ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

Water Demand –Supply Gap Prediction and Driving Factors Analysis

(The Case of Addis Ababa city)

By Genet Hailemicheal

A Thesis Submitted to

The School of Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements for the Degree of Master of Science in Civil Engineering, (Water Supply and Environmental Engineering)

Advisor: Dr. Agizew Nigussie

Addis Ababa University
Addis Ababa, Ethiopia
2024

Addis Ababa University

Addis Ababa institute of technology

School of civil and environmental engineering

Water supply and environmental stream

This is to certify that the thesis prepared by Genet Hailemicheal Ordofa, entitled **WATER SUPPLY-DEMAND GAP PERDICTION AND DRIVING FACTORS ANALYSIS** (THE CASE OF ADDIS ABABA CITY) and submitted in partial fulfillment of the requirements for the Degree Master of Science in Civil Engineering (Specialization: Water Supply and Environmental Engineering) complies with the regulation of the university and meet the accepted standards with respect to originality and quality

Approved by the Board of Examiners

Dr. Agizew Nigussie	----------------------------------	--------------------------
(Advisor)	***Signature***	***Date***
-------------------------	----------------------------------	--------------------------
(Internal Examiner)	***Signature***	***Date***
-------------------------------	--------------------------------	-------------------------
(External Examiner)	***Signature***	***Date***
Dr. Abraham Gebere	--------------------------------	----------------------
(Dean, School of Civil and Environmental Engineering)	***Signature***	***Date***

Acknowledgement

First of all, I would like to thank God, who has always been my guidance and support all these years. I would like to thank all the people who contributed in one way or the other to the successful completion of this thesis. First and foremost, I thank my academic advisor Dr Agizew Nigussie, for his encouragement, and guidance. It is a great honor to have outstanding teachers and mentors like you. Even if the machine learning project is very time-consuming and complicated, he always encouraged me to do the right thing and be professional. The technical contribution of Dr Mebruk Mohammed is truly appreciated because of the support, and information, he gave me to the specific study. Special thanks to the Water Supply and Environmental Engineering stream instructors who have made my study fruitful and particularly to Dr Geremew Sahilu, who has been patient and kind in helping me progress and learn in this field. I would like to thank my family and husband for the incredible supports, love and encouragement. I appreciate my dearest friends Terefe and Temima for being by my side in all the difficult moments, and to my kind classmates, who accompanied me on the path of success.

Abstract

This study uses FFANNs to develop a predictive model for urban water demand-supply gap forecasting. The model considers nine independent variables and uses gird search hypermeter tuning and found that Bayesian regularization back propagation training algorithm and hyperbolic tangent sigmoid transfer function (tansig) with 40 neurons in the hidden layer with seven input variables best performing model and concluded that MLP-NN is an effective tool for understanding and simulating non-linear water demand-supply gap behavior, benefiting water providers and decision-makers.In addition, this study decomposed water demand factors in Addis Ababa city using LDMI decomposition analysis methods. Ten influential factors were identified, including rainfall, temperature, sunshine hours, relative humidity, population, industrial growth rate, economic growth, tourist number, livestock number, and water tariff. The study found that except population and livestock factors contributing to total water demand change varied at different years, with industrial growth, rainfall, humidity, tourist number, and sunshine hours inhibiting change during 2021-2022. Generally, for driving factor analysis LMDI method could be a good factor decomposer provided that data interaction and independence is absent.

Key words: water demand, water supply, gap, prediction ANN, driving factors, decomposition, LMDI

Table of Contents

List of Tables

List of Figures

Abbreviations

AAWSA-Addis Ababa Water & Sewage Authority

AI -Artificial Intelligence

ANN- Artificial Neural Networks

ARIMA- Autoregressive Moving Average

CFN- Cascade Forward Network

GDP- Gross Domestic Product

GRNN- General Regression Neural Network

IDA- Index Decomposition Analysis

LMDI- Logarithmic Mean Divisa Index Method

MAE- Mean Absolute Error

MAPE-Mean Absolute Percentage Error

MBGD -Mimi Batch Gradient Descent

MLR- Multiple Linear Regression

NRMSE-Normalized Root Mean Squared Error

PDA -Production-Theoretical Decomposition Analysis

RBF-Radial Basis Function

RMSE- Root Mean Squared Error

RGD- Regular Gradient Descent

SDA- Structural Decomposition Analysis

SGD- Stochastic Gradient Descent

SOM-Self Organizing Maps

SPSS- Statistical Package for The Social Sciences

UNDP- United Nations Development Program

UNICEF- United Nation International Children's Emergency Fund

WHO- World Health Organization

1. Introduction

1.1. Background

Water is one of the most important resources for life existence. Life cannot sustain in the world without water. However, this vital natural resource is under increasing pressure due to climate change, population growth, and unsustainable water management practices. Sustainable water management is ensuring water use in a way that meets both current and future socio-economical ecological demand. In improving the development and utilization of sustainable water resources management to meet increasing water demand, water demand prediction is a profound step.

Water is used for different purpose like for drinking, irrigation, firefighting, sanitation, domestic purpose, and commercial purpose recreational activities. This water use for different reason is very greatly variable and influenced by many factor such as tariff, size of the town , land use and land cover, attribute of the population, type and size of industrial and commercial business and environmental and climate condition [1].

In designing and providing of water supply system to the town, determination of the design capacity is a prerequisite to begin. This, in turn, is a function of water demand[2]. Urban water demand forecasting will play an essential role in the planning, distribution and management of scarce water resources among competitive user especially where rapid increase in world population, per capital income, industrialization and the impacts of global warming due to climate change on the world exhibits[3]. Here in demand forecast the time scales and explanatory variables utilized are the primary factors[4].

In case of Addis Ababa city under the impact of climate change, urbanization and economic development, the water supply for the city is continually decreasing, while the water demand sharply increased with the development of the national economy, particularly for the increase of domestic water. The relationship between the water demand-supply was changed in different historical periods. Clarifying and identifying the dynamic relationship between the water supply and water demand is an important basis for the sustainable management of water resources in the city.

Urban water demand is under rapid growth the planning, designing, and administration of water supply systems must be regularly reviewed and adjusted to keep up with. According to reports by United Nations Development Programme (UNDP), World Health Organization (WHO), and other non-governmental organizations (UNDP, 2017) Water supply is not always proportionate with demand, the According to the WHO and the Nations International Children's Emergency Fund (UNICEF), 663 million people in developing countries (one in ten) lack access to adequate water supplies[4]

In spite of the progress made and different imitative taken, improved domestic water supply systems are anticipated to continue to experience water shortages. Due to population growth, economic expansion, improved living conditions, and changes in urban residents' lifestyles the issue will worsen and worsen. Therefore, creating efficient public policies and strategies for managing water in areas with access to better water supplies is urgently needed[4].

Water demand explanatory variables for predictive models are selected based on intended time horizon and data's accessibility. For short term prediction, simple and straightforward models with few variables have shown promising accuracy Weather factors, such as temperature, relative humidity, and rainfall, and socioeconomic factors, such as population and income, are the two main categories of explanatory variables that determine water demand. Most of the time Water availability and supply is subject to considerable uncertainty, as has been made clear by large global climate shifts regarding weather patterns and global warming. Therefore, it is very essential to paying more attention to how changing weather patterns affect long-term water demand[4].

Water managers and providers, uses different forecasting timeframes and different forecasting methods which are applicable depending on the specific goals. For instant a long-term forecasting with a yearly or monthly resolution with limited accuracy is required for planning a large-scale investment like dams. On other hand for optimum distribution of water supply, medium-term forecasting at a monthly or weekly resolution and somewhat higher accuracy might be required[5]

1.2. Statement of the problem

In many places of the world, currently water demand exceeded supply, and many more are predicted to face this imbalance in the near future as the global population continues to grow at unexpected rate. Regionally and seasonally, water demand varies substantially and water consumption is significantly impacted by growing urbanization, industrialization, and new problems, including changing weather patterns and population expansion[4].

Almost all towns in Ethiopia are suffering problems of increasing demand for water consumption, there is high scarcity of resources to obtain sufficient quantities and satisfy the needs of citizens. Even if these problems related to water supply are not only emanated from under capacity design which is related to failing to forecast water demand adequately for a given design period, it has its own contribution. And it could be considered as one of major problem in addition to lack of finance and unsustainable management of water resource.

In case of Addis Ababa town under the different impacts of climate change, human activity, urbanization and population growth, the water supply from the surface and ground water source couldn't cope up with the ever-increasing water demand. Figure 1 show how the water supply and demand changed in different historical periods.

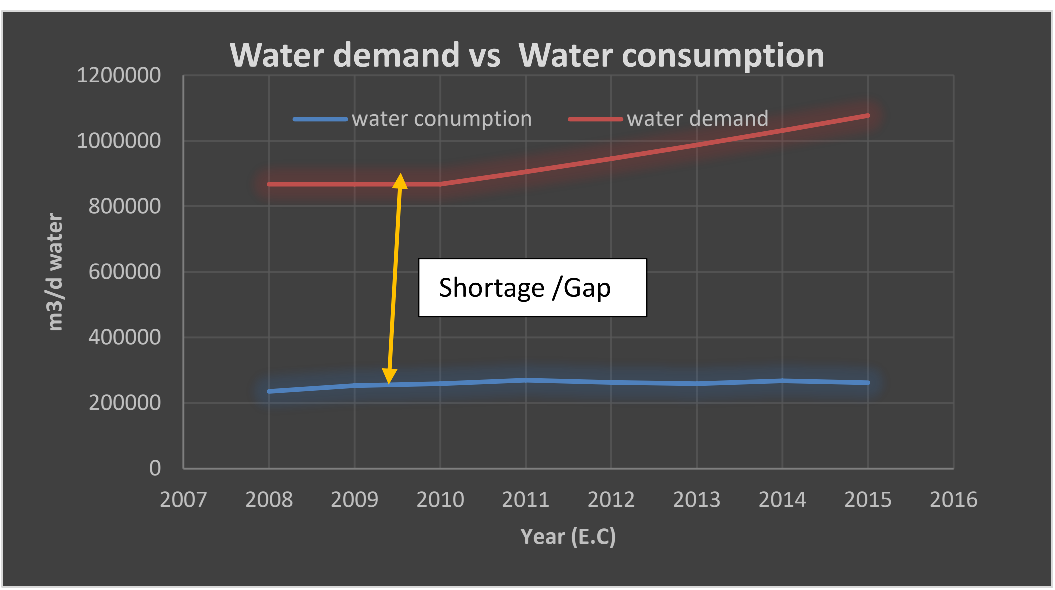


Figure 1 Addis Ababa water demand vs water supply

Global as well as local Climate change, economic expansion, population growth, the development of industrialization and urbanization are increasing the demand for water resources, but contrary to this increasing demand the water supply is limited due to different reasons. In recent years, the gap between water supply and demand which shows water shortage situation is becoming more serious[6].

Detailed understanding of water usage patterns, the amount of gap between water demand - supply and the factors that affect water use are essential for the proper management of water supply and the implementation of relevant public policies. Water usage patterns are highly complicated mechanisms which are affected by several factors, such as seasonal variations and water supply, household characteristics, water supply restrictions, tariff structure and pricing attitudes and intentions regarding water conservation. These are crucial factors to consider in the overall assessment of water management and conservation strategies. These factors influence people's water consumption and usage habits directly and indirectly, regardless of their social makeup[4].

For urban water demand forecasting, traditionally, two techniques are used either analyzing historical water consumption or using statistical models with a range of inputs such as population, price of water, income and various meteorological variables. However, these techniques have many disadvantages which include the requirement of large number of input parameters, and the fact that traditional modeling methods assume the data is linear and stationary[7].

Prediction model which consider all influential factors with their nonlinearity nature is very important. Therefore, models of ANN can provide a link between influential factors and water demand data without using a complicated type of computational method

1.3. Objective

1.3.1. General objective

The main objective of this research is to develop predictive artificial neural network model and analyze the contribution of water demand factor using Logarithmic mean divisa index method (LMDI) for the study area.

1.3.2. Specific objectives

The specific objectives of the study are the following:

- To review the existing urban water demand prediction methods.
- To review and identify influential parameters for water demand-supply gap prediction.
- To develop an ANN, model that is capable of predicting water demand supply gap.
- To decompose and evaluate driving factor of water demand using LMDI.

1.4. Research questions

- What are the trends of water consumption prediction?
- Which parameters are critical for water demand prediction considering the real contexts of the study area?
- Which ANN architecture model is efficient and accurate in predicting water demand supply gap?
- Which factor contribute more for water demand change?

1.5. Scope of the study

To meet the objective of this study different activities have been accomplished. Initially the trend of existing water supply- demand gap was assessed. Then different influential factors that affect only water demand were identified and evaluated due to data unavailability influential factor for water supply were not included. Even if there are different kind of ANN in this study only Multi-Layer Perceptron Artificial Neural Network (MLPs ANN) architecture was used for prediction purpose due to time and resource limitation. And for analysis of the water demand /consumption driving factors Logarithmic mean divisia index (LMDI) model was used.

