforts made on AMD remedial actions, this research was aimed at investigating the application of Artificial Neural Network (ANN) to predict some of AMD water properties. This would provide a tool with which unavailable data can be simulated for various use in existing processes, on-going research works as well as planning for future works around the subject. Data collected over a period of 3 year along a river within the Witwatersrand Western Basin was used to train an ANN model that would predict AMD water [SO_4^{2-}] and Total Dissolved Solids (TDS) which are in this case some of the best indicators of the level of contaminations. These were simulated using the AMD water's easy-to-measure properties, namely pH, EC and [Cl^-]. Three ANN properties, namely the number of hidden neurons, data split ratio and training algorithms were varied to work out a set of property combinations that would define the best models whilst other network variables were carefully chosen and kept constant. The analysis of different trends generated and other findings from literature have shown that the best model property combinations were determined. These are 5 hidden neurons, trainlm training algorithm and 80:10:10 data split ratio. The best model has a 10 arbitrary run average Correlation Coefficient (R) value of 0.93916 which is a good indication of the model performance. The model Mean Square Error (MSE) of 0.229 was obtained for a rescaled data. This model can reliably be used to predict Witwatersrand Western Basin AMD water properties and more.

ACKNOWLEDGEMENTS

I would like to thank Prof. Jean Mulopo for his unwavering dedication in helping me put together this dissertation even in times where I was faced with devastating personal challenges. His guidance helped me see beyond what this work is all about and to think a little differently about the life itself.

I would also like to thank my family for their unending support during the time when I was busy with this research work. The emotional support was overwhelming. Even at times where they needed me the most, they never made me feel like I was barely around.

Contents

List of Tables

List of Figures

List of Abriviations

AGI Artificial General Intelligence
AI Artificial Intelligence
AMD Acid Mine Drainage
ANN Artificial Neural Network
BFGS Broyden Fletcher-Goldfarb-Shanno
BOD Biochemical Oxygen Demand
COD Chemical Oxygen Demand
CSIR Council for Scientific and Industrial Research
DWA Department of Water Affairs
DWAF Department of Water Affairs and Forestry
EC Electrical Connectivity
ECL Environmental Critical Levels
GIS Geographical Information System
GUI Graphical User Interface
HMC Hydrological Monitoring Committee
IMC Inter-Ministerial Committee
KGR Krugersdrop Game Reserve
LMS Levenberg-Marquardt
LSMT Long-Short-Term Memory
ML Machine Learning
MLP Mutilayer Perception
MSE Mean Square Error
NN Neural Networks
PCA Principal Component Analysis
RMSE Root Mean Square Error
SVM Support Vector Machine
TCTA Trans-Caledon Tunnel Authority
TDS Total Dissolved Solids
TSS Total Suspended Solids
WTP Water Treatment Plant

CHAPTER 1. INTRODUCTION

1.1 Background

Energy and water supply are some of the most crucial aspects to humans today. Their availability or lack thereof significantly affects activities that aid to and influence survival, decent living, and socio-economic development. The currently experienced rapid global population growth and booming of electricity-powered technology have made it necessary to conserve energy and water resources, as well as increasing water and energy supply. South Africa in particular, is one of the countries experiencing both water and electricity supply challenges (Maluleke, 2021). In some parts of the country water supply challenges have been exacerbated by pollution of clean water sources.

There are many common water pollutants, but Acid Mine Drainage (AMD) has been the culprit in communities associated with mining activities (McCarthy, 2011). Although the phenomenon has been known for quite some time in various parts of the world, acidic mine water in South Africa was first discovered to be decanting from abandoned mines back in 2002 (McCarthy, 2011). This mining legacy has since been polluting water resources. This challenge is said to be common for many mining regions across the globe.

The mining sector in South Africa was once the backbone and drivers of the country's economy, with gold, iron ore, platinum and diamond as top commodities. The industry has funded the country's development to a wider extent. It is the oldest and largest industry and still plays a significant role in the country's economy today. Despite already contributing 88% of South Africa's solid wastes, the mining activities have also been linked with a drastic contribution to pollution of water sources (Manders, et al., 2009).

South African gold and coal mining, particularly the now defunct mines in the Western Basin (Krugersdorp area), the Central Basin (Roodepoort to Boksburg) and the Eastern Basin (Brakpan, Springs and Nigel area), and Witbank-Middleburg area, have been identified as source of Acid Mine Drainage water (McCarthy, 2011). AMD waters pollute and poison both surface and ground

water resources, hence putting a threat on nearby communities and ecosystems. This on its own is a devastating problem that cannot be ignored as it bares the effects that are felt on daily basis.

AMD waters come about due to the biological oxidation of metal sulfides to mental sulfates, and their acidic nature further facilitates solubilization of metals such as lead, cadmium, arsenic, and copper (Cheng, et al., 2007). It is its acidic nature, dissolved salts and the presence of dissolved heavy metals which render AMD water harmful, thus making it necessary to treat the water prior to using it or discharging it to the environment. Problems caused by AMD water have prompted research works on finding ways to treat AMD. This phenomenon is now well-understood (McCarthy, 2011), with facilities across the world already operating to remediate the AMD water-associated challenges.

It is believed that the overall impact of AMD is very much dependent on local conditions such as geomorphology, climate and to some extent, the distribution of the AMD generating deposits (McCarthy, 2011). South Africa also happens to be amongst the regions prone to AMD related challenges. South Africa's major initiative step towards finding remedial measures for AMD impacts were initiated in 2010 when the Inter-Ministerial Committee (IMC) (a team of expert researchers on AMD) was put together to investigate the problem and recommend possible solutions. Just as it is in other mining communities around the world, the study by the IMC concluded that AMD water is an issue which needed to be given a significant amount of attention, and that it has costly environmental impact (Coetzee, et al., 2010) as seen on Appendix C.

The AMD water must go through treatment processes which address the high acidity and high metal content first. Pumped AMD water is at some places diluted with water from already-scares fresh water sources before it is released into the environment as part of the efforts made towards reducing AMD overall impacts and to comply with the set regulations for industrial influent water (Bobbins, 2015). All these strategies require real-time the pumped AMD water property data for effective and efficient treatment. This data currently being acquired through various monitoring strategies. The monitoring includes gathering data on water levels in mines that are no longer operational, taking water samples for quality and compositions analysis, recording flowrate of mine water being discharged into watercourses, water ingress rates, as well as preparing for

subsidence events and seismic evens influences (Bobbins, 2015). Some of these monitoring strategies are manual and have been proven costly and time-consuming. Another way of monitoring AMD is through GIS and remote sensing which map channels with contaminated water as well as providing an idea of the level of contamination on the mapped areas (Sakala, et al., 2017). This system is however very costly and still need further work to get it working efficiently (Sakala, et al., 2017).

This study seeks to find a tool that will aid directly to the proposed measures aimed at addressing AMD waste challenges from a South African perspective as recommended on the study done by the IMC. The tool is derived from AMD water real-time data collected from the Witwatersrand Basin over a 3 and half year period. The tool will be useful in forecasting various area-specific AMD water properties from its easy-to-measure properties at different times of year. This will bring about a cost and time savings on the monitoring aspect and faster data collection for AMD water treatment requirements and assist with AMD water treatment plant designs and simulations amongst other things. The success of this study will not only bring about an additional an additional tool to AMD challenges remedial measure, but also contribute to a better understanding of the challenges for further scientific work around this subject.

1.2 Problem Statement:

Old mining regions along the Witwatersrand Basin have been experiencing a scourge of decanting AMD water from the now-defunct mines. This water flows into fresh water sources, posing serious ecological and health challenges in those areas. The situation is currently being controlled through pumping and treatment of the AMD waters at various points of decent before disposing it into fresh water sources. The processing of AMD water requires prior knowledge of the water characteristics such as levels of contaminants, salt loads, and acidity of water. This information is often not readily available, and also difficult and costly to measure. As a result, there is a need for a tool or system with which such information can be made available for a swift and appropriate response to AMD water related challenges.

1.3 Aim and Objectives

1.3.1 Aim of the Study

This study sought to find an analytical tool that would aid directly to the current and other proposed AMD water challenge remedial measures. The recent immergence of advance Artificial Intelligence (AI) was a motivating factor to the choice of solution in this case. The tool is derived from AMD water real-time data collected from the Witwatersrand Basin over a 3 and half year period. The tool will be useful in forecasting various area-specific AMD water properties from its easy-to-measure properties at different times of year. This will bring about a cost and time savings on the monitoring aspect and faster data collection for AMD water treatment requirements and assist with AMD water treatment plant designs and simulations amongst other things. The success of this study will not only bring about an additional an additional tool to AMD challenges remedial measure, but also contribute to a better understanding of the challenges for further scientific work around this subject.

1.3.2 Objectives of the Study

- To develop models that will predict, with high sense of accuracy, important AMD properties for the Witwatersrand Basin AMD water with the use of Artificial Neural Networks.
- To propose the most useful ANN model input and output parameter combination that aid to the existing AMD treatment processes.
- To identify the most suitable ANN model parameter what will maximize predicted result reliability and accuracy.
- To make recommendations on where to use the model, and on future work that can used this study as a basis to further investigate other ANN models that will be useful in this area of study.

1.4 Significance of the of Study

Various research work around this subject have been conducted lately with a special focus in coming up with tools to characterize wastewater. Wastewater characterization is undoubtedly one of the most important areas of research as it feeds directly to various waste management solutions. Wastewater can however be classified in different categories based on the source and the type of contaminants in it. This study focuses on characterization of AMD wastewater to have better knowledge of what is in the water when those easy-to-measure properties are known. AMD water is mostly variable from region-to-region due to difference in geological compositions, and the ever-changing weather patterns for different areas. There have been studies aimed at predicting AMD water in South Africa, however this work is limited because there are not enough data pools for it in South Africa. The use of historic data in this study could help determine if such data can be modelled in a certain range that is available and establish whether the models can be used in other areas or not. Having a characterizing tool for the Witwatersrand Basins AMD water will be a significant breakthrough that will help in the treatment processes there are in the areas and future design decision making tool. This tool will also help mine to comply with the legislations surround the disposal of AMD water into the surrounding environment.

The current methods of collecting AMD water related data has proven to be time consuming and expensive exercise. Various proposals have been made to address this. Others have suggested the use of drones and GIS sensors to do the monitoring, but this has proven unsustainable due to the costs involved. The water streams flow all the way from Krugersdorp area to some areas in Mpumalanga, making it increasingly difficult to monitor. It should also be noted that it is only in the last 10 years that powerful computer tools that can handle the non-linier nature of AMD were developed. This study seeks to exploit that space in order to come up with this much needed analytical tool, and also to add into scientific knowledge.

1.5 Research Questions to be Addressed

- Is it possible to develop a functional NN from the historic data provided?

- Which input-output combination make sense and show some form of correlation?
- At what confidence level does each model predict the AMD properties – Statistical Analysis
- What are the limitations of the model?
- Where can the model be applied for practical purposes, given its level of accuracy?

1.6 Structure of the Report

The report consists of the following segments, chapters, and appendices:

- Title Page
- Declaration
- Abstract
- Acknowledgements
- Abstract
- Contents
- List of Table
- List of Figure
- List of Abbreviations
- Body of the Report
  - Chapter 1: Introduction
  - Chapter 2: Literature Review
  - Chapter 3: Material and Methods
  - Chapter 4: Results and Discussion
  - Chapter 5: Conclusions
  - Chapter 6: Recommendations
  - Chapter 7: Intellectual Property
  - Chapter 8: Ethics
- References
- Appendices

CHAPTER 2. LITERATURE REVIEW

2.1 An Overview of the AMD Challenge in the Area of Focus

Acid-Mine drainage (AMD) and its associated problems have brought about concerns which have prompted research works from researchers around the globe, aimed at finding ways to remediate the associated challenges. In so doing, various research studies have used different approaches to address AMD waters related issues. Some research works were aimed solely at finding ways to desalinate and neutralise AMD water before releasing it into fresh watercourses (Skousen, et al., 2000). Some focused on both desalination and neutralisation but also went as far as recovering dissolved heavy metals from AMD water, whilst others found ways to recover different saleable by-products (Nleya, et al., 2016). Effective technological systems designed to treat AMD waters are available today, some of which can produce portable water and other products of economic value. It can however be said that very little efforts have been made in coming up with tools or systems suitable for predicting some of AMD water properties to acceptable degrees of confidence without relying on continuous stream sample analysis. Such tools could eliminate costs, save time and aids considerably to system designs and day-to-day AMD treatment process operations. This research study is focused on adding the quantifying tool to the AMD treatment measures available today, with a special focus on the Witwatersrand Western Basin.

2.2 The Witwatersrand Western Basin Case Study

South Africa's mining industry has contributed considerably on the country's economic over the years and continues to have a vital role in ensuring that the country's position in the global market is sustained. The country is graced with rich mineral resources, with reserves of gold, diamond, iron ore, platinum, manganese, chromium, copper, uranium, silver, beryllium, titanium, coal, considerable amounts of natural gas amongst others (Lowe, et al., 2020). Gold reserves in particular have been excavated and generated large revenues for various stakeholders including the government for many decades (McCarthy, 2011). South African's mineral gold deposits are the largest known repository of gold on earth (McCarthy, 2011) and South African remains amongst the largest producer of gold (Lowe, et al., 2020), having contributed over a third of the

world's gold production (Tucker, et al., 2016). Gold was mined by the indigenous people in the interior parts of south Africa long before settlers made the first discovery of gold traces in the Transvaal in the 1850s (McCarthy, 2011). The discovery of small quantities of gold dust in the Jukskei River back in 1854 led to further exploration for gold in the Witwatersrand area and across the Transvaal, with more discoveries made between 1871 and 1886 at Eersteling (south of Polokwane), Pilgrim's Rest, Sabie, Kaapsehoop, Barberton, Leydsdorp, near Malmani, Kromdraai and Wilgespruit (Beavon, 2004).

Gold in the Witwatersrand Basin was only discovered in 1886 in a farm along one of the conglomerate reefs in Langlaagte, with more gold-bearing reefs traced along the length of the Witwatersrand from Roodepoort in the west all the way to Germiston in the east (McCarthy, 2010). The known gold deposit in the Witwatersrand Basin is shown on Figure 2-1 below. These discoveries gave birth to one of the world's largest economic hubs, known then and today as the city of Johannesburg, and laid the foundation of South Africa's economy (Beavon, 2004). Over 52 000 metric tonnes of gold is said to have been extracted from the Witwatersrand Basin to the present day, with an inferred 30 000 metric tonnes remaining in the basin (Tucker, et al., 2016). Gold mining in the Witwatersrand Basin and other parts of the country started slowing down due to several factors including reduced gold ore concentrations, deeper mines making it difficult to mine, stricter mining regulations imposed, as well as relatively lower gold prices rendering some mines marginal or unprofitable. This has resulted in many gold mines closing down in the 1990s (Lowe, et al., 2020), leaving behind a legacy with a heavy price to be paid.

Underground mining operations are prone to water accumulation which require constant pumping in order to maintain good working conditions and also to avoiding flooding. Several factors were identified as sources of accumulating water in underground workings within the Witwatersrand Basin. These factors included but not limited to the following (Coetzee, et al., 2010):

- Direct recharge of rainwater falling onto open mine workings
- Percolating rainwater which seeps into the workings and cause disturbances to natural groundwater conditions in and around the mine

- Surface streams such as small river systems, storm water streams, urban pavement runoffs that lose water directly to mine openings and to the shallow groundwater systems above zones of shallow undermining and historical surface operations.
- Open surface workings that are connected directly to the underground workings provide a pathway for water to flow from the surface to the mine void.
- Mine dumps, which accelerate seepage where the volume of water entering the mine void is relatively high. This can lead to contamination of the water where there is an interconnection between water streams and tailings.
- Accidental release of water, sewage and stormwater reticulation systems are also a possible source of water into old abandoned shallow mines openings.
- Isotopic studies undertaken by the Council for Geoscience have indicated that heavy rainfall discharges result in rapid infiltrations which can make up a considerable proportion of groundwater on areas around mines.

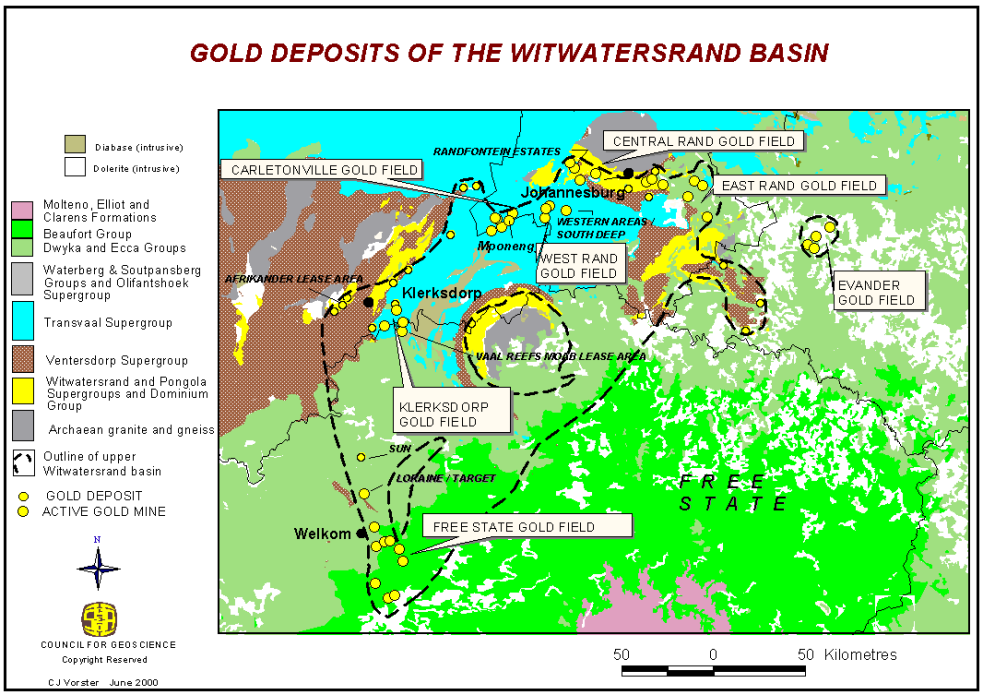

Figure 2-1: Demonstration of Gold deposits in the Witwatersrand Basin (Vorster, 2000)

The nature of gold deposits in the Witwatersrand resulted in a highly interconnected nature of the surrounding mines, and this is how the word 'basin' came about (Coetzee, et al., 2010). When several mining operations in the Witwatersrand Basin closed down in the 1950s and later thereafter

due to dropping profits, the pumping of water from underground mines also stopped (Coetzee, et al., 2010). This resulted in the accumulation of groundwater and flooding of abandoned mine shafts, the worst-case scenarios being situations where the water would find its way into neighboring mining operations (McCarthy, 2011). This water created not only huge operating expenses for active mines but also contamination and acidification of water through physical and chemical processes as explained on the following subsection 2.3. These processes lead to the formation of AMD water.

In the Witwatersrand Western Basin decanting AMD water was first noticed at defunct mine in August 2002 (Coetzee, et al., 2010), after which many other occurrences followed. The Western Basin has had pumping and treatment infrastructure, but it is said that the technology used could not treat the pumped mine water to satisfactory levels and the capacity was limited (McCarthy, 2010). Excess untreated mine water entered the downstream environment via the Tweelopie Spruit, an upper tributary of the Crocodile River System (Hobbs & Cobbing, 2007), resulting in not only ecological impact but also devastating health concerns to both human beings and other forms of lives in and around the area (McCarthy, 2010). The geological analogue between the between the Western and the Central Basin is said to be closely comparable, hence the Central Basin was therefore likely to face similar challenges to those in the Western Basin (McCarthy, 2010). Mine water from the Eastern Basin also got treated before being discharged, but the quality of water was also not at for discharge to the environment (McCarthy, 2010). The mine water situation in the Witwatersrand Basin have led to a creation of a Team of Experts on the subject, instructed by a Task Team chaired by the Directors General of Mineral Resources and Water Affairs to intervene report on its assessment and reappraisal of the situation with respect to acid mine drainage, focusing on the Witwatersrand Gold Fields (McCarthy, 2010). Some of the high-level solutions proposed thereafter included minimizing the impact of waste from mining on the water environment, minimizing AMD and other waste production in the mining sector, treatment of mining effluents and AMD, water ingress prevention, and decant management (McCarthy, 2010). These guidelines have laid down a strong foundation toward stabilizing the situation in the Basin and in other mining communities facing the same challenges. They have also probed some good grounds for research work around the subject matter for better answers and expansion of knowledge.

2.3 The Chemistry behind the Formation of Acid Mine Drainage waters

Acid mine drainage is a consequence of mining operations, particularly coal and gold mining. It is now a well-documented and thoroughly understood phenomenon. AMD water comes about when minerals found in mine tailings, rock stockpiles, mine dumps or even underground works containing metal sulfides, commonly pyrite ("fool's gold") and pyrrorite, reacts with oxygenated water (Kirby, 2014). Mineral extraction during mining operations involves an extensive fragmentation of rock masses, a process which increases the surface area of reacting rock surface, and thus favoring AMD formation process. Pyrite happens to be in much greater amount than all other metal sulfides in an ore mineral (Bejan & Bunce, 2015), making it the principal facilitator of AMD formation process. The process of AMD formation from pyrite is demonstrated by Reaction 2-1 below (Nleya, et al., 2016).

$$2FeS_2(s) + 7O_2 + 2H_2O \rightarrow 2Fe^{2+} + 4H^+ + SO_4^{2-} \qquad \textbf{Reaction 2-1}$$

This reaction is a representation of how ferrous iron and sulfuric acid, common AMD constituents form from the oxidation of pyrite with oxygen in the presence of water. This process is also dependent on several factors in the reacting environment. These include pH, availability of dissolved oxygen and bacterial activity amongst other factors (Bejan & Bunce, 2015). Under conditions where there is sufficient dissolved oxygen and pH values are greater than 3.5, ferrous iron produced by Reaction 2-1 can undergo oxidation to form ferric iron according to Reaction 2-2 below (Nleya, et al., 2016).

$$4Fe^{2+} + O_2 + 4H^+ \rightarrow 4Fe^{3+} + 2H_2O \qquad \textbf{Reaction 2-2}$$

An orange-red ferric hydroxide precipitate is likely to form at pH values between 2.5 and 3.5, causing a decline on both ferric iron concentration and pH according to Reaction 2-3 below (Nleya, et al., 2016). Pyrite further reacts with the remaining ferric iron to produce even more ferrous iron and sulfuric acid according to Reaction 2-4 below (Nleya, et al., 2016). It is the acid produced that lowers the pH of mine water, rendering it acidic. The lowered pH then makes it favorable for heavy

metals contained in mineral deposits to leach out and remain in solution, thereby increasing the concentrations of dissolved metals in AMD water formed.

$$4Fe^{3+} + 12H_2O \rightarrow 4Fe(OH)_3(s) + 12H^+ \qquad \textbf{Reaction 2-3}$$

$$FeS_2(s) + 14Fe^{3+} + H_2O \rightarrow 15Fe^{2+} + 2SO_4^{2-} \qquad \textbf{Reaction 2-4}$$

It should be noted that other metals and metalloids that can potentially contaminate AMD include Al, As, Cu, Zn, Pb and Mn (Younger, et al., 2002). Pyrite, however, is the main host of Sulfur in South African coal, and it is also associated with Witwatersrand Basin gold deposit (McCarthy, 2011). Its oxidation results in increased Fe^{2+} concentration in mine water, and it has been shown that dissolved iron concentration is typically the highest in AMD (Lottermoser, 2010). The process for iron removal could provide guidelines on how the other metals can also be removed from AMD. Conditions such as starting water pH, dissolved oxygen concentration, alternative oxidants, bacterial activities and temperature under which AMD is being formed may play a role in determining which metal minerals are oxidized and the resulting products (Kirby, 2014). Guidelines on mine water qualities and contaminants concentrations were determined by relevant authorities, with those relevant to this research work presented on Table 2-1 and Appendix B.

Table 2-1: Contaminant levels and other water quality specifications acceptable for wastewater discharge into a watercourse (Department of Water and Sanitation, 2013)

Parameter	General Limits	Special Limits
pH	5.5 - 9.5	5.5 - 7.5
Electrical Conductivity (mS/m)	70 -150	50 - 100
Calcium (mg/L Ca)	-	-
Magnesium (mg/L Mg)	-	-
Sodium (mg/L Na)	-	-
Sulphate (mg/L SO4)	-	-
Chloride (mg/L Cl)	*<0.25	*<0.25
Total Alkalinity (mg/L CaCO3)	-	-
TDS (mg/L)	*1000 - 5000	*1000 - 5000
Fe^{2+} (mg/L)	*<0.3	*<0.3

*There are no limits for stipulated on the South African Government Gazette for regulation purposes but factors such as area crop tolerance amongst others may make it necessary to consider these limits (Grewar, 2019)

Having real-time knowledge of physical and chemical properties of an AMD water stream is critical as it helps determine the water's toxicity levels, the waster's suitability for various uses including discharging it to fresh water coarse, as well as the choice of treatment processes in situations where the water does not meet relevant regulatory standards. South African has a set of compulsory guidelines the frequency at which industrial wastewater should be monitored. According to the Government Gazette No 36820 (Department of Water and Sanitation, 2013),

- Discharge quantities must be metered and the total recorded weekly.
- Wastewater discharges qualities must be monitored once every month by grab sampling.
- Such monitoring must be done at the point of discharge into a water resource.
- Results must be submitted to the responsible authorities.

Various mine water properties including some of those listed on Table 2-1 are monitored on known AMD discharge points in South African. For this research work, sample stream flowrate, pH, EC, Fe^{2+}, Na^{2+}, TDS, Cl^{-} and SO_4^{2-} concentrations were considered priority properties. The subsequent sections present brief discussions on how each of these properties is deemed important.

2.3.1 Stream Flowrate

A mine wastewater stream can either be a man-controlled or naturally decanting as shown on Figure 2-2 below. In active mines and defunct mines that are still receiving attention, AMD or wastewater is pumped out, treated in some cases, and discharged into a watercourse. Some abandoned defunct mines on the other hand get flooded to and eventually overflow due to several factors as discussed under 2.2 above. Despite the differences the nature of these two conditions, knowing the flowrate at the stream discharge point is very important for various reasons. One most important reason in this case is being able to tell whether to follow general or special limits when analyzing contaminants and other stream water properties. According to Government Gazette No 36820, general limits are only applicable to any wastewater discharge points where the flowrate

does not exceed 2000 m^3 a day into any water resource that is not part of the Listed Water Resources, shown as Appendix A of this report (Department of Water and Sanitation, 2013). If a discharge point does not meet one or both conditions, special limits automatically apply.

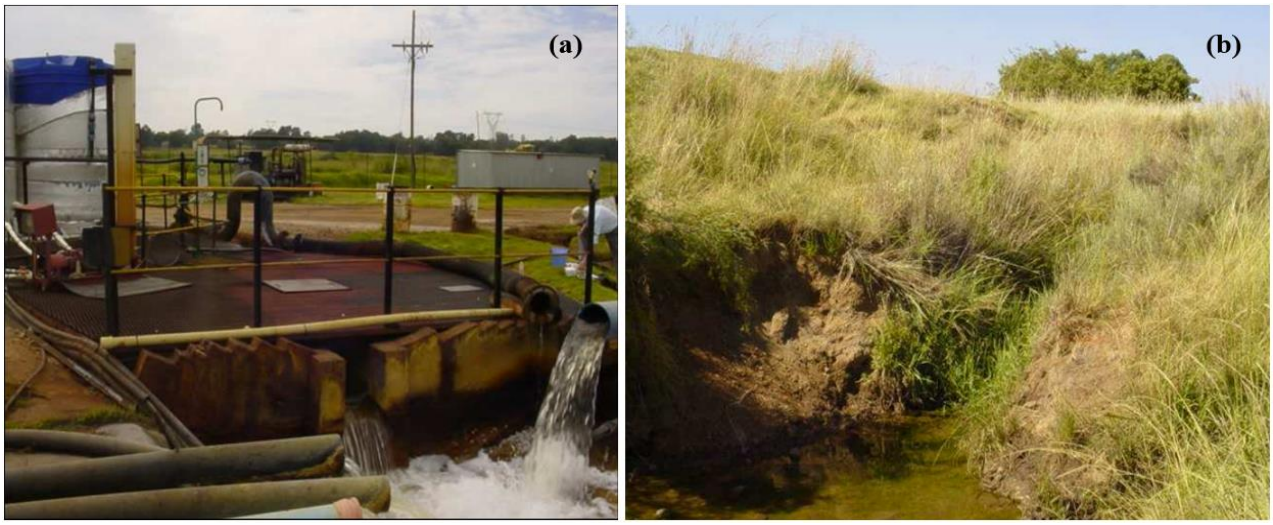

Figure 2-2: Man-controlled (a) and naturally decanting (b) AMD water discharge points where sampling takes place (Hobbs & Cobbing, 2007)

2.3.2 pH

Water pH is amongst the most important water properties. pH alone can render water toxic or unsuitable for use on various applications. AMD water is associated with low pH because of numerous oxidation reactions that take place between metals, water, and heat during the formation of AMD water. The lower pH further facilitates the dissolution of metals, rendering AMD water even more toxic as already alluded in this section. When this Acidic water finds its way into watercourses or ground surfaces, it makes the ecosystem inhabitable for flora and fauna, causing great destructions to those using water from AMD infested watercourses as irrigation water and ultimately having an impact on the overall food production value chain. The South African Government Gazette No 36820 recommends a 5.5-9.5 and 5.5-7.5 as general and special pH limits for industrial wastewater discharged onto a water resource, respectively as shown on Table 2-1. Anything outside these ranges results in the incurrence of penalties (Department of Water and Sanitation, 2013).

2.3.3 Electrical Conductivity

Electrical conductivity of water is the term used to refer to water's ability to conduct electrical current. Salts and other chemicals that dissolve in water break down into positively and negatively charged ions and these free ions conduct electricity (Mihir, et al., 2015). It can then be said that the concentration of free ions in water is the measure of water EC, and it helps indicates the level of water purity. Industrial wastewater regulations in South Africa takes EC into account because higher water EC levels can potentially affect water user depending on water applications. Cooling towers and boilers are amongst process that can be impacted negatively they make used of water with high EC levels. The use of water with high EC levels in these processes like these can result in scale build-ups in pipes and other equipment used, causing blockages and other undesired performances. The South African Government Gazette No 36820 recommends a 70–150 and 50-100 as general and special pH limits for industrial wastewater discharged onto a water resource, respectively as shown on Table 2-1.

2.3.4 Dissolved Iron Concentration

Iron is the fourth most abundant element on the earth's crust, with only oxygen, silicon and aluminum contributing higher proportion than it (Frey & Reed, 2012). It constitutes 5.6% by mass and it is found in compound form, mostly oxides (Frey & Reed, 2012). It is very likely to found iron oxides even in non-iron mines, and as already shown above, iron compounds are amongst the driving factors in the formation of AMD water. Reaction 2-1 and Reaction 2-4 are a representation of how ferrous iron, otherwise known as dissolved iron, is form during AMD water formation. Dissolved iron in the Witwatersrand Basin has much higher concentration than other metals involved. Knowing its concentration in water can give an ideal on how much dissolved solids may be in AMD water. Dissolve metals in water render that water toxic and unsuitable for human consumption. It is for these reasons that knowing dissolved iron concentration in water disposed onto water resources is important. While there are no regulations of industrial wastewater iron content in South Africa, discretion is always exercised when disposing wastewater by various industries as the dangers associated with some dissolved metals are well understood (Grewar, 2019). A concentration limit of 0.3 mg/L for industrial wastewater is recommended as shown on Table 2-1.

2.3.5 Total Dissolved Solids

Total dissolved solids refer to a combined sum of all inorganic salts and organic matter contained in water and can pass through a 2 microns filter (Weber-Scannell & Duffy, 2007). TDS is considered an important property of water because depending on its level, it provides the level of toxicity through various ways. An increase in TDS may well be an indication of an increases in salinity, a phenomenon that has been shown to cause shifts in biotic communities, limit biodiversity, exclude less-tolerant species and cause acute or chronic effects at specific life stages important to aquatic life by keeping cell density balanced (Weber-Scannell & Duffy, 2007). Changes in TDS can also be attributed to changes in ionic composition of water. Increased or decreased ionic composition in water can promote population growth of certain species while hindering those of other or endangering other species altogether (Derry, et al., 2003). TDS of industrial effluent water discharged into water resources is however not regulated in South Africa. It should however be noted that due to salinity and TDS's close relation to EC, regulating EC may well account for TDS as well.

2.3.6 Sulphate Concentration

High sulphate ion concentrations are very prevalent in mining wastewater. Various oxidation reaction that takes place during AMD water formation facilitate the formation of sulphate ions as seen in the demonstration of iron demonstration above. High sulphate ion concentration can have a negative impact on aqueous and terrestrial environments, and human and animal health (Moreno, et al., 2009). Sulphate ions also facilitates the formation of salts which can form scales and cause pipe blockages in downstream water applications. Since these ions are already accounted for in water EC, there are no regulatory standards for sulphate concentration on industrial effluent water in South Africa.

2.4 AMD Water Properties Measurements in the Witwatersrand Basin

AMD water properties are known to be well outside limits stipulated as regulatory standards for industrial effluent water. The study done by Hobbs, et al. (2007) looked at six different water sources of different geological makeup in the Witwatersrand Western Basin to provide means to

distinguish them hydrochemically. Various property analysis tests were run on the samples and the summary of results are given on Table 2-2 below. It is evident from this summary of result that the AMD-associated water properties were non-compliant with all industrial wastewater regulatory standards set out in the South African Government Gazette No 36820. This is a clear indication that AMD water required attention. Realtime properties measures should be known for treatment purposes before disposing it into the environment. It should however be noted that obtaining some of these properties can be a difficult exercise, and for this reason, a tool to project or predict some of these properties from easy-to-measure properties would come in very handy. Artificial Neural Network were chosen to come up with models for Witwatersrand Basin AMD water property predictions.

Table 2-2: Water quality data for different sources in the Witwatersrand Basin (Hobbs, et al., 2007)

Parameter	Mine Water Source					
	Pristine karst aquifer	**Fractured aquifer (quartzites)**	**Lower Riet Spruit karst groundwater**	**Tweelopie Spruit surface water**	**Acid mine drainage & associated**	**Blougat Spruit surface water**
pH	7.3	5.8	7.2	7.1	4.8	7.8
Electrical Conductivity (mS/m)	19.9	27.7	88.1	299.2	264.2	91.0
Calcium (mg/L Ca)	20.0	20.9	79.4	551.6	222.1	45.0
Magnesium (mg/L Mg)	10.1	13.3	40.4	121.6	125.0	11.0
Sodium (mg/L Na)	5.7	13.8	50.2	152.7	69.1	77.4
Sulphate (mg/L SO_4)	18.7	70.5	204.7	1997.8	1771.6	110.0
Chloride (mg/L Cl)	4.2	17.9	54.6	40.8	43.5	61.0
Total Alkalinity (mg/L $CaCO_3$)	68.0	16.7	134.8	27.2	21.5	236.0

2.5 Methods Currently Used to Remediate AMD and other Related Government Responses in South Africa

AMD water has resulted in devastating environmental impacts in mining communities around the world. Newly planned mines and those currently operating with no known AMD cases have options on approaches to take in order to prevent the formation of AMD, and to avoid adding to the persisting AMD related challenges. AMD formation prevention can be achieved through the removal of at least one of the four necessary conditions for AMD to form. These conditions include the presence of sulphide minerals, air, water, and bacterial activities (Bejan & Bunce,

2015). In some mines and mining regions, these AMD formation preventative measures cannot be opted for because damages have already been done over the years of mining. Where AMD waters have already been discovered and are affecting the environment, remedial measures become necessary. There are fully reliable remedial options in operation today, and these range from active treatment technologies to passive treatment options. A brief comparison of some of these options is shown on Table 3 below. Active treatment technologies are those that require the addition of chemicals and also energy inputs in some instances (Bejan & Bunce, 2015). Active treatments technologies follow the principles of one or more of the following processes: aeration, neutralization, reverse osmosis, ion exchange, electro dialysis, and electro zeolites (Ali, 2011). Passive treatments take form of lagoons and cascades, and these make the use of natural substances and or biological processes (Ali, 2011).

The standard way of treating for AMD in South Africa is to aerate the water to oxidise the iron (and most heavy metals simultaneously) and add lime to bring the pH up (du-Toit, 2018). The eMalahleni Treatment plant in Mpumalanga currently use reverse osmosis as an AMD metal ion removal method. In this method, water is pumped through a membrane punctured with microscopic holes that allows water through but keeps metals behind (du-Toit, 2018). Water from clean water sources is also used in some areas to dilute AMD water to acceptable levels in contaminated water bodies or before releasing it to the environment (du-Toit, 2018).

The Department of Water Affairs (DWA) has subsequently implemented some strategies to address various AMD-linked water challenges. Amongst those is the Trans-Caledon Tunnel Authority (TCTA) was instructed to undertake emergency works to protect the respective Environmental Critical Levels (ECLs) in the Eastern and Central Basins and to lower the underground mine water levels in the Western Basin; and to neutralise and remove the heavy metals from the pumped underground mine water prior to it being released to surface water resources (Vogel, 2012). Another major move by the government was the establishment of the Hydrological Monitoring Committee (HMC) under the chairmanship of DWA to monitor the re-watering and quality of water, in respect of the East, Central and West Rand (Vogel, 2012). A Feasibility Study was also initiated as a move to address underground mine water induced salt loading of major river systems in the long-term, back in February 2012 (Vogel, 2012).

Table 2-3: Comparison of different AMD remedial options (Coetzee, et al., 2010)

Characteristic	Active treatment	Passive treatment	In *situ* treatment
Application to phase of mining	Exploration and operational phases because it requires active control and management. Closure and post-closure applications mainly associated with large flows	Most attractive to the closure and post -closure phases, because it requires only intermittent supervision, maintenance, and monitoring of self-sustaining processes	Appropriate to the exploration and operational phases because it requires ongoing operation and maintenance
Operational involvement	Active and ongoing plant operations and maintenance systems and personnel	Constant operations not required, but regular maintenance essential	Active and ongoing operational personnel required, but permanent presence on site not required
Operational inputs and materials	Requires chemicals, operations staff, maintenance staff, electrical power, continuous and/or regular monitoring	Self-sustaining processes, periodic maintenance, intermittent monitoring. May require replacement or supplement of materials at low frequency	Requires chemicals, operations staff, intermittent field maintenance, electrical power and low frequency monitoring
Supply of power	Electrical and mechanical energy sources	Natural energy sources of gravity flow, solar energy and bio-chemical energy	Electrical and mechanical energy sources
Management and supervision requirements	Ongoing management engagement, constant facility supervision	Low level management engagement and low frequency intermittent supervision	High frequency supervision, but no permanent site presence required
Range of applications (flow rates and constituents)	Application to all flow rates, especially high flow rates and any constituent of interest	Mainly applied to low flow rates and acidity, metals, and sulphate removal	Large spectrum of volume and flow applications, mainly to deal with acidity and metals removal
Treated water quality	Treatment process can be purpose built to deal with spectrum of treated water requirements	Treated water quality poorer and more variable than other options	Treated water quality lower and more variable than active treatment process
Waste sludge and brine production	Waste sludge and brine are produced, depending on level of treatment, requiring disposal	No brine production, but longer-term liability to deal with accumulated pollutants in wetland sludge	Sludge and waste production accumulated in situ that may pose a long-term environmental liability.
Capital investment cost	High capital investment and periodic capital replacement required	Moderate capital investment, with periodic reinvestment to replace depleted wetlands media	Low capital investment typically to deal with a short-term problem
Operating and maintenance cost	High operating and maintenance cost, with some potential for cost recovery by sale of product water, metals and by-products.	Low operating cost.	Moderate operating costs, but chemical usage may be high owing to process inefficiency.

2.6 Artificial Neural Networks

2.6.1 Introduction to ANN

The passing of time has unveiled the need for humans to think and act faster in order to keep up with the growing need to complement the ever-growing innovation space and to sustain human survival under different environments. This then puts efficiency and productivity associated with numerous human activities on the spotlight. Humans have demonstrated their ability to work smart in order to ensure both efficiency and increased productivity in tasks that they lay their hands on, however humans are nothing short limitations. Human limitations in their abilities may include pace, strength, variation in intelligence, variation in their ability to recall, memory capacity amongst other things. It is undoubtedly evident that anything that can overcome human limitations as alluded here can bring about an increased productivity and efficiency. These are some of the reasons why Artificial Intelligence (AI) came into existence.