The past 9 years water supply and demand data of Addis Ababa was used for the study.
Predictive variables that include population growth, weather patterns, industrial activities, and urban development were considered in the development of ANN model.
Mat-lab software was used in this study though has some limited choices for modifying the number of nodes in the hidden layer and training function (learning algorithm).

1.6. Significance of the study

Accurate water demand -consumption gap prediction is an important part of smart water utilities, and is an effective way to increase the water supply system's efficiency, dependability, and safety, which has significant positive economic and social effects. Accurate water demand or gap forecasting is essential for effective resource management; mitigating water shortages and ensuring a sustainable water supply.

To minimize a city’s water supply-demand gap, identifying influential factors causing that gap is very important. Research on past water demand changes assist to formulate water-saving polices and managing existing water resources[6].

Clarifying the dynamic relationship between the water supply and demand and predicting the gap is important for the optimal allocation and sustainable management of regional water resources. Reasonable allocation of limited water resources and improvement of utilization efficiency is the key to alleviating water resources shortage and promoting social-economic development[8].

In summary, predicting or foreknowing the water supply -demand gap in the city is very crucial for resource management, planning and policy making, economic impact, and environmental sustainability\climate change adaption:

In general, predicting the water supply- demand gap is essential for effective resource management, planning, policy making, economic stability, environmental sustainability, and climate change adaption. By accurate forecasting water needs and availability, stakeholders can take proactive measures to ensure a reliable and sustainable water supply for communities, industries and ecosystems.

On top of that analyzing various factors that drive water demand in the city, such as population growth, industrial activities, agricultural practices and climatic change, water managers and policy makers can better understand the more influential water consumption factor and assist in planning for the future water management strategies. Based on the most significant driving factors it can also help to prioritize investments in water infrastructure and conservation efforts.

In addition to the above this study outcomes is useful and give insight in providing essential information about future customer demand, which factors among the given one are more influential to water demand in case of Addis Ababa, and in providing information on which modeling will be suitable for Addis Ababa case

It can also assist in optimizing resource allocation and investment in urban water supply related projects. It provides actionable insights for a specific geographic area while trying to narrow the gap and can serve as a reference for researchers, academics, and professionals working in related fields.

1.7. Organization of the research

This study is composed of five main Chapters. Chapter one introduces general about water supply and demand and their prediction methodology: which includes the reasons of this study, the research, objectives and scopes. Chapter two Review Literature: it is the foundation of this thesis; study has determined the direction of this project. A review of water demand forecasting models has been described by different researchers. Chapter three materials and Methodology used are described: describes the methodologies that are used to collect the require information. Most of raw data and summary results from different data sources are described in figures and tables. The calculation of water consumption is shown in this chapter. Chapter four Results and Discussions: includes a discussion of the reliability of the results, errors of this study and a comparison of different models. Chapter five Conclusions and Recommendations: the study findings and used methodologies have been summarized. A summary of limitations of the study and the possible future study areas are described.

2. Literature review

2.1. Water Supply- Demand and prediction

In the context of a public water supply, water demand has a basic meaning, as volume of water needed or necessary to supply customer within a certain period of time or in other word water demand is an amount of water required to fulfill the demand of customer[9].

By definition, forecasting, interchangeable word: prediction, is the rational or logical prediction of events in scientific way[10].

Water authorities and water service providers use water demand forecasts for the planning, designing the infrastructure, operating and managing water supply systems. Depending on the specific aim they can use different forecasting time horizon and different forecasting methods available[5].

Different factor should be considered in selecting forecasting methods: purpose of the forecast, data availability, requirements for accuracy of the forecast, how well the forecasting model can be explained to stakeholders in the water planning process, and the ease of updating the model[11].

Table 1. Water Demand Forecasting Types and Applications[5, 9]

Forecast Type	Forecast Horizon	Applications
Very-Short-Term	Hours, days, weeks	Optimizing, managing system operations, pumping
Short-Term	Years, 1-2 years	Budgeting, program evaluation, revenue forecasting
Medium-Term	Years to a decade, 7-10 years	Sizing, staging treatment and distribution system improvements, investments, setting water rates
Long-Term	Decades, 10-50 years	Sizing system capacity, raw water supply

2.1.1. The case of Addis Ababa city

In the capital city of Ethiopia, Addis Ababa, Addis Ababa Water and Sewage Authority (AAWSA) is the water service provider in charge of city water and sewerage business [12]

The population of Addis Ababa city was estimated around 5.2 million in 2022. The population is growing at an annual rate of about 4.4%, in the city and is expected to reach about 7.3million by 2030. The water demand is also increasing rapidly in line with the population growth [13].

One of an important development segment in the city is Water supply. However, in Addis Ababa, Ethiopia, contrary to the development the adequacy, availability, and the quality of water are in danger, and the water demand of the city is unmet for the past 10 and more years. The city has very limited resources of surface and ground water which plays an important role in the support of domestic needs in 10 sub cities. In order to address the problem inadequate effort is devoted in an integrated approach which combines the supply and demand side management in a sustainable approach. An integrated water supply planning is a significant issue for sustainable development to assure maximizes economy, secure production and minimizes environmental impacts. however the supply side of water development has been given more attention than demand side management and improvement of patterns of water use because the water managers and planners have given high priorities to locating, developing and managing new water resources [14].

Even if (AAWSA) has been undertaking a number of initiatives to address the challenge of a large supply demand gap of the city still the city is suffering from water supply demand shortage for long time past. The following figure shows the demand, supply and production amount of water of the city.

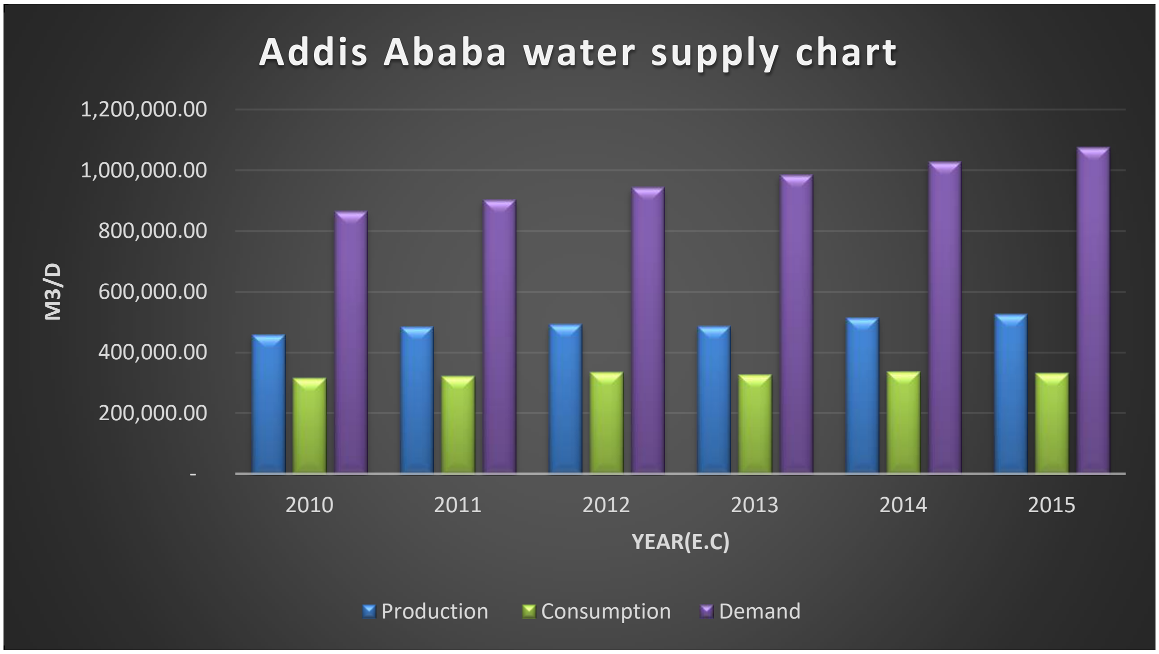


Figure 2 Discrepancy between water demand and water distribution facility capacity

This significant imbalance is recognized by the city's major water and sewerage service provider, AAWSA, as a result, water rationing delivery plan has been implemented for long time. Most parts of the city get water rationed two to three days per week. Similarly, the central part of the city, which is the downtown area, gets water supply for four to six days per week. Furthermore, the nearby area and suburban have considerable amount of water demand from new building construction The central AAWSA distribution network system has to account this new demand and the systems need to be connected[15].

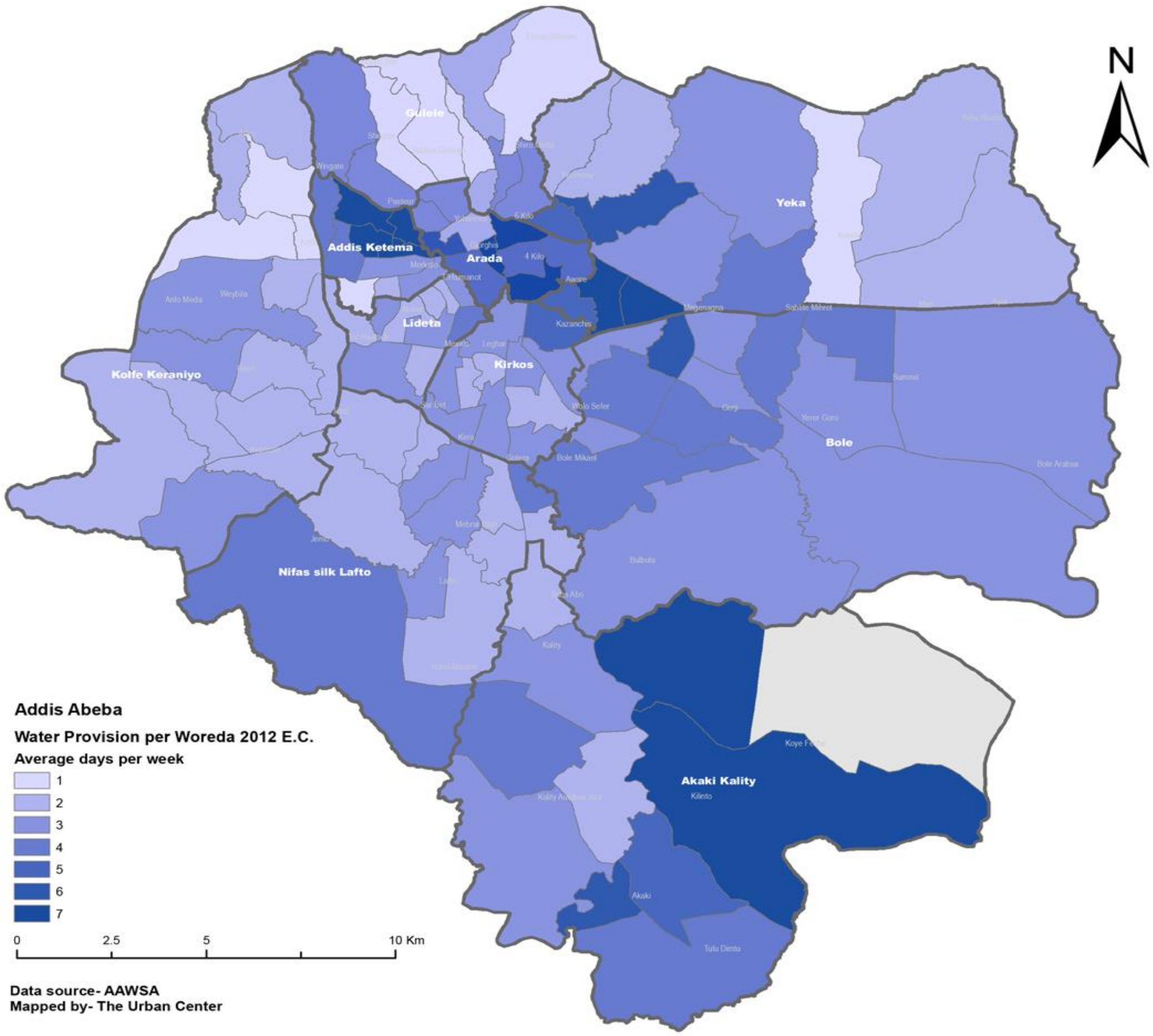


Figure 3 Regional water rationing schedule developed by AAWSA[16]

2.2. Factor that affects water demand/consumption

Number of populations, employment, economic cycles, technology, weather and climate, price, and conservation programs have significant influence on water demand[9, 11]. Other literature describes water demand influential factors for developed and developing communities separately. For developed community includes the following demographic, socio-economic and climatic factors that affect water demand[17]:

- Resident and seasonal population.
- Number, market value and types of housing units.
- Employment in service industries.

- Manufacturing employment and output (i.e. industrial water use)
- Water and wastewater prices and rate structures.
- Irrigated areas in residential and commercial/institutional use.
- Personal income.
- Climate (i.e. arid or humid)
- Weather conditions.
- Water-using appliances.
- Conservation activities

On the same literature for developing communities in addition to the above the following factors influence water use and consumption:

- The distance the consumer has to travel to the water source of the availability of water,
- Income,
- Education,
- Religion/culture, and
- Housing,

In water supply system several variables are considered influential in determining water demand that ranges from socioeconomic to various derivatives of weather-related variables. In General, Economic (water tariff), Socio demographic (population, education, income and tourisms), climate (Temperature, Rainfall, relative humidity and sunshine hour) and rule and regulations could be mentioned as influential factors for water demand.