Artificial Intelligence is an area of Computer Science that deals with the creation of machines that can mimic human behaviour. In this field computers are programmed to exhibit human traits and capabilities such as problem solving, learning, planning, reasoning, possession of knowledge, ability to logically move objects around and many more. One of the ways to impart machines with such traits is through Machine Learning (ML). By definition, Machine Learning is a field under AI which studies algorithms and techniques for automation of solutions to complex problems that would otherwise be hard to program using rule-based or conventional programming (Rebala, et al., 2019). Some of the techniques used in ML include Decision trees, Random Forest, Linear Regression, Artificial Neural Networks, Principal Component Analysis (PCA), Support Vector Machine (SVM) amongst others (Ayodele, 2010).

ANN algorithms were chosen for this research work. The choice was solely based on their robust nature in their predicting abilities without any head-to-head comparison with other known suitable techniques. This section unpacks the architect and theoretical background of ANN. It encompasses how an ANN is inspired biologically, the generic structure, the learning algorithms, limitations,

the mathematics involved, as well as how ANN complements the fast-growing field of Artificial Intelligence.

2.6.2 Why Artificial Neural Networks

ANNs came about due to a need develop artificially intelligent systems that can handle more sophisticated computations that existing computers cannot handle, in a way a human brain would. ANNs use the same analogy as that seen on biological nervous systems such as the brain, to process information (Jiang, et al., 2010). The structure of this information processing system is the most important element of the tool, and it is what defines its unique processing abilities. Just like humans, ANNs learn relationships that exist within a given set of data through examples which are typically in a form of experimental knowledge (Jiang, et al., 2010). ANNs are composed of a large number of highly interconnected processing conduits referred to as neurons, working simultaneously to handle specific problems (Jiang, et al., 2010). A Unique ANN must be configured for a specific application through a learning process.

ANNs follow the exact analogy as human brains to provide an approach that has great potential in computationally solving complex problems despite not having reached the stage where they can mimic a trivial brain function in full (Basheer & Hajmeer, 2000). They have shown many characteristics and predictive capabilities which make them attractive and appropriate for both forecasting and obtaining real-time data. More so, ANN have been widely and successfully applied in water resource engineering in recent years (Huang & Foo, 2002). This has strongly influenced the choice of ML technique to use for this research work.

2.5.2.1 Advantages of Artificial Neural Networks (Sonali & Maind, 2014) (Mijwel, 2018):

- They have ability to learn how to do tasks based on the data or events given for training or initial experience then make decisions or projections on similar events.
- They have the ability to work with incomplete knowledge after they are trained. The deterioration of performance is however dependent on the importance of the missing information.

- ANNs have the ability to create its own organization or representation of the information it receives during learning time.
- The gradual corruption of the network only slows the network over time. The corrosion does not happen immediately.
- ANNs have a parallel processing capability. Their numerical strength enables them to perform more than one computation at the same time, and special hardware devices are being designed and manufactured to make use of these capabilities.
- ANNs store information on the entire network and not on a database. Loss of pieces information does not prevent the network from functioning.

2.5.2.2 Disadvantages of Artificial Neural Networks (Sonali & Maind, 2014) (Mijwel, 2018):

- The structure of network requires a processor with parallel processing power.
- ANNs do not give exact solutions, predictions or projections.
- Large complexity of the network structure, with no specific rule to determine the structure. Determining the best structure requires experience and rial and error.
- The duration of the network is not known. The network simply reduces the value of the error on the samples and stops to signify the completion of training.
- All problems shown to the network must be translated into numerical values before being introduced to ANN.

2.5.2.3 Neural Networks vs Conventional Computers

Neural Networks process information the same way human brain does. The network is composed of a large number of highly interconnected processing conduits (neurons) working in parallel to solve a specific problem (Grosan & Abraham, 2011). Neural networks learn by being trained using examples relating to the kind of problems the NN is intended to solve. They cannot be programmed to perform a specific task. Examples used to train NNs must be selected carefully otherwise the network might function incorrectly. The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable (Grosan & Abraham, 2011).

On the other hand, conventional computers use a cognitive approach to problem solving. The reasoning behind the way the problem is to be solved must be known, stated in small unambiguous instructions and confirmed to work satisfactorily (Haung, 2017). These instructions are then converted to a high-level language program and then into machine code that the computer can understand. The functioning of these machines is totally predictable. Should anything go wrong, the problem is due to a software or hardware fault (Haung, 2017). Neural Networks and conventional algorithmic computers are not in competition but complement each other. There are tasks that are more suited to an algorithmic approach like arithmetic operations and tasks that are more suited to Neural Networks. Even more, a large number of tasks require systems that use a combination of the two approaches (normally a conventional computer is used to supervise the Neural Network) in order to perform at maximum efficiency (Haung, 2017). Presented on Table 2-4 below is a head-to-head comparison between neurocomputers (NN-based computers) and conventional algorithmic computers at a more detailed level.

Table 2-4: Comparison between NN-based computers and conventional computers (Haung, 2017)

	Conventional computer	**Neurocomputer**
Components	Transistor to construct switching circuits	Neuromorphic devices to imitate biological neurons and synapses
Architecture	Turing model Von Neumann architecture	Based on biological neural networks
Hardware	Arithmetic logic unit Control unit Registers Data and control bus	Neuron array Synapse array Routing bus
Basic software	Operation system Complier	Neuromorphic array configuration Mapping neural networks to neuromorphic array Dynamic behaviors management
Application software	Data structure + algorithm = (artificial) program	(Big) data + training algorithm = (autonomous) learning
Typical functions	Calculation Logical and structural data processing	Structure discovery from non-structural sensing and perception
Target intelligence	Artificial intelligence	Autonomous intelligence Artificial General Intelligence (AGI)
Science base	Cognitive science	Neuroscience

2.6.3 Neuron Network Architecture

2.5.3.1 Biological Neuron

As already mentioned in 2.6.1 ANN are based on natural human nervous system analogy to carry out their function. The human nervous system is made up of hundreds of billions of neurons of different types and lengths depending on their location in the body (Haung, 2017). A simplified schematic diagram of a biological neuron is shown as Figure 2-3. There are three main functional units in a biological neuron. The first unit is the cell body which houses the nucleus that contains hereditary information, and plasma that houses necessary molecular equipment used for producing material a neuron needs (Jaing, et al., 1996). The second main unit is the cluster of dendrites whose function is to receive electrical signals from other neurons and carry them over to the cell body. The third unit is the axon which receives signals from the cell body and transfer them to various dendrites belonging to other neurons close by through synaptic terminals. The amount of strength of the electrical signals that passes through a receiving neuron is determined by the strength of the signal released by each feeding neuron, the feeding neurons' synaptic strength, and the receiving neuron's capacity or threshold (Jaing, et al., 1996). Signals received can either be activate or inhibit the activation of a neuron (Haung, 2017). A neuron can receive signals from various feeds, combined them in ways learned through experience and transfer output signals to other receptors for further actioning until a familiar message is put together.

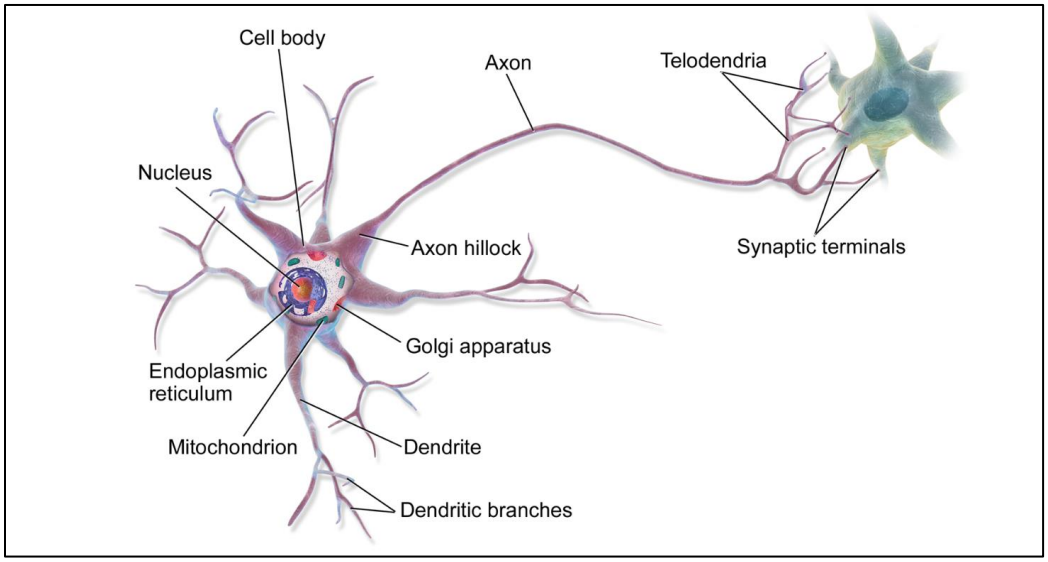

Figure 2-3: Parts of a Biological Neuron (Blaus, 2013)

2.5.3.2 Artificial Neuron

Artificial neurons are analogous to biological neurons in they are structurally perceived and the functions of corresponding components. They are to Neural Networks what biological neurons are to a human nervous system, a fundamental processing element whose function is to collect information and construct outputs for the neurons in the succeeding layer of the network. A schematic diagram of an artificial neuron is given as Figure 2-4. A neuron receives its inputs either from other neurons or the user program, combines them in a special way to form a single input *u*, pass it through an activation function with a threshold, and sends its output *y* to other neurons or the user program. A neuron consists of several elements called nodes, denoted as x_i and analogous to axons-dendrites combination. Each node receives signals for the neuron. The connection weights denoted as w_i are analogous to synapses.

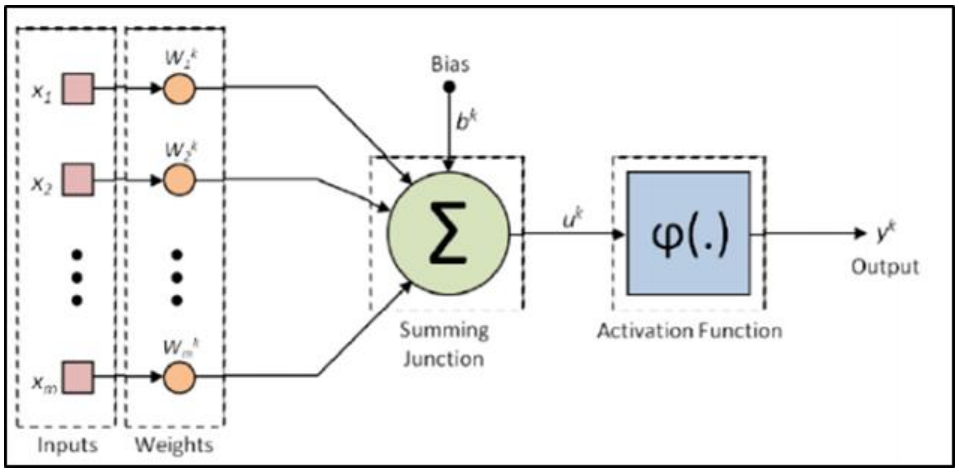

Figure 2-4: Artificial neuron model (Tumer & Edebali, 2015)

Figure 2-4 is also referred to as a perceptron. A perception can be trained by passing special learning rules, and this process is analogous to human learning process. An ANN is formed when artificial neurons are lined up in a user defined manner to bring about a common output. These networks are characterized by their pattern of connections between the neurons, the method with which connection weights are determined (training or learning algorithm), and their transfer function.

2.6.4 Connections

As already mentioned, ANN is artificial neurons connected to form a network. The paths between neurons are called connections and they play a vital role in defining an ANN. The function of a connection in an ANN is to create a link through which an input can be sent from on neuron to another. It is very common that neurons on adjacent layers are fully interconnected and all the information flows through these connections in one direction. Networks with this type of connections are referred to as feedforward networks. There are also networks with nodes in one layer connected to nodes in the same layer, the next layer, or the previous layer (Maier & Dandy, 2000). Those networks are referred to as recurrent or feedback networks depending on how the connections within them are configured. Depending on the on the ANN architecture and nature of input data, there might exist additional connections going to further layers or even missing connections between certain layers (Zhang, et al., 1998). There are various types of ANN classifications based on the type of connections involved. Feedforward and Recurrent/Feedback are the two major categories under which every other known network fall as depicted on Figure 2-5.

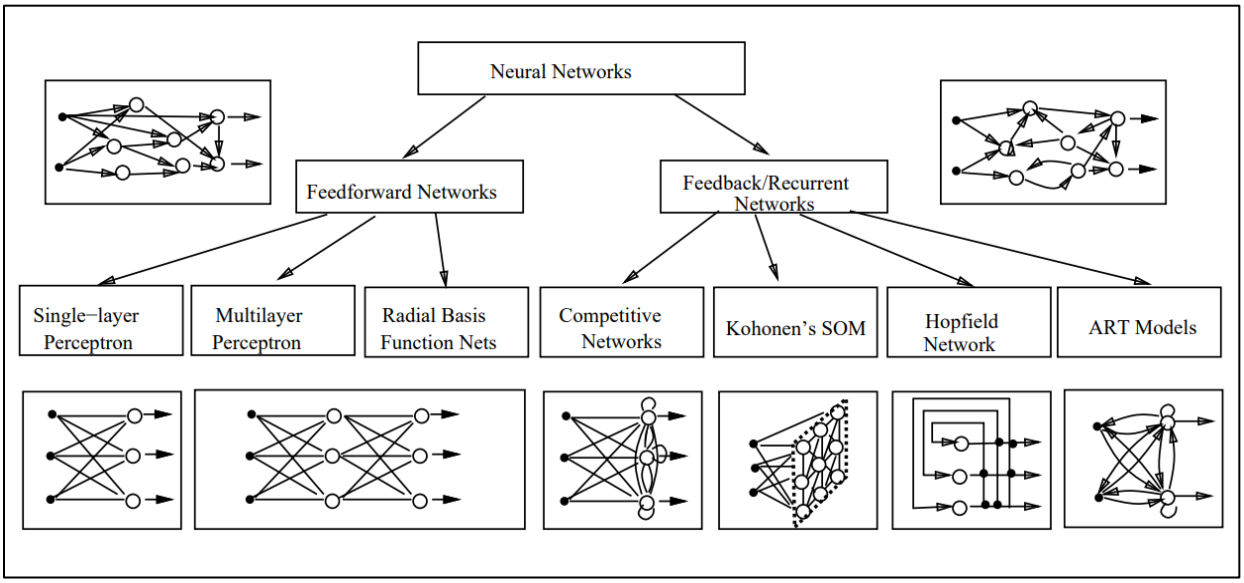

Figure 2-5: Various ANN classifications based on network architecture (Jaing, et al., 1996)

2.6.5 Weights and Biases

Connections within a network are each equipped with an individual weight and bias that modifies the signal flow on the respective connection. Weights represents connection strength between neurons and gets adjusted during learning algorithm. Each input is multiplied by a weight factor based on the experience a network gained during learning. The weight factor then determines the amount of impact the associated connection will have on a neuron's overall activation function. The bias neuron on the other hand works as a fine-tuning nob. In a multilayer ANN, each layer has a single bias and connected to all the neurons in the next layer. Final refinement of neuron's output is made through the addition of a bias to a successive layer. As the information is stored and distributed through weights and bias neurons in a neural network, a sight destruction on these parameters can result into a large effect on the neuron's ability to recall learnt functions or relationships.

Depending on the influence of the input, the value of the weight of an artificial neuron could be higher or lower. Weights can sometimes be negative, which means that the signal is suppressed by the negative weight. Using the artificial neural model given as Figure 2-4, an output of a neuron is given by Equation 2-1 (Jaing, et al., 1996), where there are n inputs, y is the output, x_i is an input, w_i is the corresponding input weight, b is the bias for that layer and f is the chosen activation function.

$$y = f\left(\sum_{i=1}^{n} x_i w_i - b\right) \qquad \text{Equation 2-1}$$

2.6.6 Transfer Functions

Transfer functions are applied to net input received by a node to adjust the inputs into a scale and data type the network can best work with. Having large numbers as input values can result in poor convergence indication. Inputs in a form of images may also need to be transferred to what the learning functions can easily work with, and numerical figures usually works well. These functions can be linear or non-linear depending on the user's choice. The output or range of the transfer function is usually 0 to 1 or -1 to 1. Choosing an appropriate transfer function is a very important

step when working with ANNs. Considering the nature of network input data is very important in choosing a transfer function to use for that specific network. Some of the most common and useful transfer functions are pure linear, log-sigmoid and tan-hyperbolic functions (Nasr, et al., 2011). For non-linear datasets log sigmoid and hyperbolic tangent are the most suitable transfer function choice, whilst purelin works well on linear datasets. These three transfer functions are defined by Equation 2-2, Equation 2-3 and Equation 2-4 respectively (Nasr, et al., 2011), where *x* is a every discrete network or neuron input value.

Log-sigmoid function:

$$f(x) = \frac{1}{1 + e^{x}} \qquad \text{Equation 2-2}$$

Tan-sigmoid function:

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \qquad \text{Equation 2-3}$$

Linear function:

$$f(x) = x \qquad \text{Equation 2-4}$$

2.6.7 Training of ANNs

During the training stage a neural network is given a sample from a set of the input data. Weights and biases of the neural network are adjusted are adjusted based on the learning function applied during training (Fausett, 1994). There are two learning or training procedures for ANNs. The first procedure is supervised learning, where an external source is required to guide the progression of the learning process. Mean Square Error (MSE) is an example of supervised learning algorithms (Fausett, 1994). A portion of the overall dataset is randomly selected and an input-output pairs from the selected portion are used during supervised learning. All inputs from this subset are propagated through the ANN and the model outputs are compared with the target data. With LMS

algorithms it is assumed that an optimal solution is reached during the learning process when the mean square error between the corresponding model output and target value is at its lowest. At this point the learning process gets terminated.

The second learning procedure is the unsupervised learning. During unsupervised learning by ANN the system organizes itself using its internal criteria and the training gets terminated by itself when these internal criteria are satisfied (Fausett, 1994). The risk with unsupervised learning is the possibility of over-training, where a neural network gets overexposed to an example to a point where it has fully memorized the example together with the example's inherent noises (Cybenkot, 1989). At this stage the trained network will fail to generate reliable outputs if a different feed is introduced.

2.6.8 Learning Algorithms in Neural Network

The purpose of the training process is to optimize the ANN to minimize the differences between ANN generated output and target data values provided. Such is achieved by adjusting the weights between nodes of a network. For this research work supervised learning methods was chosen as it provides a better control on how an ANN model turns out. This this reason, only the supervised training functions will be discussed in this report. The three widely used training algorithms of multilayer feedforward networks are back-propagation, Conjugate Gradients and the Levenberg-Marquardt algorithm (Haykin, 1994). Each of these algorithms is briefly discussed below.

2.5.8.1 Back-propagation algorithm

Back-propagation is the most used training algorithm for feedforward ANNs and is a gradient descent method (Haykin, 1994). When this algorithm is used during ANN training, an input vector is introduced and processed in a forward propagation mode to produce a corresponding output vector (Zhang, et al., 1998). Elements of the output vector generated are then compared with the given corresponding actual output vector elements. The error of network is then given as the sum of the squared partial errors of network as shown on Equation 2-5. The cycle continues while weights are adjusted, and training stops when this error is below the tolerance value. Each cycle run is referred to as an epoch.

$$E = \sum_{k=1}^{p} E_k = \sum_{j} \sum_{i} (d_{ji} - y_{ji})^2$$ Equation 2-5

Where E_k is partial network error, j is the input number (position) in the training set, i is the output node number (position), and d_{ji} and y_{ji} are the target and actual output values for the i^{th} node on the j^{th} pattern.

2.5.8.2 Conjugate Gradient algorithm

Conjugate gradient algorithms use the same analogy as the back-propagation technique. The different between the two approaches is that a basic conjugate gradient algorithm adjusts the weights in the steepest negative direction. The direction in question here is one in which the performance function is decreasing most rapidly.

2.5.8.3 Levenberg-Marquardt algorithm

The Levenberg-Marquardt (LMS) algorithm is based on two algorithms, namely steepest descent algorithm and Newton's method (Beale, et al., 2013). Out of these two optimization methods, the primary is based on first order Taylor series and following is on second order Taylor series. It is known for its speed and efficient implementation in MATLAB, as well as a generally good performance on moderately sized neural networks (Beale, et al., 2013). For these reasons, this algorithm is used in this research work.

2.6.9 ANN Simulation and Generalization

When coming up with an ANN model, original dataset is divided into three sub-sets, namely the training set, the validation set and the testing set. The split ratios can be decided upon arbitrarily, but it is always best to apply a few split ratios and choose one where performance is the highest (Mavani, 2014). The validation data set is for model performance validation as the name suggests, and it is applied on the trained ANN to confirm that the model can indeed be trusted, before subjecting it to a large dataset for robust testing. It is also necessary to put the trained ANN model

to a robust performance test before applying the model. In the simulation process, an ANN is subjected to input data never used in the training and validation process, then the ability of the model to match the target output values is measured. The stopping criteria during ANN training is satisfied when either,

- the network reaches a minimum error as specified by the modeler (Beale, et al., 2013),
- a maximum runtime/number of epochs specified by the modeler is reached (Beale, et al., 2013)

For the simulation purposes, two types of model performance assessment measures are commonly used, and these are:

- R value: This is referred to as a coefficient correlation, used in this research work as Pearson Correlation (Schober, et al., 2018). It is a widely used statistical tool used present a measure of association between two variables. The higher the R-squared value, the better the model performance in terms of explaining variability in observed output (Beale, et al., 2013). Programs such as MATLAB have built-in functions which automatically generate R values for each data set presented. The function used to calculate these R values is given by Equation 2-6 below (Ba, et al., 2019). This performance measure was the main determining factor when ranking the modules developed in this research work from best to worse. The logic that this measure follows is that the strength of correlation increases as R increases from 0 to 1, hence higher model absolute R magnitudes indicate better model performance. Presented on Table 2-5 below are some guidelines used to interpret the value of R.

Table 2-5: Conventional Approach to interpreting a Correlation Coefficient value (Schober, et al., 2018)

Absolute Magnitude of the Observed Correlation Coefficient	Interpretation
0.00 – 0.10	Negligible correlation
0.10 – 0.39	Weak correlation
0.40 – 0.69	Moderate correlation
0.70 – 0.89	Strong correlation
0.90 – 1.00	Very strong correlation

- MSE or RMSE: It is another method to quantify model bias and precision. It is a way to aggregate the model residuals into a single value. The lower the value of the MSE or

RMSE, the smaller the differences between observed and predicted values (Helsel, et al., 2002). These two performance indicators are given by Equation 2-7 and Equation 2-8 respectively (Zhang, et al., 1998).

For this research work, these performance indicators were merely used to compare different models that were developed and not necessarily for comparison with other studies. Models with lower MSE values were regarded as better performing models irrespective of the MSE magnitudes.

$$R = \frac{\sum_{k=1}^{p}(d_k - d_{av}) \times (y_k - y_{av})}{\sqrt{\sum_{k=1}^{p}(d_k - d_{av})^2 \times \sum_{k=1}^{p}(y_k - y_{av})^2}} \qquad \text{Equation 2-6}$$

$$MSE = \frac{1}{P}\sum_{k=1}^{p}(d_k - y_k)^2 \qquad \text{Equation 2-7}$$

$$RMSE = \sqrt{MSE} \qquad \textit{Equation 2-8}$$

Where P is the total number of inputs-output pairs in the set, d_k and y_k are the k^{th} target and actual output values. Symbols d_{av} and y_{av} are averages of the target and actual output values respectively.

Generalization is the ability of an ANN model to correctly approximate target values for given inputs that are not part of the training and validation sets (Mavani, 2014). Good generalization ability typically requires the following (Mavani, 2014):

- Inputs which contain enough information about the target that it is possible for the ANN to develop a functional relationship between inputs and outputs with an adequate degree of accuracy.
- The function which model is trying to learn is at least somewhat smooth, i.e., a small change in inputs produces a small change in outputs.

- The training cases are sufficiently large and representative of the subset or sample of the larger population of data that the model is required to be able to generalize.

Sometimes a model shows poor performance due to model overfitting. Overfitting occurs when the model learns too many specific input-output relationships (Haykin, 1994), making it important to always have less training data than the testing subset when deciding on the split rations.

2.7 Related Research Work

Artificial Neural Network have been tried and tested in various research work and they have been put in various applications today. Some of these studies were focused on wastewater property predictions, and treatment process performance predictions and optimization model development. It was through those studies and many other research works that the ANN tool earned its merits and has been proven useful.

2.7.1 Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT

In the study done by Nasr, et al., (2012), ANNs were used to predict EL-AGAMY wastewater treatment plant performance. It is known that the operational control of a biological Wastewater Treatment Process can be complicated because of variations in raw wastewater compositions, strengths and flow rates owing to the changing and complex nature of the treatment process. This paper focused on applying an ANN approach with a Feed-Forward Back-Propagation to predict the performance of EL-AGAMY WWTP-Alexandria in terms of Chemical Oxygen Demand (COD), Biochemical Oxygen Demand (BOD) and Total Suspended Solids (TSSs) data gathered during a research over a one-year period. COD, BOD and TSSs in the feed into the plant were used as network inputs, whilst the levels of the same water properties in the product were used as network output. Some of the model network properties used in the study included TRAINLM as a training function, LEARNGDM as an adoption learning function, MSE as a performance function, and the 3 hidden layers.

It is the nature of complexity and behavior of the wastewater streams that the study focused on, the choice of machine learning program and its properties that makes the study suitable for

comparison with the current study. The study Achieved an MSE value of 15 and a coefficient correlation of 0.90317 and seen on Figure 2-6 below. These performance indicators were rendered the model generalizing capabilities suitable for its intended application as proposed in the study. Whilst the model works best for the type of wastewater it is intended for, it might not be very useful in AMD water application.

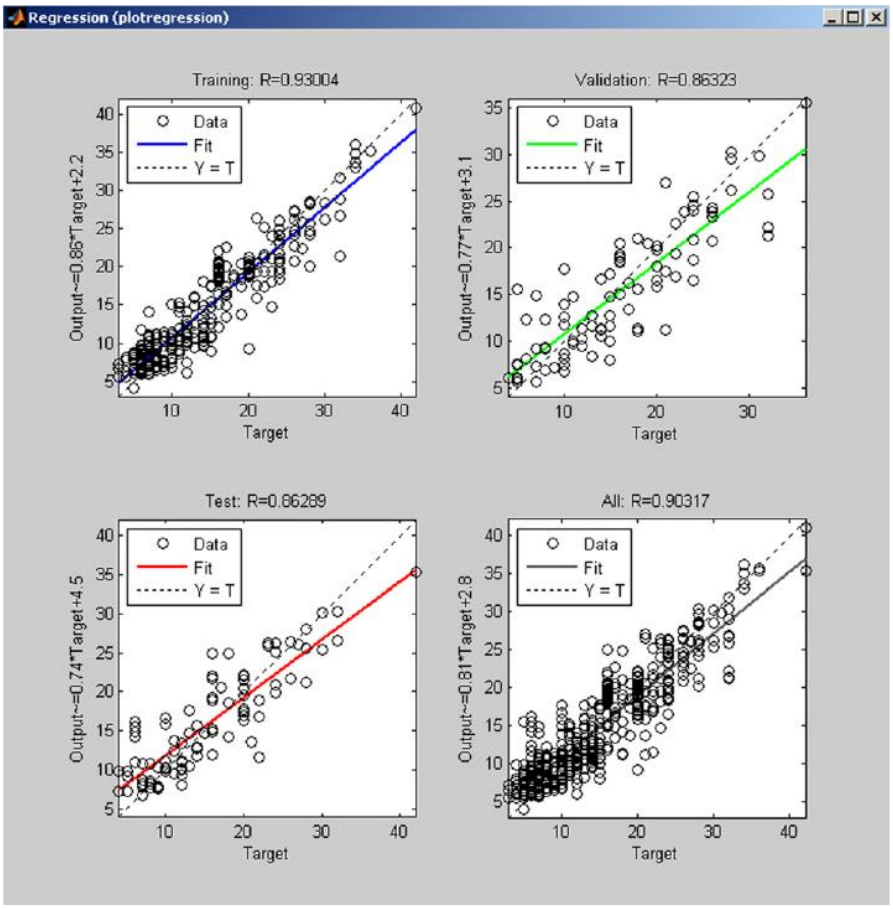

Figure 2-6: Network regression for the ANN modelled the EL-AGAMY wastewater treatment plant performance prediction (Nasr, et al., 2011)

2.7.2 The Application of Artificial Neural Network to Predicting the Drainage from Waste Rock Storages

A study done by Ma, et al., (2021) focused on outlining an approach to use ANN to predict the drainage flow rate and drainage chemistry based on weather monitoring data collected at mine sites. The study added to the efforts channeled towards the minimization of the impact that waste rock storage drainage, or simply AMD has on the environment. The added advantage of the

approach included doing away with the need for additional characterization make prediction because the relevant geo-bio-chemical mechanisms are embedded naturally in the monitored data, which can be captured through machine learning process (Ma, et al., 2021). Average temperatures, amounts of precipitation as well as days accumulation from the first day of monitoring were chosen as inputs into a feedforward NN. The output of the model were stream flow rates and chemical composition (acidity), both of which are influenced by weather, chemical reactions and time allowed for the process to take place. Such information come in handy in determine the wastewater treatment requirements any point in time.

The proposed model was trained using Levenberg–Marquardt backpropagation algorithm. The number of hidden neurons tested for an optimal size were 5, 10 and 20. The data split used was 80% for training, 20% for validation and 0% for testing. Key model performance indicators chosen were coefficient correlation and MSE. It was found that the proposed model could capably capture the underlying correlation between the drainage and weather, as not only seasonal fluctuations in a year but also long-term evolution across years (Ma, et al., 2021). Th model performed better when 20 hidden neurons were used over weekly average data. The MSE and R values for flowrates and acidity were 0.18 and 0.95, and 0.28, 0.84, in that respective order. The study presents some important breakthroughs in a form of a useful tool. However, its shortcomings are that the models can only be applied in the area where model training data was collected. The chemical nature of the Witwatersrand AMD may not be compatible with this model.

2.7.3 Application of Artificial Intelligence (AI) to predict mine water quality, a case study in South Africa

The Ermelo Coalfield study investigated the application of a hybrid system, which combined long-short-term memory nets LSTM and ANN to predict AMD water quality from an abandoned underground coal mine in Carolina Town, in Mpumalanga. The water properties of the Witkranz discharge site were used to develop AMD water characteristics predicting model. The inputs chosen for this purpose were rainfall, soil temperature, depth to water table and AMD pH. The model put a special emphasis on determining sulphate concentration as this is a good indicator of the level of pollution associated with AMD water (Sakala, et al., 2019). The model made use of a Gradient Descent Optimizer with only one hidden layer. It also made use of MSE and R as model

key performance indicators. Over 99% training accuracy was achieved, with an MSE of 0.054 (Sakala, et al., 2019). It can be deduced from these figures that the predicting capabilities of the model designed are not to be doubted.

The study presents some important breakthroughs in a form of a useful tool. However, its shortcomings are that the models can only be applied in the area where the geological structure is coal based. The chemical nature of the Witwatersrand AMD may not be compatible with this model.

2.7.4 The Use of Artificial Neural Networks to Predict the Physicochemical Characteristics of Water Quality in Three District Municipalities, Eastern Cape Province, South Africa

The study by Setshedi, et al., (2021) used data collected from three district municipalities and ANN technique to develop the best model fits to predict water quality parameters by employing multilayer perceptron (MLP) neural network. Two input combination models were tested, with both Combination 1 and Combination 2 comprising of temperature, chloride, sulphate, and phosphate as input, and pH, electrical conductivity, turbidity, and dissolved solids as outputs. The difference between the two combinations was the number of hidden neurons. The two models were trained, verified, and tested for their predictive capability, and their physicochemical prediction accuracy was compared by using each model's sample data points with the predicted figures.

The proposed networks were trained using MLP with 3 to 10 hidden neurons, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) training algorithm, and feedforward backpropagation to determine a gradient needed in the computation of the weights for the network (Setshedi, et al., 2021). The data split ratio was 80% for training, 10% for validation, and 10% for testing. Combination 1 model showed a better training of the data sets and water quality predictive ability, retaining an MSE of 39.06589 and an R value 0.994678 compared to the 0.99676 R value and 39.03087 MSE value for the second combination in question (Setshedi, et al., 2021). Presented on Table 2-6 below are some comparisons between the two models. It is of no that that the two models developed are both suitable to predict physicochemical compositions from the three municipalities. This will however not be as useful in mine wastewater due to the differences in the chemical nature of the water.

Table 2-6: Comparison of Combination 1 and 2 model architectures and performance parameters (Setshedi, et al., 2021)

Network Name	R	MSE	Training Algorithm	Hidden Activation	Output Activation
Combination 1	0.994677	39.03087	BFGS	Logistic	Logistic
Combination 2	0.996761	39.06589	BFGS	Tanh	Exponential

2.7.5 Application of artificial neural network (ANN) for the prediction of water treatment plant influent characteristics

The paper focused on applying an ANN approach with a feed-forward back-propagation non-linear neural network to predict water quality for Sanandaj WTP for use prior to treatment. Wastewater characteristic data gathered over a 2-year period were used to building the prediction model. The prediction models developed the study were for Alkalinity, pH, CO_2, temperature, total hardness, turbidity, TDS, and EC. Each model had the fresh plant feed value for a tested parameter as and input into the model, and the corresponding effluent stream value as and output. The choice was considered useful because the model would provide an effective analysing and diagnosing tool to understand and simulate the non-linear behaviour of the influent water characteristics. It will also come in handy to WTP operators and decision makers (Solaimany-Aminabad, et al., 2014).

The generic model network from which all the other models were derived was trained using the Levenberg-Marquardt algorithm. The network had non-linear sigmoid transfer function for the hidden layer and a linear transfer function for the output layer neurons, like the setup employed in this research work. The number of hidden layers tested were 1 to 20, but 1 was proven optimal (Solaimany-Aminabad, et al., 2014). The results of the study displayed a good correlation coefficient between the measured and predicted output variables, reaching up to 0.93. Presented on Table 2-7 below is a summary of the models' performance matrices.

The models developed in this study are only useful in assessing the efficiency of the treatment plant. They do not predict unknown concentrations in the fresh feed into the wastewater treatment plant.

Table 2-7: Performance matrix for each parameter tested (Solaimany-Aminabad, et al., 2014)

Parameter	Training phase		Validation phase		All phases		Testing phase	
	RMSE	R	RMSE	R	RMSE	R	RMSE	R
Cl	0.68	0.59	0.89	0.37	0.74	0.53	0.87	0.42
EC	12.81	0.55	12.41	0.55	12.69	0.53	12.43	0.49
TDS	6.70	0.72	6.96	0.61	7.20	0.67	9.34	0.55
Tur	1.76	0.72	1.97	0.69	1.91	0.71	2.41	0.69
TH	3.61	0.86	4.77	0.71	4.07	0.83	5.19	0.77
T	1.20	0.86	2.19	0.71	1.53	0.83	2.01	0.77
CO_2	0.72	0.76	0.77	0.71	0.76	0.72	0.92	0.64
Ca	2.27	0.71	2.03	0.72	2.18	0.71	1.87	0.74
pH	0.14	0.76	0.14	0.79	0.14	0.74	0.17	0.62
Alk	4.79	0.88	4.75	0.86	5.33	0.86	7.74	0.72

2.7.6 Other ANN Applications and Corresponding Performance Matrices

ANN application is compatible with various industries. The tool has a considerable impact in the environmental sustainability space. It is widely used as a tool that aids to waste management process and has shown acceptable predicting abilities in various cases. Presented on Table 2-8 below is a summary of some of ANN applications their corresponding performance matrices per case. While these studies have all developed models that have commendable predicting capabilities, none of them is focused specifically on AMD water predictions. There is however a lot to benchmark them with the models from the current study as the network architecture share a lot in common.

Table 2-8: ANN applications and performance matrices

Case study	**Application**	**Performance Matrices**	**Reference**
Japan	Predicting the performance of CO2 enhanced oil recovery and storage in residual oil zones	R = 0.989949; MSE = 0.02	(Thanh, et al., 2020)
India	Forecasting groundwater level using artificial neural networks	R= 0.964365; MSE = 20.25	(Sreekanth, 2009)
India	Application of ANN for Water Quality Index	R = 0.89	(Gupta, et al., 2019)
Istanbul, Turkey	Leachate flow control in landfill sites	R = 0.916515; MSE = 0.00168	(Karaca & Özkaya, 2006)
Logan city, Australia	Waste generation forecast	R = 0.678233; MSE = 84419.30	(Abbasi & El-Hanandeh, 2016)
Mashhad, Iran	Weekly waste generation forecast	R = 0.894427	(Noori, et al., 2010)
Faridabad city, India	Waste generation forecast	R = 0.915969; MSE= 0.0000037	(Singh & Satija, 2018)
China	Prediction of heating value of waste	R = 0.932899	(Ozveren, 2016)
Alborz Industrial City, Iran	Identification of the effect of MSW compost and phytoremediation of the contaminated soil	R = 0.9989949	(Roohi, et al., 2020)

2.8 Neural Networks on the Prediction of AMD Properties

ANNs are known for their robustness in their predictability. For this research work, data that was collected on actual AMD-containing stream every week for over 4 years will be utilized to model an algorithm which can be used to predict various properties of AMD water, from other available or pre-defined properties. This model will come in handy in some of the on-going government initiatives aimed at remediating AMD related challenges, as well as other scientific studies.

CHAPTER 3. METHODS AND MATERIAL

3.1 Data Acquisition

The modelling work for this research study was conducted as outlined in the sub-sections below. Experimental data applied in this research was collected by the Natural Recourses and Environment Resource Group from Council for Scientific and Industrial Research (CSIR) and used for model fitting and testing purposes. Samples were sampled at one spot where untreated decanting water flowed into Krugersdorp Game Reserve (KGR) in Gauteng South Africa, once a week for three and half years. Sample pH, EC as well as the concentrations of Fe^{2+}, Na^{2+}, TDS, Cl^- and SO_4^{2-} were determined. The decanting water stream flowrates were also taken each time a sample was collected. This data characterizes the stream flow and gives an indication of the level of contamination for various purposes.

3.2 Building of a Suitable Analytical Tool

One of the most significant use for such data would be to set up AMD water treatment plants with a better understanding of the process inputs irrespective of the change of seasons. This would supplement a significant portion of the design phase. It could also be used to determine the amounts of reagents needed to neutralize AMD water or precipitate out dissolved solids in AMD water before releasing it into the environment. An automated and faster predictive approach for generating such data would however be a good idea instead of collecting and analyzing manual samples for future works as well as day-to-day use of such data. Due to the size of the data collected and the unclear nature of the correlation which may exist amongst the measured properties, ANN models were identified as some of the best tools to unpack and define existing correlations. ANN provides an effective analyzing and diagnostic tool to understand and simulate the non-linear behavior of the elements of stored data through learning the recalling patterns (Mavani, 2014).

3.2.1 Choosing the Most Efficient Artificial Neural Network

This section outlines the analogy followed when choosing ANN parameters to consider over others when building models from which the final model was chosen. The rationale behind each selection or omission is provided.

3.2.1.1 Choosing Model Inputs and Outputs

As discussed in the Literature Review section ANN have both inputs and outputs. The choice of inputs and outputs is a very important in making sure that the predictability of the model is satisfactory and that it serves the purpose that it is intended for. In this research the ANN model input choice is based on a combination of factors, namely the ease of measurement and the known relationship with the other water qualities. Presented in Table 3-1 are some of the combinations explored when deciding which inputs combination were more suitable under the given setup and with the available data. Each measured variable was assessed for its suitability for use as an input or an output.

Input combination 3, comprising of pH, EC and Cl^- as given in Table 3-1 was chosen for model building. Water pH is an easy-to-measure variable and can be used to idealize neutralization requirements. SO_4^{2-} as well as Cl^- concentration also have strong link to solution pH. Electrical conductivity of a solution is equally a measure of ionic substances. EC together with the amount of Cl^- can be used to give an idea of how much dissolved solids there is in a solution, hence also chosen alongside pH model input to quantify TDS and SO_4^{2-}as model outputs as defined by combination 1 in Table 3-2. It should be noted that these are not the only possible combinations but combinations that would make practical sense to look at.

3.2.1.2 Choosing a Program to Model with

MATLAB was chosen for ANN model building and simulation purposes. It has a built-in NN function (nnstart), a Graphical User Interface (GUI) deep learning toolbox fully developed for neural networks. MATLAB provides a platform to build from scratch or edit the default models using the program's well-defined tools. The biggest advantage of using MATLAB for NN models

is that it allows for a fast and efficient implementation of well-specified models (Beale, et al., 2013). There are several other programs which could also be used but the MATLAB platform advantages were enough for the program to be considered and chosen.