A study done in Australia took different factors associated with water demand and found that community intervention factors like use of water efficient appliances and rainwater tanks has most significant influence on water demand. The study also shows community intervention program with water pricing policy plays grate role in reducing water demand. On the other hand, rainfall found to have limited influence on water demand, while temperature shows some degree of correlation with water demand [18].

Generally, the population growth, climate change, industrial development and urbanization are creating high stress on water demand but water pollution is reducing the limited water supply. Nowadays, water scarcity that indicated by water supply and demand gap is becoming more serious. Therefore, clear knowing this gap and its main driving factors could help to take and put forward water protection measures correctly[19].

2.3. Water supply- demand forecasting methods

For any forecasting technique to use accurate information on current water sales & revenues and current population and trends are very critical[9].

Physical based modeling is traditional modeling of physical processes also called knowledge driven modeling because it tries to explain the underlying processes. On contrary the so-called data-driven models, borrowing heavily from Artificial Intelligence (AI) techniques, are based on a limited knowledge of the modeling process and rely on the data describing input and output characteristics. This method often plays a complementary role to physically-based models because they can make abstractions and generalizations of the process. One concludes that, Data-driven modeling uses results from such overlapping fields as data mining, artificial neural networks (ANN), rule-based type approaches such as expert systems, fuzzy logic concepts, rule-induction and machine learning systems. Sometimes "hybrid models" are built combining both types of models[20].

Traditional way of urban water demand forecasting was either by analyzing historical water use or by using statistical models with a range of inputs such as population, price of water, income, and various meteorological variables. However, drawback or disadvantages with this traditional urban water demand forecasting models include the requirement of a large number of input parameters, and the fact that most traditional modeling methods assume the data is linear and stationary[7]. Among the statical methods multiple linear regression (MLR) and autoregressive moving average (ARIMA) models have been used for short-term urban water demand forecasting.

According to West Virginia water resource document for forecasting average annual demand different methods data needs and pro and cons are summarized as below.

Table 2 Summery of water demand forecasting and their advantages.[21]

Method	How it works	Data needs	Advantages	Disadvantage
Extrapolation	Predicate future water use by extrapolating past data	10 to 20 average annual demand	• Minimal data need • Uses simple statical methods	• Assume indefinite growth (decline) • Doesn't identify
Per capita use	Multiply future population by per capita use	Population projections Historical billing and population data	• Simple data • Simple calculations	• Assumes factors driving demand are well correlated with population
Unit use per customer type	Project growth in demand by customer category	Water billing by customer type Demographic data by customer type Projected growth for each customer	• Disaggregation addresses more trends in water use	• Require more detailed data & analyses

There are two mostly and widely used particular types of data-driven models, namely artificial neural networks (ANN) and fuzzy logic-based models, to modeling in the water resources management field are considered[20].

Recently, artificial neural networks (ANNs) have been introduced in urban water demand forecasting. Two of the main advantages of ANNs over other methods are that their application does not require a priori knowledge of the process, and they are effective with nonlinear data[7].

2.4. Artificial neural network fundamental

Artificial neural networks (ANNs) are branch of machine learning models that are built or operate based on principle of biological neural networks consisting in human brain.

A neural network is an information processing paradigm that is inspired by the way biological nervous systems like the human brain, process information. It is a computer based machine (artificial intelligence) that is designed to model the manner in which the brain performs a particular task or function of interest[22]. It operates by creating connections between mathematical processing elements, called neurons. Knowledge (data) is encoded into the network through the strength of the connections between different neurons, called weights, and by creating groups, or layers, of neurons that work in parallel. Even if ANN are empirical by nature they can provide practically accurate solutions for precisely or imprecisely formulated problems and for phenomena that are only understood through experimental data and field observations [23].

From mathematical point of view ANN can be defined as it is a complex non-linear function with many parameters that are adjusted (calibrated, or trained) until output becomes similar to the measured output on a known data set[20]. Or it tries to find a mathematical relation between input and output. This mathematical relation is modeled in the form of neurons, in which the input values are multiplied by weights and added together. The weights are in these neurons until this multiplication and summation accurately predicts the corresponding output. This updating of the weights is called training. After a model has been trained, it can predict the output based on the input up to certain accuracy. This predicting is called inferring. Training a neural network might take hours or days depending on its complexity and its training settings, however, inferring the output from some input may only take millisecond[24].

ANN brings many advantages in modeling amount which is model building don't require physics based algorithm and this makes faster and flexible, nonlinear relationship can be handled and user knowledge and experience can be incorporated in building a model[25].

A relationship hidden in the historical data can be revealed by using data driven -ANN which assist forecasting water demand of an area[25]. Other literature describe this as without knowing the exact relationships, ANN is suitable to perform a kind of function fitting by using multiple parameters on the existing information and predict the possible relationships in the coming future, if significant variables are known[20].

2.4.1. Architecture and Elements of Artificial Neural Network

An ANN is based on collection of connected units or nodes called artificial neurons, which loosely model the neurons in biological brain.

The basic building block of a (artificial) neural network (ANN) is the neuron and it is processing unit which have some (usually more than one) inputs and output. Neuron receives signals then processes them and can signal neurons connected to it. First each input x_i is weighted by a factor w_i and the whole sum of inputs is calculated as

$$\text{Total input} \quad = \quad \sum_{i=1}^{n} \boldsymbol{xi} * \boldsymbol{wi} = a \qquad (1)$$

Where n is number of input and a is sum of weighted input, after activation function is applied to the result **a** and then the output of each neuron is computed by some non-linear function of the sum of its inputs which is taken to be f(a). Generally the ANN are built by putting the neurons in layers and connecting the outputs of neurons from one layer to the inputs of the neurons from the next layer[26].

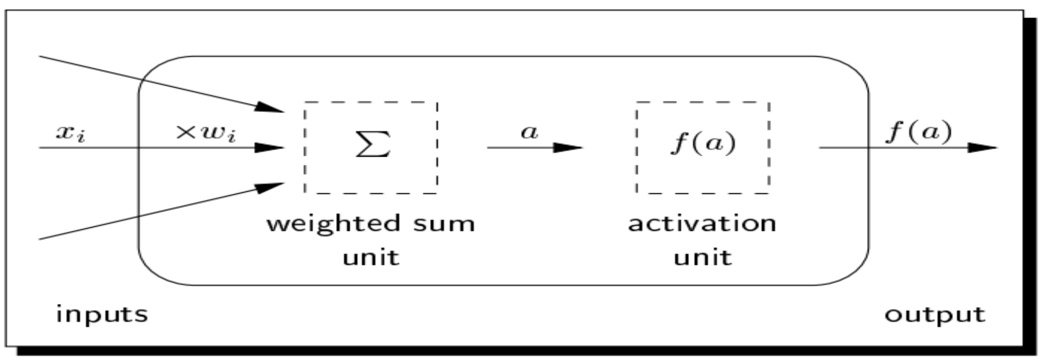


Figure 4 The neuron

Finally, to compute the output, an activation function is applied on the weighted sum of inputs which is a:

$$\text{Output= activation function } (\sum_{i=1}^{n} \boldsymbol{xi} * \boldsymbol{wi}) = f(a) \qquad (2)$$

Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer), to the last layer (the output layer), possibly after traversing the layers multiple times.

Input layer: - A first layer, usually the inputs distributed to all neuron of this layer or in other word they are neuron that receive information from external sources, and passes this information to the network for processing.

Hidden layer: - A layer between input layer and output layer and they receive information from the input layer and process them in a hidden way.

Output layer: - Last layer, which is a neuron commonly the desired output of the network and are neuron that receives processed information and sends output signals out of the system.

Bias: Acts on a neuron like an offset. The function of the bias is to provide a threshold for the activation of neurons. The bias input is connected to each of the hidden and output neurons in a network.

The number of input neurons corresponds to the number of input variables into the neural network, and the number of output neurons is the same as the number of desired output variables. The number of neurons in the hidden layer(s) depends on the application of the network. A network is typically called a deep neural network if it has at least 2 hidden layers[27].

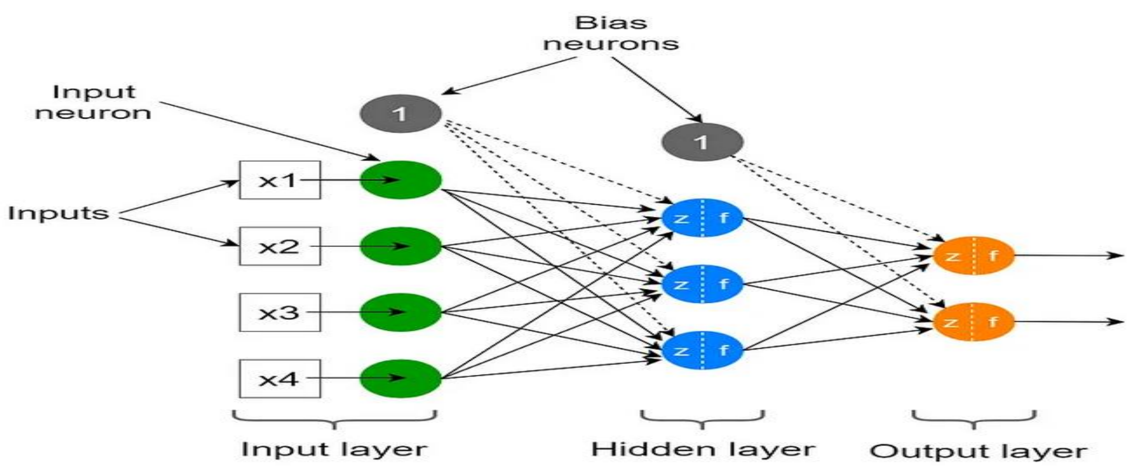


Figure 5 Basic of Artificial Neural Network Architecture

The arrangement of neural processing units and their interconnections can have a profound impact on the processing capabilities or on the performance of the model. As a result, there could be different kind of connections how the data flow between the input, hidden and output layers. Multi-layer perceptron (MLP), the radial basis function (RBF),General regression neural network(GRNN), Cascade forward network(CFN) and Koho-nen's self-organizing maps (SOM) are some example of the arrangement type[22].

The dataset must be split into three sets: training (70% of the available data), validation (15% of the available data), and testing (15% of the available data) in order to create the ANN model

2.4.2. Types of ANN

Based on one or more of their relevant features ANNs may be classified in many different ways Generally, the classification of ANNs may be based on (i) the function that the ANN is designed to serve (e.g., pattern association, clustering), (ii) the degree (partial/full) of connectivity of the neurons in the network, (iii) the direction of flow of information within the network (recurrent and nonrecurrent)[28].

Again, based on the model architecture which refers to the overall structure and manner of how information flows from one layer to another the three model architectures include feedforward, recurrent networks, and hybrid models[29].

2.4.2.1. Feedforward Architectures

A neuron connection is said to be "feed-forward" if it's only between neurons in the input layer and other neurons in the hidden layer, or between neurons in the hidden layer and neurons in the output layer. however, a layer's neurons are not related to one another. In many types of feedforward ANNs Multi –layer perceptron's (MLPs) with only three layers are the most widely used architectures followed by BPNNs which use the back-propagation algorithms to train networks[29].

2.4.2.2. Recurrent Architectures (RNNs)

RNNs differs from feedforward ANNs, in that neurons within a layer are interconnected and allow feedback. Various RNN types are created to improve neural networks' memory performance[29].

2.4.2.3. Hybrid Architectures

Recently there is a growing tendency to use hybrid ANNs models, which play a huge role in modeling, for their ability to integrate with other conventional and more advanced modeling techniques to create flexible and efficient models. These Hybrid models are divided into three categories, namely model-intensive, technique-intensive, and data-intensive. Model-intensive techniques works by modelling the sub-components of the entire physical system and then aggregate each model's total response. The core of the technique intensive methods is to develop a modeling framework that is able to take advantage of different technologies like combine ensemble approaches or time series models that remove trends or periodicities like Autoregressive Integrated Moving Average-Radial Basis Function neural networks (ARIMA-RBFNNs) or ARIMA-ANN whereas data-intensive approaches are to combine different technologies to preprocess the data. Wavelet analysis approaches such as WANN can provide some useful information about the physical structure of the data[29].

2.4.2.4. Emerging Methods

Cconvolutional neural network (CNN) , Deep belief network (DBN) and self-organizing Deep belief network (SODBN) are some of emerging methods. CNN is a feed-forward neural network, primarily used in the image field and can be implemented more than one time to reveal the relationship between the parameters hidden in the input matrix. Deep belief network (DBN) is a kind of neural network based on deep learning and several visible and hidden layers, stacked in order to make up the DBN. To overcome these limitations, a SODBN has been proposed. The structure of the SODBN is not determined by artificial experience but the automatic growing and pruning algorithm[29].

Table 3 Advantage and disadvantage of different type of ANN[30-32]

Types of ANNs	Advantages	Disadvantages
Multilayer Perceptron (MLP)	Versatility: Can model complex, non-linear relationships. High Accuracy: Often provides accurate predictions when properly trained.	Computationally Intensive: Requires significant computational resources and time for training. Overfitting: Prone to overfitting if not regularized properly.
Radial Basis Function (RBF) Networks	Fast Training: Generally faster to train compared to MLPs. Good for Small Datasets: Effective with smaller datasets.	Limited Capacity: May struggle with very large or complex datasets. Hyperparameter Sensitivity: Performance heavily depends on the choice of parameters.

Recurrent Neural Networks (RNNs)	Temporal Data Handling: Well-suited for time-series data, capturing temporal dependencies. Flexibility: Can handle variable-length input sequences.	Vanishing Gradient Problem: Training can be difficult due to gradients diminishing over time. Complexity: More complex to implement and train compared to simpler models.
Convolutional Neural Networks (CNNs)	Feature Extraction: Excellent at automatically extracting features from data. Parallel Processing: Efficient processing due to parallelism in convolutional layers.	Data Requirements: Requires large amounts of data to perform well. Interpretability: Models can be difficult to interpret and understand.
Hybrid Models (ANN + Genetic Algorithms)	Optimization: Genetic algorithms can optimize ANN hyperparameters for better performance. Improved Accuracy: Combining methods can lead to more accurate predictions.	Complexity: Increased complexity in model design and implementation. Computationally Demanding: Requires significant computational resources

2.4.3. Multi-layer perceptron (MLP)

There is various ANN model, Multi –layer perceptron's (MLPs) are the most commonly used feed-forward neural network. Typical feed forward network architecture contains input layer, hidden layer and output layer. Most common training algorithm used with MPLs is back-propagation algorithm.[25].

The MLP (modern name feed forward) neural network consists of multiple layers of computational units, usually interconnected in a feed forward way that approximates a relationship between sets of input data and a set of appropriate output. It is based on standard linear perceptron and it makes use of three or more layers of neurons (nodes) with nonlinear activation functions, and is more powerful than the perceptron and mathematically expressed as follow[22].

$$\boldsymbol{y_k = f_{outer}\left[\sum_{j=1}^{M} w_{kj}^{(2)} f_{inner}\left[\sum_{i=1}^{d} w_{ji}^{(1)}\ x_i + w_{jo}^{(1)}\right] + w_{ko}^{2}\right]} \quad (3)$$

where y_k represents the k-th output, f_{outer} represents the output layer transfer function, f_{inner} represents the input layer transfer function, w represents the weights and biases, (i) represent the i-th layer[20].

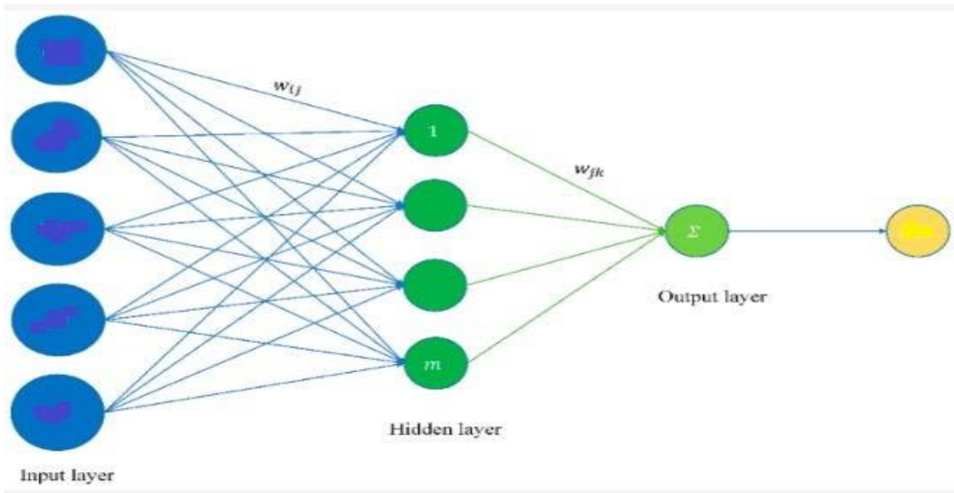


Figure 6 A simple feed forward three-layer networks

2.4.4. Training of Artificial Neural network (learning)

Beside data collection and data analysis, training of ANN is one of the step-in modeling ANN. During training of the model actual and target output data are compared, in such a way best network parameters which will minimize the error are defeminized [25]. The error terms are then used to adjust the weights in the hidden layers so that, hopefully, the next time around the output values will be closer to the "correct" values after much iteration. This iterative learning process is called back-propagation algorithm in which error resulting from the comparison are propagated backwards through the network.

Once a network has been structured for a particular application, that network is ready to be trained. To start this process, the initial weights are chosen randomly. Then the training, or learning, begins. During the training of a network the same set of data is processed many times as the connection weights are continually refined. Mathematically learning process can be expressed as follow.

$$\boldsymbol{w_{new}} = \boldsymbol{w_i} \pm \Delta \boldsymbol{w} \quad (4)$$

2.4.5. Testing of Artificial neural network

The test data set is the data remaining after taking the values in the training process. The purpose of this set is to evaluate the accuracy of the network. It can consist 15% of the data. The validation set is basically used for the conclusive verification stage of the network, where the data are the consequent values to those of the training which can represent 15% of the total data set [33]. The ratio can vary depending on the purpose of model.

The purpose of testing is to compare the outputs from the neural network against targets in an independent set (the testing instances). The neural network can move to the so-called deployment phase if all the testing metrics are considered okay. The validation methods that need to be used depending on the application type: Approximation testing methods, Classification testing methods and Forecasting testing methods

The most used testing methods in approximation applications are the following: Testing errors, Errors statistics, Errors histogram, Maximal errors and Linear regression analysis.

To test a neural network and most critical errors for measuring the accuracy of a neural network are:

- Sum squared error.
- Mean squared error.
- Root mean squared error.
- Root mean squared logarithmic error.
- Normalized squared error.
- Minkowski error.

The performance of ANN models to predict outcomes is evaluated using the correlation coefficient (R) for the training set and all datasets, mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE)[34]. However, the data set that was used as an input for the models has different ranges and means, the normalized root-mean squared error (NRMSE) is included to syncretize those differences.

$$\boldsymbol{MAPE = \left\{ \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \check{y}_i}{y_i} \right| \right\} * 100} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y}_i)^2} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \overline{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{7}$$

$$NRMSE = \frac{RMSE}{\check{y}_t} \tag{8}$$

where n = number of data points; y_i and $\overline{\mathbf{y}}_{\mathbf{i}}$ = actual and forecasted water demand, respectively; and $\check{y}_t$ = data set mean.[35]

2.5. Hyperparameter tuning for Artificial Neural Networks

Hyperparameters of ANN are parameters that are set before the training process begins and affect the learning process and model performance. These play a crucial role in determining the performance and generalization ability of neural network model so tuning these hyperparameters is an essential step in optimizing the performance of ANN for specific task[36].

There are two types of hyperparameters that determine the structure of a neural network; these are the **number of layers** in a network and the **number of neurons per layer**. Basically, a neural network has three layers: input layer, hidden layer and output layer. The number of neurons in the input layer is always equal to the number of input variables. The number neurons in the hidden layers can be arbitrarily chosen per layer. The number of neurons in the output layer is always equal the number of output variable. As the number of neurons increased in the hidden layer more multiplications and summations are performed, which enables the network to model more complex relations between the input and the output[24].There are different hyperparameter options among which are described below.

2.5.1. Optimizer

This has role during back propagation in which networks weights are update in such a way that the network best predicts the output values, or in other words, minimizes the error based on training data. They determine how updated of weight and bias are performed. Commonly used optimizers are Regular Gradient Descent (RGD), Stochastic Gradient Descent (SGD), Mini-Batch Gradient Descent (MBGD) and Adam, adaptive moment[24].

2.5.2. Loss function

Loss functions play a role in back propagation like optimizer and they define the training error that the optimizer is trying to minimize. Mean squared error, Mean squared logarithmic error and Huber are among the loss function.

2.5.3. Learning rate

This is also a hyperparameter used in back propagation in that it will determines by how much weights are changed at each weight update. Choosing a too low learning rate or too high learning rate will have its own drawback for instance choosing too low rate will create very low training network process in which it will take a long time before the network reaches a minimum of its loss function. And choosing a too high learning rate will make the network jump around at the minimum, which may prevent the network from ever reaching a minimum, i.e., it may lead to divergence. As a result, finding a good balance is thus crucial for achieving good results in any neural network[24].

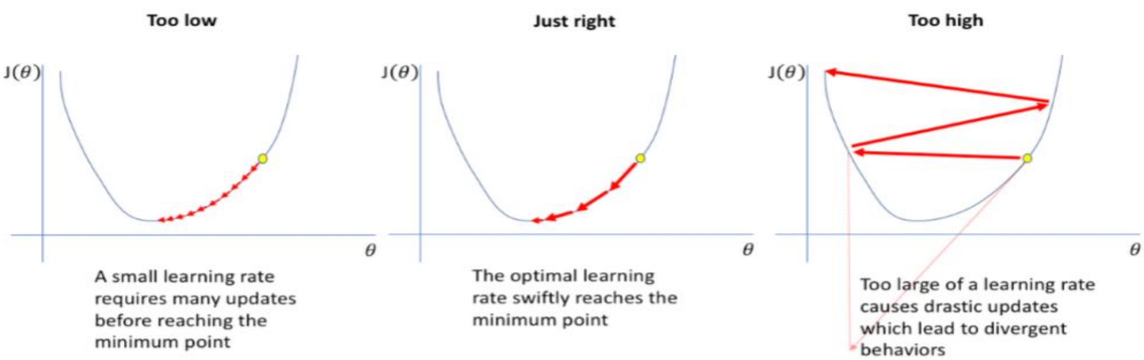


Figure 7 different learning rates effects. θ are the weights and J(θ) is the loss function

2.5.4. Activation function

Activation function is applied in hidden layer and output layer and their purpose is to add nonlinearity which allows the neural network to estimate nonlinear relationship. There are a lot of activation functions but commonly used activation functions in hidden layers are Sigmoid and Relu. Every layer and even every neuron can have a different activation function but most of the time only a single activation function per layer is used to avoid over complexity[24].

Based on the kind of the problem different activation function can be used for instance in the output layer for regression problems, usually a linear activation function is selected, due to values are unbounded, which is beneficial for extrapolating outside the bounds of the (normalized) training data[24].

2.5.5. Neural network structure

Here the structure of a neural network means the number of hidden layers and number of neurons in each individual layer, which has influences on how well the neural network is able to learn some mapping. So far there are no strict rules on how to choose these parameters. However, some rules of thumbs for the number of neurons in the first hidden layer are provided by some literature.

2.5.6. Automatic hyperparameter tuning methods

There are no strict rules or rule of thumb on how to set the hyperparameters in building neural network, and sometimes one hyperparameter options might behave poorly with certain other hyperparameter options, whereas they might behave well in combination with other options. Manual hyperparameter tuning can be carried out via a trial-and-error approach but the result of this approach is greatly affected by the skill and experience of the engineer, as well as some luck.t therefore automatic hyperparameter tuning is preferred. There are three automatic hyperparameter tuning methods grid search, random search and Bayesian search[24].

2.6. Logarithmic Mean Divisia Index (LMDI) method

Driving force for water consumption can be quantified using decomposition methods and generally there are three basic decomposition methods: structural decomposition analysis (SDA), index decomposition analysis (IDA) and production-theoretical decomposition analysis (PDA). IDA is an analytical tool originally for energy studies but extended to other areas including CO_2 emission analysis, environmental management, and sustainable use of natural resources. Based on IDA, many specific decomposition methods were developed and Logarithmic Mean Divisia Index (LMDI) method is the one and the most widely used. Changes in water and energy consumptions can be studied by quantifying the impacts of changes in several factors[19].

Factor decomposition technique commonly applied in energy consumption and carbon emission are are structure decomposition approach (SDA) and index decomposition approach (IDA) which include the Laspeyres Index Decomposition, Divisia Index Decomposition, and the Logarithmic Mean Divisia Index (LMDI)[37].

LMDI is used to learn more about the contribution of people's activities to pollution discharge. For problems in factor decomposition and in analyzing forces LMDI is very suitable and widely used[38].

Shortly Logarithmic Mean Divisia Index (LMDI) model can be defined as a method used to decompose changes in an aggregate variable into the effects of various factors. It is commonly applied in the analysis of energy consumption, economic growth, and environmental impact. Logarithmic Mean Divisia Index (LMDI) method is the most widely used one. Changes in water consumptions can be studied by quantifying the impacts of changes in several factors.

And LMDI model is useful for understanding the drivers of changes in an aggregate variable and can help policymakers and analysts identify opportunities for improving efficiency and sustainability. It is a powerful tool for analyzing complex systems and can provide valuable insights into the underlying dynamics of a given phenomenon.