Table 3-1: Possible model input combinations

	AMD Water Property						
COMBINATION 1	pH	EC					
COMBINATION 2	pH	EC	$[Na^+]$	$[Fe^{2+}]$	$[Cl^-]$		
COMBINATION 3	pH	EC	$[Cl^-]$				
COMBINATION 4	pH	EC	$[Fe^{2+}]$	$[Cl^-]$			
COMBINATION 5	pH	EC	$[Na^+]$				
COMBINATION 6	pH	EC	$[Na^+]$	$[Cl^-]$			
COMBINATION 7	pH	EC	$[Fe^{2+}]$				
COMBINATION 8	pH	EC	$[Na^+]$	$[Fe^{2+}]$			
COMBINATION 9	pH	EC	$[SO_4^{2-}]$	$[Na^+]$	$[Fe^{2+}]$	$[Cl^-]$	
COMBINATION 10	EC	$[SO_4^{2-}]$	$[Na^+]$	$[Fe^{2+}]$	$[Cl^-]$		
COMBINATION 11	EC	$[Na^+]$	$[Fe^{2+}]$	$[Cl^-]$			
COMBINATION 12	EC	$[SO_4^{2-}]$					
COMBINATION 13	pH	EC	$[SO_4^{2-}]$	$[SO_4^{2-}]$/TDS	$[Na^+]$	$[Fe^{2+}]$	$[Cl^-]$
COMBINATION 14	pH	$[SO_4^{2-}]$					

Table 3-2: Possible model output combinations

	AMD Water Property								
COMBINATION 1			TDS	$[SO_4^{2-}]$					
COMBINATION 2			TDS	$[SO_4^{2-}]$		$[Na^+]$	$[Fe^{2+}]$	$[Cl^-]$	
COMBINATION 3			TDS	$[SO_4^{2-}]$			$[Fe^{2+}]$		
COMBINATION 4			TDS	$[SO_4^{2-}]$				$[Cl^-]$	$[SO_4^{2-}]/[Cl^-]$
COMBINATION 5	pH		TDS	$[SO_4^{2-}]$					
COMBINATION 6	pH			$[SO_4^{2-}]$			$[Fe^{2+}]$		
COMBINATION 7	pH	EC	TDS	$[SO_4^{2-}]$	$[SO_4^{2-}]$/TDS	$[Na^+]$	$[Fe^{2+}]$	$[Cl^-]$	$[SO_4^{2-}]/[Cl^-]$

3.2.1.3 Choosing Input and Output Pre/Post-Processing Functions

3.2.1.3.1 Transfer functions

Since the MATLAB built-in NN model is used for this research, *mapminmax* are the transforming functions for both inputs and outputs. It transforms input data so that all values fall into the interval [−1, 1], a feature associated with a hyperbolic tangent transfer function (Beale, et al., 2013). By default, *removeconstantrows* is also used for removing the rows of the input vector that correspond to input elements (Beale, et al., 2013). It is also said that using linear functions in multilayer

network is pointless because the biological correspondence is nonlinear (Mavani, 2014), hence linear functions were not chosen. Since the data used for this research is non-linear and multiple layers are explored, it is advisable to pick one between log sigmoid and hyperbolic tangent which are the most used functions for such a setup. For these reasons, the default transfer functions were used as they are for building networks in this research since they correspond to a *tansig* function that is among the most suitable function for the available data.

3.2.1.3.2 Model Supervision

The ANN models tested in this research all rely on supervised learning. The training stops when the performance grader given by LMS algorithm reaches the lowest value. No human intervention necessary in the termination of the training.

3.2.1.4 Choosing the Most suitable Training Functions

There are several training functions an ANN can be trained with. For this research however only three most commonly used were considered for model building purpose. The best of the three is chosen based on the performance outcomes. These functions are (Beale, et al., 2013):

- The *trainlm* network training function that updates weight and bias values according to Levenberg-Marquardt optimization
- The *traingd* is network training function that updates weight and bias values according to gradient descent
- The *trainrp* network training function that updates weight and bias values according to the resilient backpropagation algorithm

3.2.1.5 Choosing a Performance Function

In order to tell which model performs better compared others when applied on the same set of data, there should be some sort of indication. As discussed under literature review, multiple performance indicators are recommended. For this research, two performance statistic indicators namely correlation coefficient (R) and MSE were used. Coefficient correlation was used to determine each

model's generalization ability and to benchmark the models with models developed by other researchers. The MSE were used to compare models developed in this study only. This figure is not appropriate for use to compare with model networks from other research work as it is depended on the developer's choice of data scales and normalization approach.

3.2.2 Data Preparation

Before model training can take place, raw data needs to be processed to ensure that it is in the correct format to be read by MATLAB and to get rid of biased measurements if there are any. The following procedure was used to prepare raw data acquired:

- All samples which had at least one measurement missing among the five properties considered were removed. i.e., pH, EC, $[Cl^-]$, $[SO_4^{2-}]$ and TDS
- Box and whisker pot for the three inputs were produced, and samples showing outliers were discarded.
- The fraction of samples removed were determined in order to tell whether the remaining set is still representative or not.

3.2.3 Model Building and Running

3.2.3.1 Data Importing and ANN Code Writing

After achieving the data preparation, neural network models were created in MATLAB software. On MATLAB Toolbox, open Network/Data Manager window which enable data importing, ANN creation, use, and exporting. Different ANN models with different network properties were created as follows:

Step 1:

A dummy ANN model code with the following properties was created:

- Network inputs: pH, EC and $[Cl^-]$
- Network outputs: TDS and $[SO_4^{2-}]$
- Network type: Feed-Forward Back-Propagation
- Input and Output Processing Functions: REMOVECONSTANTROWS, MAPMINMAX

- Data Division function: DIVIDERAND
- Training function: TRAINLM
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 3
- Training/Validation/Testing data %split: 60:20:20

Step 2:

A preliminary test run using the dummy model created was performed to determine the number of hidden neurons, training algorithm, and data split ratio to keep constant when comparing various model properties. The following procedure was followed:

a) The dummy model as specified above was run while varying the number of hidden neurons at 3, 4, 5, 10, 20, and 50. Records of overall MSE and R value were kept.
b) Step a) above was repeated using the training function trainrp.
c) Step a) above was repeated using the training function traingd.
d) Steps a) - c) above were repeated using 70:15:15 data split ratio.
e) Steps a) - c) above were repeated using 80:10:10 data split ratio.
f) The combination with both lowest MSE and highest R values was identified (These were kept the same where some properties had to stay constant during model performance comparisons in the succeeding sections)

Step 3:

The purpose of this step was to determine various models from which the best performing ANN model were be selected. Different models were generated through systematically differing model properties then assigning model numbers as outlined on Table 3-3 below. Where there was a need to keep certain properties constant, the best performing combination identified as alluded under f) of step 2 above was used for parameters selection. This process has also been presented schematically on Figure 3-1 below.

Table 3-3: Guideline for coming up with ANN models that were developed

Model	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
Hidden Nodes	5	10	20	50	20	20	20	20	20
Algorithm	Trainlm	Trainlm	Trainlm	Trainlm	Traingd	Trainrp	Trainlm	Trainlm	Trainlm
Data Split	70:15:15	70:15:15	70:15:15	70:15:15	70:15:15	70:15:15	80:10:10	90:5:5	60:20:20

Step 4:

This step was meant for evaluating the performance evaluation for each of the 9 models developed. This was achieved in the following manner:

a) Model 1 was run, and the predicted training, validation, and testing output together with their corresponding measured outputs were exported.
b) Measured and predicted outputs were plotted linearly on the same set of axes to evaluate the model's ability to forecast the outputs. Training, Validation and Testing plots were drawn separately.
c) Model 1 was run ten time and determine the averages of Overall MSE, Training MSE, Validation MSE, Testing MSE, Overall R, Training R, Validation R and Testing R.
d) Steps a) – c) above were repeated for Models 2 to 9.
e) The performance indicators (Overall R and Overall MSE) of the 9 models developed were compared.
f) The best performing model was identified, then it was determined if its performance was satisfactory based on what the statistical evaluation of its result mean in terms of AMD water property prediction.
g) The best performing model was used to validate the claim that MSE varies based on the scale of data used, by converting the 'milli' on the units of the parameters involved into a numerical value. i.e. ($x10^3$).

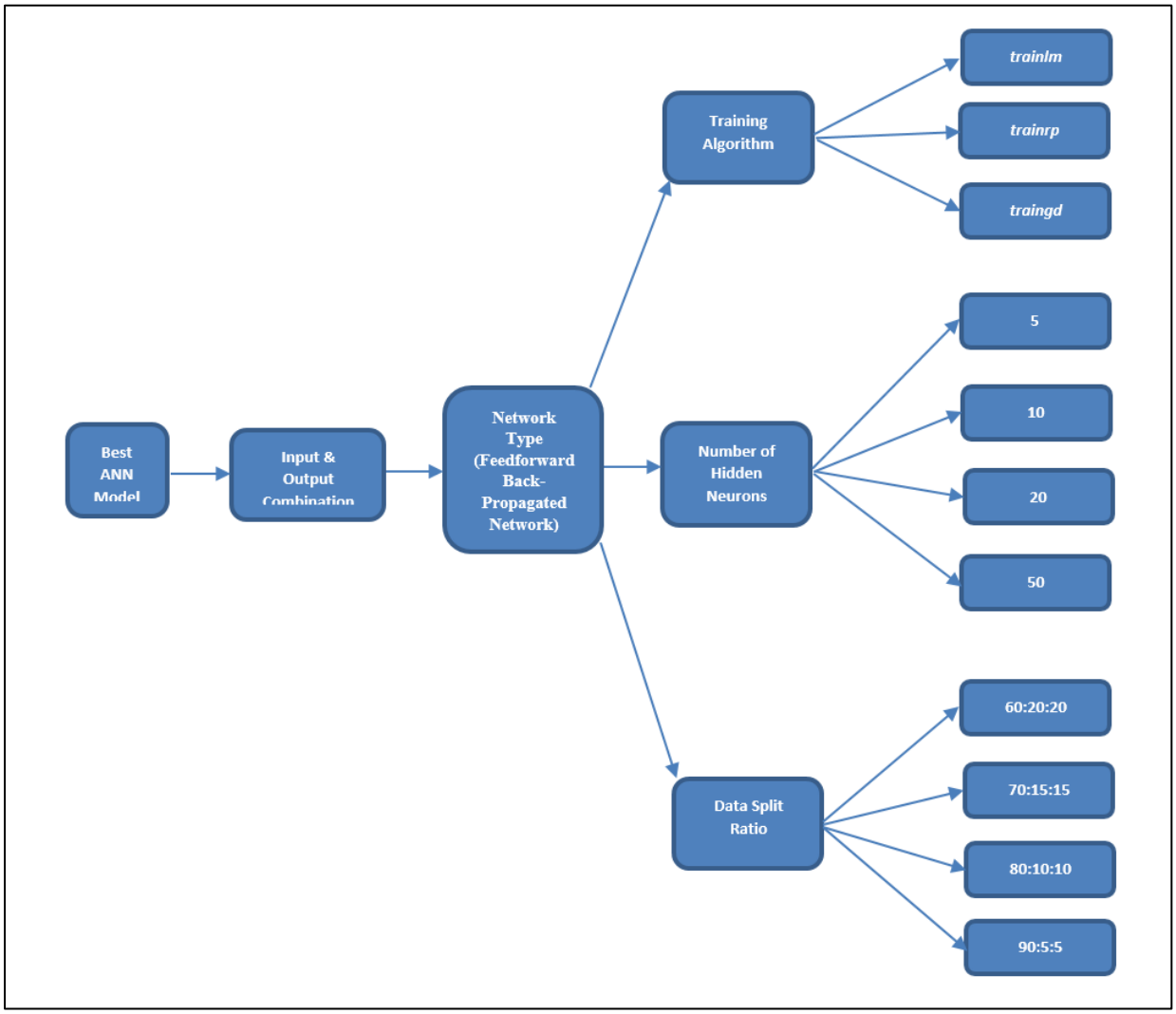

Figure 3-1: Building ANN model

CHAPTER 4. RESULTS AND DISCUSSION

4.1 Data Preparation Outcomes

Data collection for environmental settings such as the one considered in this research normally take longer periods. This is to allow for a well representative data as they cover overlapping or more years of varying seasonal weather patterns. They also enable the capturing of data from season to season which may differ significantly due to a number of factors. Factor such as rainfall and evaporation rates which are influenced by heat experienced are important determinants of stream flowrates as well as the concentrations of various stream water contaminants (Merz, 2013). The data used for was collected weekly at the same spot over a period of three and half years. This period can be deemed long enough to cover different possible conditions and variations, with a reduced probability of producing outlier readings.

The data preparation step was conducted as outlined under the Methodology section of this report. The box and whisker plots given as Figure 4-1, Figure 4-2 and Figure 4-3 below show the statistical analysis of the raw data. Weekly samples that did not have corresponding measurements for components considered in the building of the model were discarded. These only amounted to less than 0.5% of the total data set. This data exclusion rate is too small to render the entire data set unrepresentative, particularly for application in ML (Deshpande, 2021). Sample pH measures in this context are a reflection of the amount of acid bearing contaminants in the AMD water and also as one of the main defining features of the samples. For these reasons, pH measurements were also considered in determining which other samples to disregard from the original data set. Looking at Figure 4-1 below, it can be seen that the measured sample pH is uniformly distributed across the pH scale with no outlier. Such an observation is a result of variation in AMD water concentration due to different steam flowrates that differ from season to season. The absence of outliers led to a decision to retain all remaining samples despite having outliers on Electrical Conductivity and Chlorine concentrations as seen on Figure 4-2 and Figure 4-3 respectively. These outliers may have been due special instances including excessive weather such as heavy rains or extreme heat, and even increased timeous stream polluting activities on the weeks or days the samples were collected.

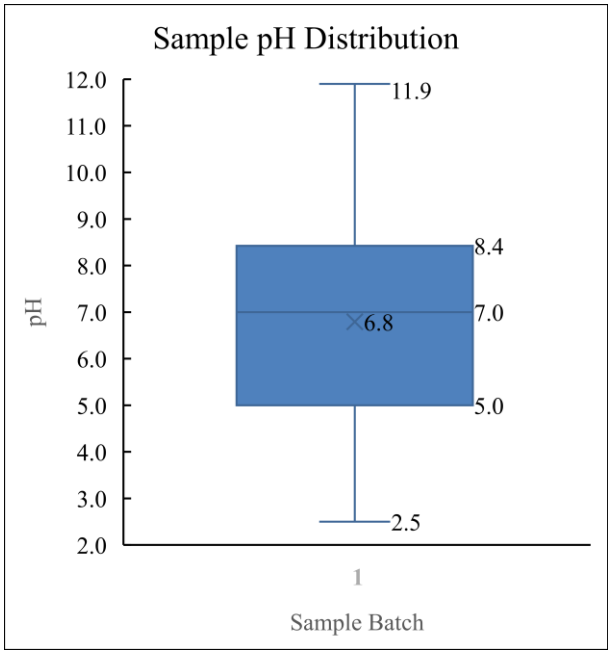

Figure 4-1: Raw data pH distribution

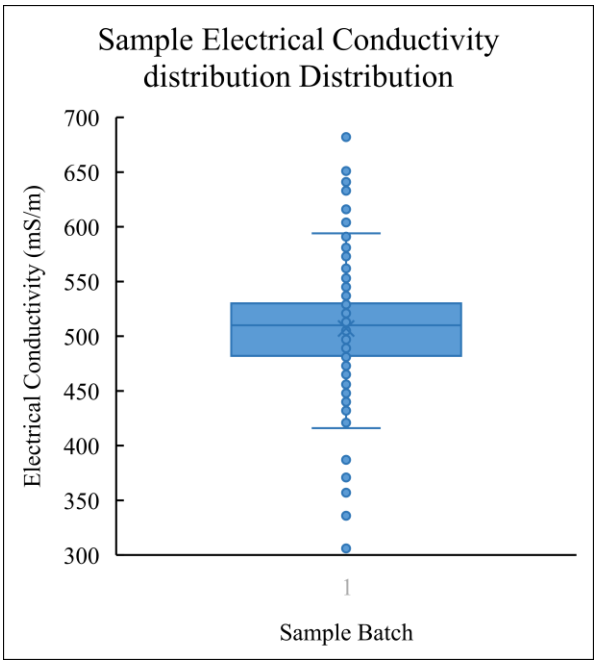

Figure 4-2: Raw data Electrical Conductivity distribution

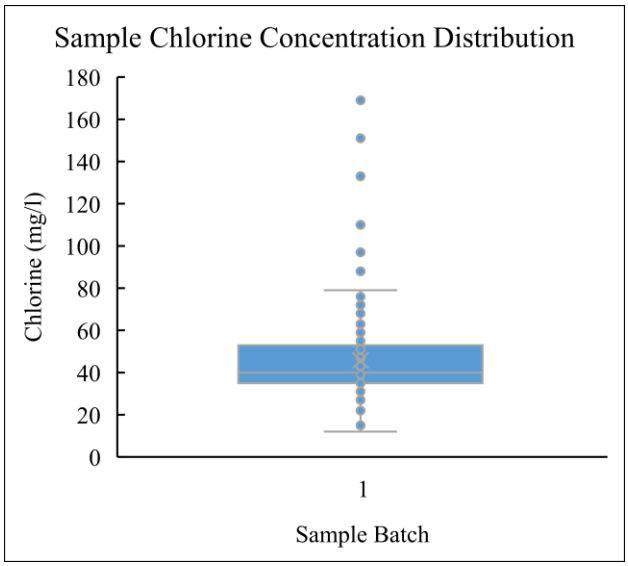

Figure 4-3: Raw data Chlorine concentration distribution

4.2 Preliminary Run for Split and Training Algorithm to Keep Constant

It was shown with the preliminary test runs that a combination of settings that comprised of 20 hidden neurons, 75:15:15 data split ratio and trainlm training algorithm showed the best

performance as shown on Table 4-1 below. This step was conducted solely for getting and idea of what to adopt as a generic network model on which further refinements were done. An overall R value of 0.985 and MSE value of 1 860 000 were achieved with this combination. This R value was the highest of these achieved amongst the 9 models and signifies a very strong correlation between the input and output data supplied to the model network (Schober, et al., 2018). The corresponding MSE value is the lowest among those obtained through the 9 models. The implication of this is that the cumulative sum of differences between the actual and the predicted values was the lowest for the chosen model. With this observation, it was decided that:

- The *trainlm* algorithm would be used as a default training algorithm when simulating to compare model properties other than the training algorithm.
- Twenty hidden neurons would be used as a default hidden neuron when simulating to compare model properties other than the number of hidden neurons.
- A 70:15:15 data split ratio for training, validation and testing would be used as a default split ratio when simulating to compare model properties other than the data split ratio.

Table 4-1: Preliminary model run results

Number of Neurons	SPLIT	Overall Error			Overall R		
		trainlm mse	trainrp mse	traingd mse	trainlm R	trainrp R	traingd R
3	**60:20:20**	2.32E+05	2.43E+05	5.17E+05	0.9811	0.9796	0.9685
4	**60:20:20**	2.04E+05	2.75E+05	6.43E+05	0.9828	0.9767	0.9674
5	**60:20:20**	4.23E+05	3.12E+05	7.02E+05	0.9655	0.9737	0.9532
10	**60:20:20**	2.00E+05	2.65E+05	8.90E+05	0.9831	0.9775	0.9586
20	**60:20:20**	2.25E+05	2.22E+05	1.28E+06	0.9821	0.9803	0.97
50	**60:20:20**	2.82E+05	9.74E+05	1.45E+06	0.9764	0.9209	0.936
60	**70:15:15**	2.22E+05	2.88E+05	3.56E+05	0.9812	0.9763	0.9754
70	**70:15:15**	2.07E+05	2.53E+05	3.93E+05	0.9825	0.9786	0.9665
5	**70:15:15**	4.34E+05	2.52E+05	3.15E+05	0.9773	0.9786	0.975
10	**70:15:15**	2.25E+05	2.15E+05	1.87E+06	0.981	0.9818	0.9362
20	**70:15:15**	1.86E+05	2.35E+05	1.52E+06	0.985	0.9814	0.9662
50	**70:15:15**	1.45E+06	3.29E+05	2.39E+06	0.8879	0.9722	0.834
3	**80:10:10**	2.15E+05	3.22E+05	9.57E+05	0.9818	0.973	0.9697
4	**80:10:10**	2.08E+05	2.17E+05	5.32E+05	0.9826	0.9817	0.9657
5	**80:10:10**	2.26E+05	2.24E+05	1.83E+06	0.9818	0.981	0.8613
10	**80:10:10**	1.92E+05	3.14E+05	9.55E+05	0.9838	0.9733	0.9514
20	**80:10:10**	5.34E+05	2.70E+05	1.28E+06	0.9555	0.9771	0.9463
50	**80:10:10**	3.82E+05	3.32E+05	3.00E+06	0.9689	0.9724	0.8795
	Min	1.86E+05	2.15E+05	3.15E+05	0.8879	0.9209	0.834
	Max	1.45E+06	9.74E+05	3.00E+06	0.985	0.9818	0.9754

4.3 The Generic Simulation Code

Presented Appendix D is the code generated for simulation purposes. It was used to simulate all the models discussed on this research. Data files, model inputs, model outputs, input globalization, output globalization, input/output preprocessing functions, input/output post-processing functions, data division functions, performance functions, plot functions, as well as the overall structure all remained the same for all models simulated. The only model network properties that were varied to create different models were data split ratios, number of hidden layers and training functions. The model runs without errors and generates outputs well within specified epoch range at reasonable run times.

The nature of the model and the simulation platform used present a built-in function which does the normalization process by itself as discussed under subsection 2.6.6. In the study by Nasr, et al., (2012), raw data had to normalized separately first before it was subjected to the network. This is seen as additional effort by user, hence it was opted to rely on the built-in functions for roughly the same generalization capabilities shown by R values involved, as well as the same error performance. MES performance. In the studies by Sakala, et al., (2017), ANN had to be coupled with LSTM nets in order to to deal with the vanishing gradient problem that can be encountered when training traditional NNs where gradient descent algorithms are used. This is seen as an additional network on top of the ANN produced. It can introduce a wide range of complexities for user, hence not considered, even though a gd algorithm-based network model was also tested in this study.

4.4 Individual Simulated Model Performance Results Evaluation

The presented in the succeeding sub-sections (4.4.1 – 4.4.9) are the performance outcomes of the nine models built. These outcomes were amongst the indicators used to determine the best performing model of the nine tested models, and subsequently the final model that was recommended for predicting AMD properties from the Witwatersrand Western Basin. These subsections merely outline the performance of each described model. A collective result

presentation, critical analysis and results comparison with previous study results were done on subsection 4.6.

4.4.1 Simulated Model 1 Network Properties

- Network inputs: pH, EC and [Cl^-]
- Network outputs: TDS and [SO_4^{2-}]
- Network type: Feed-Forward Back-Propagation
- Data Division function: DIVIDERAND
- Training function: TRAINLM
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 5
- Training/Validation/Testing data %split: 70:15:15

Network Model 1 once off simulation performance summary is presented on the figures below. It can be seen on Figure 4-4 that the simulation ran to completion, with all variables within healthy ranges. The model's coefficient correlation R is within a very high range at 0.9217 as shown on Figure 4-5. The model was trained over a total of 6 iterations as shown on Figure 4-6. Its error distribution is concentrated around the mean error of 0, with a narrow range as shown on the histogram given by Figure 4-7. The overall model multiple simulation performance is shown as Figure 4-8. It can be seen from this figure that the model performance is consistent by less variation on R and MSE, and this renders the model reliable for use in predicting AMD water characteristics as prescribed in this study.

The model's reliability is further demonstrated by how the simulated outputs compare with the measured outputs. The comparisons done on the training and validation data outputs (simulated and measured) demonstrate how the learning process improved as the training progressed from one input value to the next as seen on Figure 4-9 to Figure 4-12. This is an indication that the learning algorithm chosen works well in this particular situation. The comparisons seem to be slightly poor for TDS than it is for [SO_4^{2-}] but also. The comparison done on the testing data

output further cements the claims that this network model has reliable predicting abilities as seen on Figure 4-13 and Figure 4-14.

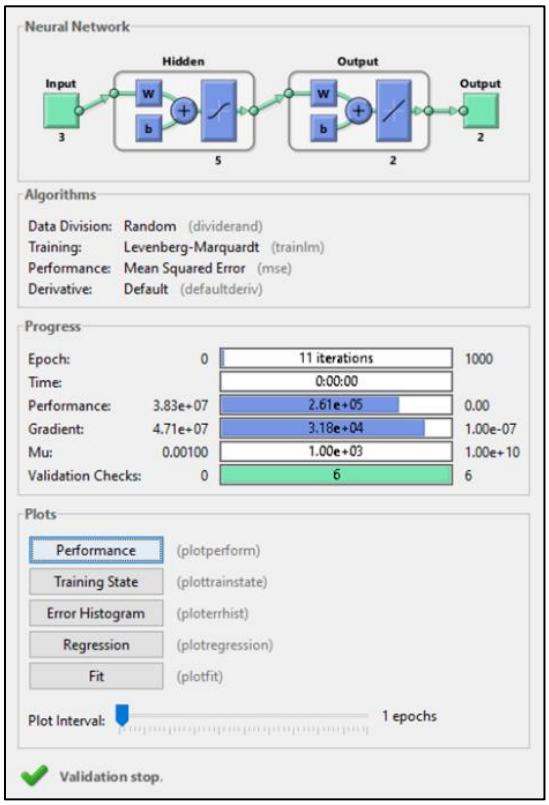

Figure 4-4: Model 1 simulation faceplate during training

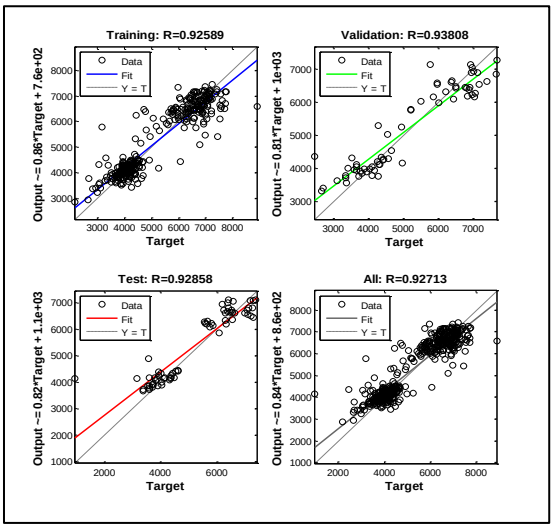

Figure 4-5: Model 1 network regression

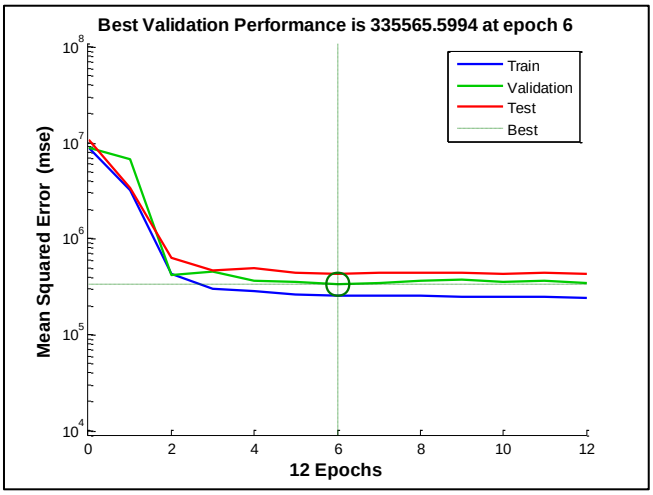

Figure 4-6: Model 1 supervised learning stopping criteria

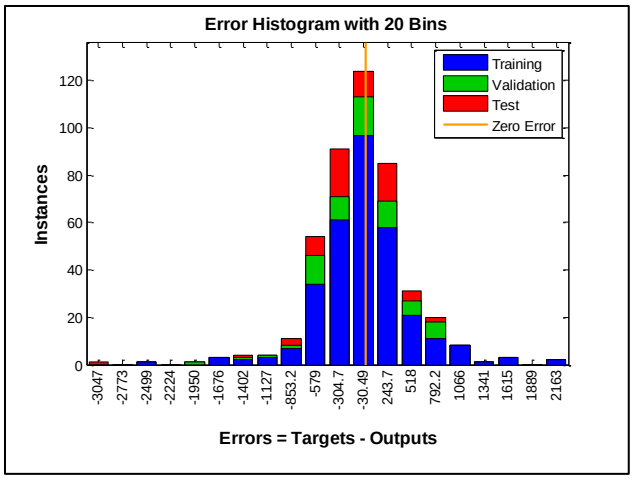

Figure 4-7: Model 1 error histogram

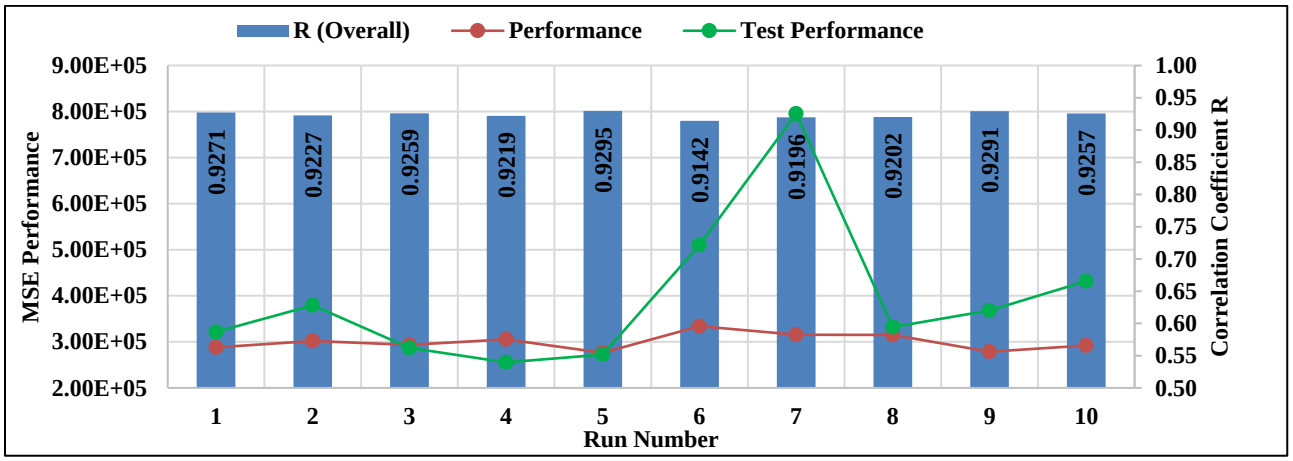

Figure 4-8: Model 1 performance variation and consistency demonstration

4.4.1.1 Model 1 Training Data Simulation

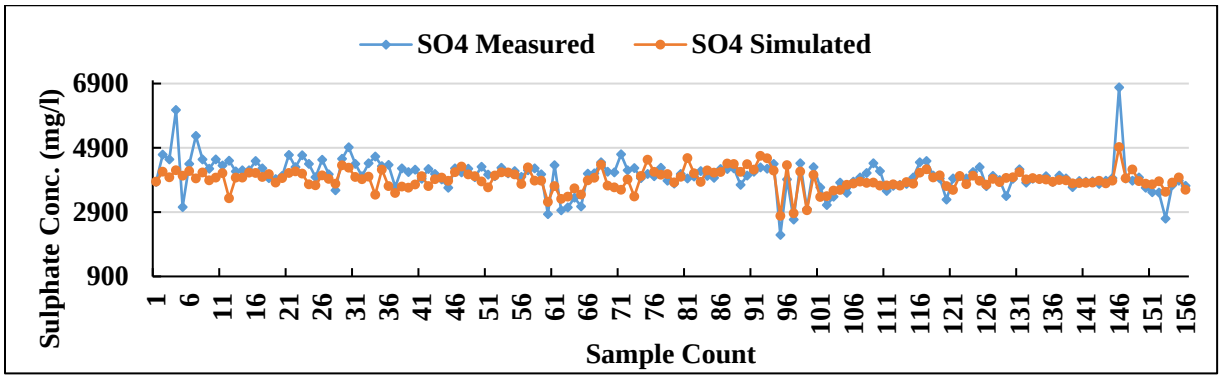

Figure 4-9: Model 1 measured and simulated [SO_4^{2-}] comparison for the training set

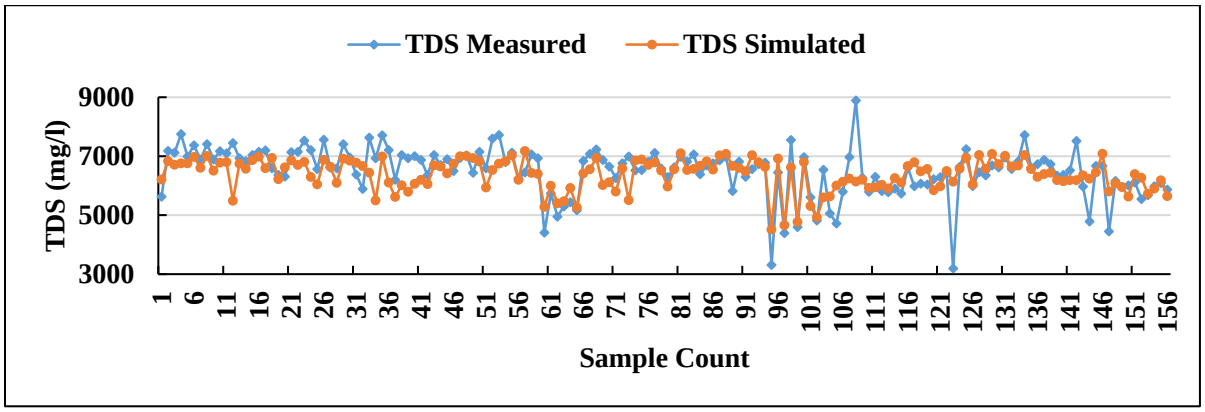

Figure 4-10: Model 1 measured and simulated TDS comparison for the training set

4.4.1.2 Model 1 Validation Data Simulation

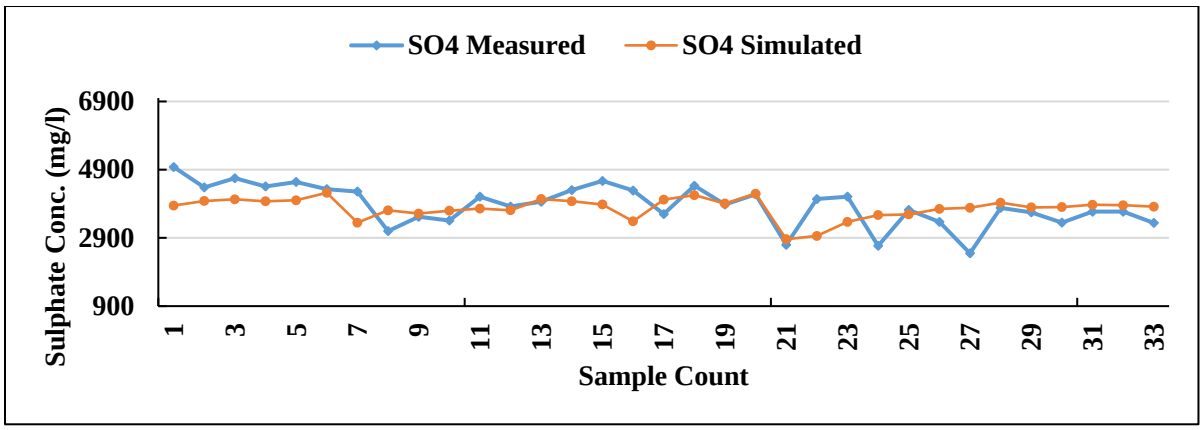

Figure 4-11: Model 1 measured and simulated [SO_4^{2-}] comparison for validation data set

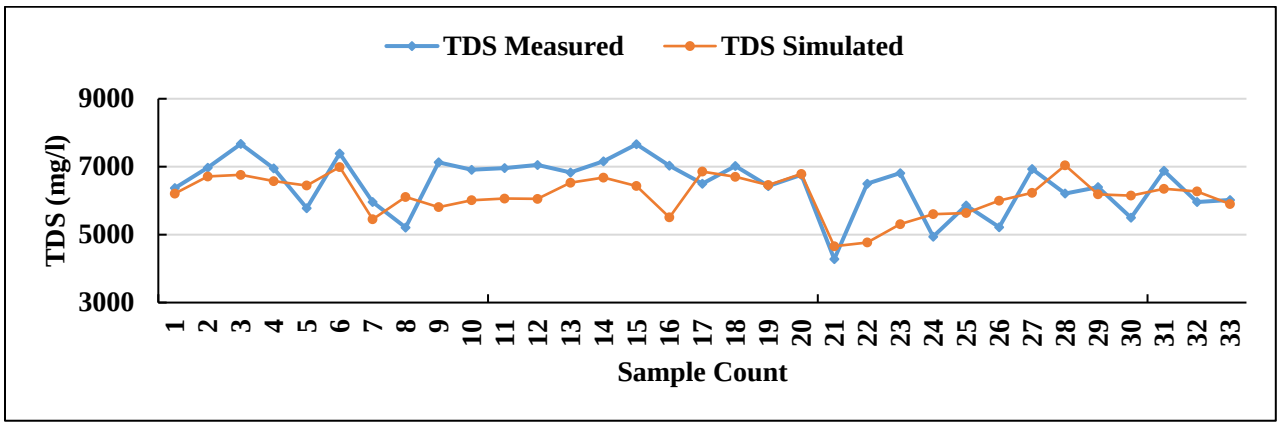

Figure 4-12: Model 1 measured and simulated TDS comparison for validation data set

4.4.1.3 Model 1 Testing Data simulation

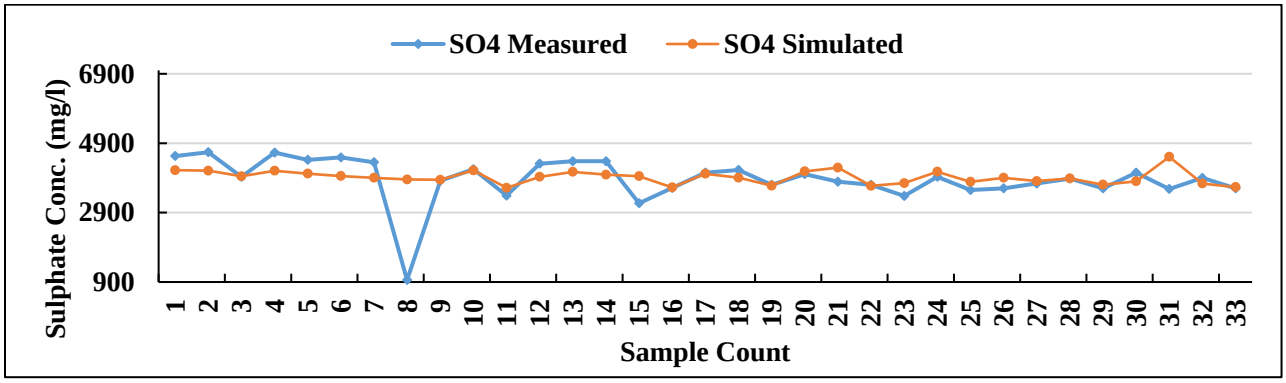

Figure 4-13: Model 1 measured and simulated [SO_4^{2-}] comparison for testing data set

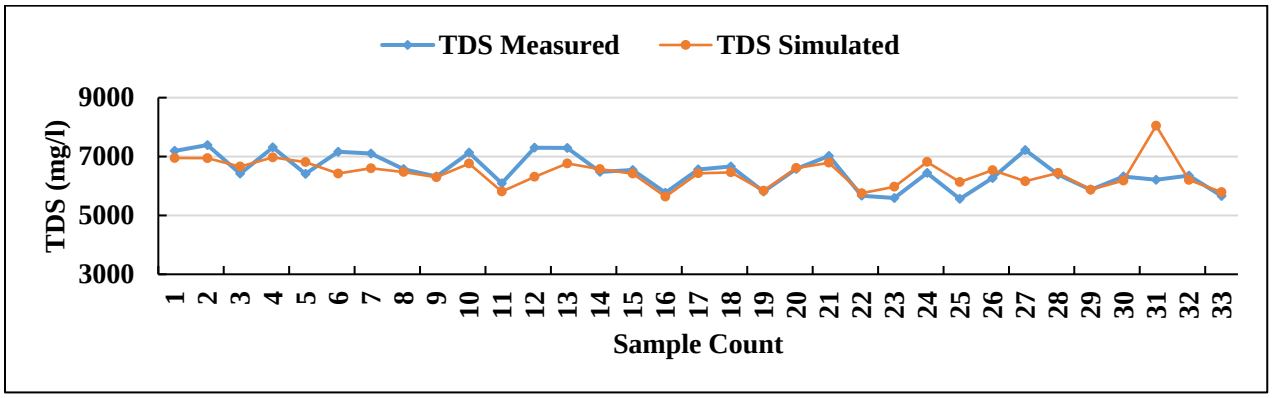

Figure 4-14: Model 1 measured and simulated TDS comparison for testing data set

4.4.2 Simulated Model 2 Network Properties

- Network inputs: pH, EC and [Cl^-]
- Network outputs: TDS and [SO_4^{2-}]
- Network type: Feed-Forward Back-Propagation
- Data Division function: DIVIDERAND
- Training function: TRAINLM
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 10
- Training/Validation/Testing data %split: 70:15:15

Network Model 2 once off simulation performance summary is presented on various figures below. It can be seen on Figure 4-15 that the simulation ran to completion for this model also, with all variables within healthy ranges. The model's coefficient correlation R is also high at 0.9041 as shown on Figure 4-16Figure 4-5. The model was trained over a total of 14 iterations as shown on Figure 4-17. Its error distribution is also centered around the mean error of 0 as shown on the histogram given by Figure 4-18. The overall model multiple simulation performance is shown as Figure 4-19. It can be seen from this figure that the model performance is consistent since the variation on R is takes above the 0.9 mark while the MSE varies just slightly. This renders the model's predicting abilities reliable even though the testing MSE is fluctuates slightly.