Log Mean Di's Decomposition Method is the most widely used model and has the characteristics of not producing residual value and allowing zero in the data, which is clear and objective for the results of influencing factor analysis. Ang., Liu and Ang both think it is the Optimal Decomposition Method. Many scholars use the LMDI to study the energy consumption structure, energy consumption carbon emissions and other issues. So, it gives us a lot references to use LMDI Model to explore and analyze the deep-seated problems of water resources in Huaihe River Basin[6].

To analyze and understand historical changes in economic, environmental, employment or other socio-economic indicators, it is very important to assess the driving forces or determinants that underlie the changes[39].

2.6.1. Kaya Equation and Decomposition analysis

A Japanese scholar Kaya originally proposed Kaya's identity to break down greenhouse gas emissions into four driving factors: energy carbon intensity, energy consumption per unit of GDP, GDP per capita, and population[38].

To analyses and understand the changes in socio-economic indicators such as the economy, environment, and employment, among others it is very important to evaluate potential drivers or decisive factors. Therefore, because of not only listing the influencing factors of research objectives but also calculates their overall contribution, factor analysis or decomposition analysis approach is widely used in the field of resources and the environment[37].

According to the equation of kaya, domestic water is decomposed as follows:

$$DW = \frac{DW}{TW} x \frac{TW}{GDP} x \frac{GDP}{P} xP \tag{9}$$

Where: DW is domestic water consumption, TW is the total quantity of water consumption, GDP is the gross regional product, P is the population of a region, and represents the population effect on domestic water use; and S $=\frac{DW}{TW}$, is defined as the structural factor, use the proportion of domestic water consumption in total water consumption to indicate; T$=\frac{TW}{GDP}$, is the technological factor, expressed by the volume of water resources consumed per unit GDP; E $=\frac{GDP}{P}$, which means that GDP per capita represents an economic factor [37].

The above formula can further express as

$$DW = SxTxExP \tag{10}$$

To consider the change in time again the above formula can be defined as follow

$$\Delta DW = DW_0 - DW_t = \Delta Sx\Delta Tx\Delta Ex\Delta P \tag{11}$$

Where DW_0 and DW_t are the domestic water consumption at the beginning and end of the period, respectively, and ΔDW represents the change in domestic water use during this period. And ΔS, ΔT, ΔE, ΔP represent the structure, technology, economy, and population effects on domestic water use, respectively.

The sum of the above four effects or a factor over a certain period is the change in domestic water consumption. If the effect is positive within a certain period, it led to an increase in domestic water consumption and promoted the increase of domestic water consumption; otherwise, it had an inhibitory effect on the increase in domestic water consumption[37].

According to the LMDI decomposition model, the driving factor or contribution rate of each factor to water consumption change can be expressed as follows in Equation 12-15 where S_0 and S_t, T_0 and T_t, E_0 and E_t, P_0 and P_t represent, respectively, the value of S, T, E, P at the beginning and the end of the period.

$$\Delta S = \left[\frac{DW_t - DW_0}{lnDW_t - lnDW_0}\right] ln\left(\frac{S_t}{S_o}\right) \tag{12}$$

$$\Delta T = \left[\frac{DW_t - DW_0}{lnDW_t - lnDW_0}\right] ln\left(\frac{T_t}{T_o}\right) \tag{13}$$

$$\Delta E = \left[\frac{DW_t - DW_0}{lnDW_t - lnDW_0}\right] ln\left(\frac{E_t}{E_o}\right) \tag{14}$$

$$\Delta P = \left[\frac{DW_t - DW_0}{lnDW_t - lnDW_0}\right] ln\left(\frac{P_t}{P_o}\right) \tag{15}$$

Similar equation was used in the study made in chain and titled as driving force analysis of the consumption of water and energy in China based on LMDI method. In this study they calculate economic scale effect, intensity effect, structure effect and population on water consumption[19].

2.6.2. Some related works

A study carried out in chani, Hubei Province took the data of Huaihe River Basin from 2001 to 2016 as an example and use ecological water footprint to describe the demand, with the water carrying capacity representing the supply and analyze the water supply-demand situation of Huaihe River Basin and its five provinces from footprint view in time and space. Then they apply the Logarithmic Mean Divisia Index model to analyze the driving factors of the ecological water footprint. As result of the decomposition of water demand drivers, they reach on conclusion that economic development is the most important factor, with an annual contribution of more than 60%. The study finalize with optimal allocation of water resources in Huaihe River basin management suggestion and provides reference for the formulation of water-saving policies in the world[6].

A study made by He He and Rupert J analyze the demand for building material using log mean divisa index decomposition method. They use the LMDI method combined with the concept of dynamic material flow analysis, and physical and monetary flows to decompose the demand for building materials into the following six effects: material intensity, floor area shape, dwelling type, dwelling intensity, economic output, and population and found that of these six effects, the material intensity effect is the most important, overall contributing to increasing concrete demand by +79 Mt from 1950 to 2014, while the dwelling intensity effect plays an opposite role, overall reducing concrete demand from 1950 to 2014 by−56 Mt. The economic output effect is also significant (+38 Mt from 1950 to 2014)[40].

Another study in china carried out to investigates the spatial-dynamic driving forces governing industrial water pollutant emission. By applying the LMDI decomposition method heterogeneous effects of different drivers, that is, technology, energy consumption, and economic size distribution are decomposed and quantified. The result indicates that the most important contribution to pollutant abatement is the development of technology, followed by energy consumption, and the least affected is the distribution of economic scale.

A study by Yajing Li and colleges investigate critical driving forces of water and energy consumption in chain using a modified logarithmic mean divisia index (LMDI) method. in this study the decompose the total effect of water and energy consumption to different driving factors with no residues. Total water consumption change is decomposed into four driving factors which include industrial water consumption, industrial scale, economic scale and population scale. The result of this study shows that the growth of population and GDP per capita were influential factors that drive water and energy consumption, while the improvement of consumption intensity was inhibiting factor on water and energy consumption[19].

3. Material and Methodology

3.1. Study area

3.1.1. Background

Addis Ababa lies at an elevation of 2,355 meters (7,726 ft) and is a grassland biome, located at 9°1′48″N 38°44′24″E. The city lies at the foot of Mount Entoto and forms part of the watershed for the Awash [41].

Addis Ababa the capital city of Ethiopia’s is one of the most dynamic and fastest growing cities in Africa. The city is the seat to headquarters of the African Union and considered to be the diplomatic capital of Africa due to a significant presence of foreign embassies and United Nations offices. because of this the water supply system should be stabilized not only for its year-round residents but also for seasonal visitors that call the city home anywhere from several weeks to many months[15].

As of the 2007 population census conducted by the Ethiopian national statistics authorities, Addis Ababa has a total population of 2,739,551 urban and rural inhabitants. For the capital city 662,728 households were counted living in 628,984 housing units, which results in an average of 5.3 persons to a household[42].

The economic activities in Addis Ababa are diverse. According to official statistics from the federal government, some 119,197 people in the city are engaged in trade and commerce; 113,977 in manufacturing and industry; 80,391 homemakers of a different variety; 71,186 in civil administration; 50,538 in transport and communication; 42,514 in education, health and social services; 32,685 in hotel and catering services; and 16,602 in agriculture. In addition to the residents of rural parts of Addis Ababa, the city dwellers also participate in animal husbandry and the cultivation of gardens[42].

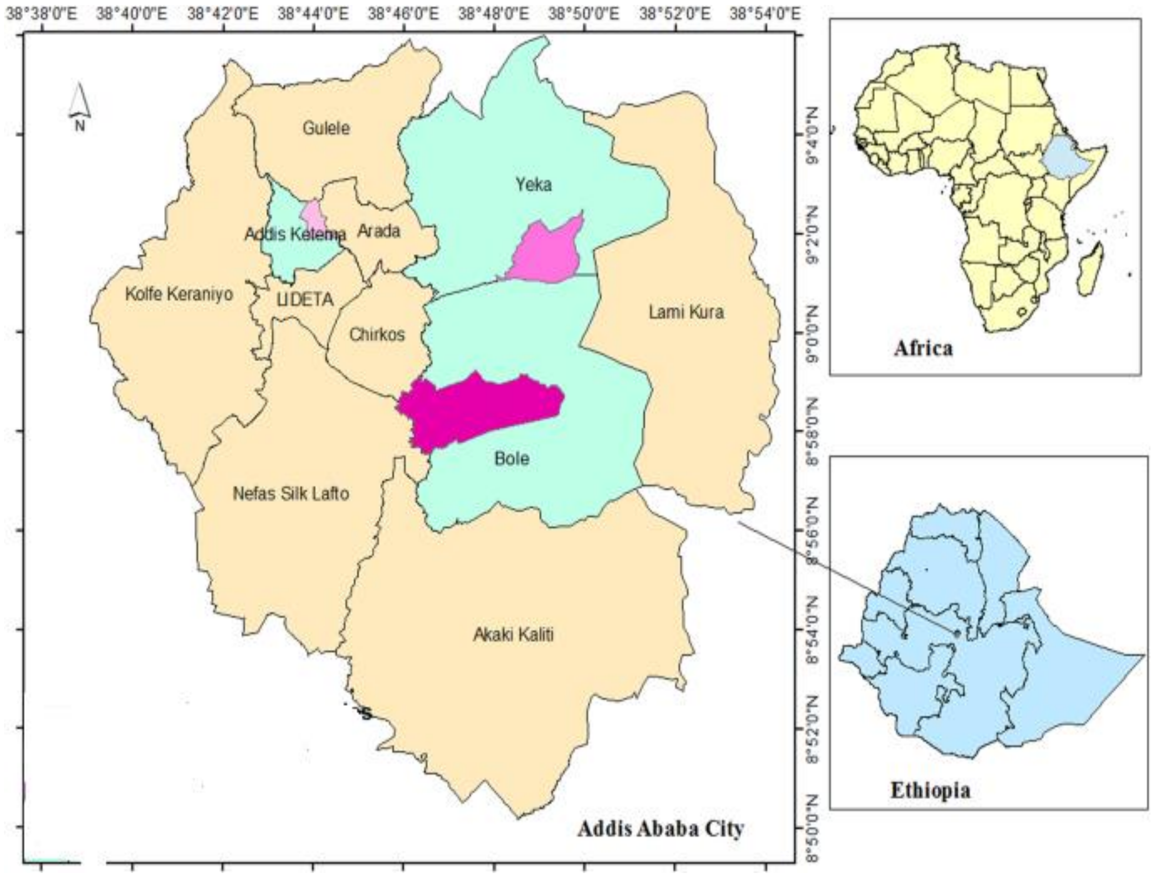


Figure 8 Location Map of the study Area

3.1.2. Existing water supply situation

Addis Ababa Water and Sewerage Authority is currently providing water for the City from Legedadi, Dire and Gefersa dams with additional supplies from ground water pumped from Akaki well field to the south of Addis Ababa, and other wells and springs within the City[43]. The total production by the end of 2010 was 103,923 Mm3/year 301.597m3/day). 65% of the water source is surface water from the two dams and the remaining 35% is from ground water[35].

Addis Ababa has grown from a small village to become the predominant and largest urban center in Ethiopia since being selected as the capital in 1886. The city has grown from an area of some 16 Km^2 with an estimated population of about 100,000, prior to the Italian occupation, to its current size of about 532 km^2 and Millions of population (1997)[44].

In the recent years Addis Ababa's city population has been increased by more than 4% over year with the current estimated city's population standing at about 5.2 million[16]. This rate of growth is expected to continue with a significant gap in supply and demand for water, a problem expected to linger for the foreseeable future. If the current water demand was to be fully met, it would require an estimated supply quantity of 1.2million m3/day. This discrepancy is in stark contrast with the current water production rates from both surface water and groundwater estimated to be at an annual average of 0.48 million m3/day, a coverage of only about 40% of the daily requirement.

3.2. Dataset

Billed Water consumption data were obtained at monthly time interval from Addis Ababa water and sewerage authority in the period from the 30th of September 2015 till the end of August 2023. In addition, the forecasted demand of the city is taken from the Addis Ababa water and sewerage authority business plan 2011 – 2020 final report and የአዲስ አበባ ውኃና ፍሳሽ ባለስልጣን የ10 ዓመት የልማት ዕቅድ (2013-2022 ዓ.ም)[35].

Information for the study area is gathered in four major categories: water consumption, census data, economic growth data and climate data. All collected secondary data sources and their respective organizations are summarized below.

Table 4 detailed description of the data.

Data Types	Period Data	Sources/Description
Billed Water consumption	2015-2023	AAWSA
Water demand and water production data and tariff	-	AAWSA
Meteorological (rainfall, minimum and maximum temperature, sunshine hours and relative humidity)	2002-2023	National meteorology agency
Economic growth rate	2015-2023	Addis Ababa City Administration Plan and Development Commission

Number of tourists	2015-2023	Addis Ababa cultural and tourism office
Number of livestock	2015-2023	
Addis Ababa Population	2015-2023	CSA

3.2.1. Water demand driving factors

Domestic water consumption is highly affected by water supply patterns, the characteristics of heads of households, vegetable gardening, and the use of water appliances. In addition to that Water demand mainly depends on the size of the city, quality of water, metering system, pressure in the pipeline. Generally, there are various factors affecting water demand but the major factors that play a vital role are listed as follows[45].