The model's reliability is further demonstrated by how the simulated outputs compare with the measured outputs on Figure 4-20 - Figure 4-23. The comparisons seem to be slightly poor for TDS than it is for [SO_4^{2-}] but still reliable. The comparison done on the testing data output further cements the claims that this network model has reliable predicting abilities as seen on Figure 4-24 and Figure 4-25.

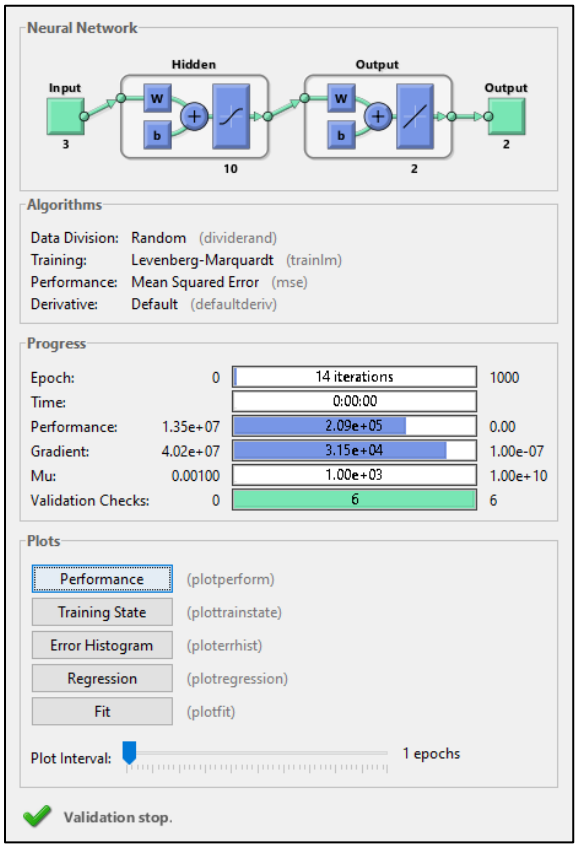

Figure 4-15: Model 2 simulation faceplate during training

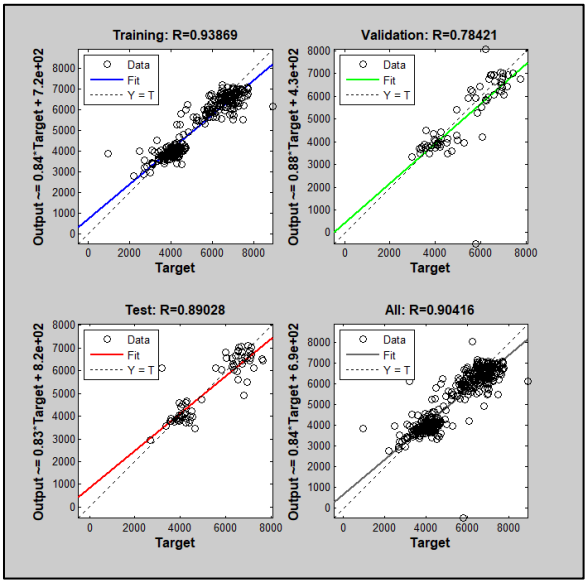

Figure 4-16: Model 2 network regression

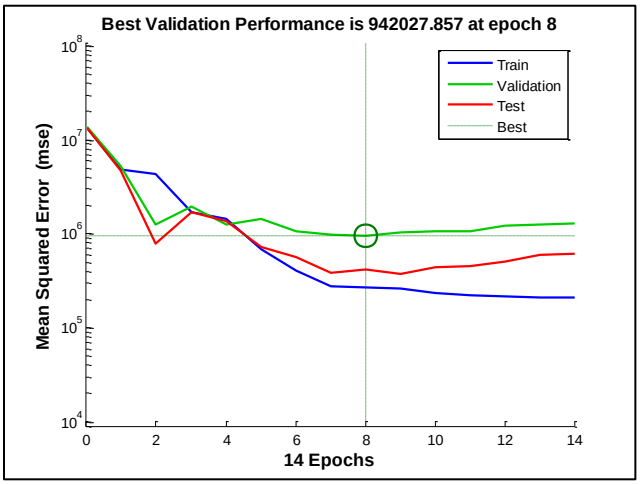

Figure 4-17: Model 2 supervised learning stopping criteria

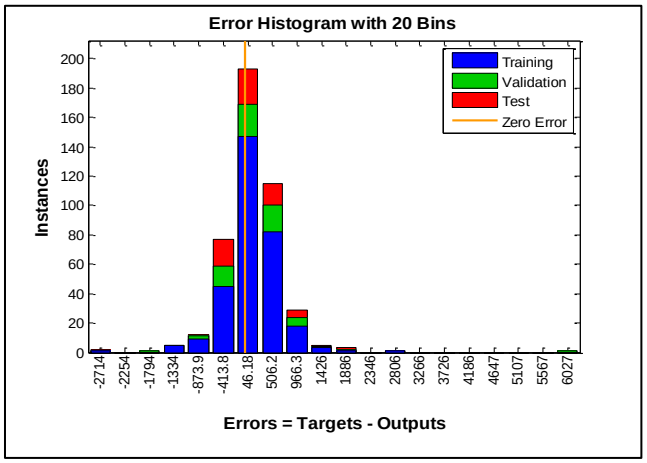

Figure 4-18: Model 2 error histogram

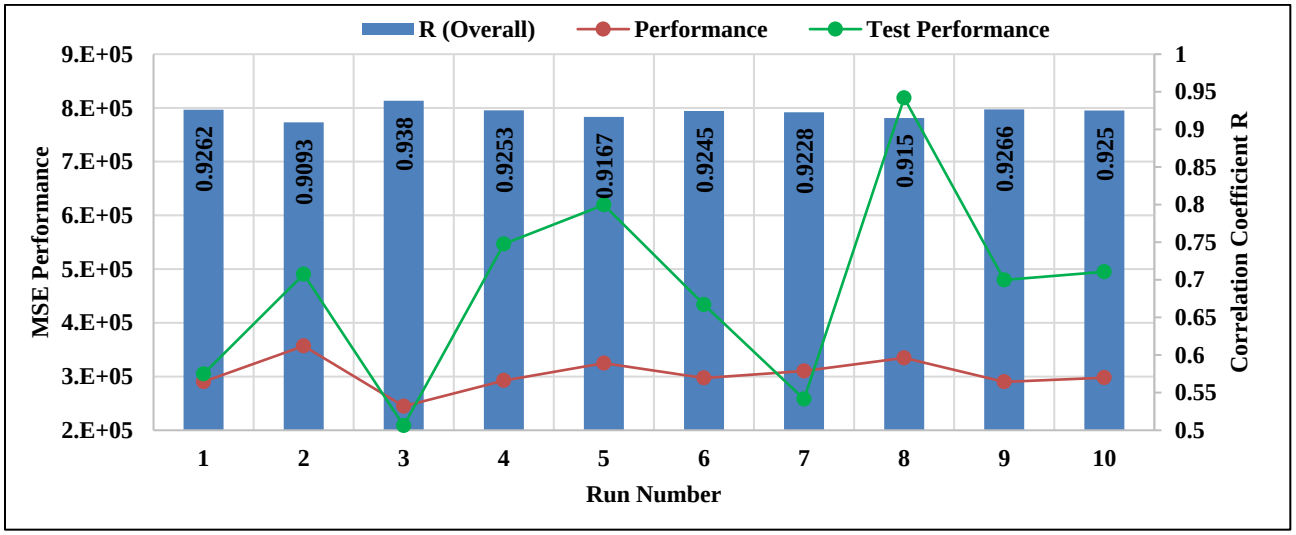

Figure 4-19: Model 2 performance variation and consistency demonstration

4.4.2.1 *Model 2 Training Data Simulation*

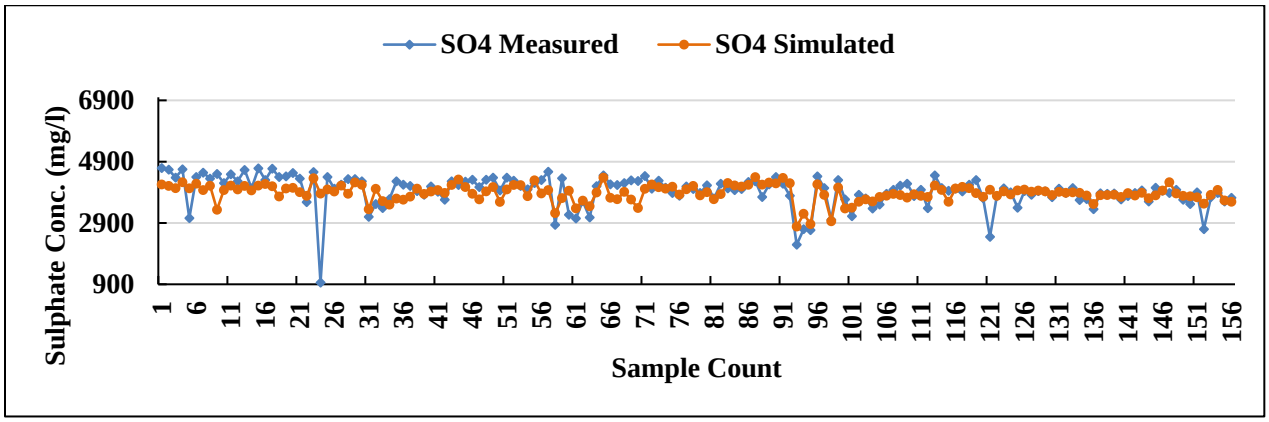

Figure 4-20: Model 2 measured and simulated [SO_4^{2-}] comparison for the training set

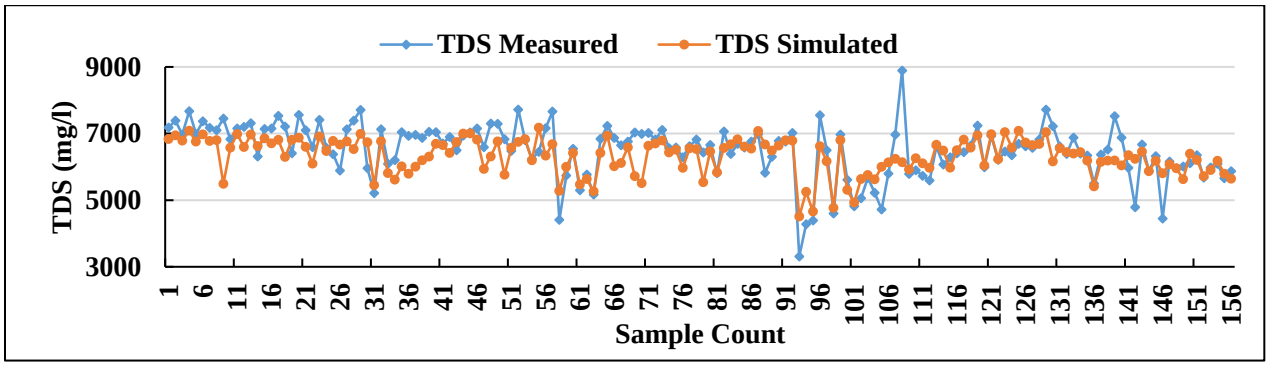

Figure 4-21: Model 2 measured and simulated TDS comparison for the training set

Model 2 Validation Data Simulation

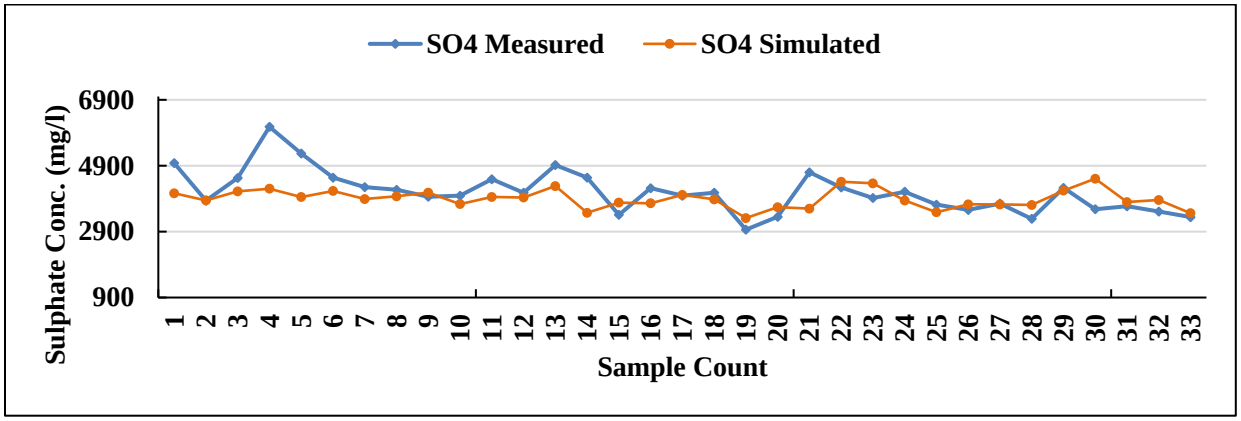

Figure 4-22: Model 2 measured and simulated [SO_4^{2-}] comparison for validation data set

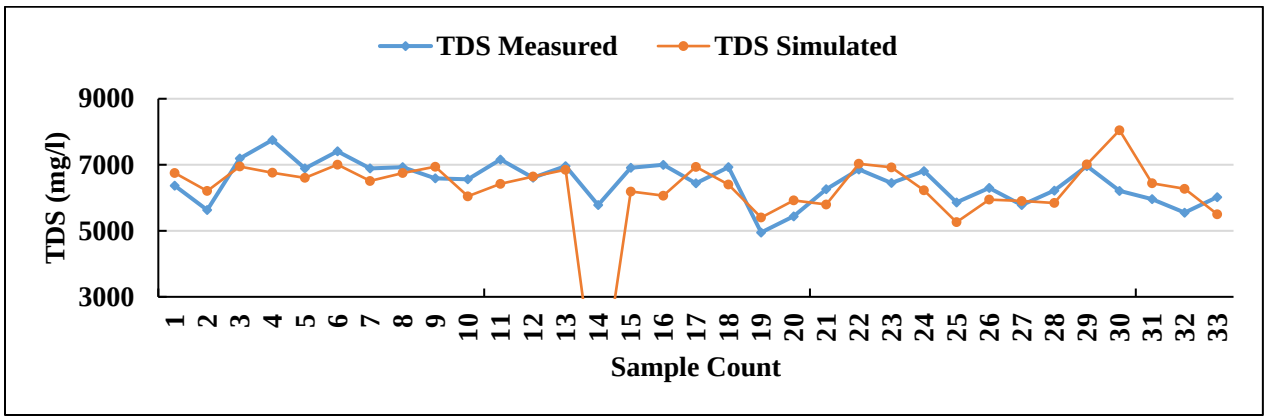

Figure 4-23: Model 2 measured and simulated TDS comparison for validation data set

4.4.2.2 Model 2 Testing Data simulation

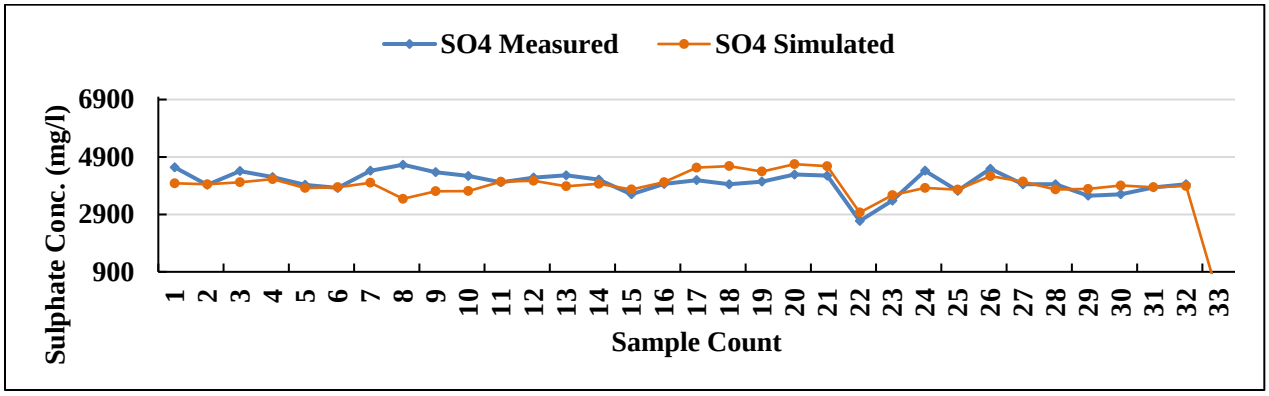

Figure 4-24: Model 2 measured and simulated [SO_4^{2-}] comparison for testing data set

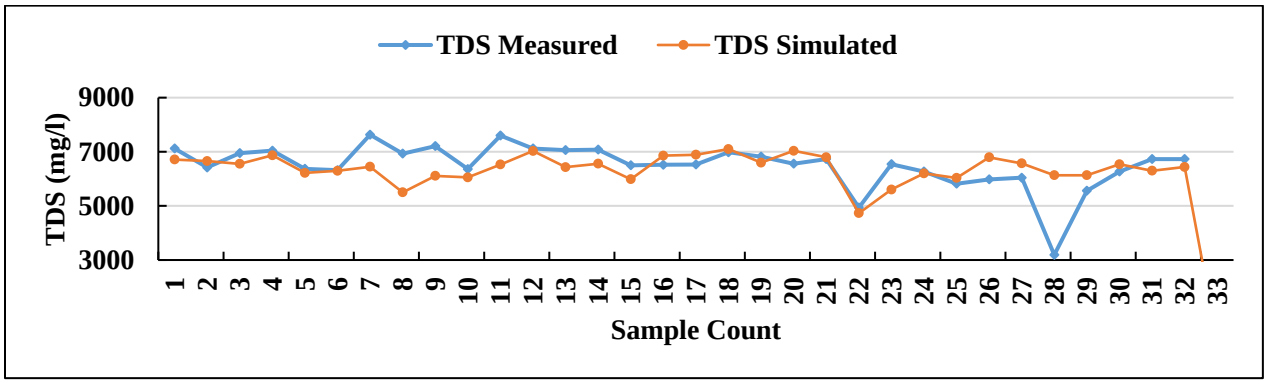

Figure 4-25: Model 2 measured and simulated TDS comparison for testing data set

4.4.3 Simulated Model 3 Network Properties

- Network inputs: pH, EC and [Cl^-]
- Network outputs: TDS and [SO_4^{2-}]
- Network type: Feed-Forward Back-Propagation
- Data Division function: DIVIDERAND
- Training function: TRAINLM
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 20
- Training/Validation/Testing data %split: 70:15:15

Network Model 3 once off simulation performance summary is presented on figures below. It can be seen on Figure 4-26 that the simulation ran to completion for this model also, with all variables within healthy ranges. The model's coefficient correlation R is also high at 0.9432 as shown on Figure 4-27Figure 4-5. The model convergence is shown on Figure 4-28. Its error distribution is also centered around the mean error of 0 as shown on the histogram given by Figure 4-29. The overall model multiple simulation performance is shown as Figure 4-30. It can be seen from this figure that the model's coefficient correlation is constantly above 0.9 whilst the MSE varies just slightly on the training data and quite significantly for the test data set. This renders the model's predicting abilities reliable despite the testing MSE slightly fluctuating.

The manner in which the simulated outputs compare with the measured outputs as shown on Figure 4-31 - Figure 4-34 further demonstrates the acceptability of the model's predicting abilities. The comparisons seem to be slightly poor for TDS than it is for [SO_4^{2-}] but still reliable. There is also a good correlation on the testing data output for this network model as seen on Figure 4-35 and Figure 4-25.

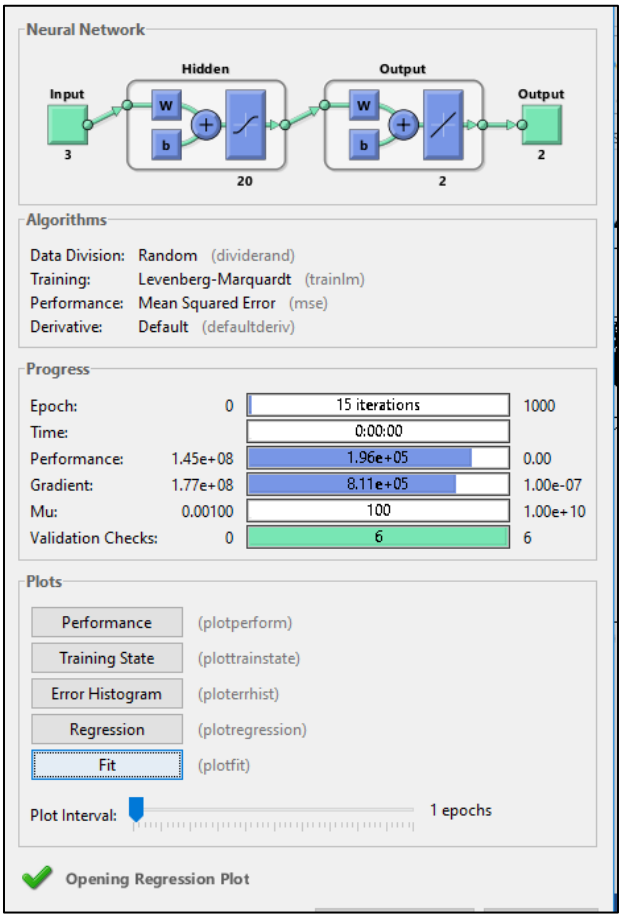

Figure 4-26: Model 3 simulation faceplate during training

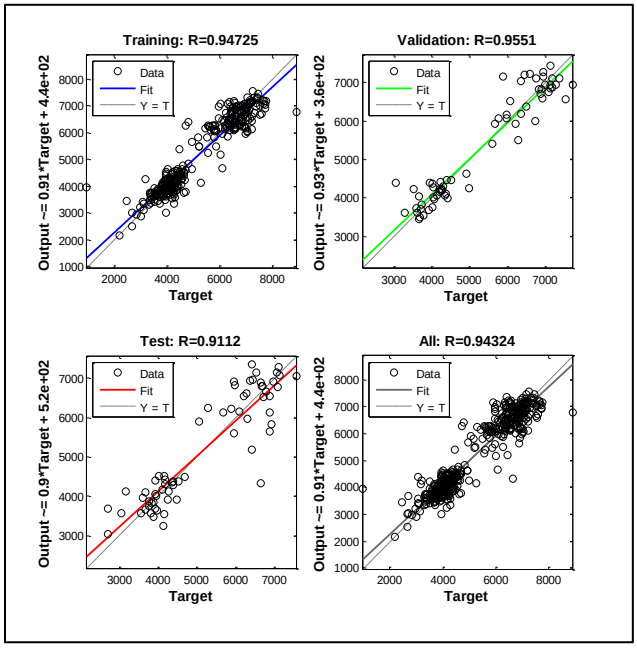

Figure 4-27: Model 3 network regression

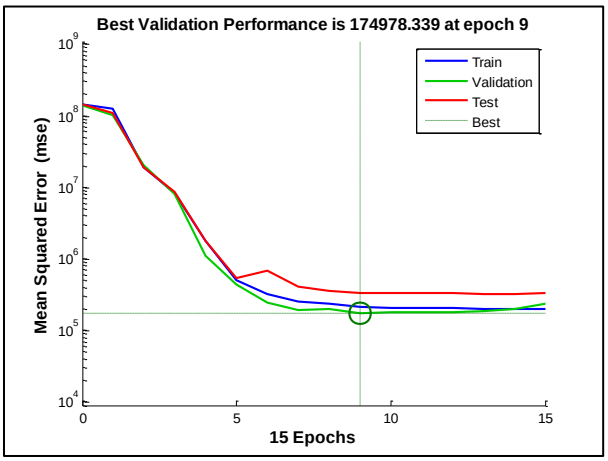

Figure 4-28: Model 3 supervised learning stopping criteria

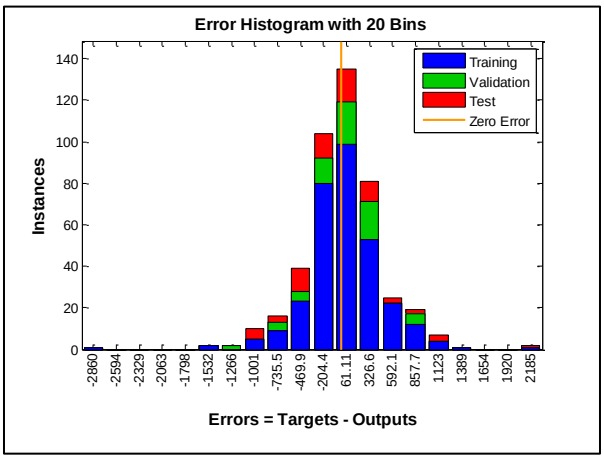

Figure 4-29: Model 3 error histogram

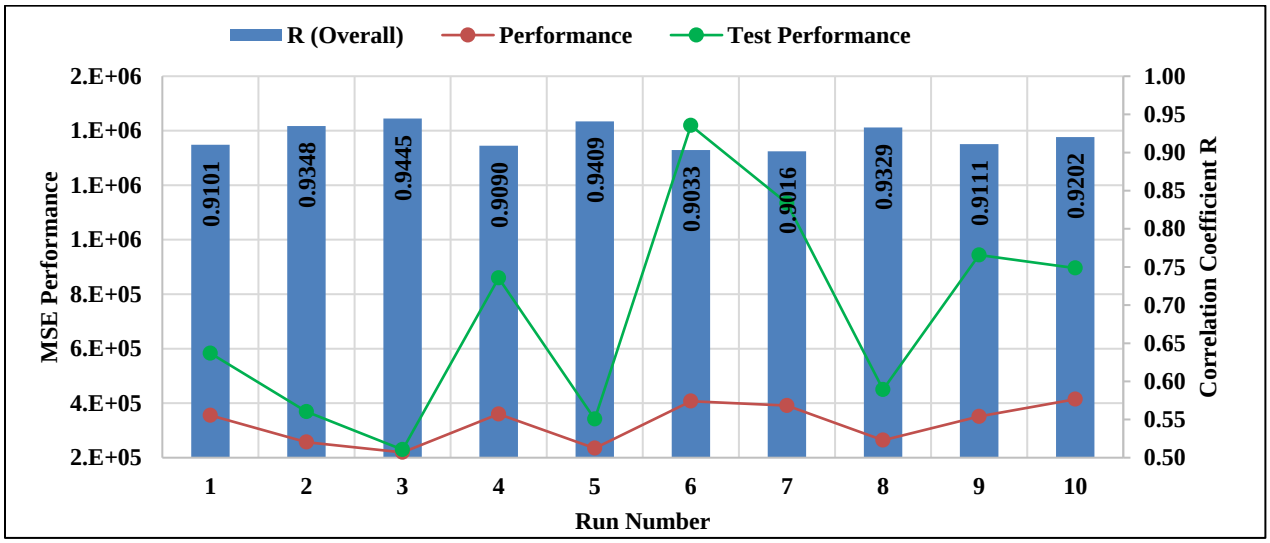

Figure 4-30: Model 3 performance variation and consistency demonstration

4.4.3.1 Model 3 Training Data Simulation

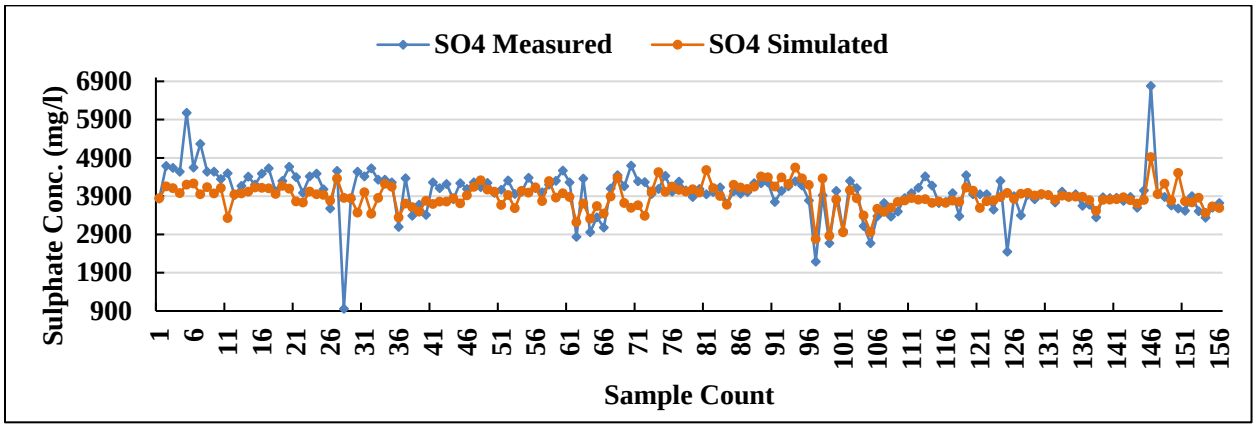

Figure 4-31: Model 3 measured and simulated [SO_4^{2-}] comparison for the training set

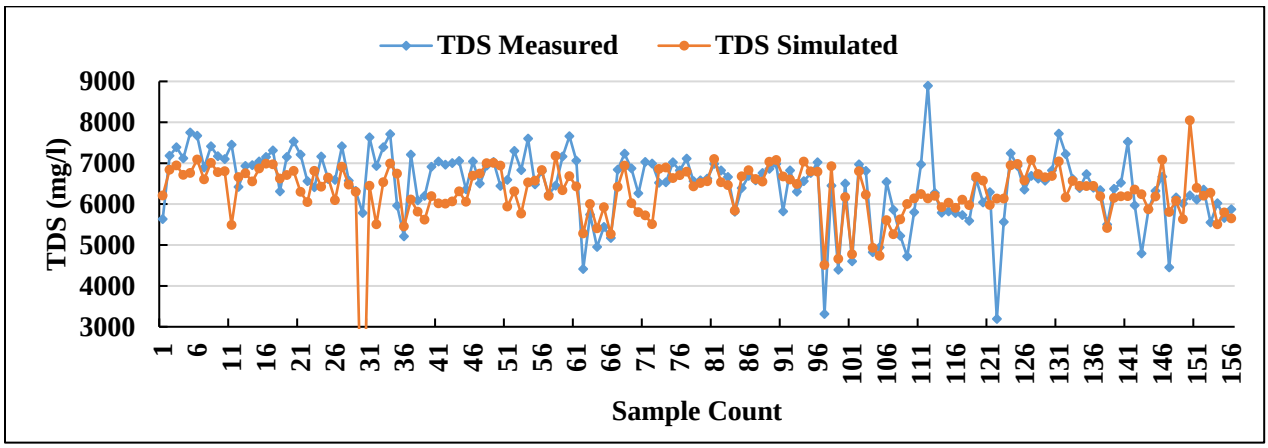

Figure 4-32: Model 3 measured and simulated TDS comparison for the training set

4.4.3.2 Model 3 Validation Data Simulation

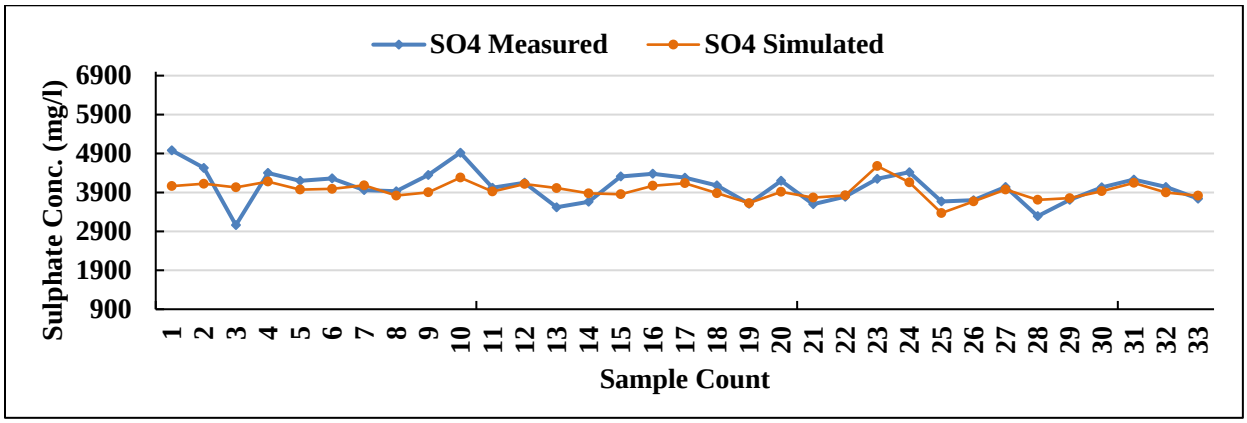

Figure 4-33: Model 3 measured and simulated [SO_4^{2-}] comparison for validation data set

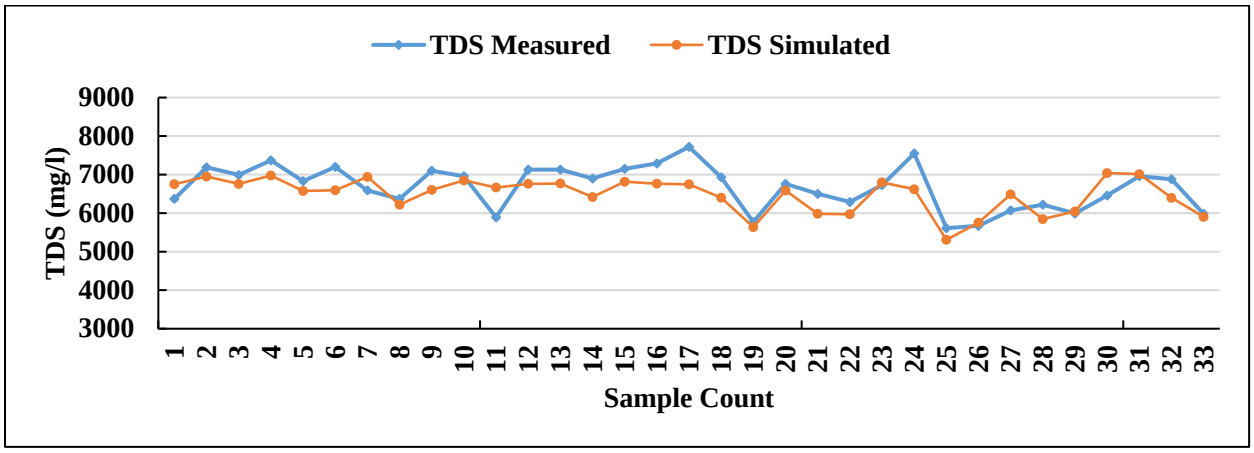

Figure 4-34: Model 3 measured and simulated TDS comparison for validation data set

4.4.3.3 Model 3 Testing Data simulation

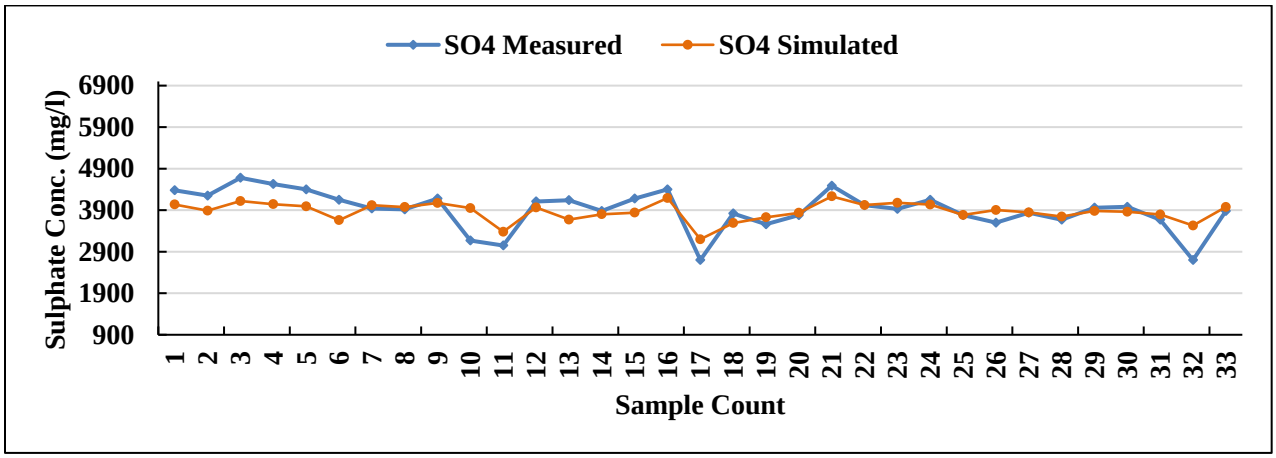

Figure 4-35: Model 3 measured and simulated [SO_4^{2-}] comparison for testing data set

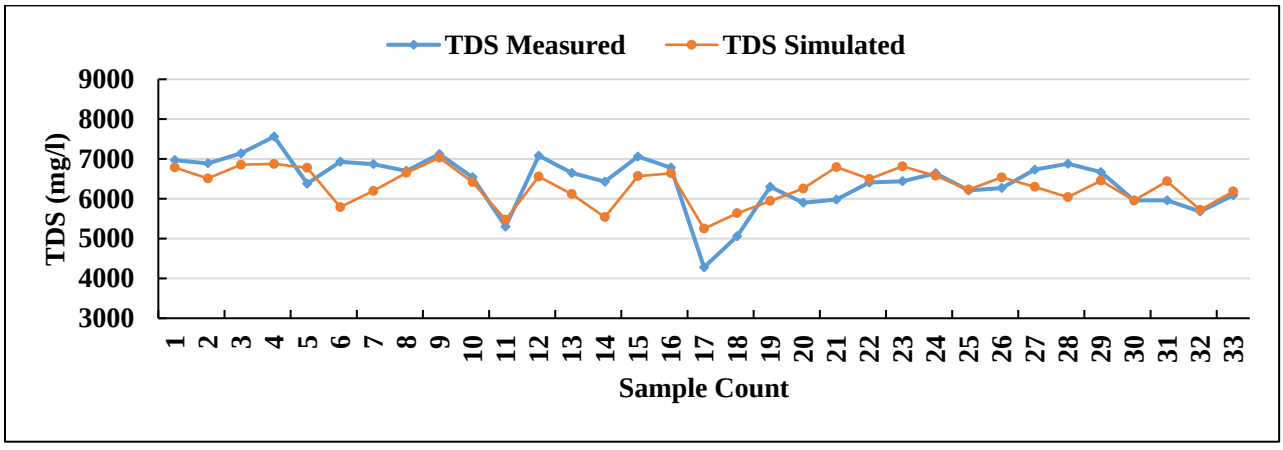

Figure 4-36: Model 3 measured and simulated TDS comparison for testing data set

4.4.4 Simulated Model 4 Network Properties

- Network inputs: pH, EC and [Cl^-]
- Network outputs: TDS and [SO_4^{2-}]
- Network type: Feed-Forward Back-Propagation
- Data Division function: DIVIDERAND
- Training function: TRAINLM
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 50
- Training/Validation/Testing data %split: 70:15:15

Network model 4 was also trained successfully with 15 iterations as seen on Figure 4-37 and Figure 4-39. Its one-time arbitrary run R performance is 0.87708 with a very narrow error distribution around zero mean error as seen on Figure 4-38 and Figure 4-40. From the multiple run outcomes shown as Figure 4-41, it can be seen that the model does not maintain the R value at a higher, and the testing error performance fluctuates as well. The actual and simulation outcome comparisons were done and given as Figure 4-42 - Figure 4-47. It can be seen from these figures that there is a good correlation, particularly on the [SO_4^{2-}] comparisons.

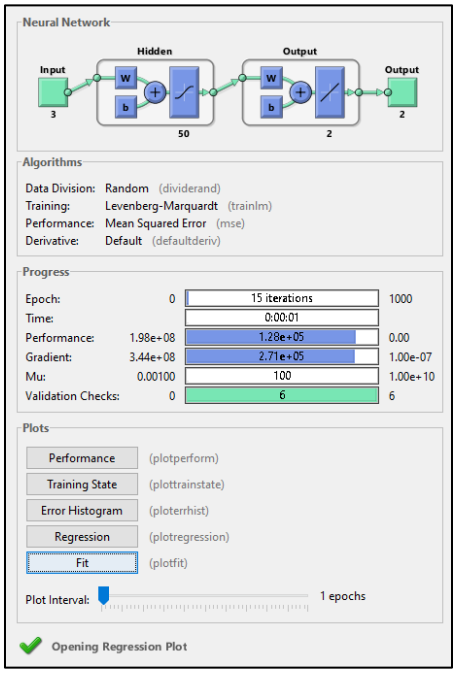

Figure 4-37: Model 4 simulation faceplate during training

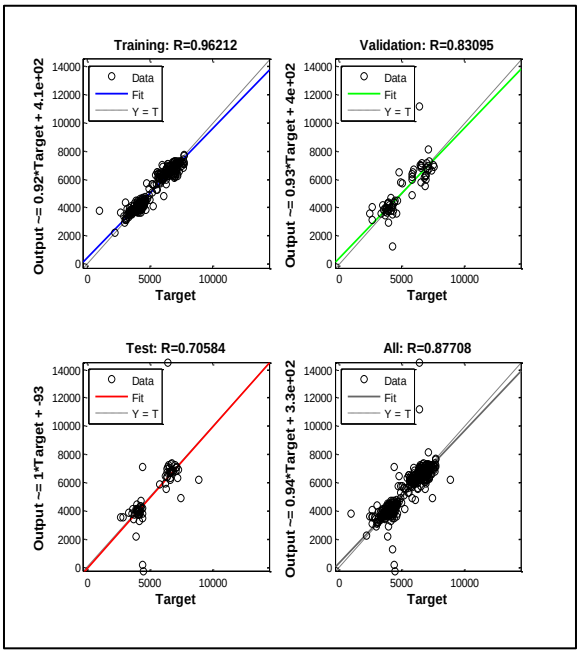

Figure 4-38: Model 4 network regression

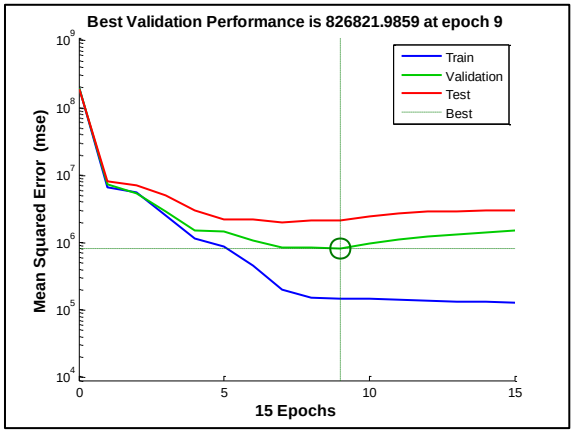

Figure 4-39: Model 4 supervised learning stopping criteria

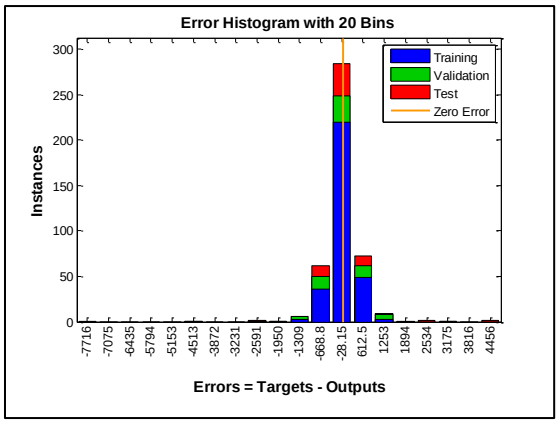

Figure 4-40: Model 4 error histogram

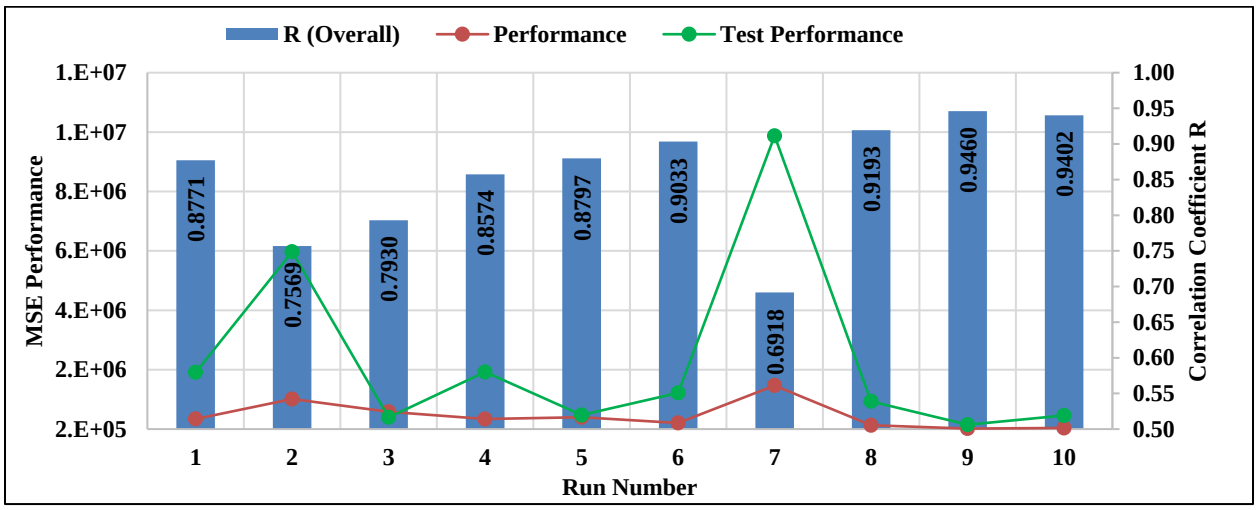

Figure 4-41: Model 4 performance variation and consistency demonstration

4.4.4.1 Model 4 Training Data Simulation

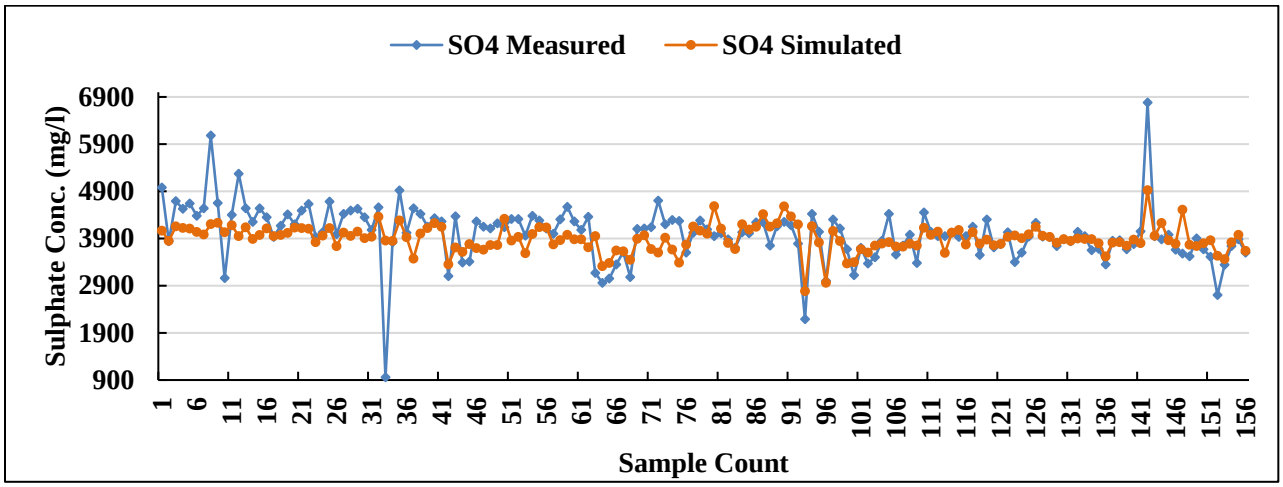

Figure 4-42: Model 4 measured and simulated [SO_4^{2-}] comparison for the training set

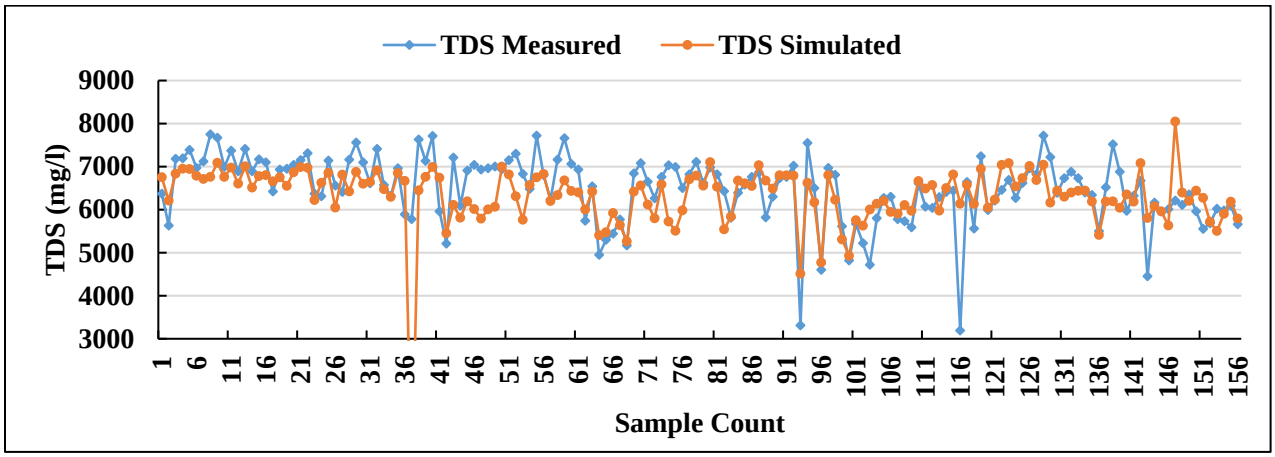

Figure 4-43: Model 4 measured and simulated TDS comparison for the training set

4.4.4.2 Model 4 Validation Data Simulation

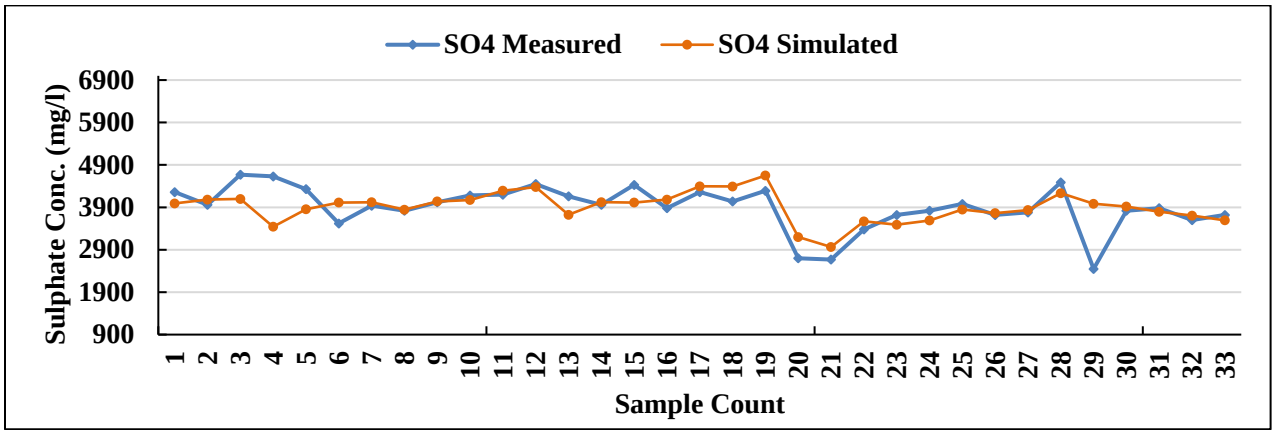

Figure 4-44: Model 4 measured and simulated [SO_4^{2-}] comparison for validation data set

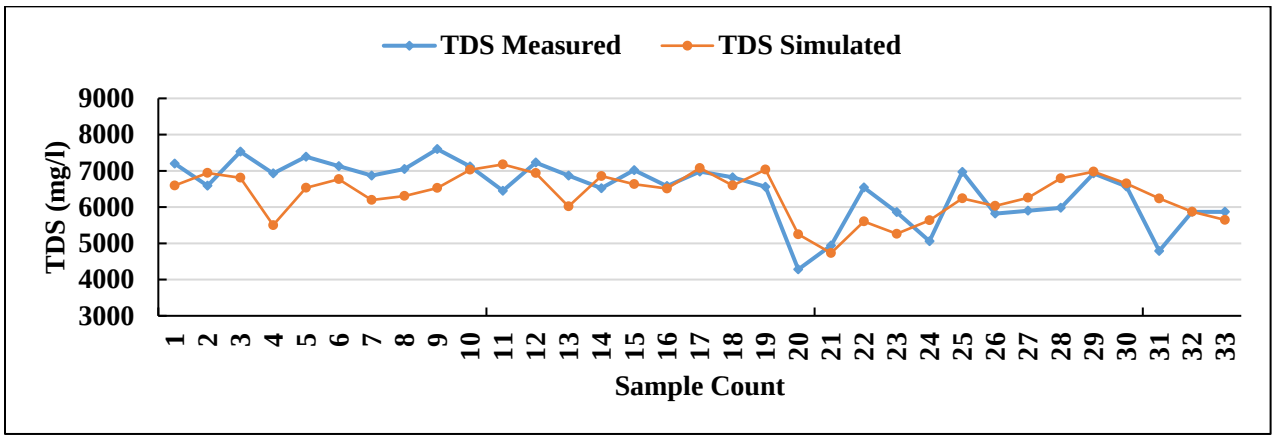

Figure 4-45: Model 4 measured and simulated TDS comparison for validation data set

4.4.4.3 Model 4 Testing Data simulation

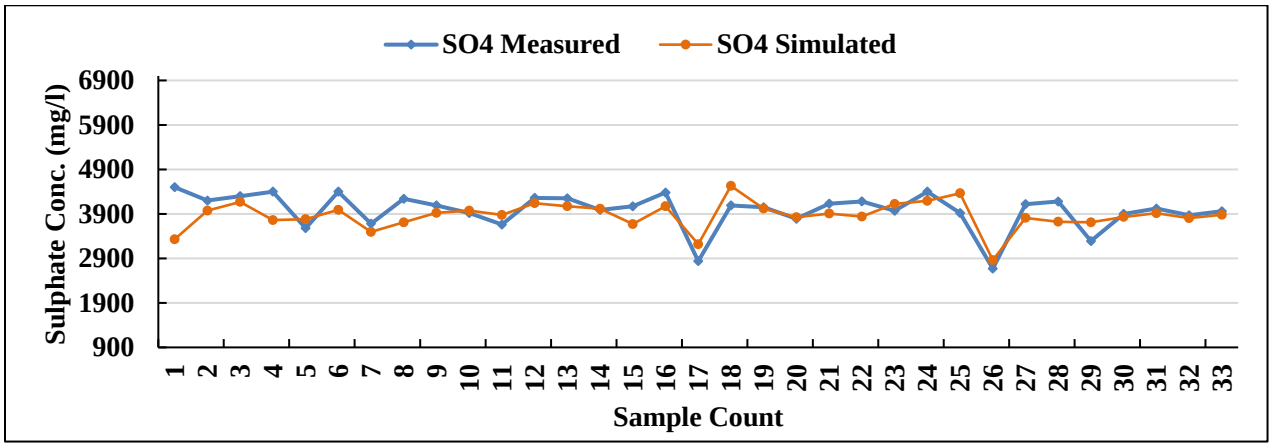

Figure 4-46: Model 4 measured and simulated [SO_4^{2-}] comparison for testing set

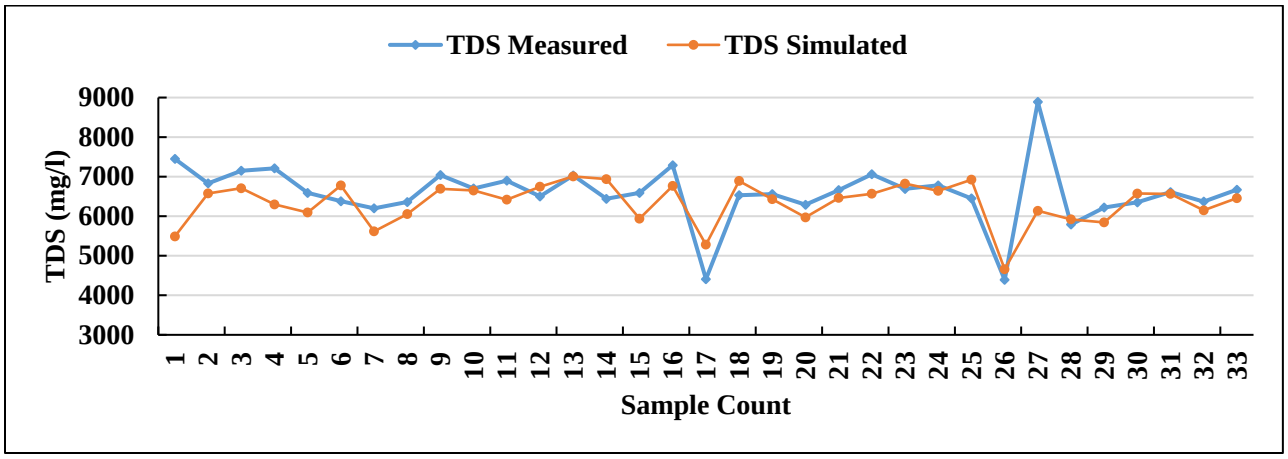

Figure 4-47: Model 4 measured and simulated TDS comparison for testing data set

4.4.5 Simulated Model 5 Network Properties

- Network inputs: pH, EC and [Cl^-]
- Network outputs: TDS and [SO_4^{2-}]
- Network type: Feed-Forward Back-Propagation
- Data Division function: DIVIDERAND
- Training function: TRAINGD
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 20
- Training/Validation/Testing data %split: 70:15:15

Model 5 was built to introduce a variety in training algorithms, but otherwise similar to Model 3 in all other properties. This model also ran successfully to convergence and seen on Figure 4-48. Its regression performance however seems very poor on training, validation, testing, and consequently overall data set simulation which was only 0.74864 as seen on Figure 4-49. The error distribution was very broad as seen on Figure 4-51. This means that it generates both small and large positive and negative errors. This observation helps explains the totally skewed learning curve on Figure 4-50 and also the negative and positive MSE and R performances obtained during the multiple runs as shown on Figure 4-52. The simulations outcomes for comparisons with the actual outcomes are also slightly distorted for training, validation and testing data sets as seen on Figure 4-53 - Figure 4-58.

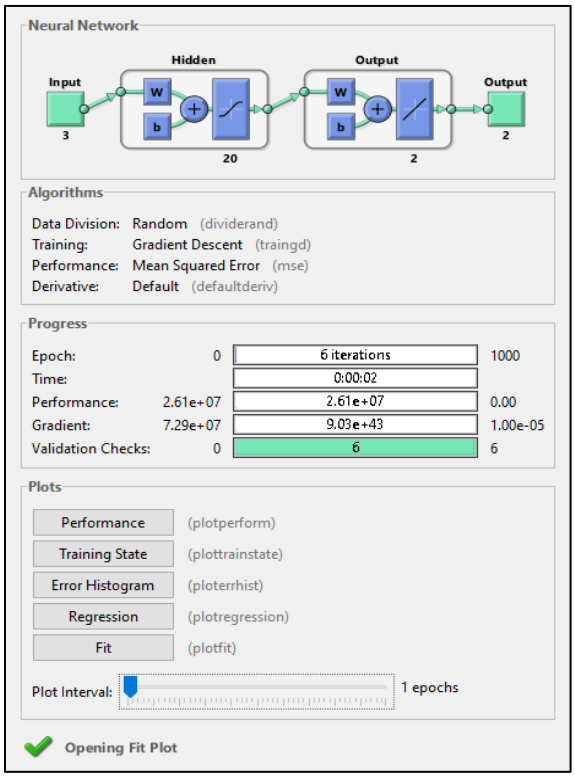

Figure 4-48: Model 5 simulation faceplate during training

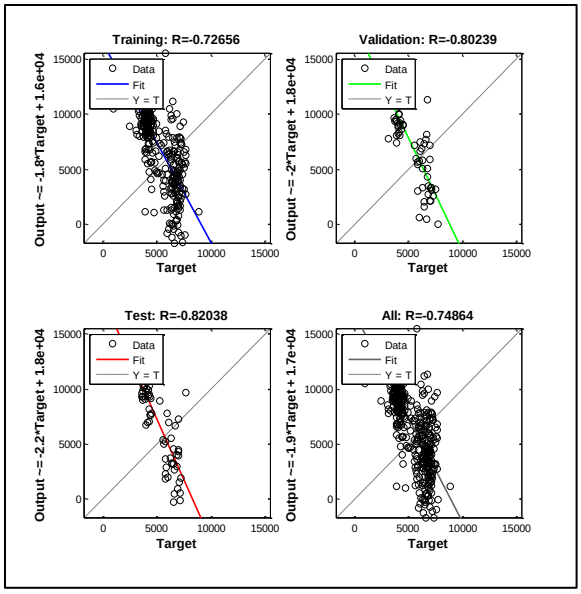

Figure 4-49: Model 5 network regression

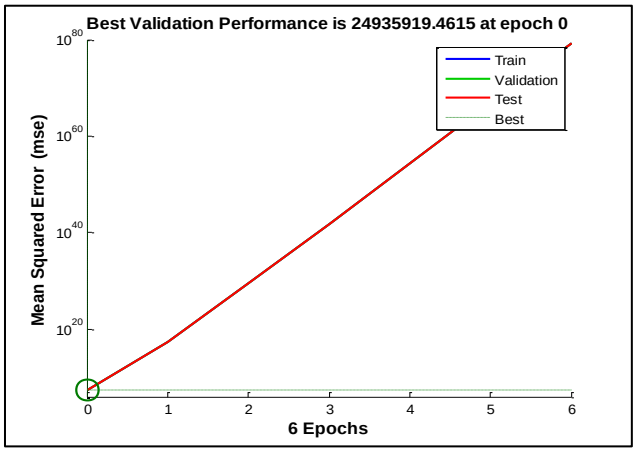

Figure 4-50: Model 5 supervised learning stopping criteria

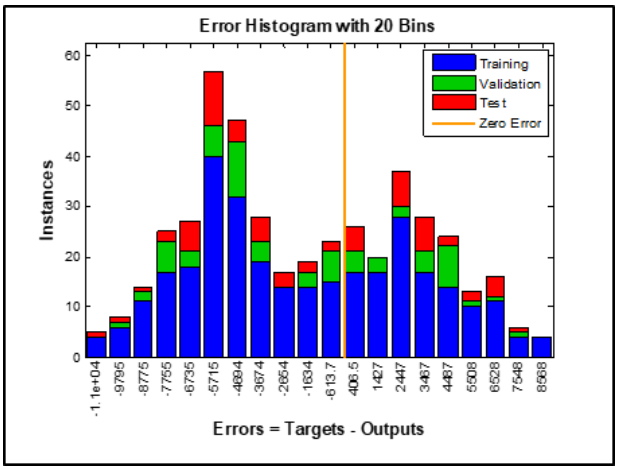

Figure 4-51: Model 5 error histogram

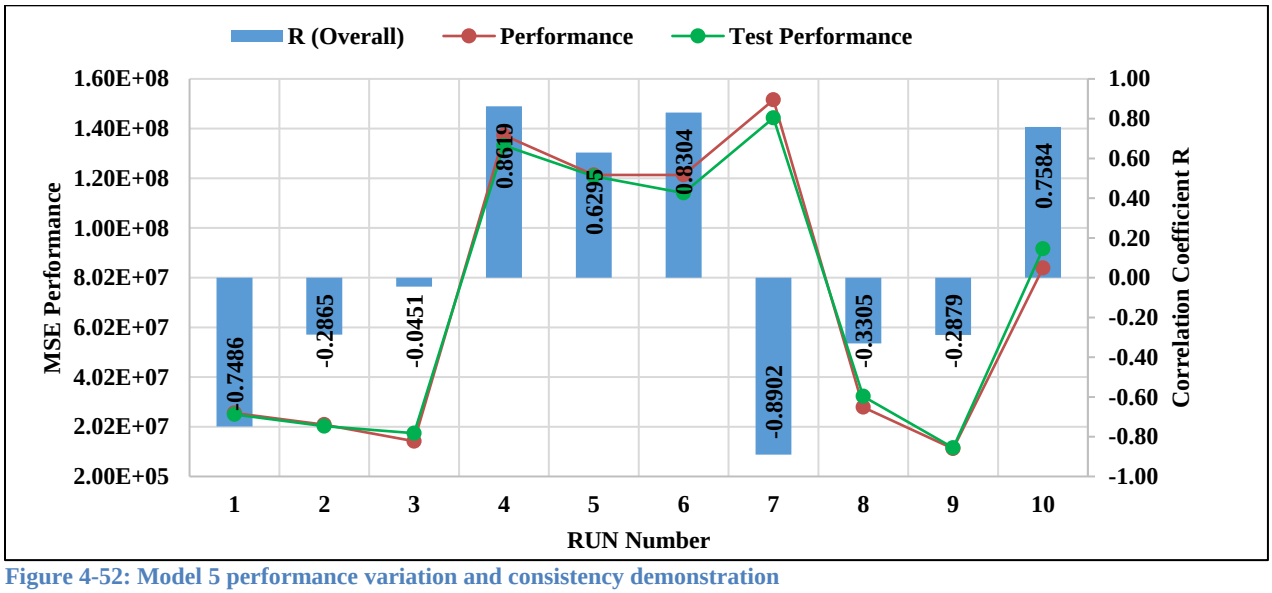

Figure 4-52: Model 5 performance variation and consistency demonstration

4.4.5.1 Model 5 Training Data Simulation

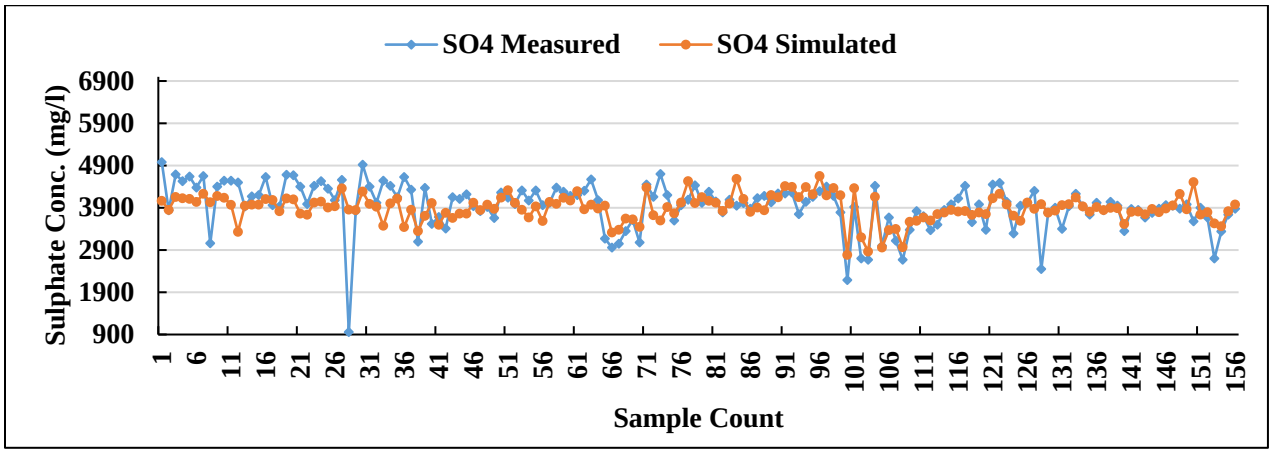

Figure 4-53: Model 5 measured and simulated [SO_4^{2-}] comparison for the training set

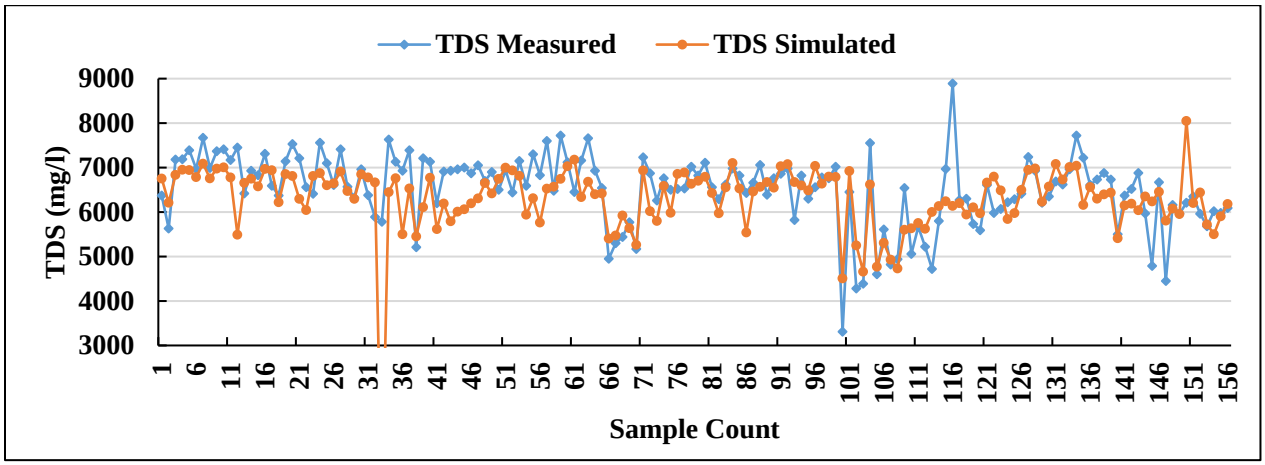

Figure 4-54: Model 5 measured and simulated TDS comparison for the training set

4.4.5.2 Model 5 Validation Data Simulation

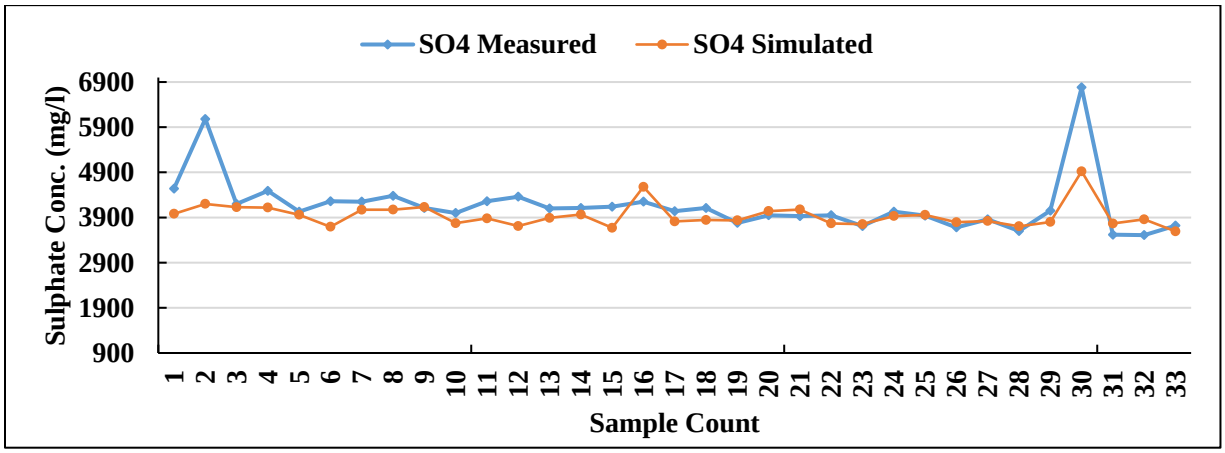

Figure 4-55: Model 5 measured and simulated [SO_4^{2-}] comparison for validation data set

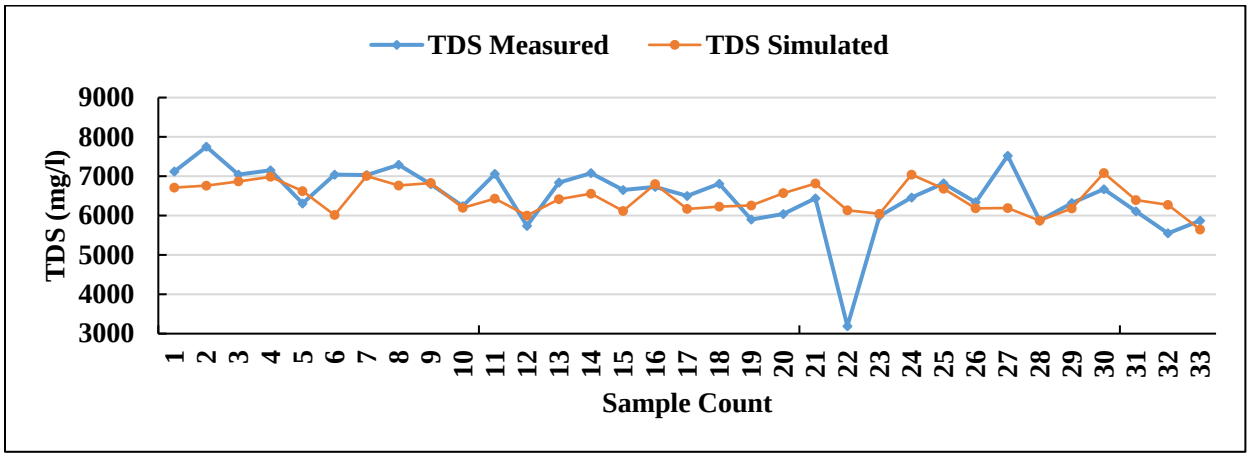

Figure 4-56: Model 5 measured and simulated TDS comparison for validation data set

4.4.5.3 Model 5 Testing Data simulation

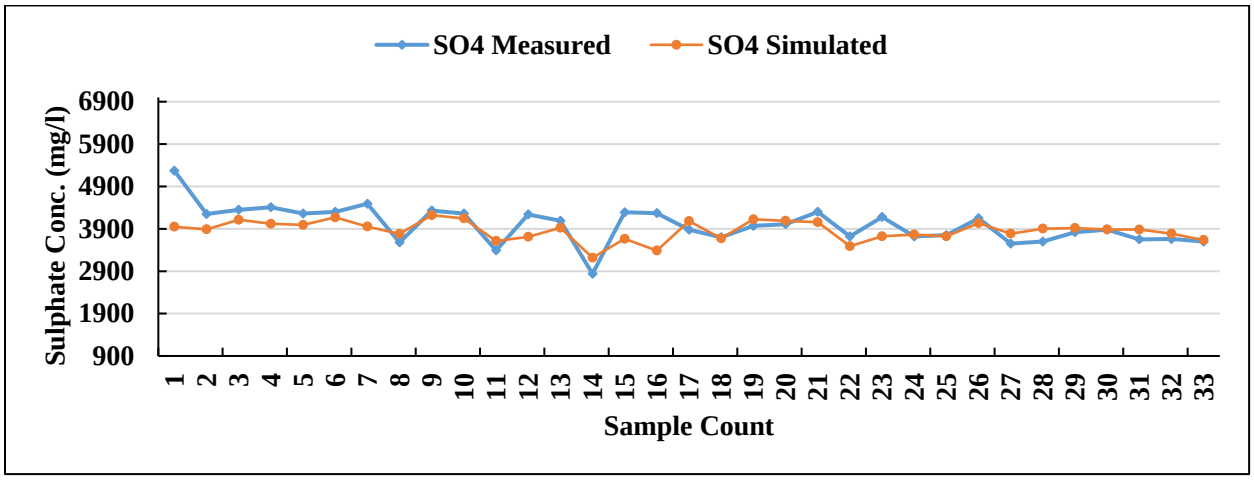

Figure 4-57: Model 5 measured and simulated [SO_4^{2-}] comparison for testing data set

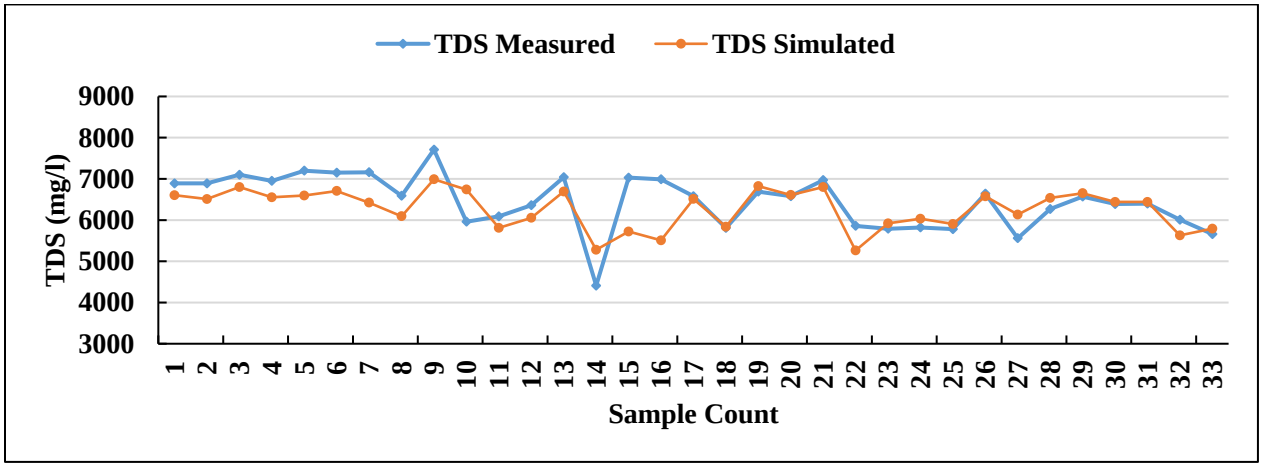

Figure 4-58: Model 5 measured and simulated TDS comparison for testing data set

4.4.6 Simulated Model 6 Network Properties

- Network inputs: pH, EC and [Cl^-]
- Network outputs: TDS and [SO_4^{2-}]
- Network type: Feed-Forward Back-Propagation
- Data Division function: DIVIDERAND
- Training function: TRAINRP
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 20
- Training/Validation/Testing data %split: 70:15:15

Model 6 was also built to bring about variety in terms of training algorithms used on the models from which the best model was chosen. It is configured exactly the same way as Model 3 and Model 5, with the algorithm being the only thing that differs. Model 3 was also trained, validated and tested successfully as seen on Figure 4-59, with an overall R performance of 0.91268 from its once off run as seen on the regression plots given as Figure 4-60. There seems to be some form of convergence when this model runs as seen on Figure 4-61 and the error distribution seems to be a lot better than when traingd is used (Model 5) but not as good as when trainlm is used (Model 3) as seen on Figure 4-62.

The multiple run performance of this model shows that even though the model's correlation coefficient is predominantly above 0.9, it does sometimes go below this mark as seen on Figure 4-63. Its MSE performance for both the training and testing data sets also fluctuates a bit as seen on Figure 4-63, which brings about questions on its performance consistence in this case. The comparisons between the model's simulated outputs to the actual outputs were quite good as seen on Figure 4-64 - Figure 4-69 below, but the properties discussed above are more indicative than what these figures reveal to a mere eye. By looking at the three models built for the purpose of comparing algorithms performance, Model 3 has done a lot better than the Model 5 and Model 6 based on its consistency.