- Size of town or city
- Living Standard (economic status of consumer)
- Climatic Condition
- Industrial and Commercial Activities
- Quality of Water
- System of Water Supply (intermittently or continuously)
- Metering System
- System of Sanitation
- Cost of Water
- Age of Community
- Pressure in Pipeline

Considering data availability and resource limitation the following variables are considered as water demand driving or influential factors for the selected study area.

Independent Variables:

- Weather parameters (temperature, sunshine hours, humidity and Rainfall).
- Population

- Tourist number
- Industrial growth
- Economic growth
- Livestock number
- Water tariff

Dependent Variable:

- Water supply- demand gap in Addis Ababa, Ethiopia

3.3. Data quality analysis and preparation

After the relevant data collection, three data preprocessing procedures are conducted to train the ANNs more efficiently. These procedures are: (1) solve the problem of missing data, (2) normalize data and (3) randomize data. The missing data are replaced by the average of neighboring values. Normalization procedure before presenting the input data to the network is generally a good practice, since mixing variables with large magnitudes and small magnitudes will confuse the learning algorithm on the importance of each variable and may force it to finally reject the variable with the smaller magnitude.

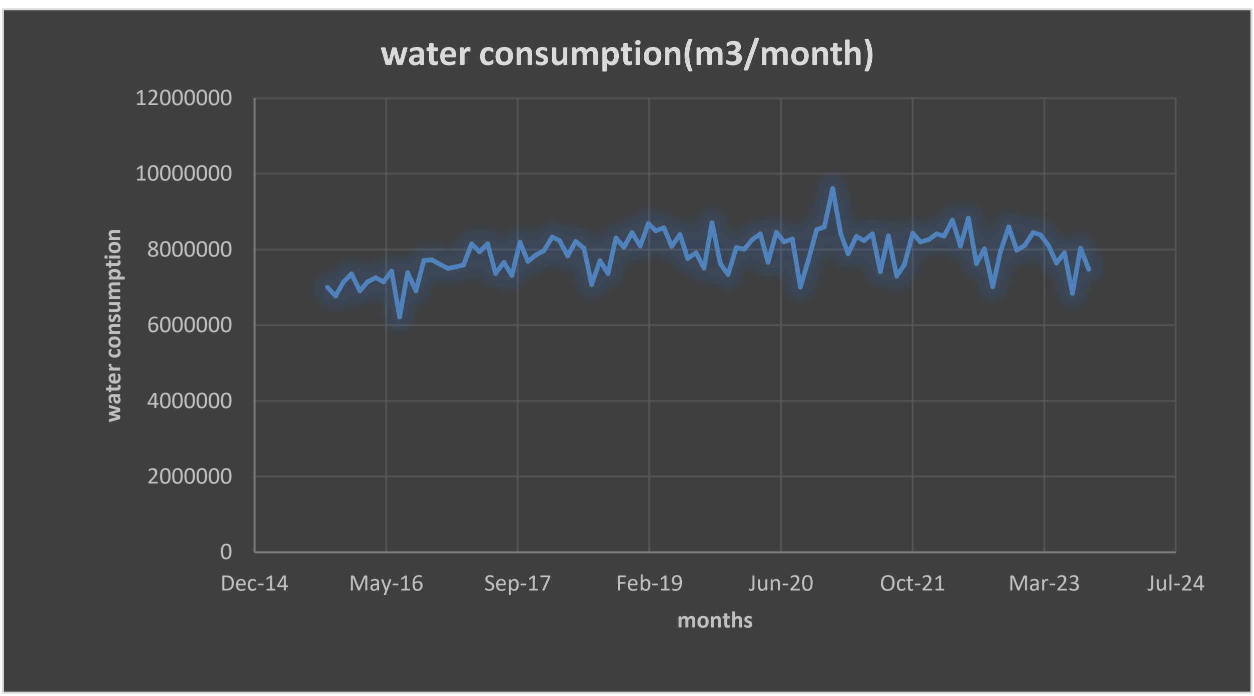


Figure 9 Water consumption profile of Addis Ababa city

Therefore, effective data cleaning and preprocessing of raw data is very important to achieve accurate water demand/ gap forecasting model because most of the time water consumption data are imperfect with frequent erroneous time steps, missing data, seasonality, and trends. It is necessary to pre-process the data before entering them into the software so that they are in a noise-free format (without null or inconsistent values) in order to prepare them for the forecast.

3.3.1. Outlier detection and missing value

During data analysis frequently two problems encounter which are outlier and missing value. Outliers are values imbedded in the data series that do not seem to make sense and analyst shall identify such events so that they can be included in the model or the data can be adjusted to remove their effects. Similar consideration will apply for missing values too. Analyst can fill in a missing data point by interpolating from neighboring[9].

Among the different method Simple Arithmetic Mean Method/ Local Mean Method, the simplest method commonly used to fill in missing meteorological data in meteorology and climatology is used. Depending on the data distribution there are different type of outlier detection and to determine the distribution normality test has been carried out in SPSS software and found that the data distribution is not normal therefore, any statical data test should follow non-parametric.

There are two way of normality test graphical and numerical. The former one requires extensive experience but the second one don’t. SPSS determine the analysis based on both ways and the result is as follow.

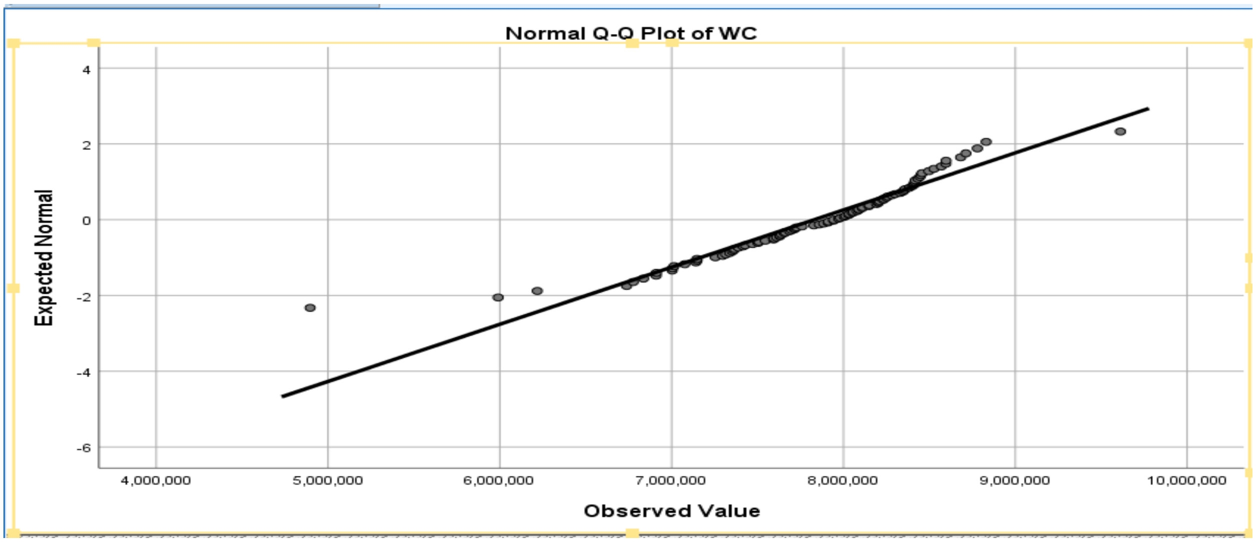


Figure 10 Graphical result of normality test

Significance value of Shapiro-wilk becomes zero that means the data is not normal or there is significant difference in the distribution of the data.

There are several ways to treat outliers in a dataset, depending on the nature of the outliers and the problem being solved. Trimming, Capping and Treating outliers as a missing value are some of the ways to treat outliers in data sets. Z-score and Interquartile Range (IQR), are some of the most popular methods used in outlier detection. The technique to be used depends on the specific characteristics of the data, such as the distribution and number of variables, as well as the required outcome. For this study further data refining was performed by excluding any extreme that are not within the 1.5 interquartile ranges (IQRs) of (Q1-1.5*IQR) and (Q3+1.5*IQR)[46].

3.3.2. Data preparation

After the necessary data cleaning the input dataset is then divided into three sets: 70% for training, 15% for validation and the remaining 15% corresponds to the testing set. The input and target variables (water supply -demand gap) are normalized in the range of 0 and 1 according the following equation:

$$\bar{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{16}$$

where $\bar{X}$ is the normalized input or target variables (water supply -demand gap) X, X_{min} and X_{max} represent the minimum and the maximum of observation values respectively. The aim of normalization is to avoid any prioritization of the variables and to remove any arbitrary effect of similarity between the objects.

As water demand is function of two or more variable associated with water use in this paper ten factors identified as a dependent variable for water demand in Addis Ababa and a correlation analysis was carried out between the factors to determine the degree of relation between the factors by constructing correlation matrix.

Ten combinations of weather predictor variables, economic variable, and population variable were considered, as shown in Table 4, in order to investigate their effect on water supply-demand gap. The MLP neural network methods are applied for predicting the water supply-demand gap in Addis Ababa city based on the combinations shown in Table 4.

Table 5 input parameters combinations for different model

Model	RF	T	*SS*	RH	Pop	To	IG	GR	LN	WT	Gap
1	√	√	√	√	√	√	√	√	√	√	√
2		√	√	√	√	√	√	√	√	√	√
3			√	√	√	√	√	√	√	√	√
4				√	√	√	√	√	√	√	√
5					√	√	√	√	√	√	√
6						√	√	√	√	√	√
7							√	√	√	√	√
8								√	√	√	√
9									√	√	√
10										√	√

3.4. Artificial neural network (ANN) for water supply-demand gap prediction modeling

Comparing MLP with six other models for the short-term prediction of water needs while considering the climatic conditions the rresults showed that MLP provides the best performance[47]. Due to MLP has got common preference than other type of ANN and simplicity in this study it was choose for prediction model.

In designing ANN models for prediction purposes, it follows a number of systemic procedures. In general, there are five basics steps: (1) collecting data, (2) preprocessing data, (3) building the network, (4) train, and (5) test performance of model[48].

3.4.1. Data collection

In designing ANN models Collecting and preparing sample data is the first step. Meteorological data (rainfall. sunshine hours, relative humidity and temperature), number of populations, tourist number, industrial growth rate, economic growth rate, livestock number and water tariff of Addis Ababa city were collected. Detail about data set used is described under section 3.2.

3.4.2. Data pre-processing

After relevant data collection, data preprocessing procedures were conducted in order to train the ANNs more efficiently. These procedures can be summarized as three steps (1) filling the missing data; (2) data normalize and (3) randomize data.

The procedure used in the development of the MLP models starts with input data normalization in the range of 0 to 1 followed by the dataset matrix size identification. After being normalized, sub-datasets are created and prepared for training and testing. The training set is first randomized before creating and training the ANN. The output values are generated and de-normalized, and finally the performance of the ANN is verified by comparison of output and target values. Data cleaning was carried out in excel manually and other data pre-processing, normalization and data division done using MATLAB tool.

3.4.3. Building the network architecture

At this stage, the number of neurons in hidden layers, transfer function in each layer, training algorithm, and performance method were specified using one of hypermeters tuning method which is gird search method.

There are many ways to establish the architecture of a neural network; in most cases back propagation training algorithms are used. In this study, three network training function tested and the one with lost MSE has been taken as training function.

3.4.4. Network Training

During network training, the weights are adjusted in order to make the actual outputs (predicated) close to the target (measured) outputs of the network. The available data (96 months data from September, 2015 to August 2023) are randomly divided into training, testing and validation subsets. The training set is used to estimate the unknown connection weights, the testing set is used to decide when to stop training in order to avoid over fitting and/or which network structure is optimal, and the validation set is used to assess the generalization ability of the trained model.

In order to determine the optimal MLP architecture with the highest coefficient of determination, and the lowest mean square error some of training algorithms (Levenberg-Marquardt (LM) Algorithm (trainlm), Bayesian regularization back propagation (trainbr) and scaled conjugate gradient back propagation (trainscg) , transfer functions (tangent sigmoid, log sigmoid and linear functions) and epoch numbers are tested and selected..

3.4.5. Validation the neural network

In order to validate the model, the same data set as used for training is given as the input, during this time, the targets are not provided and the trained network now predicts the outputs according to the weight matrix formed during the training process. Here the errors are not calculated and back propagation is not occurring. Back propagation occurs only during the training process.

3.4.6. Testing the neural network

With testing data set the model is tested and the overall model performance is measured by coefficient of determination (R^2), and mean squared error (MSE). In this paper, the MSE and coefficient of determination (R2) are used to evaluate the prediction performance of the model.

$$\boldsymbol{MSE} = \left(\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y}_i)^2\right) \tag{17}$$

$$\boldsymbol{R^2} = 1 - \frac{\sum_{i=1}^{N}(y_i-\bar{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{18}$$

Where: - n = number of data points; y_i and $\bar{\mathbf{y}}_\mathbf{i}$ = actual and forecasted water demand, respectively; and $\check{y}_t$ = data set mean.