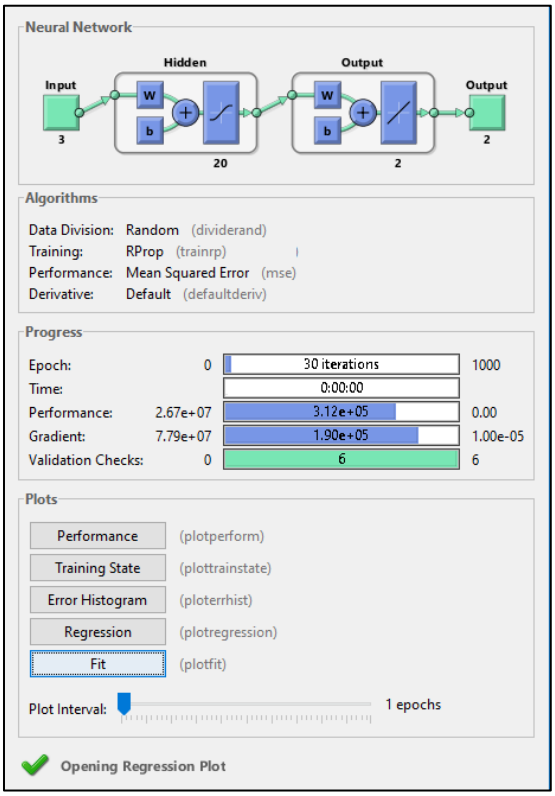

Figure 4-59: Model 6 simulation faceplate during training

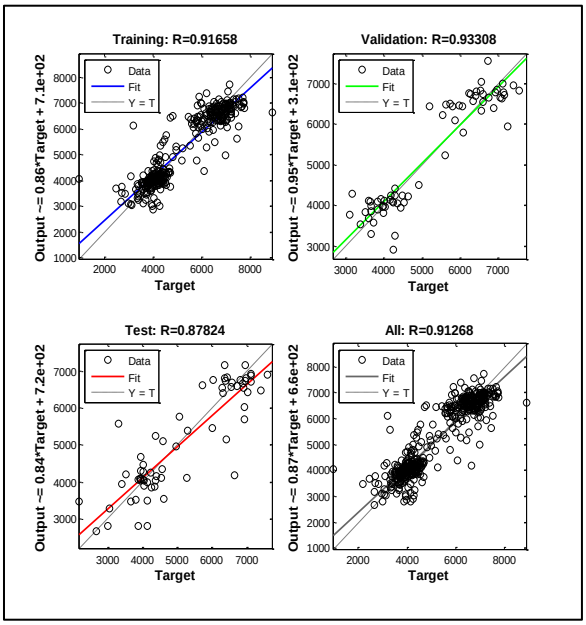

Figure 4-60: Model 6 network regression

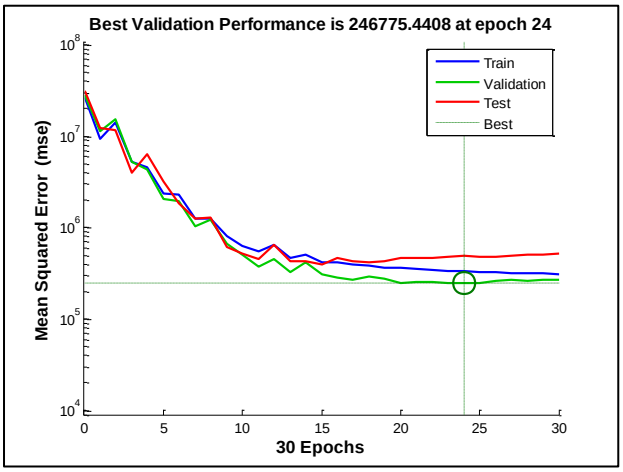

Figure 4-61: Model 6 supervised learning stopping criteria

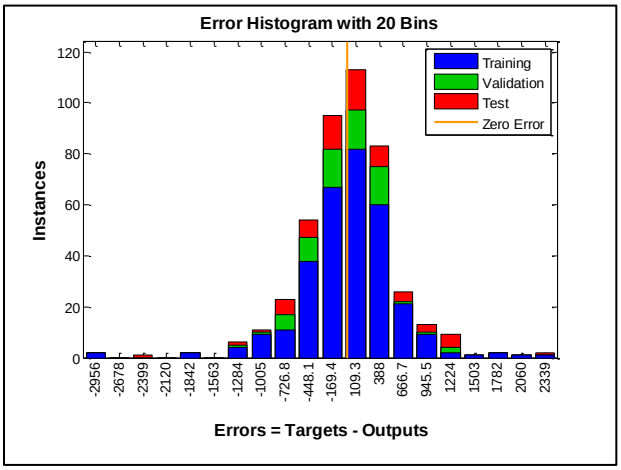

Figure 4-62: Model 6 error histogram

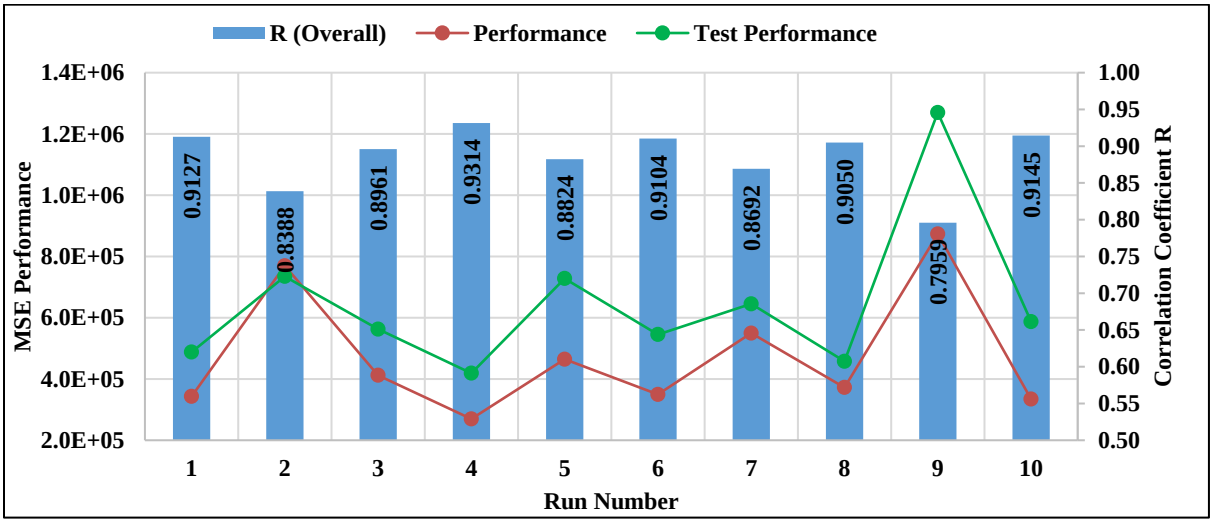

Figure 4-63: Model 6 performance variation and consistency demonstration

4.4.6.1 Model 6 Training Data Simulation

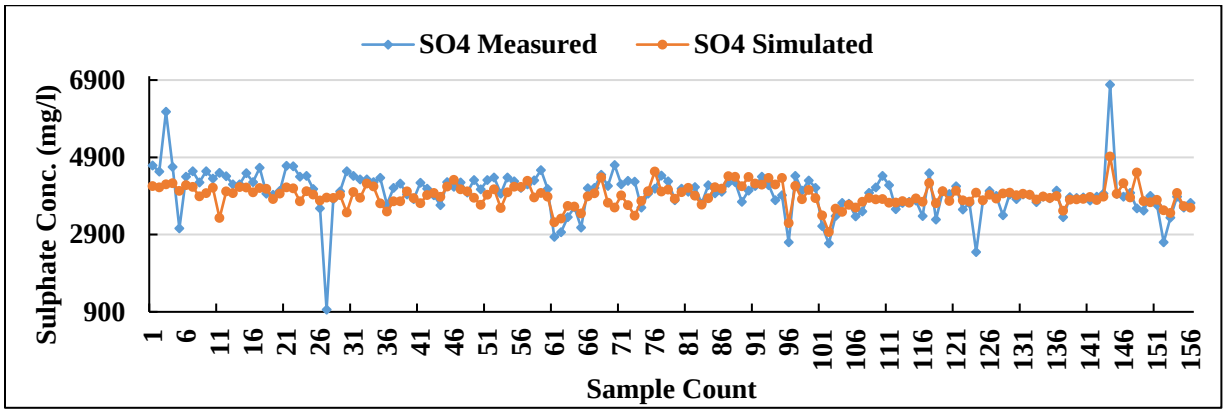

Figure 4-64: Model 6 measured and simulated [SO_4^{2-}] comparison for the training set

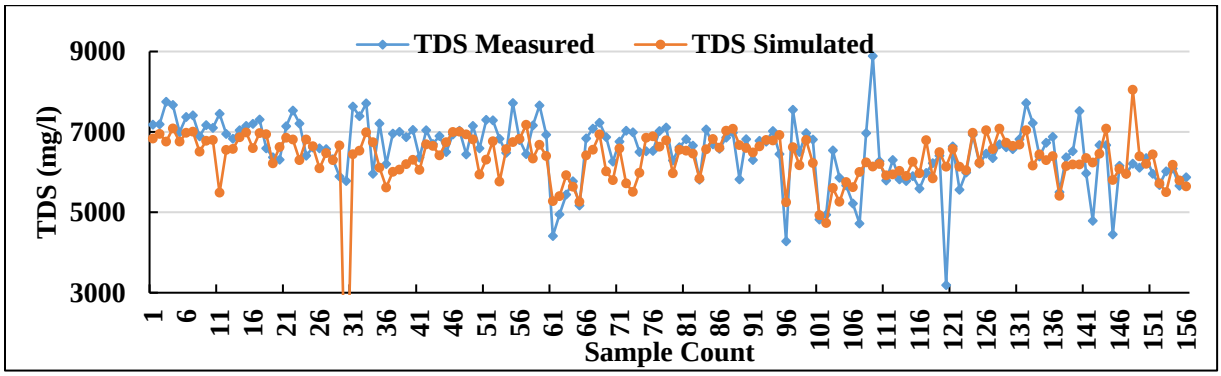

Figure 4-65: Model 6 measured and simulated TDS comparison for the training set

4.4.6.2 Model 6 Validation Data Simulation

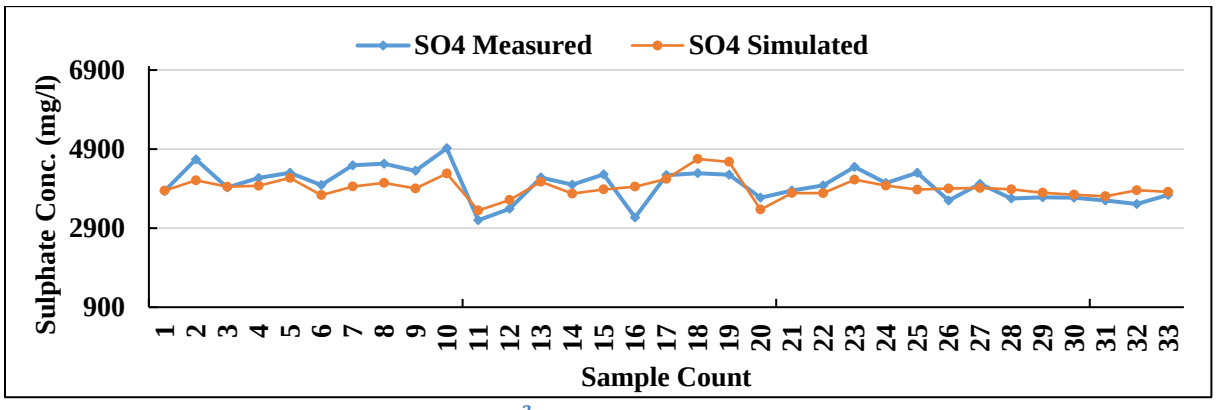

Figure 4-66: Model 6 measured and simulated [SO_4^{2-}] comparison for validation data set

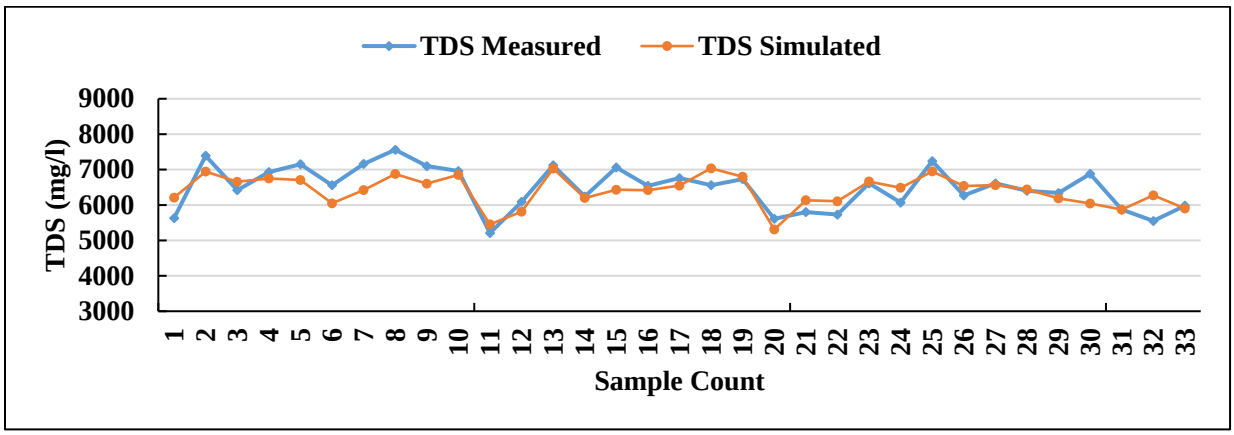

Figure 4-67: Model 6 measured and simulated TDS comparison for validation data set

4.4.6.3 Model 6 Testing Data simulation

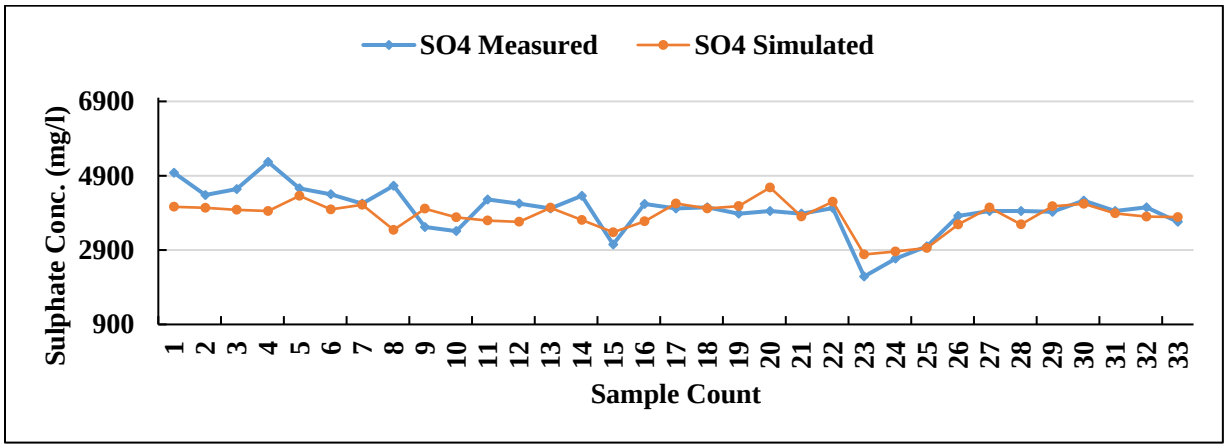

Figure 4-68: Model 6 measured and simulated [SO_4^{2-}] comparison for testing data set

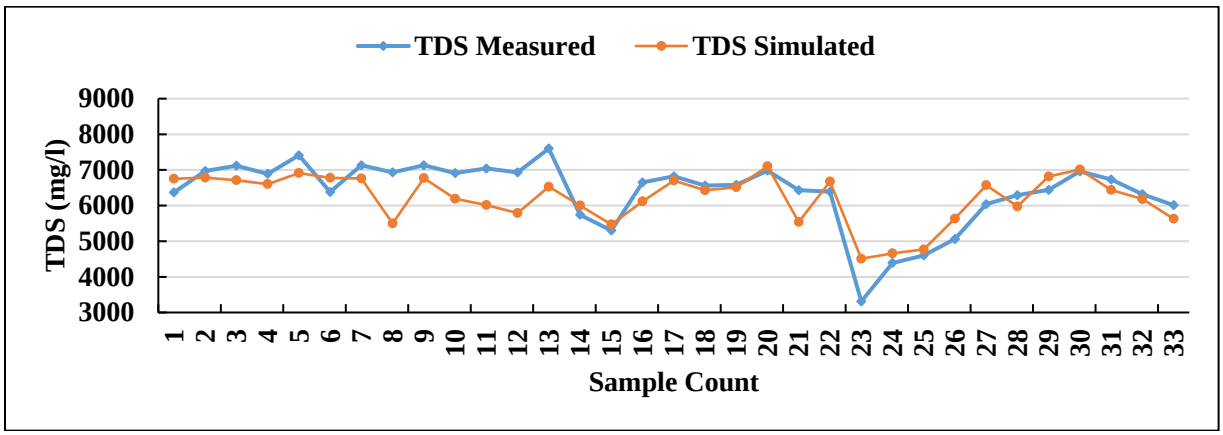

Figure 4-69: Model 6 measured and simulated TDS comparison for testing data set

4.4.7 Simulated Model 7 Network Properties

- Network inputs: pH, EC and [Cl^-]
- Network outputs: TDS and [SO_4^{2-}]
- Network type: Feed-Forward Back-Propagation
- Data Division function: DIVIDERAND
- Training function: TRAINLM
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 20
- Training/Validation/Testing data %split: 80:10:10

Model 3, 7, 8 and 9 were built specifically to determine the effect of that the data split ratio has on model performance so that the split ratio choice made when choosing the best model in this research would be an informed on. Model 7 run successfully with no errors as seen on Figure 4-70 below. The model's once off run coefficient correlation performance 0.93502 as shown on Figure 4-71, and the model least error was achieved after 4 epochs as seen on Figure 4-72. It also have a good error distribution across the zero mark as seen on Figure 4-73. The multiple run performance revealed that the model R performance is consistently above 0.92, and the MSE performance on the training data does not fluctuate much as seen on Figure 4-74. The simulation outcome comparisons with the actual outcomes for trailing, validation and testing data are also quite good as seen on Figure 4-75 - Figure 4-80.

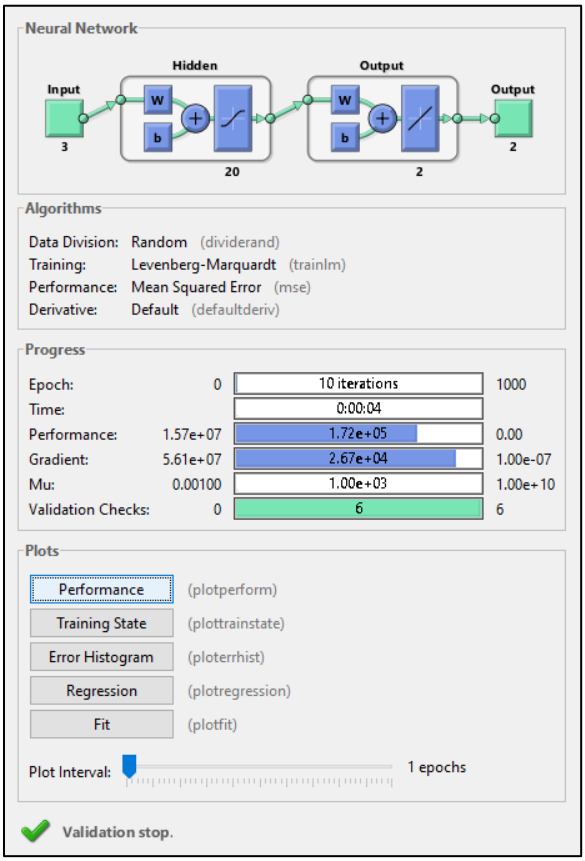

Figure 4-70: Model 7 simulation faceplate during training

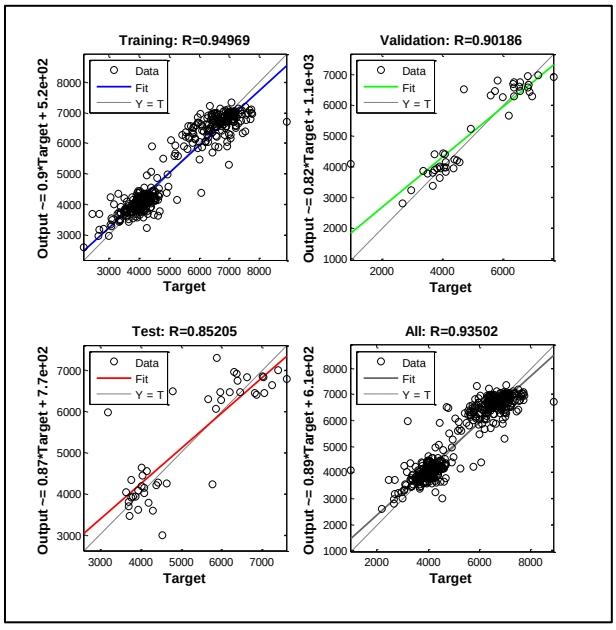

Figure 4-71: Model 7 network regression

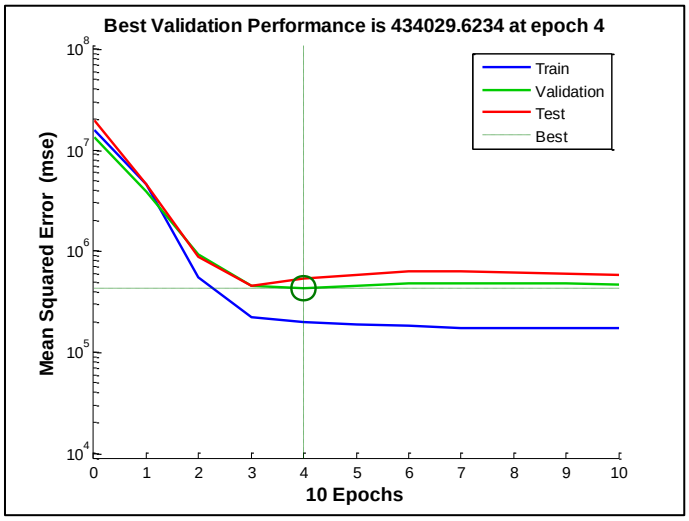

Figure 4-72: Model 7 supervised learning stopping criteria

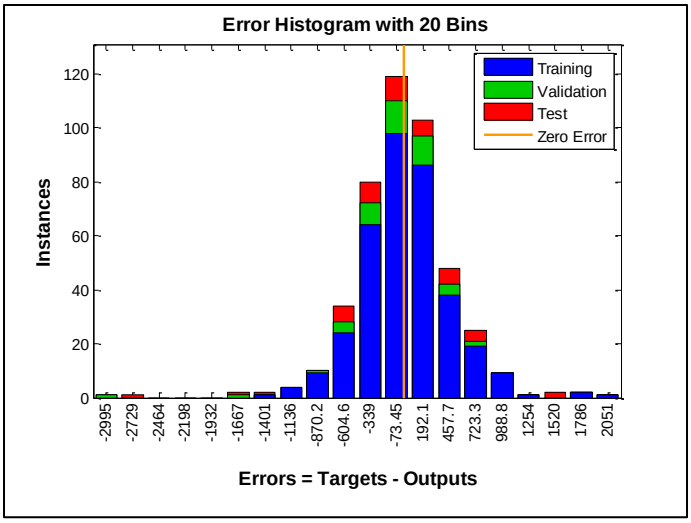

Figure 4-73: Model 7 error histogram

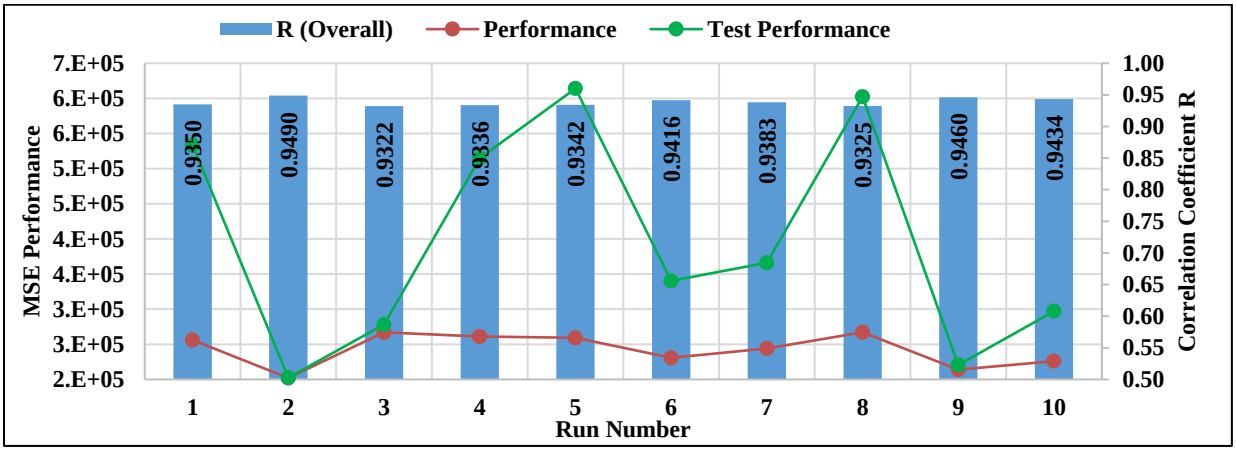

Figure 4-74: Model 7 performance variation and consistency demonstration

4.4.7.1 Model 7 Training Data Simulation

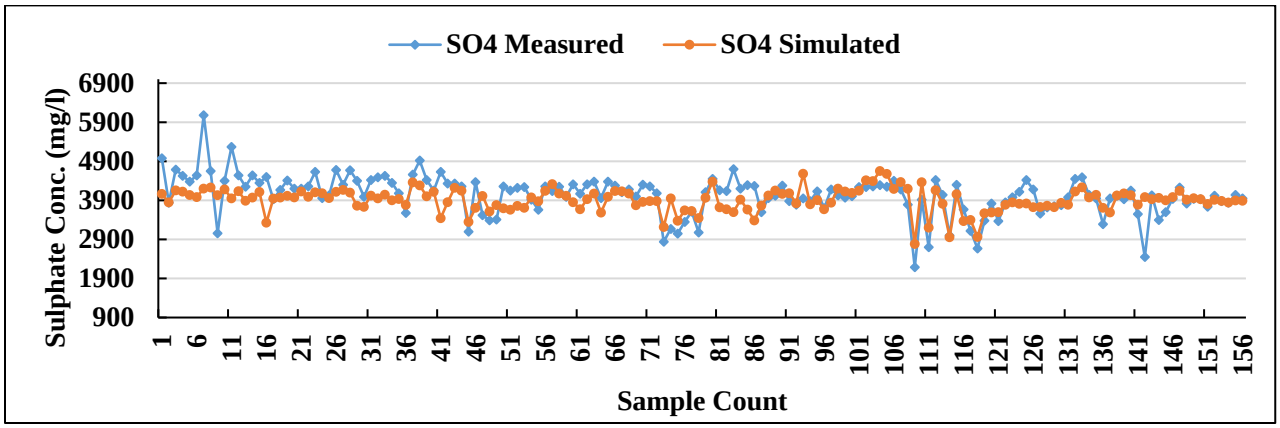

Figure 4-75: Model 7 measured and simulated [SO_4^{2-}] comparison for the training set

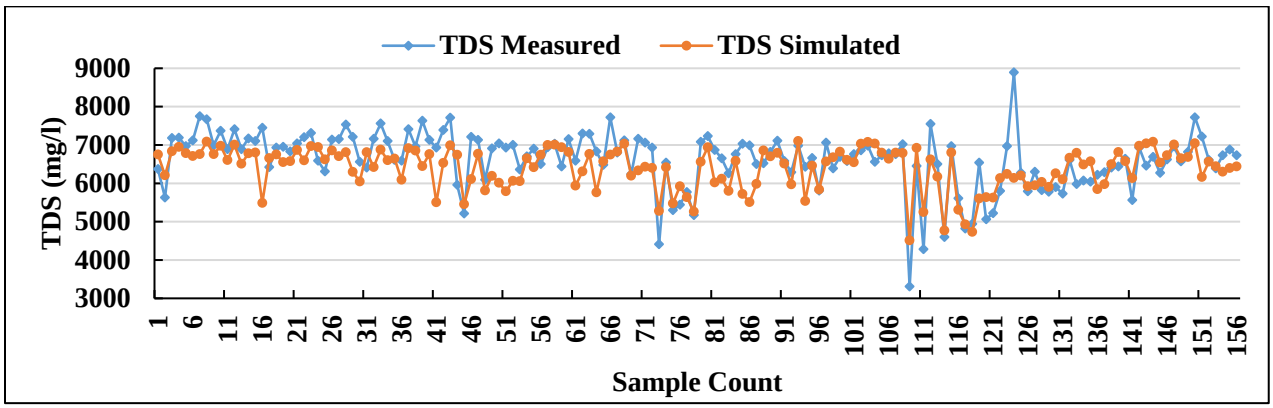

Figure 4-76: Model 7 measured and simulated TDS comparison for the training set

4.4.7.2 Model 7 Validation Data Simulation

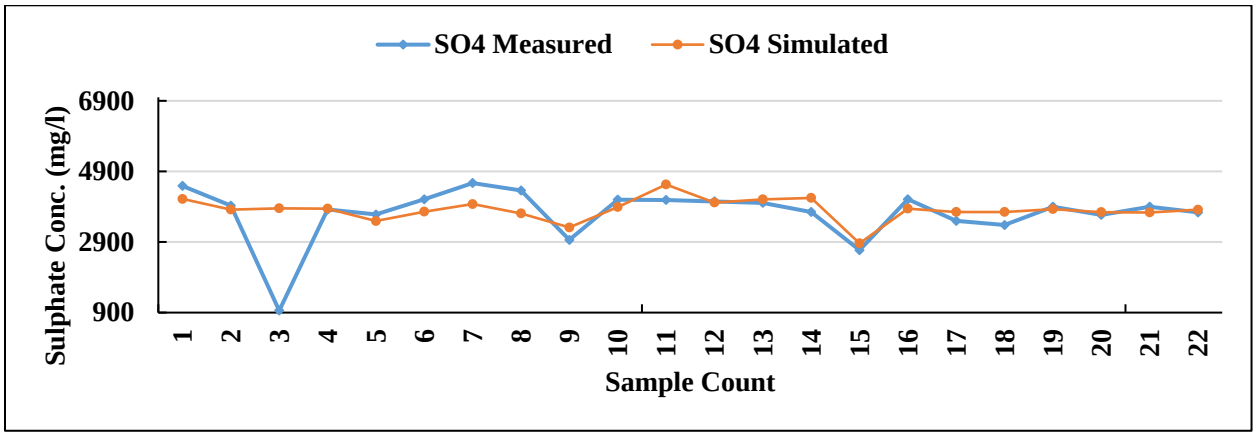

Figure 4-77: Model 7 measured and simulated [SO_4^{2-}] comparison for validation data set

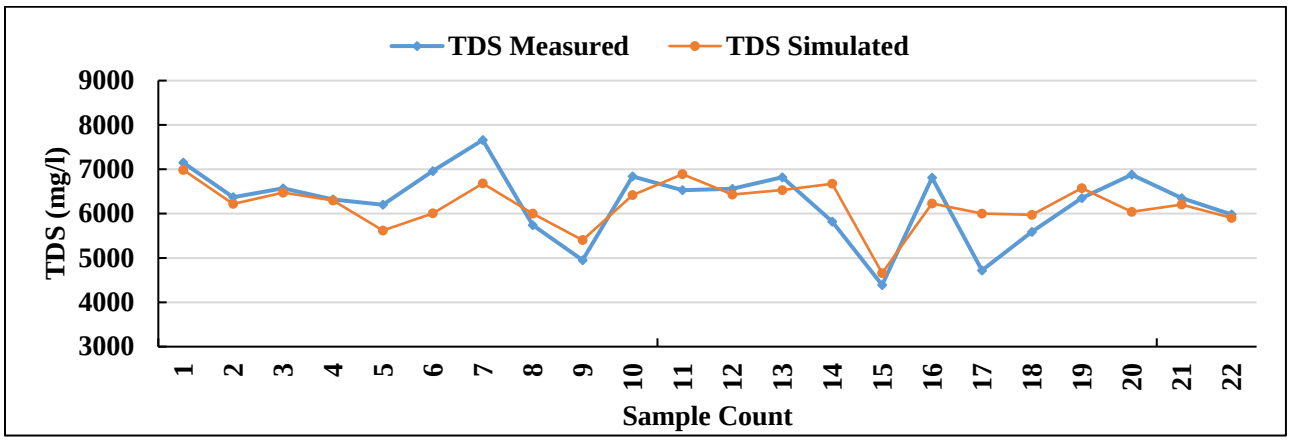

Figure 4-78: Model 7 measured and simulated TDS comparison for validation data set

4.4.7.3 Model 7 Testing Data simulation

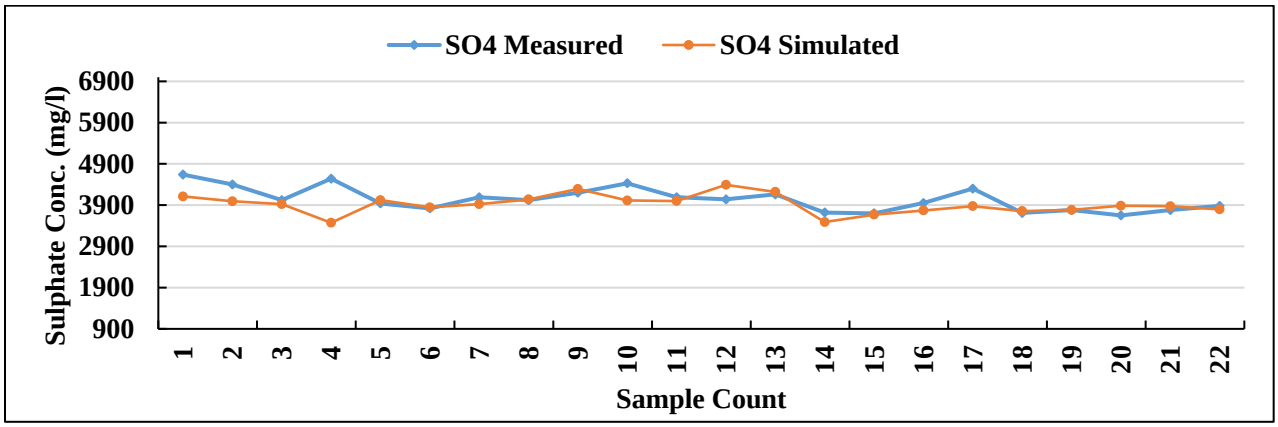

Figure 4-79: Model 7 measured and simulated [SO_4^{2-}] comparison for testing data set

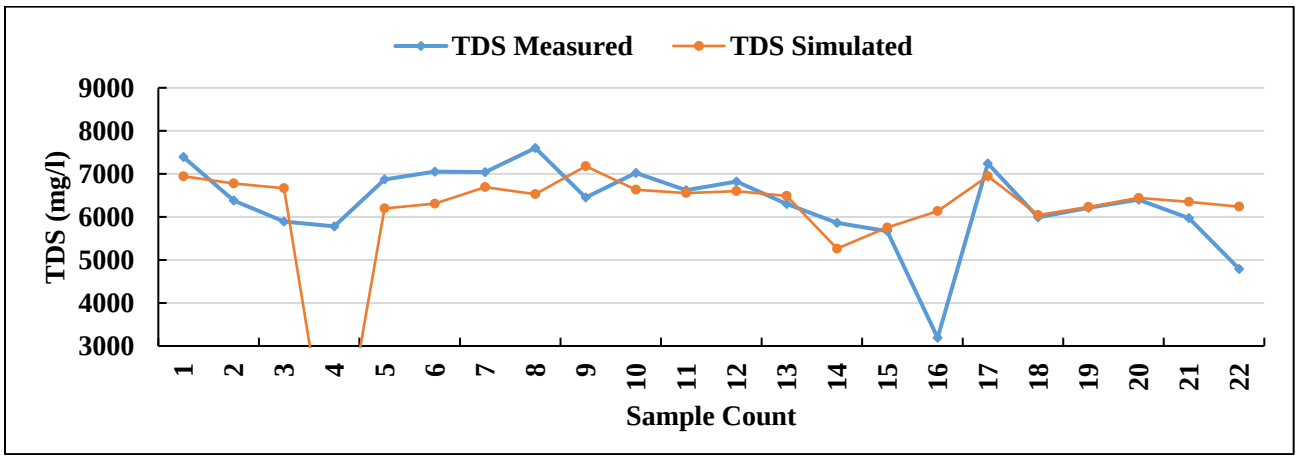

Figure 4-80: Model 7 measured and simulated TDS comparison for testing data set

4.4.8 Simulated Model 8 Network Properties

- Network inputs: pH, EC and [Cl^-]
- Network outputs: TDS and [SO_4^{2-}]
- Network type: Feed-Forward Back-Propagation
- Data Division function: DIVIDERAND
- Training function: TRAINLM
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 20
- Training/Validation/Testing data %split: 90:5:5

Model 8 also ran successfully as seen on Figure 4-81 below. The once of run R performance of 0.9491 was achieved as shown on the regression plot given as Figure 4-82. The lowest MSE was achieved after 22 epochs and the error distribution seems to be very narrow as seen on Figure 4-83 and Figure 4-84 below. The multiple run results revealed that the model's MES performance trends at the lowest value of around 2 000 000, much lower than the rest of the models built during this research. The R performance also trends above 0.925, which again is seen as the highest of all 9 models. The testing data MSE performance also did not fluctuate but has just one spike. The multiple runs results are shown as Figure 4-85 below.

Its ability to predict outcomes to a desirable level is shown with the use of simulated outcome and actual outcome comparison graphs given as Figure 4-86 - Figure 4-91 below. Better correlations are seen on both [SO_4^{2-}] and TDS plots. This greatly adds to the confidence levels associated with the ability of Model 8 to simulate to acceptable levels.