3.5. Neural network model programming

A feed-forward back-propagation artificial neural network is used in the numerical modeling using a Mat lab script. The MATLAB® (R2018b) (MATLAB, 2018) software neural network toolbox, a high-performance interactive software program for scientific and engineering computation, is used to design, build, train, and test the MLP neural networks.

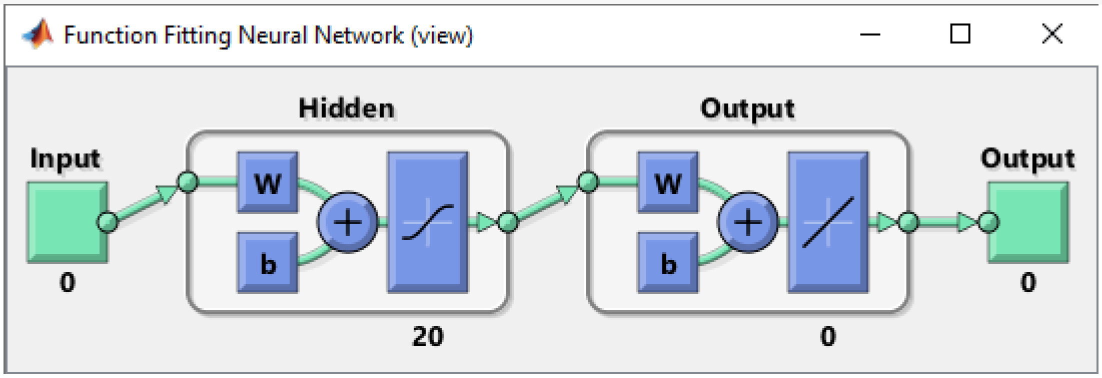
The MLP program starts by reading data from an Excel file and the data are organized into input matrix ‘x’ and the target matrix ‘wc’ (water supply-demand gap. Each i^{th} column of the input matrix will have 96 elements representing the input data value. Each corresponding column of the target matrix will have one element, representing the water supply-demand gap as shown in the Mat Lab program below.

```
% the input data of for the model
 %1. Rainfall (mm)
 %2.Averag temperature (^C)
 %3.Sunshine hours (hr)
 %4.Relative Humidity
 %5. Population (number)
 %6.Tourist number (number)
%7. Industrial growth (%)
 %8.economic growth (%)
 %9.Water tariff (Birr)
 %10.Water consumption-demand gap (%)
%input data
%% load data
Data=xlsread('ANN','sheet1');
inputvariable=Data(:,1:9); % input variables used for water supply-demand gap modeling
gap=Data(:,10);% Target variables which going to be modeled
 x=inputvariable';
g=gap';
```

Three-layer (i.e. one hidden layer and one output layer) feed forward back propagation neural networks can fit any input-output relationship given enough neurons in the hidden layer. Numbers of hidden layer investigated for this study were from 1 to 300.

```
%% Creating the ANN model /feed forward neural network with 20 hidden numbers
setdemorandstream(15) % set the training and testing randomly
```

```
hiddenLayerSize=20;
net = fitnet(hiddenLayerSize);
view(net)
```



After preparing the number neurons in the layer and testing parameters, the network is ready to be trained and the overall training is done as the scripts below.

```
% Set training parameters (optional)
net.trainFcn = 'trainlm';
net1.layers{1}.transferFcn='logsig'; %hidden layer activation function
net1.layers{2}.transferFcn='purelin';% output layer activation function
net.performFcn = 'mse';
net.trainParam.show=50;
net.trainParam.epochs=1000; %Number of training times
net.trainParam.goal=1e-5; %accuracy of training
net.trainParam.lr=0.01; %learning rate

%% Network configuration and training the Neural Network
[xn,xs]=mapminmax(x);% Normalization of Data(scale inputs and targets so that they fall in
the range [-1,1]).
[gn,gs]=mapminmax(g);
net=train(net,xn,gn);
[net,tr] = train(net,xn,gn); %train the network nntraintool
plotperform(tr);
%% simulate the network
an=sim(net,xn);
```

```
W=mapminmax('reverse',an,gs);
```

A neural network training process involves tuning and revising the values of the weights and biases of the network to optimize network performance, as defined by the network performance function net.performFcn. For feed forward networks default performance function is mean square error (MSE) which is the average squared error between the network's outputs and the target outputs.

Whenever a neural network is trained, it can result in a different solution due to different initial weight and bias values and different training, validation, and test data division sets. As a result, for the same problem and for the same input different outputs can be achieved. To ensure that a model of good accuracy has been found, several times retrain is recommended.

With respect to the testing samples the mean squared error of the trained neural network can be measured. And it will give us a sense of how well the network will do when applied to data from the real world.

```
%% Evaluating the model /testing the Neural Network
%the mean squared error of the trained neural network can now be measured with respect to the testing samples.
[net,tr] = train(net,xn,gn);
testX = xn(:,tr.testInd); % separates input into three sets: training, validation, and testing
testT = gn(:,tr.testInd);% separates targets into three sets: training, validation, and testing
testg = net(testX);
perf = mse(net,testT,testg);
%% Testing the Neural Network A measure of how well the neural network has fit the data is the regression plot.
plotregression(g,W)
%% Fitting Error Histogram
% another third measure of how well the neural network has fit  data is the error histogram. This shows how the error sizes are distributed.
error = g - W; % 'W' is Forecasted water supply demand gap
ploterrhist(error)
```

Finally, the ANN model is ready to perform the required water supply- demand gap forecasting with desired time step.

```
%% Simulate a time-series model.
init_sys = idpoly([1 -0.99],[],[1 -1 0.2]);
e = iddata(randn(76,1),[]);
data = sim(init_sys,e);  % Estimate an ARMAX model for the simulated data.
na = 1;
nb = 4;
sys = armax(data(1:76),[na nb]);
% Obtain a 4 step-ahead prediction for the estimated model.
K = 12;
WD = predict(sys,data,K);  % Analyze the prediction.
figure;
plot(g,'g')
hold on
grid on
plot(error,'b')
plot(W,'r')
ylabel('%');
xlabel('Time (month)');
legend('g (Actual gap)','error (Forecasting Error)','G (Predicted data)')
grid on
% Create title
title({'Comparison between actual and predicted (%)'});
```

3.5.1. Hyper parameter tuning

One of neural network parameter that needs to be set before the training process begins is number of neurons per layer. For this study number of neurons in input layer set based on the number of available input variables (9 variables) and output layer is set based on problem outcome which is one in number (water-supply demand gap). But number of neurons in the hidden layer should be searched and set for better model performance. Consequently, the following mat lab script was used to determine the optimum number of neurons in the hidden layer.

```
%% load data
Data=xlsread('ANN','sheet1');
inputvariable=Data(:,1:9); % input variables used for water demand modeling
gap=Data(:,10);% Target variables which going to be modeled
x=inputvariable';
g=gap';
%% Define grid search parameters
neurons_grid = [1, 20,30,40,50,60,70,80,90 100,150,200,250,300]; % Possible number of
neurons in hidden layer
num_epochs = 100; % Number of training epochs (randomly selected)
num_folds = 5; % Number of folds for cross-validation
% Initialize variables to store results
results = zeros(length(neurons_grid), 2); % [Neurons, MSE]
% perform grid search
for i = 1:length(neurons_grid)
    num_neurons = neurons_grid(i);
    % Create a feedforward neural network
    net = feedforwardnet(num_neurons);
    net.trainParam.epochs = num_epochs;
    % Perform cross-validation
    cv_indices = crossvalind('Kfold', g, num_folds);
    mse_values = zeros(num_folds, 1);
    for fold = 1:num_folds
        % Split data into training and validation sets
        train_idx = (cv_indices ~= fold);
        val_idx = (cv_indices == fold);
        % Train the network
        net = train(net, x(:, train_idx), g(:, train_idx));
        % Validate the network
```

```
        predictions = net(x(:, val_idx));
        mse_values(fold) = mean((g(:, val_idx) - predictions).^2);
    end
    % Average MSE over all folds
    avg_mse = mean(mse_values);
    % Store results
    results(i, :) = [num_neurons, avg_mse];
end
% Find the best configuration
[best_mse, best_index] = min(results(:, 2));
best_neurons = results(best_index, 1);
% Display results
disp('Grid Search Results:');
disp(array2table(results, 'VariableNames', {'Neurons', 'MSE'}));
fprintf('Best Number of Neurons: %d with MSE: %.4f\n', best_neurons, best_mse);
```

In the same manner activation function that applied in the layer has to be set before training begins. Again, gird search MATLAB script (shown below) is used to choose among commonly used activation function in the hidden layer.

```
clc
clear
%% load data
Data=xlsread('ANN','sheet1');
inputvariable=Data(:,1:9); % input variables used for water supply-demand gap modeling
gap=Data(:,10);% Target variables which going to be modeled
x=inputvariable';
g=gap';
% Define grid search parameters
activation_functions = {'tansig', 'logsig', 'purelin'}; % Possible activation functions
num_neurons = 10; % Fixed number of neurons in hidden layer
num_epochs = 100; % Number of training epochs
```

```matlab
num_folds = 5; % Number of folds for cross-validation
% Initialize variables to store results
results = zeros(length(activation_functions), 2); % [ActivationFunctionIndex, MSE]
% Perform grid search
for i = 1:length(activation_functions)
    activation_function = activation_functions{i};
    % Create a feedforward neural network
    net = feedforwardnet(num_neurons);
    net.layers{1}.transferFcn = activation_function; % Set the activation function
    net.trainParam.epochs = num_epochs;
    % Perform cross-validation
    cv_indices = crossvalind('Kfold', g, num_folds);
    mse_values = zeros(num_folds, 1);
    for fold = 1:num_folds
        % Split data into training and validation sets
        train_idx = (cv_indices ~= fold);
        val_idx = (cv_indices == fold);
        % Train the network
        net = train(net, x(:, train_idx), g(:, train_idx));
        % Validate the network
        predictions = net(x(:, val_idx));
        mse_values(fold) = mean((g(:, val_idx) - predictions).^2);
    end
    % Average MSE over all folds
    avg_mse = mean(mse_values);
    % Store results
    results(i, :) = [i, avg_mse]; % Store index and MSE
end
% Find the best configuration
[best_mse, best_index] = min(results(:, 2));
best_activation_function = activation_functions{best_index};
```

```
% Display results
disp('Grid Search Results:');
disp(array2table(results, 'VariableNames', {'ActivationFunctionIndex', 'MSE'}, ...
        'RowNames', activation_functions));
fprintf('Best Activation Function: %s with MSE: %.4f\n', best_activation_function, best_mse);
```

Again, the possible training algorithm and number of epochs is searched using the above mat lab script just by substituting activation function with training function and epoch number.

3.6. Logarithmic Mean Divisia Index (LMDI) method

The analytical tool Index decomposition analysis (IDA) is used to measure different driving factors and their environmental side effects. The method LMDI allows to break down the aggregate indicator into a combination of several factors and to determine the influence of a single factor on the aggregate indicator.

In other words, the IDA method allows us to measure the impact of the main factors influencing changes over time in the water demand produced by different driving factors. The method also makes it easy to explain and compare the results of the study.

To determine the impact of individual factors on total water demand change from different influential factor in the study area, here LMDI method based on Kaya identity is used. Kaya identity was first introduced at the Intergovernmental Panel on Climate Change (IPCC) in 1989 and presented by Kaya.

Kaya identity is commonly used to determine total (CO2 equivalent) emissions from anthropogenic activities and is expressed as the product of four factors: human population, GDP per capita, energy intensity (per unit of GDP), and carbon intensity (emissions per unit of energy consumed) but some country also used to decompose driving factor of water demand for different purpose[19].

The study used the following formula (22) in order to decompose water demand influential factors in Addis Ababa city.

$$WD = RFxTxSSxRHxPopxTNxIGxEGxLNxWT. \qquad (19)$$

Where: -

WD represents water demand,

RF rainfall

T temperature

SS:- sunshine hours,

RH:- relative humidity,

Pop:- population number,

TN:- tourist number,

IG:- industrial growth rate,

EG:- economic growth rate,

LN:- livestock number,

WT: water tariff

In additive decomposition for the model in the above Equation, the effect of various driving factors from the baseline year 0 to the final year t were set as WD_0 and WD_t, respectively, can be expressed as follows:

$$\Delta \mathrm{WD(tot)} = \mathrm{WD}(0) - \mathrm{WD}(t)$$

$$= \Delta RF+\Delta T+\Delta SS+\Delta RH+\Delta Pop+\Delta IG+\Delta EG+\Delta To+\Delta LN+\Delta WT \quad (20)$$

Where $\triangle WD(tot) =$ represents the total change of water Consumption. Each effect is further expressed in equations below.