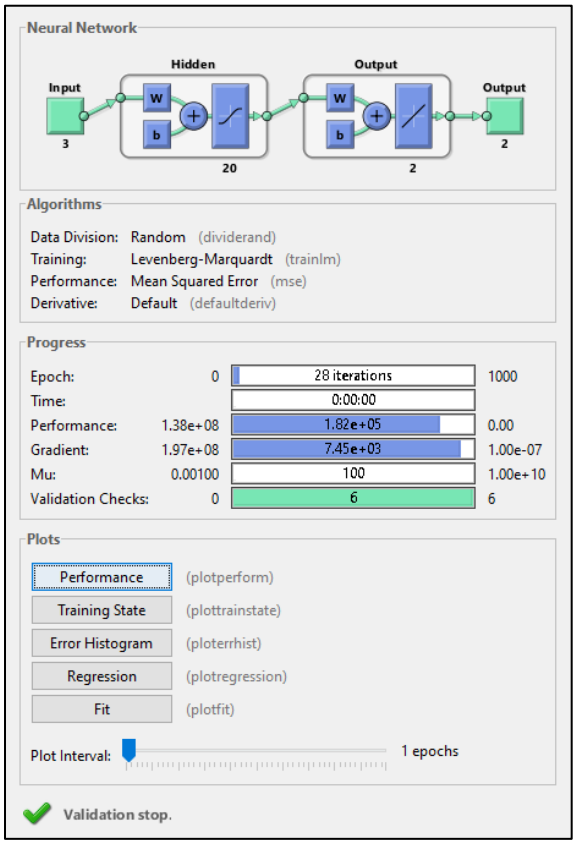

Figure 4-81: Model 8 simulation faceplate during training

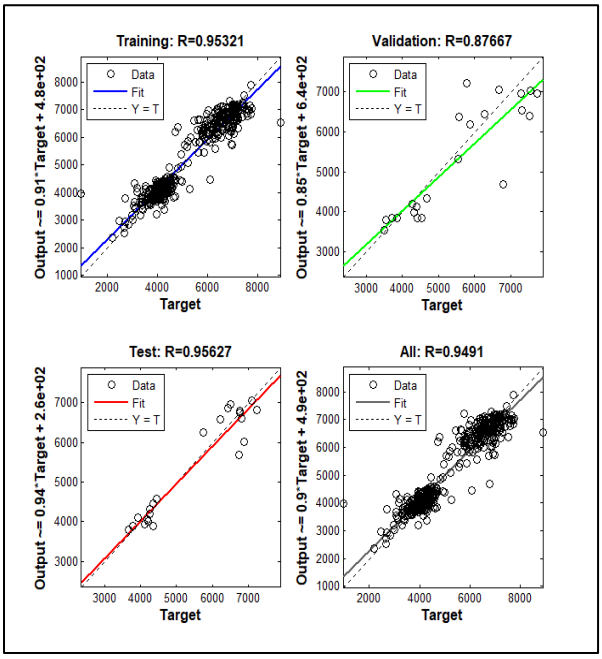

Figure 4-82: Model 8 network regression

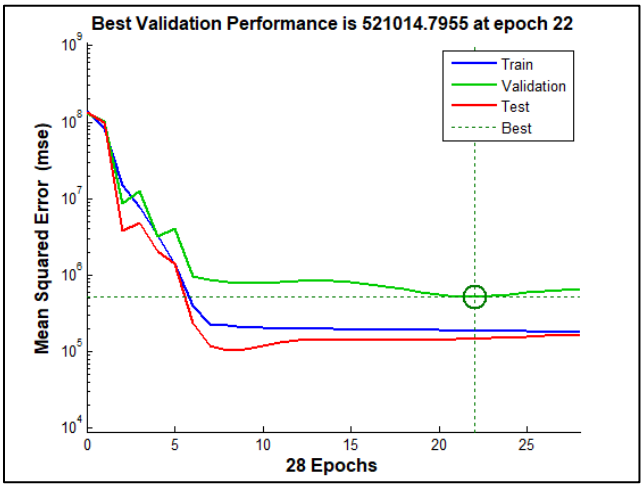

Figure 4-83: Model 8 supervised learning stopping criteria

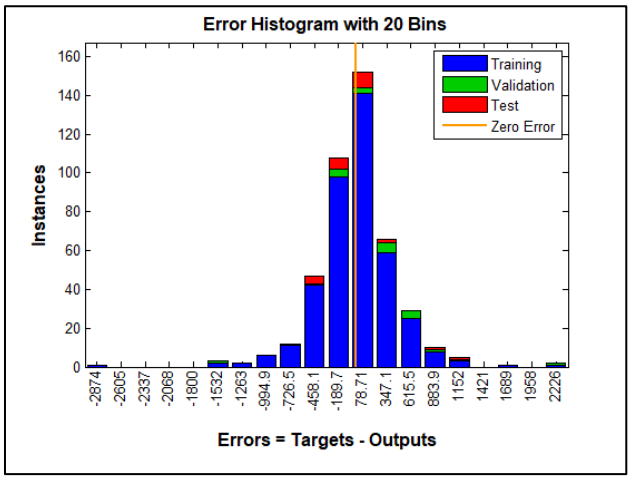

Figure 4-84: Model 8 error histogram

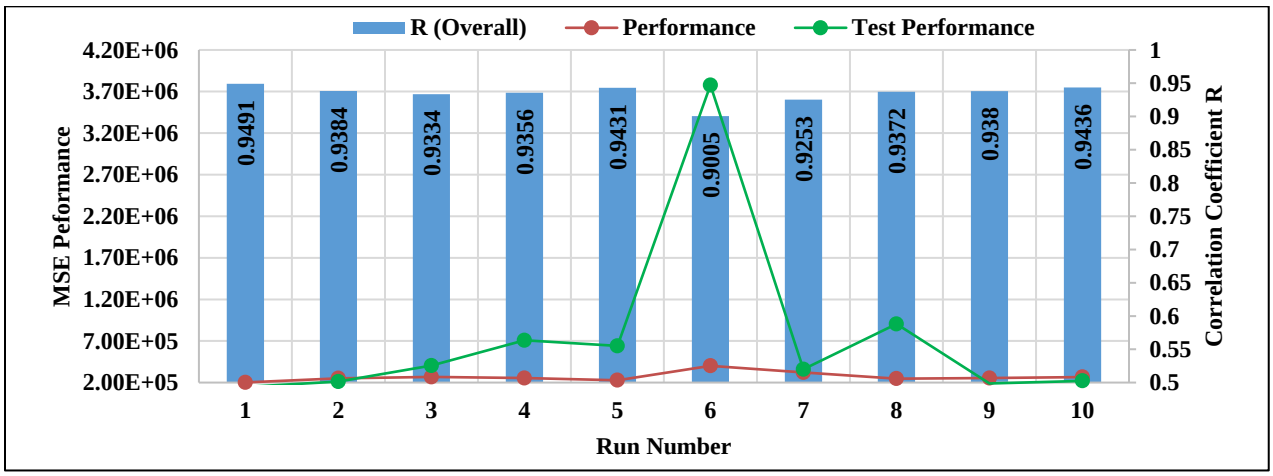

Figure 4-85: Model 8 performance variation and consistency demonstration

4.4.8.1 Model 8 Training Data Simulation

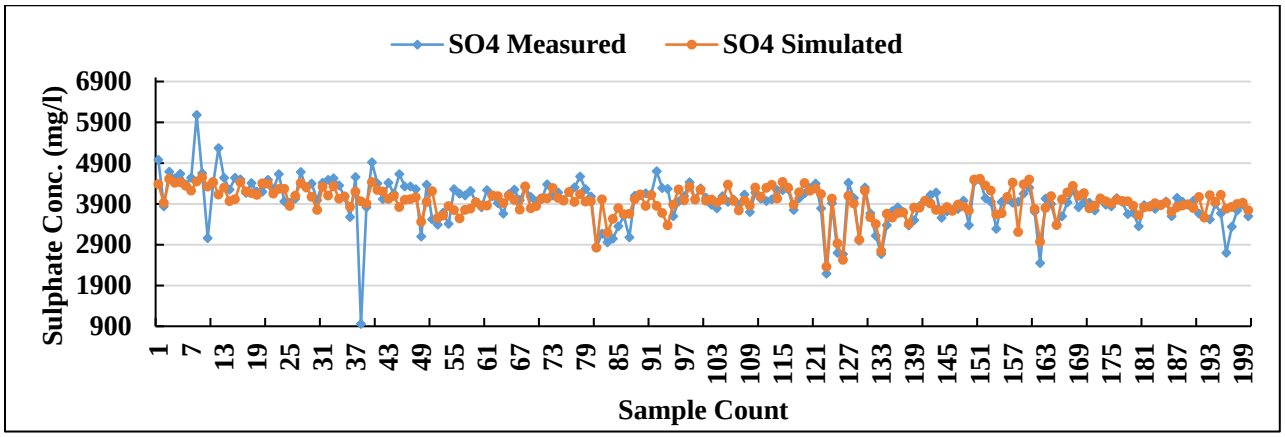

Figure 4-86: Model 8 measured and simulated [SO_4^{2-}] comparison for the training set

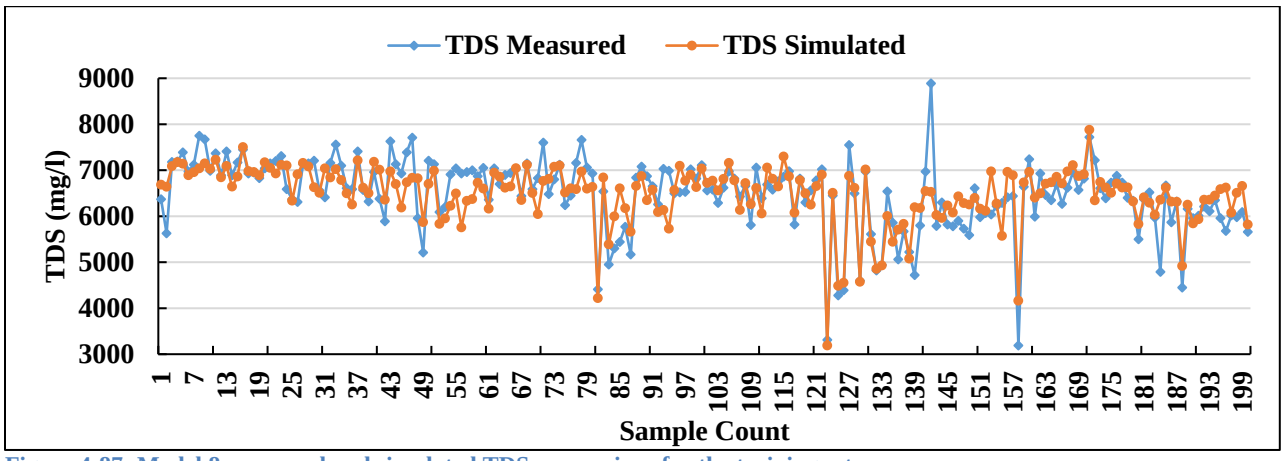

Figure 4-87: Model 8 measured and simulated TDS comparison for the training set

4.4.8.2 Model 8 Validation Data Simulation

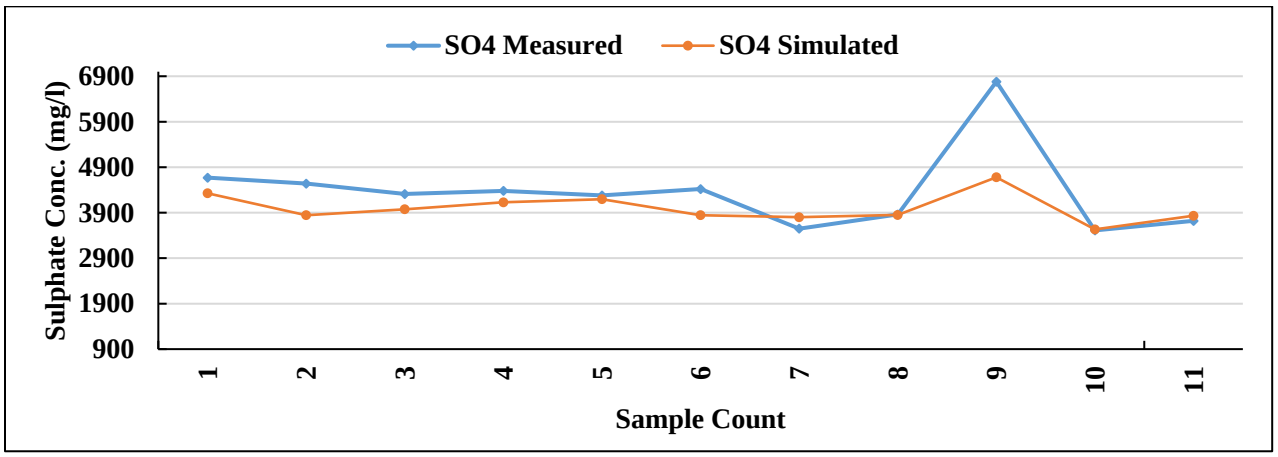

Figure 4-88: Model 8 measured and simulated [SO_4^{2-}] comparison for validation data set=

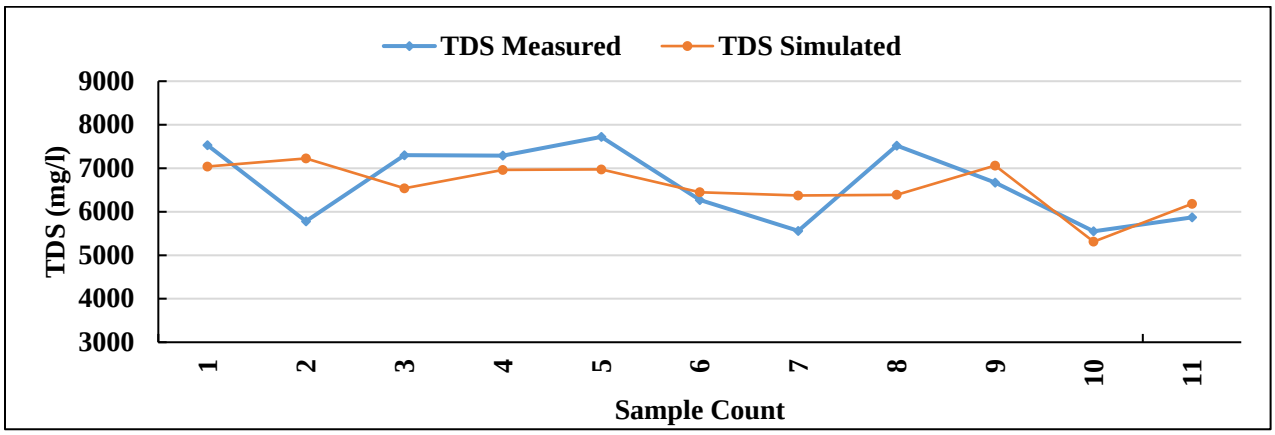

Figure 4-89: Model 8 measured and simulated TDS comparison for validation data set

4.4.8.3 Model 8 Testing Data simulation

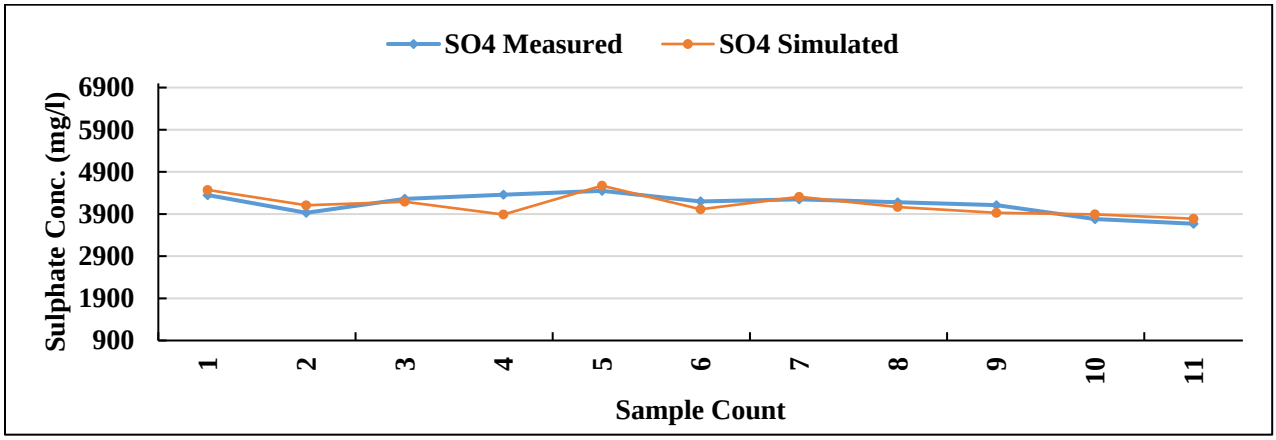

Figure 4-90: Model 8 measured and simulated [SO_4^{2-}] comparison for testing data set

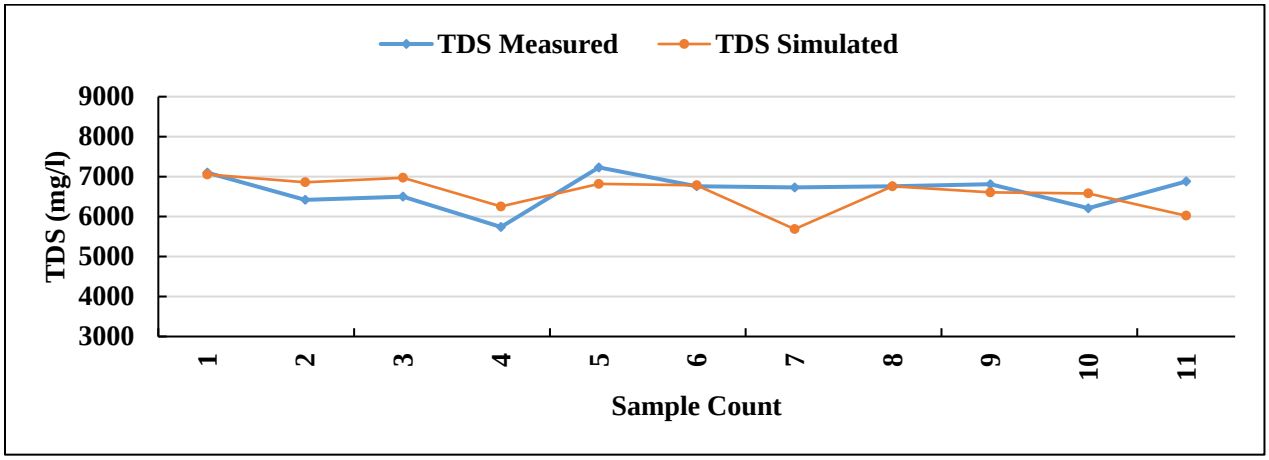

Figure 4-91: Model 8 measured and simulated TDS comparison for testing data set

4.4.9 Simulated Model 9 Network Properties

- Network inputs: pH, EC and [Cl^-]
- Network outputs: TDS and [SO_4^{2-}]
- Network type: Feed-Forward Back-Propagation
- Data Division function: DIVIDERAND
- Training function: TRAINLM
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 20
- Training/Validation/Testing data %split: 60:20:20

Model 9 also ran successfully with no associated errors as seen on Figure 4-92. It recorded a dismally low R performance of 0.76548 as seen on the regression plot shown as Figure 4-93 below. The model does not seem to converge well but it did stop after 3 epochs where the MSE was at this lowest as seen on Figure 4-94. The error distribution is skewed toward the negative end of the histogram, meaning that the model is associated with more cases where it under predicts as seen on Figure 4-95. Such is seen by the shift of the mean error away from the zero mark and toward the negative axis.

The multiple run performance outcomes revealed that the model is not consistent. The R value performance were as low as 0.82, and the worst was demonstrated by the initial once off run already mentioned above. The MSE performance on both training and testing data sets also show a great deal of fluctuations which indicates a great deal of a concern regarding the model's consistency. It can however be alluded that the comparisons done between the actual and simulated outcomes reveal a fair performance of the model as some points seem to compare well as seen on Figure 4-97 - Figure 4-102 below.

It can therefore be said the Model 8 with a split ratio of 90:5:5 has done a lot better than all the models built for the purpose of determining the effect of split ration.

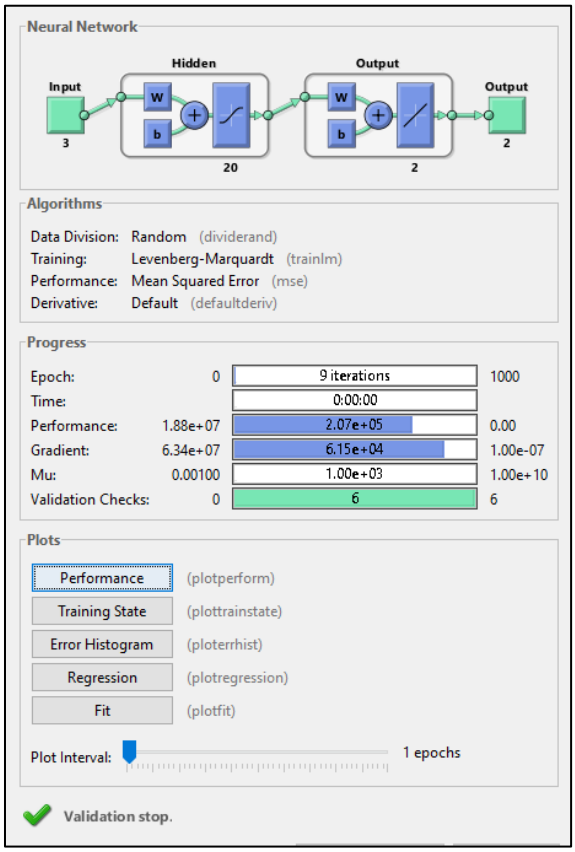

Figure 4-92: Model 9 simulation faceplate during training

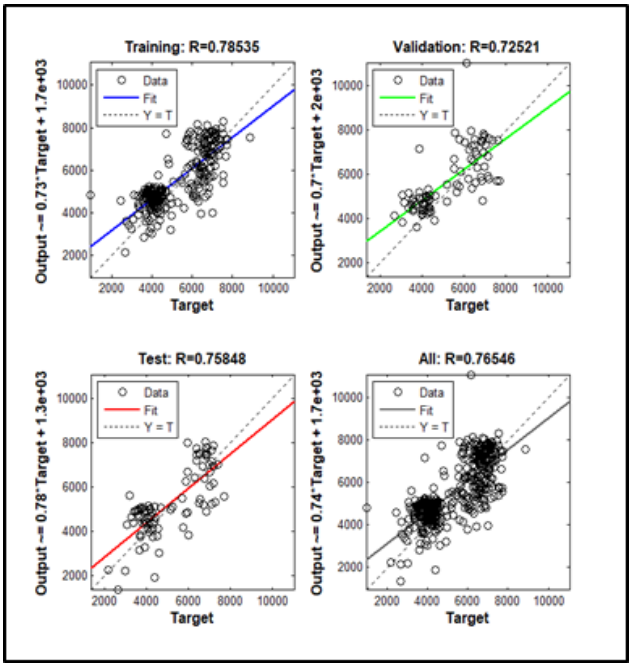

Figure 4-93: Model 9 network regression

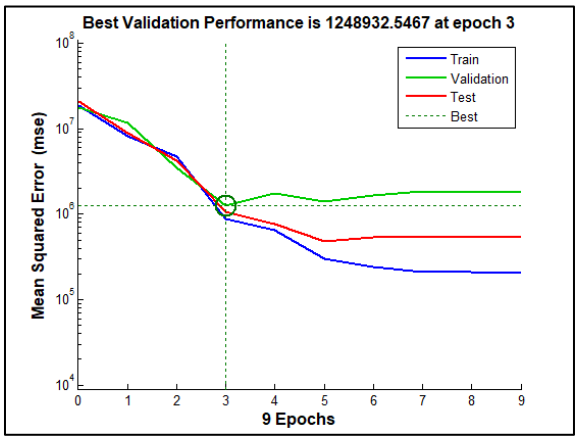

Figure 4-94: Model 9 supervised learning stopping criteria

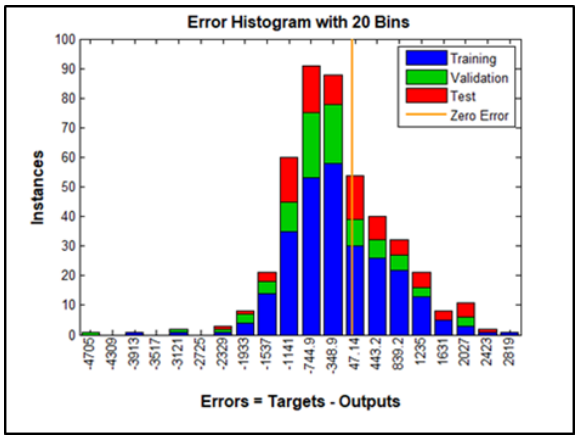

Figure 4-95: Model 9 error histogram

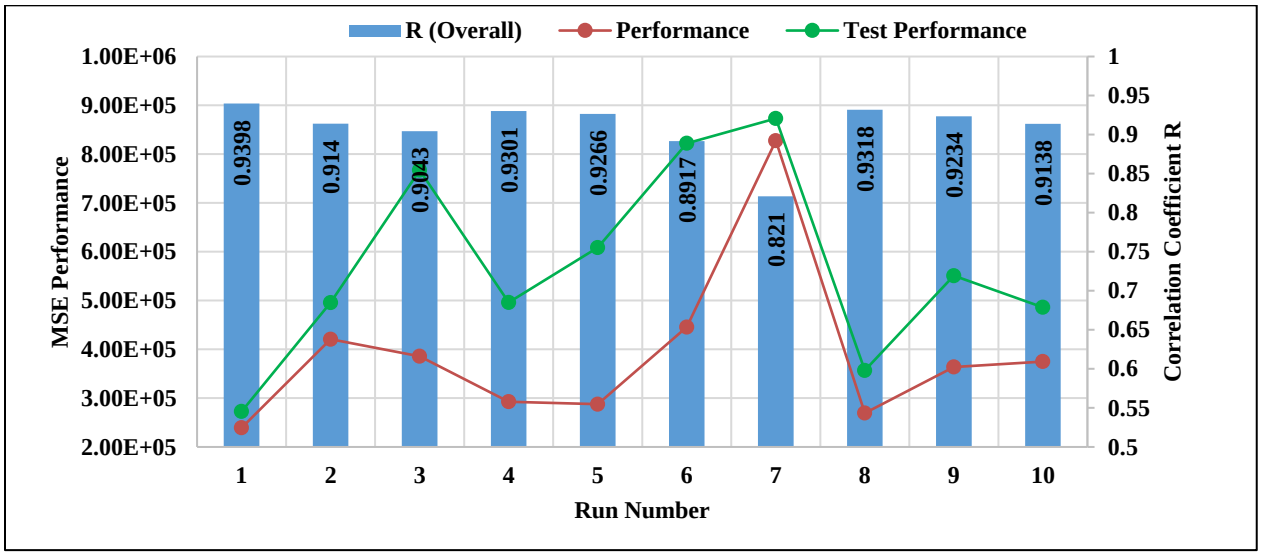

Figure 4-96: Model 9 performance variation and consistency demonstration

4.4.9.1 Model 9 Training Data Simulation

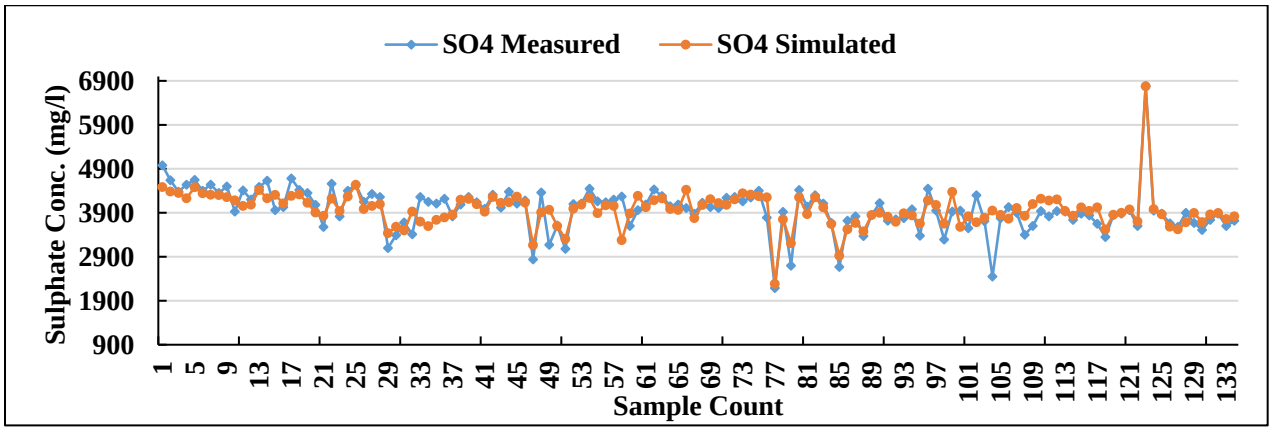

Figure 4-97: Model 8 measured and simulated [SO_4^{2-}] comparison for the training set

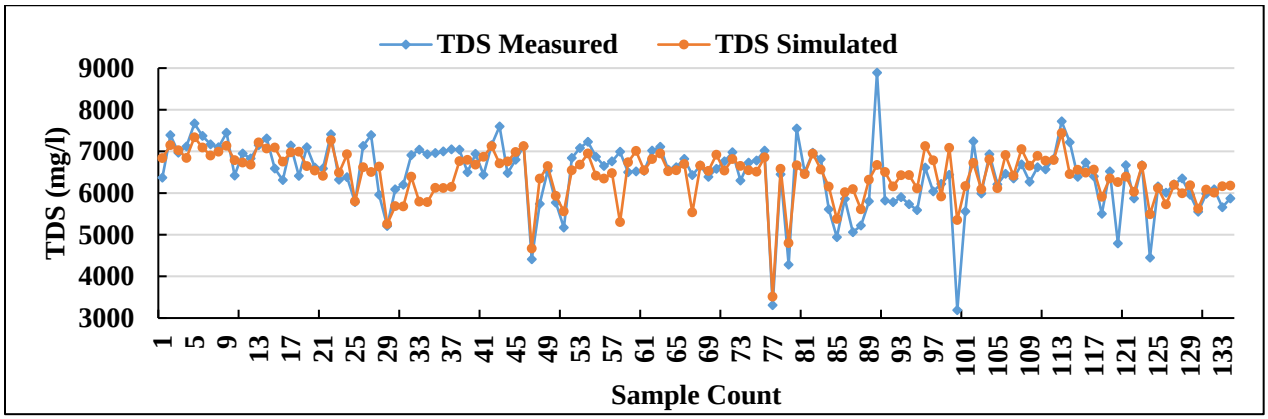

Figure 4-98: Model 9 measured and simulated TDS comparison for the training set

4.4.9.2 Model 9 Validation Data Simulation

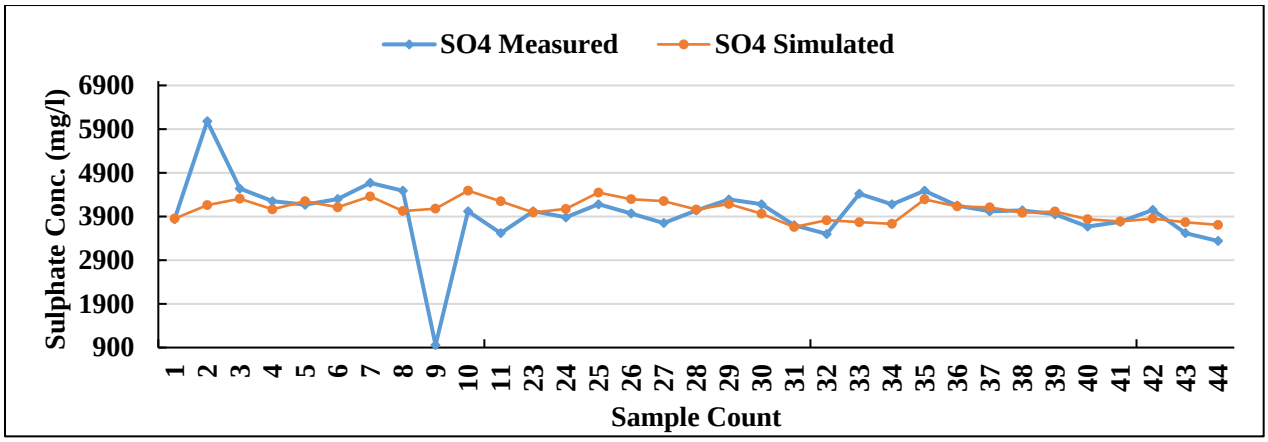

Figure 4-99: Model 9 measured and simulated [SO_4^{2-}] comparison for validation data set

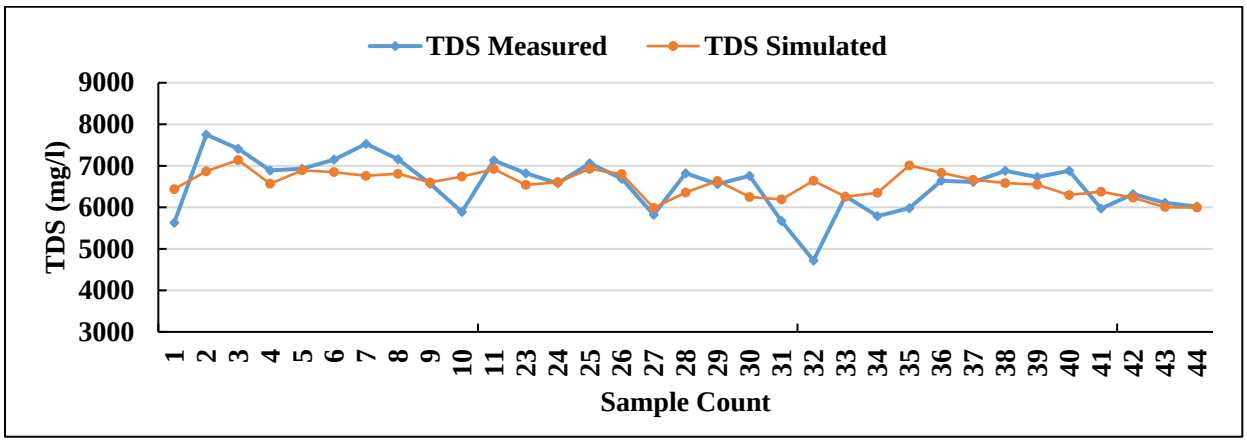

Figure 4-100: Model 9 measured and simulated TDS comparison for validation data set

4.4.9.3 Model 9 Testing Data simulation

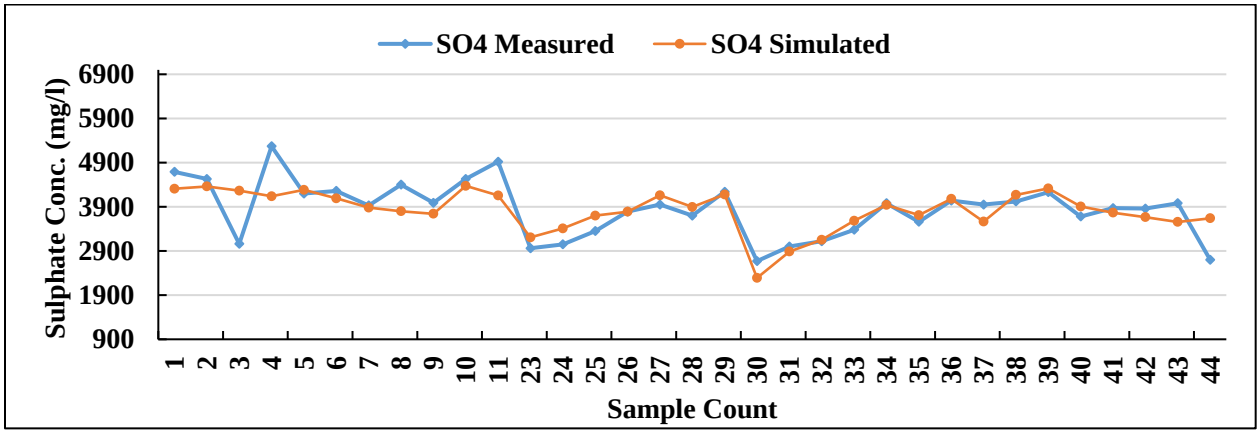

Figure 4-101: Model 9 measured and simulated [SO_4^{2-}] comparison for testing data set

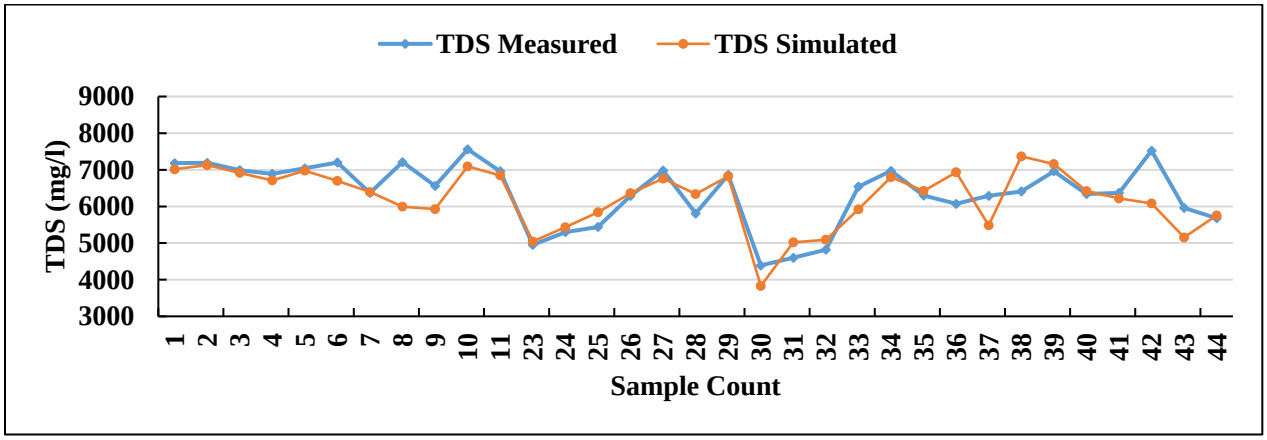

Figure 4-102: Model 9 measured and simulated TDS comparison for testing data set

4.5 Re-scaled Data Testing Results

The parameters of the model regarded as the best model were used to test if the model performance will vary significantly if the input data is to be used. The pH figure remained the same as this is a universal way of presenting it with very limited alternative ways to present them. EC values were converted to S/m from mS/m. Sulphate concentrations, dissolved calcium concentrations, and TDS figures were converted to g/L from mg/L. The significance of this part of the study is to test the reliability of the model when subjected to similar nature of data but given in different scales. There might be standard ways of presenting sample results, but these are not compelling, meaning that lab data can be present in whatever scale deemed fit. The model retained it is robustness in this case as seen on Figure 4-105 below, with R values of over 0.9 which is an indication of a strong correlation.

This section also sought to verify if the high MSE values seen are an indication of model errors, poor performance or both. It can be seen from Figure 4-103 that the MSE obtained when using scaled data which is relatively smaller in magnitudes that the original sample data is small. The MSE came out as 0.229. The MSEs obtained when utilizing the original data as are in the margins of hundreds of thousands, or even millions in some instances. Such observations should not be interpreted as errors, but merely used to compare different models. These figures are this big because the data itself was in the margins of hundreds and thousands. Error margins between actual values and model predicted values, which are above one will result in enormous cumulative square errors according to Equation 2-7 which is used to calculate MSE.

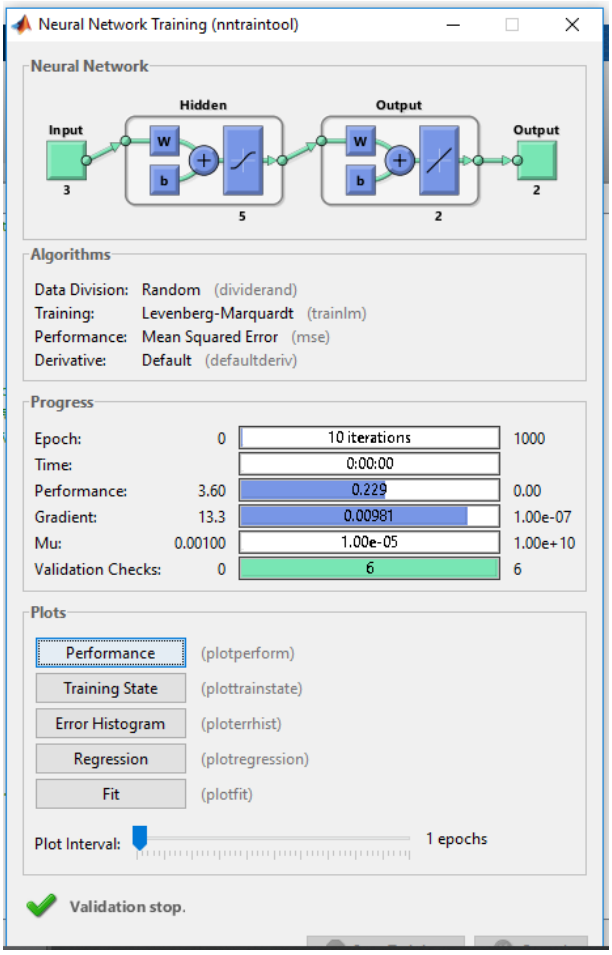

Figure 4-103: Best model network performance result when retested with re-scaled data

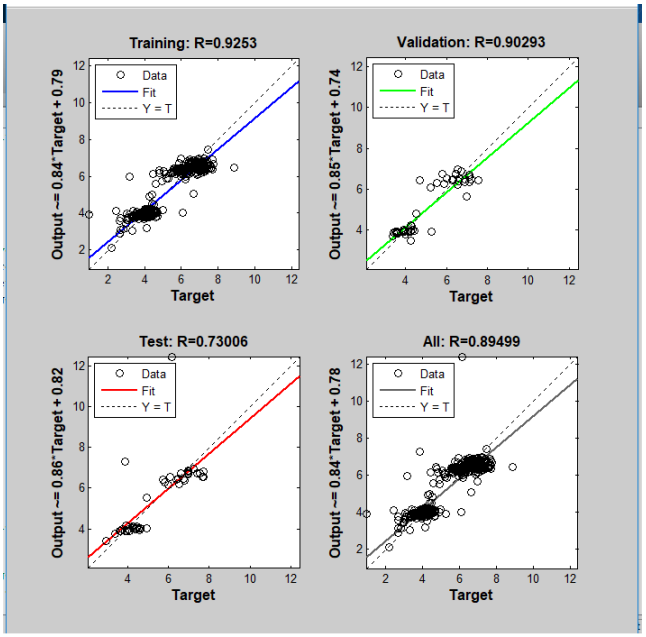

Figure 4-104: Best model network regression result when retested with re-scaled data

4.6 Combined Results Presentation and Discussion

4.6.1 Simulated Models Key Performance Indicators

Each of the 9 models built were tested for their robustness and consistency on their simulation capabilities. Model key performance indicators in this case are both R value and MSE, with the R value giving an indication of how closely correlated the simulated values are to the corresponding measured figures whereas the MSE give an error margin associated with a model's predicting abilities. Higher R values signify better correlations, and lower MSE values signify better performance as well (Schober, et al., 2018). It should however be noted that the MSE value in this case were solely used to compare individual model performance for models developed in this study based only on which one is higher or lower. Their magnitudes were not used in isolation to assess a model's capabilities because unlike R values there is no scale to provide guidelines for when a performance poor or good enough. One would expect an error that approaches zero as an indication of a good MSE performance, however the output measures involved on the data used for building the involved models are in order of thousands. A 0.1% deviation between an actual and a simulated data point may seem reasonably small but results in big MSE due to the nature of the formula given as *Equation 2-8*, hence the magnitudes of MSE seen in this study. Each model was run 10 times to get averaged performance indicators. The results thereof are presented on Table 4-2 below. A decision on the best model for this research was drawn from these outcomes. For a clearer representation of this information, different graphs were plotted.

Table 4-2: Ten runs average performance summary of all 9 models developed

Model number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
Hidden Nodes	5	10	20	50	20	20	20	20	20
Algorithm	Trainlm	Trainlm	Trainlm	Trainlm	Traingd	Trainrp	Trainlm	Trainlm	Trainlm
Data Split	70:15:15	70:15:15	70:15:15	70:15:15	70:15:15	70:15:15	80:10:10	90:5:5	60:20:20
Overall Performance	300055	303958	325480	652477	71706500	473770	242671	269228	390759
Train Performance	279292	262546	226452	230657	71225500	430713	199700	233018	300982
Validation Performance	302872	338166	396626	768964	74446390	507338	436125	440646	482022
Test Performance	395390	465505	722480	2530081	71236900	643758	396891	756210	572914
Average R (Overall)	0.92359	0.92294	0.92084	0.85647	0.04914	0.88564	0.93858	0.93442	0.90965

4.6.2 Selecting the Most Suitable Training Algorithm

Model performance based on training algorithm were compared on Figure 4-105. The Model where trainlm algorithm was used has shown better perform with the lowest overall MSE of 303 958 and the highest overall R value of 0.92294. This further supports the findings that this training algorithms performs better on environmental data (Beale, et al., 2013), hence commonly used in as discussed under literature review. The second-best training algorithm was trainrp which has also shown a good correlation with an R value of 0.88564. The Conjugate Gradient algorithm (traingd) performed dismally in this case. It was used in the study by Sakale, et al., (2019) in an ANN property predicting network that was coupled with LSTM. LSTM overcome the error backflow error problems associated with conjugate gradient RNN (Hochreiter & Schmidhuber, 1997). Had the model in this study incorporated LSTM, the performance would have also been as good. From these observations, it was decided that the model that will be chosen as the best model must make use of trainlm as its training algorithm for modeling AMD water properties.

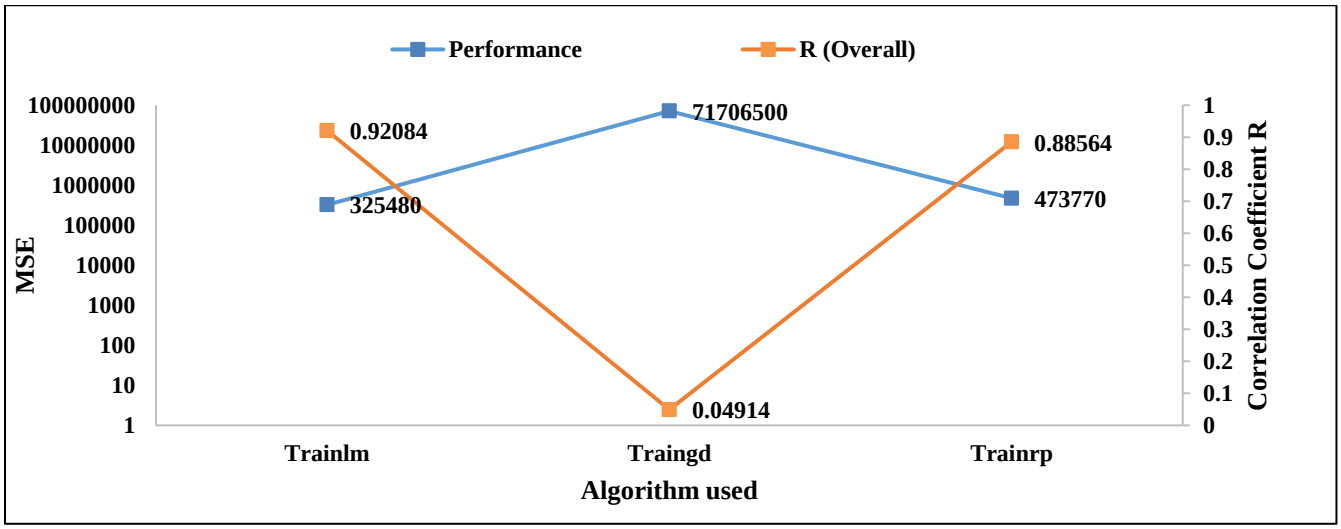

Figure 4-105: 10 runs average performance comparison for the three algorithms tested

4.6.3 Selecting the Most Suitable Number of Hidden Layers

Model performance based on the number of hidden neurons used were compared on Figure 4-106. It can be seen from the figure that there is a very clear relationships between the number of layers and MSE. The same can be said with hidden layers and R. These relationships were defined using

the model generated R and MSE at various hidden layers using the best fit function on Microsoft Excel. It can be seen from this figure that Both R and MSE performance increases and then suddenly drop exponentially as the number of hidden neuron increase. The most suitable number of hidden neurons could then be determined by a definite integral of the derived equations given by Equation 4-1 and Equation 4-2 as shown below, where the best suitable number of hidden neuron is seen where R is at its maximum, and where MSE is at its minimum. From the R and Hidden neuron relationship (Equation 4-1), the optimal number of hidden neurons was found to be 11.25. The optimal number of hidden neurons based on the MSE equation (Equation 4-2) was found to be 8.44. From these derivations and simple analogy, it can be deduced that the best theoretical number of hidden neurons lies between the two values determined. To simplify the process, a straight average of these two figures were taken, and it amounted to 9.845. Since the number of hidden neurons can only be a whole number, the value was rounded off to the nearest whole number. The most suitable theoretical number of hidden layers was then found 10.

The model where 10 hidden neurons were applied had an average overall R and MSE of 0.92084 and 325 480, respectively. It is also interesting to note from Table 4-2 that Model 1 with 5 hidden neurons perform better than Model 2 with 10 hidden neurons, where model 1 overall R and MSE averaged at 0.92359 and 300055, respectively. This then called for an in-depth head-to-head model comparison to establish conclusive grounds. From the individual model evaluations done in 4.4.1 and 4.4.2, particularly on
Figure 4-8 and Figure 4-19, it can be seen that the model with 5 hidden neurons is more consistent in its predicting abilities than the one with 10. If the total number of runs from which the averages were taken was increased to a much higher value, the exponential functions seen on Figure 4-106 would have moved slightly to the left in favor of a smaller number of hidden neurons. From this analogy it was concluded that the model with a smaller number of hidden neurons, 5 hidden neurons in this particular case performs better than the rest of the models tested for the purposed of determining the optimal number of hidden neurons.

In the study done by Nasr, et al., (2012), three hidden layers were used, and this is close enough to the one this study has found to be the most suitable. The study done by Ma, et al., (2021) showed that the best performing model had 20 hidden neurons, a value slightly higher than what this

current study suggests and achieve a lower but good correlation coefficient of 0.84 as seen on Table 4-3 below. It can be seen from that table that lower numbers of hidden layers are preferrable by simply looking at the number of hidden neurons used in studies that had a higher R value than that of the current study.

Optimal number of hidden neurons based on R:

$$y = -0.00004x^2 + 0.0009x + 0.9194 \qquad \text{Equation 4-1}$$

$$\frac{dy}{dx} = -0.00008x + 0.0009$$
$$0.00008x = 0.0009$$
$$x = 11.25$$

Optimal number of hidden neurons based on MSE:

$$y = 204.05x^2 - 3442.4x + 314328 \qquad \text{Equation 4-2}$$

$$\frac{dy}{dx} = 408.1x - 3442.4$$
$$408.1x = 3442.4$$
$$x = 8.44$$

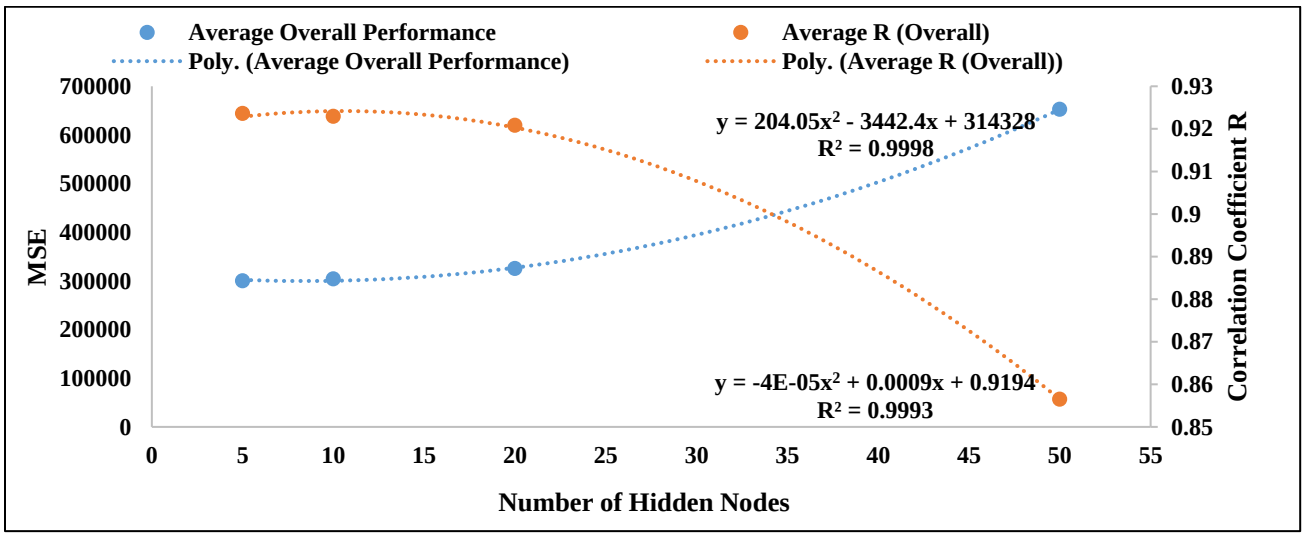

Figure 4-106: 10 runs average performance comparison for the four sets of hidden neurons tested

4.6.4 Selecting the Most Suitable Data Split Ratio

Model performance based on data split ratio were compared on Figure 4-107. The Model where a split ratio of 80:10:10 was used has shown better perform with the lowest overall MSE of 242 671 and the highest overall R value of 0.93858. This further supports the theory a split ratio where 70 % of the data or more is good for training an ML tool (Rácz, et al., 2021). More training data in this case enables the models to be better knowledgeable about the data as it learns numerous combinations than it would with lesser data. This is also seen from performance of the models with 70:15:15 (Model 3) and 90:5:5(Model 8) data split ratio which was fairly comparable to the best performing ratio and better than the model 9 which was also built for the purpose of determining a suitable data split ratio. In the study by Nasr, et al., (2012), a split ratio of 60:20:20 was used. The R value for that study is still good but lower compared to that of the current study.

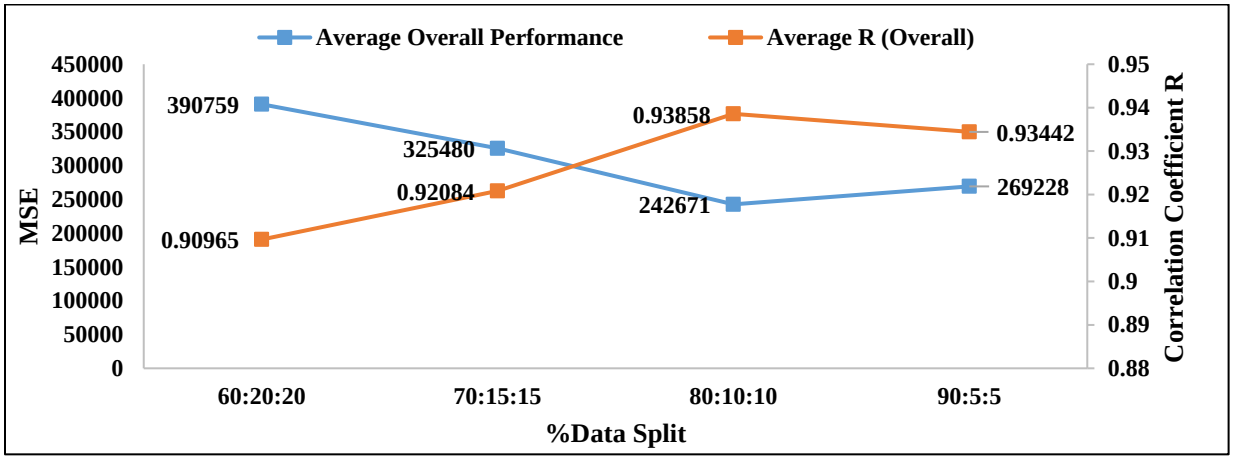

Figure 4-107: 10 runs average performance comparison for the four data split ratios tested

4.6.5 Best Model Summary

Model 1 – 9 built on this research were used to analyse different sets of conditions under which ANN can be used to simulated outcome given AMD water properties as their inputs. The analysis of different trends generated and the application of theory as well as other findings from literature have shown that the best model property combinations are as follows:

- Network inputs: pH, EC and [Cl^-]
- Network outputs: TDS and [SO_4^{2-}]

- Network type: Feed-Forward Back-Propagation
- Data Division function: DIVIDERAND
- Training function: TRAINLM
- Adaption learning function: LEARNGDM
- Performance function: MSE
- Number of hidden layers: 5
- Training/Validation/Testing data %split: 80:10:10

This is not any of the nine models tested but a remake from the conditions that were seen as the most suitable based on the analysis done. It has a 10 arbitrary run average R value of 0.93916 which is a good enough indication of simulated and actual data combination. Closest to the best model is Model 7 with R value of 93858. Model 8 properties only differ from those of the best model with the number of hidden layers.

Table 4-3: Current study best performing model performance comparison with previous study model performances

Reference	Application	Network type	Training Function	No. hidden Layers	Data Split Ratio	Correlation Coefficient (R)	MSE
Current Study Best Model	AMD water Property Predictions	Feed-Forward Back-Propagation	Levenberg-Marquardt	5	80:10:10	0.93916	0.229
Nasr, et al., (2012)	Wastewater treatment plant performance prediction	Feed-Forward Back-Propagation	Levenberg-Marquardt	3	60:20:20	0.90317	15
Ma, et al., (2021)	Prediction of properties of drainage water from waste rock storage	Feed-Forward Back-Propagation	Levenberg-Marquardt	20	80:20:00	0.84	0.28
Sakala, et al., (2019)	Predicting water quality from an abandoned underground coal	LTSM and Back-propagation algorithm	Back-propagation algorithm/Conjugate Gradients	1		0.99	0.054
Setshedi, et al., (2021)	Prediction of Physicochemical Characteristics of Water Quality in Three District Municipalities	Feed-Forward Back-Propagation	Broyden–Fletcher–Goldfarb–Shanno	3	80:10:10	0.994677	39.03
Solaimany-Aminabad, et al., (2014)	prediction of water treatment plant influent characteristics	Feed-Forward Back-Propagation	Levenberg-Marquardt	1	70:10:20	0.93	28.41

4.6.6 Short-Comings of the Models Developed

The models may have been proven useful in several ways but are also linked with some disadvantages. The following were identified as disadvantages associated with the models developed on this research:

- The tested inputs and output combination may not need be needed for some operations. It has however been shown that there exists some form of correlation between the properties considered in this research. For this reason, the input-output may be reworked from to operation to operation, based on what is readily available to that operation and what the water properties of interest are. The model can still be used but it is uncertain how well the model will perform because existing correlations may or may not be as good as they are on the tested combinations.
- Some easy to measure variable may not be available on days when weather conditions make it impossible to access sampling points. The model will not be useful for those operation that require real-time input data under these conditions.

CHAPTER 5. CONCLUSION

This research work has studied the application of artificial neural networks as an alternative for Witwatersrand Western Basin Acid Mine Drainage characteristics determining tool. The current methods used to get this information is through manual measurements which is time consuming, costly and not useful in forecasting works which may deem necessary for future projects as well as AMD treatment reagent requirements. ANN models were developed for use in simulating AMD water Sulphate ion and Total Dissolved Solids concentrations. The former gives negative impact on aqueous and terrestrial environments, and human and animal health, and also facilitates the formation of salts. The later indicates the level of toxicity and it is closely regulated in many countries including South Africa. The models were developed bearing in mind that all the input data (pH, Electrical Conductivity and Chlorine ion concentration), which are easy to measure variables, should be measured or be made available for it to be used for to simulate what the outputs are, whether real time or for seasonal forecasting. These were chosen because they are easier to measure.

The studied models in this research work belong to the feedforward back-propagation neural network structure. Nine ANN models were developed with the purposed of working out the best model properties which would be suitable for simulating AMD water properties to a satisfactory level. The simulation results obtained for 8 of these models including the best model proved that satisfactory performance has been achieved. The best network model developed uses Levenberg-Marquardt training function. It uses the MATLAB built-in *nnstats* function is an execution platform. It has 5 hidden layers and uses a split ration of 80:10:10. Its performance indicators are R and MSE. The rest of the unspecified parameters are default functions of the *nnstats* model.

The best performing model achieved very good R value of 0.93916 and an MSE performance of 0.229 when input data is scaled by a factor of 10^{-3}. It should be noted that the models remain equally reliable irrespective of the scales of the input data. Only consistency in the use of units of choice is important when using the model. The model performance is highly comparable to the performance of ANN used in other wastewater property prediction tools. The unique thing about

the model developed in this study is that is not as complex as those developed in other studies, and easy to apply.

The model described here can be relied on in predicting sulphate ion concentration and total dissolved solids in the Witwatersrand Wester Basin AMD water with a great sense of certainty. These predictions can be applied directly to the existing treatment processes to provide guidance on the quantities of reactants to use at t given point in time.

CHAPTER 6. RECOMMENDATIONS

It is recommended that future works focus on exploring the feasibility of other input-output combinations. This may come in handy depending on current process requirements, as well as future AMD water treatment process design modeling.

Looking at how this model has performed, tests to determine if the Eastern Basin and coal-based AMD water properties can be applied to this model and still achieve satisfactory results is recommended. This can be achieved by testing just about enough sample data collected from those respective area to confirm its applicability.

Continuous monitoring can of streams can also be made possible through the integration of field monitoring devices that will provide continuous model input data and neurocomputers. Such integrations will provide live AMD water properties which can be used to optimize the existing and future treatment plants.

CHAPTER 7. INTELLECTUAL PROPERTY

The simulation work done on during this research was carried out on MATLAB 2016 version.

CHAPTER 8. ETHICS

The research to be undertaken does not involve animal experimentation or human participants as defined in A.4 of the Senate Standing Orders on Higher degrees.

REFERENCES

Abbasi, M. & El-Hanandeh, A., 2016. Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Management,* Volume 56, pp. 13-22.

Ayodele, O. T., 2010. Types of Machine Learning Algorithms. In: Z. Yagang, ed. *New Advances in Machine Learning.* Rijeka, Croatia: InTech, pp. 19 - 48.

Ba, Ł., Szela, B., Sałata, A. & Studzin, J., 2019. Modeling of HeavyMetal (Ni,Mn, Co, Zn, Cu, Pb, and Fe) and PAH Content in Stormwater Sediments Based onWeather and Physico-Geographical Characteristics of the Catchment-Data-Mining Approachof the Catchment-Data-Mining Approach. *Water,* 11(626), p. 7.

Basheer, I. A. & Hajmeer, M., 2000. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods,* Volume 43, pp. 3-31.

Beale, M. H., Hagan, M. T. & Demuth, H. B., 2013. *Neural network toolbox-user's guide.* Natick, MA, USA: The MathWorks, Inc.

Beavon, K. S. O., 2004. *Johannesburg : the making and shaping of the city.* 1st ed. Pretoria: University of South Africa Press.

Bejan, D. & Bunce, N. J., 2015. Acid mine drainage: electrochemical approaches to prevention and remediation of acidity and toxic metals. *Journal of Applied Electrochemistry,* 45(12), pp. 1239-1254.

Blaus, B., 2013. *Blausen 0657 MultipolarNeuron.png.* [Art] (Creative Commons 3.0).

Bobbins, K., 2015. *Acid mine Drainage and its Governance in the Gauteng City-Region,* Jahannesburg: Gauteng City-Region Observatory.

Cheng, S., Dempsey, B. A. & Logan, B. E., 2007. Electricity Generation from Synthetic Acid-Mine Drainage (AMD) Water Using Fuel cell Technologies. *Environmental Science & Technology,* 41(23), p. 8149–8153.

Coetzee, H. et al., 2010. Mine water management in the Witwatersrand Gold Fields with special emphasis on acid mine drainage. *Report to the inter-ministerial committee on acid mine drainage,* pp. 1-128.

Cybenkot, G., 1989. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems,* Volume 2, pp. 303-314.

Department of Water and Sanitation, A. S., 2013. *Revision of General Authorisation in terms of Section 39 of the National Water Act, 1998 (Act No. 36 of 1998). Government Gazette No.36820.* Pretoria, Government Printer.

Derry, A., Prepas, E. & Hebert, P., 2003. A comparison of zooplankton communities in saline lakewater with variable anion composition. *Hydrobiologia,* 505(1), pp. 199-215.

Deshpande, S., 2021. *Learning Curve Graphs Part 1: Countering the Data Requirement Curse.* [Online]
Available at: https://towardsdatascience.com/learning-curve-graphs-part-1-countering-the-data-requirement-curse-6bdeb7750edf
[Accessed 8 October 2021].

du-Toit, D., 2018. *www.wits.ac.za.* [Online]
Available at: https://www.wits.ac.za/news/latest-news/research-news/2018/2018-05/the-heat-of-acid-mine-drainage.html
[Accessed 17 October 2018].

Fausett, L. V., 1994. *Fundamentals of Neural Networks: Architectures, Algorithms And Applications.* 1st ed. Englewood Cliffs, NJ: Prentice-Hall.

Frey, P. & Reed, G., 2012. The Ubiquity of Iron. *ACS Chem. Biol.,* 7(9), pp. 1477-1481.

Grewar, 2019. South Africa's Options for mine-impacted water re-use: A review. *The Journal of the Southern African Institute of Mining and Metallurgy,* Volume 119, pp. 321-331.

Grosan, C. & Abraham, A., 2011. Artificial Neural Networks. In: *Intelligent Systems.* Berlin, Heidelberg: Springer, pp. 281-323.

Gupta, R., Singh, A. N. & Singhal, A., 2019. Application of ANN for Water Quality Index. *International Journal of Machine Learning and Computing,* 9(5), pp. 688 - 693.

Haung, T.-J., 2017. Imitating the Brain with Neurocomputer A “New” Way Towards Artificial General Intelligence. *International Journal of Automation and Computing,* 14(5), pp. 520-531.

Haykin, S., 1994. *Neural Networks: A comprehensive foundation.* New York: Prentice Hall PTR.

Helsel, D. R. et al., 2002. *Statistical Methods in Water Resources.* 1.1 ed. Reston, Virginia: Elsevier.

Hobbs, P. & Cobbing, J., 2007. *A hydrogeological assessment of acid mine drainage impacts in the West Rand Basin, Gauteng Province,* Pretoria: Council for Scientific and Industrial Research.

Hobbs, P. & Cobbing, J., 2007. *A Hydrogeological Assessment of Acid Mine Drainage Impacts in the West Rand Basin, Gauteng Province,* Pretoria: CSIR/THRIP.

Hobbs, PJ & Cobbing, J., 2007. *The hydrogeology of the Krugersdorp Game Reserve area and implications for the management of mine water decant,* Pretoria: Council for Scientific and Industrial Research.

Hochreiter, S. & Schmidhuber, J., 1997. Long Short-term Memory. *Neural Computation,* 9(8), pp. 1735-1780.

Huang, W. & Foo, S., 2002. River, Neural network modeling of salinity variation in Apalachicola. *Water Research,* 36(1), pp. 356 - 362.

Jaing, A., Mao, J. & Mohiuddin, K., 1996. Artificial Neural Networks: A Tutorial. *IEEE Computer,* Volume 29, pp. 31-144.

Jiang, J., Trundle, P. & Ren, J., 2010. Medical Imaging Analysis with Artificial Neural Networks. *Computerized Medical Imaging and Graphics,* 34(8), pp. 617-631.

Karaca, F. & Özkaya, B., 2006. NN-LEAP: A neural network-based model. *Environmental Modelling and Software,* Volume 21, pp. 1190-1197.

Kirby, D., 2014. *Effective Treatment Options for Acid Mine Drainage in the Coal Region of West Virginia,* WEST VIRGINIA: Marshall Digital Scholar.

Lottermoser, B., 2010. *Mine Waste: Characterisation, Treatment and Environmental Impacts.* 3rd ed. Townsville: Springer Science & Business Media.

Lowe, C. C. et al., 2020. *Encyclopædia Britannica.* [Online]
Available at: https://www.britannica.com/place/South-Africa
[Accessed 08 November 2020].

Maier, H. & Dandy, G., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software,* Volume 15, pp. 101-124.

Ma, L., Huang, C. & Liu, Z.-S., 2021. The Application of Artificial Neural Network to Predicting the Drainage from Waste Rock Storages. In: *Artificial Neural Networks and Deep Learning - Applications and Perspective.* Canada: Intechopen, pp. 1-13.

Maluleke, R., 2021. *Electricity, gas and water supply industry, 2019,* Pretoria: Statistics South Africa.

Manders, P., Godfrey, L. & Hobbs, P., 2009. *Acid Mine Drainage in South Africa*, PRETORIA: CSIR.

Mavani, j., 2014. *Artificial Neural Network based nowcasting model for beach water quality,* Toronto, Ontario, Canada: Ryerson University.

McCarthy, T., 2011. The Impact of Acid Mine Drainage in South Africa. *South African Journal of Science,* 107(5/6), pp. 01 -07.

McCarthy, T. S., 2010. *Decanting of acid mine water in the Gauteng City-Region - Analysis, Prognosis and Solutions,* Johannesburg: GCRO.

Merz, S. K., 2013. *Characterising the relationship between water quality and water quantity.* [Online]

Available at: https://www.waterquality.gov.au/sites/default/files/documents/characterising.pdf [Accessed 08 October 2021].

Mihir, P., Nihar, R. S., Pankaj, K. R. & Malabika, B. R., 2015. Electrical Conductivity of Lake Water as Environmental Monitoring – A Case Study of Rudrasagar Lake. *IOSR Journal of Environmental Science, Toxicology and Food Technology,* 9(3), pp. 66-71.

Mijwel, M. M., 2018. *Artificial Neural Networks Advantages and Disadvantages.* [Online] Available at: https://www.researchgate.net/publication/323665827_Artificial_Neural_Networks_Advantages_and_Disadvantages [Accessed 7 November 2020].

Moreno, P., De Ingeniería, F. & Portales, U., 2009. *Environmental Impact and Toxicology of Sulphate. In Proceedings of the EnviroMine 2009, 30 September–2 October 2009.* Santiago, s.n.

Nasr, M., Moustafa, M., Seif, H. & El Kobrosy, G., 2011. Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT. *Alexandria Engineering Journal,* Volume 53, p. 37–43.

Nleya, Y., Simate, G. S. & Ndlovu, S., 2016. Sustainability assessment of the recovery and utilisation of acid from acid mine drainage. *Journal of Cleaner Production,* Volume 113, pp. 17-27.

Noori, R., Karbassi, A. & Sabahi, M., 2010. Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste prediction. *Journal of Environmental Management,* Volume 91, pp. 767-771.

Ozveren, U., 2016. An artificial intelligence approach to predict a lower heating value of municipal solid waste. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects,* Volume 38, p. 2906–2913.

Rácz, A., Bajusz, D. & Héberger, K., 2021. Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification. *Molecules,* Volume 29.

Rebala, G., Ravi, A. & Churiwala, S., 2019. *Introduction to Machine Learning.* Cham: Springer.

Roohi, R., Jafari, M. & Jahantab, E., 2020. Application of artificial neural network model for the identification the effect of municipal waste compost and biochar on phytoremediation of contaminated soils. *Journal of Geochemical Exploration,* Volume 208.

Sakala, E., Fourie, F., Gomo, M. & Coetzee, H., 2017. Mapping surface sources of acid mine drainage using remote sensing: case study of the Witbank, Ermelo and Highveld coalfield. *Mine water and Circular Economy,* pp. 1246-1253.

Sakala, E., Novhe, O. & Vadapalli, V. R. K., 2019. *Application of Artificial Intelligence (AI) to predict mine water quality, a case study in South Africa.* Pretoria, s.n., pp. 140-145.

Schober, P., Boer, C. & Schwarte, L., 2018. Correlation Coefficients: Appropriate Use and Correlation Coefficients: Appropriate Use and. *Special Article,* 126(5), pp. 1763-1768.

Setshedi, K. J., Mutingwende, N. & Ngqwala, N. P., 2021. The Use of Artificial Neural Networks to Predict the Physicochemical Characteristics of Water Quality in Three District Municipalities, Eastern Cape Province, South Africa. *Int J Environ Res Public Health,* 18(10).

Singh, D. & Satija, A., 2018. Prediction of municipal solid waste generation. *for optimum planning and management with artificial neural network - case study: Faridabad City in Haryana State (India),* Volume 9, pp. 91-97.

Skousen, J., Sexstone, A. & Ziemkiewicz, P., 2000. Acid Mine Drainage Control and Treatment. In: *Reclamation of Drastically Disturbed Lands* . West Virginia: West Virginia University and the National Mine Land Reclamation Center.

Solaimany-Aminabad, M., Maleki, A. & Hadi, M., 2014. Application of artificial neural network (ANN) for the prediction of water treatment plant influent characteristics. *Journal of Advances in Environmental Health Research,* 1(2).

Sonali, M. & Maind, B., 2014. Research Paper on Basic of Artificial Neural Network. *International Journal on Recent and Innovation Trends in Computing and Communication,* 2(1), p. 96 – 100.

Sreekanth, P. D., 2009. Forecasting groundwater level using artificial neural networks. *Current Science Association,* 96(7), pp. 933-939.

Thanh, H. V., Sugai, Y. & Sasaki, K., 2020. Application of artificial neural network for predicting the performance of CO2 enhanced oil recovery and storage in residual oil zones. *Scientific Reports,* Volume 10.

Tucker, R. F., Viljoen, R. P. & Viljoen, M. J., 2016. A Review of the Witwatersrand Basin - The World's Greatest Goldfield. *Jounal of International Geoscience,* 1 June, pp. 104-133.

Tumer, A. E. & Edebali, S., 2015. An Artificial Neural Network Model for Wastewater Treatment Plant of Konya. *International Journal of Intelligent Systems and Applications in Engineering,* 3(4), pp. 131-135.

Vogel, F., 2012. *Feasibility Study for a Long-Term Solution to address the Acid Mine Drainage Associated with the East, Central and West Rand Underground Mining Basins. Study Report No. 1: Inception Report,* Pretoria, South Africa: The Department of Water Affairs.

Vorster, C., 2000. *Gold Deposits of the Witwatersrand.* [Online]
Available at: https://www.geoscience.org.za/images/Maps/golddeposits.gif
[Accessed 10 November 2020].

Weber-Scannell, P. K. & Duffy, L. K., 2007. Effects of Total Dissolved Solids on Aquatic Organisms: A Review of Literature and Recommendation for Salmonid Species. *American Journal of Environmental Sciences,* 3(1), pp. 1-6.

Younger, P. L., Banwart, S. A. & & Hedin, R. S., 2002. *Mine water: hydrology, pollution, remediation.* 18th Edition ed. Dordrecht: Kluwer Academic Publishers.

Zhang, G., Patuwo, B. E. & Hu, M. Y., 1998. Forecasting with artificial neural networks:The state of the art. *International Journal of Forecasting*, 14(1), pp. 35-62.

APPENDICES

Appendix A

Primary drainage region	Secondary/Tertiary/Quaternary drainage region and excluded resources	Description of main river in drainage region for information purposes
A	All catchments	Limpopo River
B	All catchments	Olifants River
C	C10, C20, C40, C50, C60, C70, C80 & C90	Vaal River
D	Orange River downstream of Gariep Dam	
	D13	Kraai River
E	Olifants River	
	E10A to K	Olifants River above the confluence with the Doring River
	E21	Groot River
G	G10	Berg River
	G21	Diep River
	G22	Eerste River
	G30	Verlorevlei River
	G40A to E	Bot River
	G40L & M	Klein River, Uilkraals River
	G70H	Onrus River
H	H20A, B, D, E, F, G & H	Hex River
	H30	Kingna River
	H40	Breede River downstream of Brandvlei Dam to confluence with the Kingna River
	H60A, B & C	Tributaries of the Sonderend River upstream of the Theewaterskloof Dam
	H80A to E	Duivenhoks River
J	J12	Touws River
	J25	Gamka River
	J31 to 35	Olifants River
	J40C	Langtou and Weyers Rivers
K	K10	Little Brak River
	K20	Great Brak River
	K40C	Karatara River
	K50 & K60	Knysna, Keurbooms Rivers

Primary drainage region	Secondary/Tertiary/Quaternary drainage region and excluded resources	Description of main river in drainage region for information purposes
	K70B	Bloukrans River
	K80A to F	Lottering, Storms, Sandrif, Groot, Tsitsikamma, Klippedrift Rivers
	K90A to G	*Kromme, Seekoei, Kabeljous Rivers*
L	L81	Baviaanskloof River
	L82	Kouga River
	L90	Lower Gamtoos River
M	M10	Swartkops River
	M20	Maitland River
	M30	Van Stadens River
N	Sundays River downstream of Darlington Dam	
	Skoenmakers River downstream of Skoenmakers Canal Outlet	
	N11, N12	Sundays River upstream of Vanrynevelds Pass Dam
	N12A, B & C	Gats River
P	P10	Bushmans River
	P30	Kowie River
	P40	Kariega River
Q	Great Fish River, Little Fish River	
	Q41A, Q41B, Q41C, Q41D, Q44A, Q44B	Tarka River
	Q42A & B	Elands River
	Q43A & B	Vlekpoort River
	Q44A & B	Lake Authur
R	R20	Buffels River
	R30A, B, C & D	*Kwenxura, Kwelera, Gonubie Rivers*
	R30E & F	Nahoon River
S	S10 A to E	White Kei River upstream of the Xonxa Dam
	S20A to C	Indwe River upstream of the Lubisi Dam
	S32A to C	Swart Kei River upstream of the Klipplaat confluence
	S32D & E	Klipplaat River upstream of Waterdown Dam
	S32F	Oxkraal River upstream of the Oxkraal Dam
	S40A, B & C	Thorn, Thomas Rivers
	S50A, B & C	Tsomo, Kwa-Qokwama and Mbokotwa Rivers
	S60A & B	Kubusi River upstream of Wriggleswade Dam
	S60C & D	Toise River
	S70C	Xilinxa River upstream of the Xilinxa Dam
T	T11A & B	Slang, Xuka Rivers
	T20A & B	Mtata River upstream of Mtata Dam
	T35A, B, C, D, F & G	Tsitsa, Pot, Mooi, Inxu, Wildebees, Gatberg Rivers
U	U20 & U40	Mgeni, Mvoti Rivers
V	V11	Upper Thukela River
	V20	Mooi River
	V60	Sundays River
	V70	Bushmans River
W	W12	Mhlatuze River
	W21A	White Mfolozi River upstream of Klipfontein Dam
	W30	Hluhluwe and Mkuzi Rivers
X	All catchments	Nkomati River

TABLE 1.2 Subterranean government water control areas excluded from General Authorisation for groundwater abstraction

Primary drainage region	Tertiary/ Quaternary drainage region	Description of subterranean government water control area	Government Notice No.	Government Gazette Date
H	H30	Baden	136	1967-06-16
A	A30	Bo-Molopo	1324	1963-08-30
C	C30	Bo-Molopo	1993	1965-12-17
D	D41	Bo-Molopo	R634	1966-04-29
A	A24	Crocodile River Valley	208	1981-10-23
A	A21	Crocodile River Valley	18	1983-02-18
A	A21, A22	Kroondal-Marikana	180	1963-06-17
G	G10,G30	Lower Berg River Valley/Saldanha	185	1976-09-10
A,B	A60,B50,B31	Nyl River Valley	56	1971-03-26
G	G30	Strandfontein	2463	1988-12-09

Primary drainage region	Tertiary/ Quaternary drainage region	Description of subterranean government water control area	Government Notice No.	Government Gazette Date
M	M10,M20,M30	Uitenhage	260	1957-08-23
G	G30	Wadrif	992	1990-05-11
G	G20	Yzerfontein	27	1990-02-09
G	G30	Graafwater	1423	1990-06-29
A	A70	Dendron-Vivo	813	1994-04-29
A	A60	Dorpsrivier	312	1990-02-16
C	C24	Ventersdorp	777	1995-06-02

TABLE 1.3 Areas excluded from General Authorisation for storage of water

Primary drainage region	Secondary/Tertiary/Quaternary drainage region	Description of main river in drainage region for information purposes
X	X11, X12	Komati River Catchment upstream of Swaziland
	X21A, B, C	Crocodile River Catchment upstream of Kwena Dam
	X21F,G	Elands River Catchment upstream of Waterval Onder
B	B1	Olifants and Klein-Olifants River
	B2	Wilge River
	B3	Elands River
	B4	Steelpoort River

Appendix B

DEPARTMENT OF WATER AFFAIRS – GENERAL AND SPECIAL AUTHORISATION

Discharge limits and conditions set out in the National Water Act, Government Gazette No. 20526, 8 October 1999

Wastewater limit values applicable to discharge of wastewater into a water resource

SUBSTANCE/PARAMETER	GENERAL LIMIT	SPECIAL LIMIT
Faecal Coliforms (per 100 ml)	1 000	0
Chemical Oxygen Demand (mg/l)	75*	30*
pH	5,5-9,5	5,5-7,5
Ammonia (ionised and un-ionised) as Nitrogen (mg/l)	3	2
Nitrate/Nitrite as Nitrogen (mg/l)	15	1,5
Chlorine as Free Chlorine (mg/l)	0,25	0
Suspended Solids (mg/l)	25	10
Electrical Conductivity (mS/m)	70 mS/m above intake to a maximum of 150 mS/m	50 mS/m above background receiving water, to a maximum of 100 mS/m
Ortho-Phosphate as phosphorous (mg/l)	10	1 (median) and 2,5 (maximum)
Fluoride (mg/l)	1	1
Soap, oil or grease (mg/l)	2,5	0
Dissolved Arsenic (mg/l)	0,02	0,01
Dissolved Cadmium (mg/l)	0,005	0,001
Dissolved Chromium (VI) (mg/l)	0,05	0,02
Dissolved Copper (mg/l)	0,01	0,002
Dissolved Cyanide (mg/l)	0,02	0,01
Dissolved Iron (mg/l)	0,3	0,3
Dissolved Lead (mg/l)	0,01	0,006
Dissolved Manganese (mg/l)	0,1	0,1
Mercury and its compounds (mg/l)	0,005	0,001
Dissolved Selenium (mg/l)	0,02	0,02
Dissolved Zinc (mg/l)	0,1	0,04
Boron (mg/l)	1	0,5

- **After removal of algae**

Appendix C

Table 3.3: Summary of capital costs to maintain water levels (DWA, 2010).

Item	Description	Western Basin	Central Basin	Eastern Basin
Pumping costs to maintain water levels at ECL	Pumps and pump installation	R2.8 million*	R5-9 million* (dependent upon number of pump sites)	
Treatment costs for neutralisation plants	For treatment plants at R2 million each	R40 million* (ensuring existing plants are continued)	R120 million* (ensuring Grootvlei infrastructure is maintained)	Assumes Grootvlei mine will be maintained
Total		R42.8 million	R129 million	
Total (all basins)				R171.8 million

** Notation as used by ToE in the DWA (2010) report*

Appendix D:

Code generated for simulation purposes

```
% Solve an Input-Output Fitting problem with a Neural Network
% Script generated by NFTOOL
% Created Tue Feb 19 17:57:24 CAT 2019
%
% This script assumes these variables are defined:
%
clc; clear; close all;

FileName = 'dataset.xlsx';
A = xlsread(FileName);

%Individual inputs, where x1=pH,x2=EC,x3=Cl
x1 = A(:,2);
x2 = A(:,3);
x3 = A(:,9);

%Individual outputs, where y1=TDS,y2=SO4
y1 = A(:,4);
y2 = A(:,5);

%ANN globalised inputs and output
inputs = [x1,x2,x3].';
targets = [y1,y2].';
```

```
% Create a Fitting Network
hiddenLayerSize = 20;
net = fitnet(hiddenLayerSize);

% Choose Input and Output Pre/Post-Processing Functions
% For a list of all processing functions type: help nnprocess
net.inputs{1}.processFcns = {'removeconstantrows','mapminmax'};
net.outputs{2}.processFcns = {'removeconstantrows','mapminmax'};


% Setup Division of Data for Training, Validation, Testing
% For a list of all data division functions type: help nndivide
net.divideFcn = 'dividerand';  % Divide data randomly
net.divideMode = 'sample';  % Divide up every sample
net.divideParam.trainRatio = 80/100;
net.divideParam.valRatio = 10/100;
net.divideParam.testRatio = 10/100;

% For help on training function 'trainlm' type: help trainlm
% For a list of all training functions type: help nntrain
net.trainFcn = 'trainlm';  % Levenberg-Marquardt

% Choose a Performance Function
% For a list of all performance functions type: help nnperformance
net.performFcn = 'mse';  % Mean squared error

% Choose Plot Functions
% For a list of all plot functions type: help nnplot
net.plotFcns = {'plotperform','plottrainstate','ploterrhist', ...
  'plotregression', 'plotfit'};


% Train the Network
[net,tr] = train(net,inputs,targets);

% Test the Network
outputs = net(inputs);
```