To assess the contribution to the water demand change, ten indexes computed during every interval in this way:

$$\Delta RF = \sum \frac{\mathrm{WD_t - WD_0}}{\mathrm{lnWD(t) - lnWD(o)}} (\ln \frac{\mathrm{RF_t}}{\mathrm{RF_0}}) \quad (21)$$

$$\Delta T = \sum \frac{WD_t - WD_0}{lnWD(t) - lnWD(o)} (ln \frac{T_t}{T_0}) \quad (22)$$

$$\Delta SS = \sum \frac{WD_t - WD_0}{lnWD(t) - lnWD(o)} (ln \frac{SS_t}{SS_0}) \quad (23)$$

$$\Delta RH = \sum \frac{WD_t - WD_0}{lnWD(t) - lnWD(o)} (ln \frac{RH_t}{RH_0}) \tag{24}$$

$$\Delta Pop = \sum \frac{WD_t - WD_0}{lnWD(t) - lnWD(o)} (ln \frac{Pop_t}{Pop_0}) \tag{25}$$

$$\Delta To = \sum \frac{WD_t - WD_0}{lnWD(t) - lnWD(o)} (ln \frac{To_t}{To_0}) \tag{26}$$

$$\Delta IG = \sum \frac{WD_t - WD_0}{lnWD(t) - lnWD(o)} (ln \frac{IG_t}{IG_0}) \tag{27}$$

$$\Delta EG = \sum \frac{WD_t - WD_0}{lnWD(t) - lnWD(o)} (ln \frac{EG_t}{EG_0}) \tag{28}$$

$$\Delta LN = \sum \frac{WD_t - WD_0}{lnWD(t) - lnWD(o)} (ln \frac{LN_t}{LN_0}) \tag{29}$$

$$\Delta WT = \sum \frac{WD_t - WD_0}{lnWD(t) - lnWD(o)} (ln \frac{WT_t}{WT_0}) \tag{30}$$

The sum of the annual changes in the individual factors is equal to in total water demand change. Generally, if the value of the effect is positive, it means that the factor contributes to increases the total water demand change; if the value is negative, it reduces total change.

In this study because of available data limitation, only the above-mentioned water demands driving factors were considered.

4. Result and discussion

Under this section analysis done and results obtained with ANN model for water supply - demand gap forecasting with a monthly base and contribution of water demand driving factors analyzed using LMDI method are summarized and presented.

In this study the number of input variables used for the prediction of urban water supply – demand gap was nine which gradually decreased in subsequent models as shown in Table 4.

4.1. Correlation matrix

A correlation analysis was carried out between water demand and its driving factors to determine the degree of relation between the factors by constructing correlation matrix. The developed correlation matrix, summarized in Table 5, indicates that, except population and livestock number there was no a high correlation between the factors that means there was no multi-collinearity, but between population and livestock number there was high correlation (r=0.935). Therefore, in such a situation where highly correlated variables existed either removing or combining would address multi-collinearity. As a result livestock number is removed from water supply –demand gap prediction modeling.

Table 6 Nonparametric correlation result /spearman

	RF	***T***	***SS***	***RH***	***Pop***	***To***	***IG***	***GR***	***LN***	***tariff***	***Gap***
RF	1										
T	-0.19	1									
SS	-0.84	0.04	1								
RH	0.80	-0.16	-0.8	1							
Pop	0.04	0.09	0.03	0.04	1						
To	0.07	-0.09	0.02	-0.01	0.06	1					
IG	0.012	0.09	-0.2	0.12	0.055	-0.16	1				
GR	-0.03	0.09	-0.04	0.11	0.086	0.505	0.215	1			
LN	0.007	-0.02	0.1	-0.05	**0.935**	0.194	0.103	-0.108	1		
WT	-0.03	0.07	0.05	0.08	0.888	-0.13	-0.007	-0.043	0.72	1	

Gap	**0.136**	**0.08**	**-0.1**	**0.11**	**0.884**	**0.127**	**0.25**	**0.034**	**0.88**	**0.69**	**1**

In addition to that, input variables can be prioritized based on their relationship strength which was determined by correlation analysis. As shown in the above table in terms of the strength of relationship, the value of the correlation coefficient varies between +1 and -1. A value of ± 1 indicates a perfect degree of association between the two variables. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker. Therefore, an input parameter which has strong association with the output can be selected for further analysis.

Based on the above result even if number of population, tourist number and water tariff had strong relationship with the water supply-demand gap all factors except livestock number used for further analysis. .

4.2. ANN modeling and prediction

The development of the ANN model for water supply-demand gap prediction was realized using various structures to obtain the optimal ANN performance. A mat lab script was used for gird search of best parameters (number of neurons in hidden layer, activation function, training algorithm and number of epoch). Other parameters important for model optimization were set default by the application. Finally, a model with best parameters was used for water supply-demand gap prediction.

Figure 11 shows an opened window of the network during training. This window displays training progress and allows the user to interrupt training at any point by clicking stop training. The regression button in the training window performs a linear regression between the network outputs and the corresponding targets. The application randomly divides input parameters and target parameter into three sets as follows: 70% were used for training; 15% were used to validate that the network is generalizing and to stop training before over-fitting; the last 15% were used as a completely independent test of network generalization.

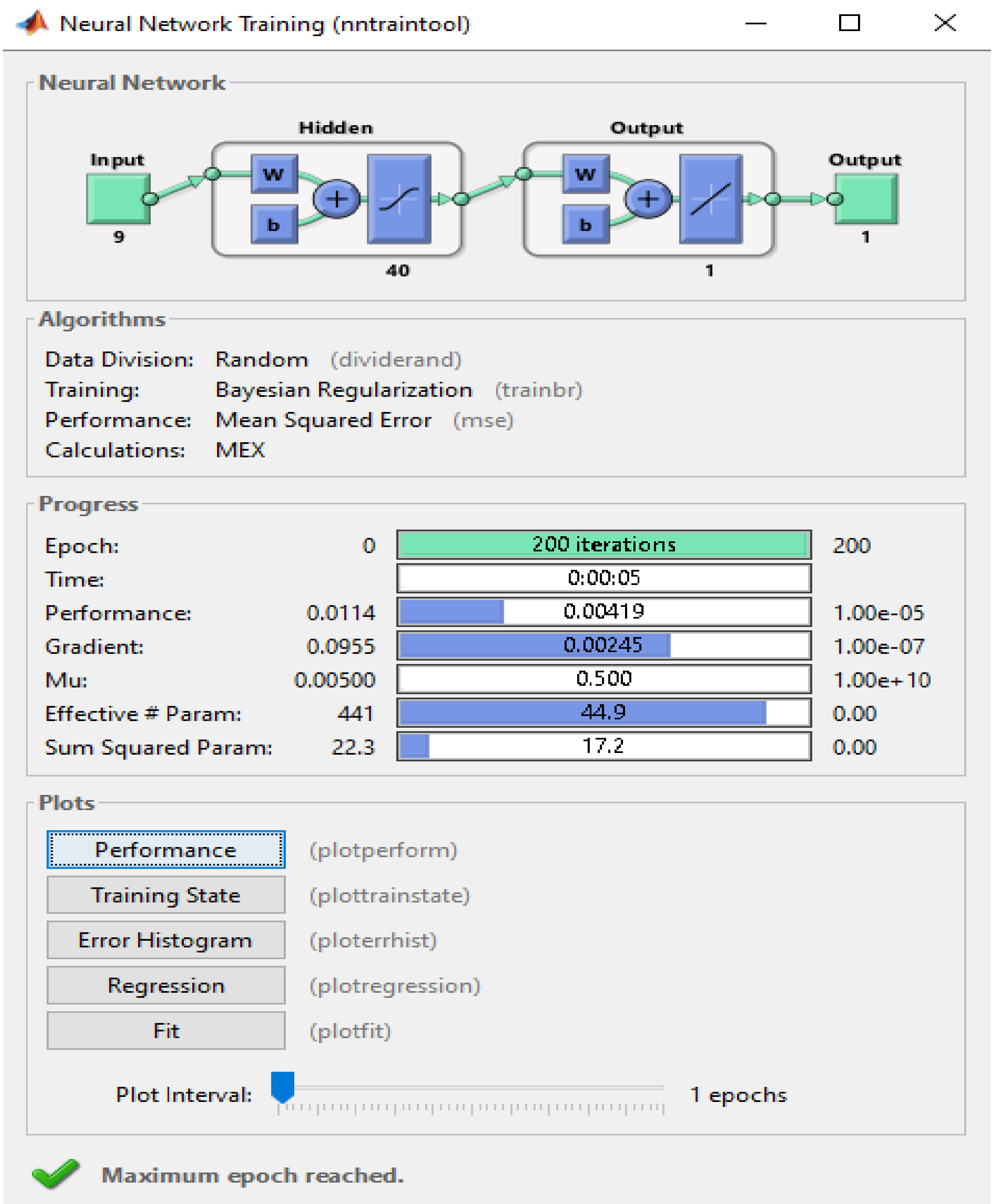


Figure 11 A Neural network windows during training

Before using a model for prediction some neural network parameter has to be set and to set optimum parameters, possible options should be searched for the given data set. For this study gird search hyper parameterization was used and presented as below.

4.2.1. Number of neurons in hidden layer

Even though adding more neurons to hidden layers increase the model's capacity to learn more complex patterns and relationship among data, in other hand finding optimal number of hidden layers for particular problem is more realistically and important in order to avoid both under and over fitting condition.

In this study initially the standard neural network structure of single input, hidden and output layer was used with the number of neurons in hidden layer varied from 1 to 300 to identify the best performing configuration. Here, MSE value was used to measure the performance of the model with different number of neurons in hidden layer and the result is shown below (Table 6).

Table 7 MSE result for possible number of neurons in hidden layers

Neurons	1	10	20	30	40	50	60	70	80	90	100	150	200	250	300
MSE	6.9	11.19	16.82	8.22	5.90	15.28	28.25	18.43	18.43	24.08	29.96	15.91	59.44	136.75	257.05

As shown above in the table the as the number of neurons in hidden layer increase the MSE value oscillates up to some point but above some large number 200 in this case shows linear increment as displayed below in the figure. The Best number of neurons for the given data set is 40 with MSE value of 5.90.

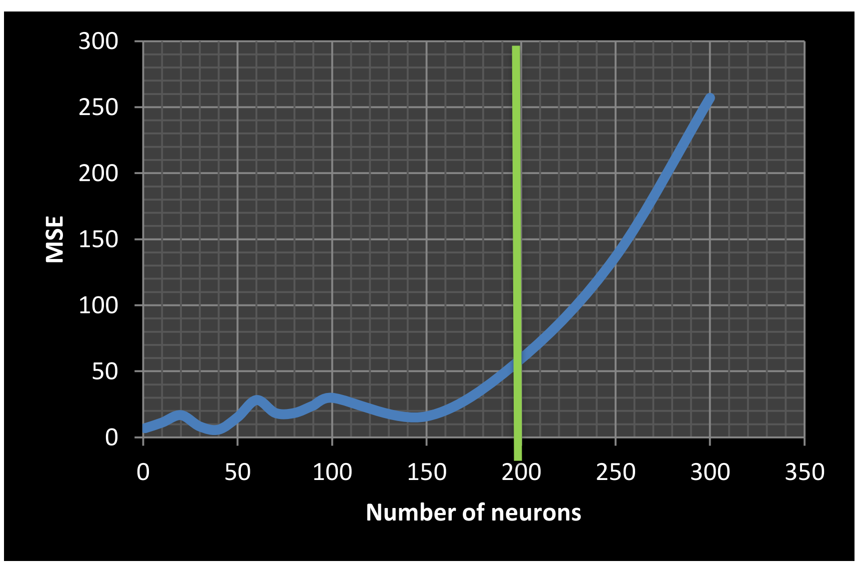


Figure 12 MSE value for possible number of neurons

4.2.2. Learning algorithm of neural network

Taking the best number of neurons, the possible neural network learning algorithm (Levenberg-Marquardt (LM) Algorithm (trainlm), Bayesian regularization back propagation (trainbr) and scaled conjugate gradient back propagation (trainscg) were searched and the following MSE values were achieved.

Table 8 MSE value for possible training algorithm

Training algorithm	MSE value
trainlm	15.904
trainscg	23.081
trainbr	10.471

As shown in Table 7 Bayesian regularization back propagation training algorithm has the lowest MSE value than the two for the given data set and 40 neurons in hidden layer.

4.2.3. Activation function

Transfer function in hidden layer and output layer is the key for capturing nonlinearity nature of the data in the model. In this regard 3 activation function hyperbolic tangent sigmoid transfer function (tansig), Log-sigmoid transfer function (logsig) and linear transfer function (purline) were examined for hidden layer and the following result was found.

Table 9 MSE value for possible activation function

Activation function	MSE
tansig	14.446
logsig	15.524
purelin	14.927

Based on the above table hyperbolic tangent sigmoid transfer function (tansig) activation function in the hidden layer better perform than the two for the given data set and Bayesian regularization back propagation training algorithm.

4.2.4. Epoch number

The number of complete passes through entire training data set which is number of epoch is crucial to control how long the model will learn from the given training data set. For this matter the optimum number was searched from different possible numbers and found the following result.

Table 10 MSE value for possible number of epoch

Epochs	MSE
50	**10.212**
100	15.229
150	**16.969**
200	9.0198

Based on the above table at 200 epoch number the model has better performance with MSE value of 9.01.

4.2.5. Water supply-demand gap prediction model

For the prediction purpose the result of hypermeters tuning using gird search from section 4.2.1-4.2.4 was taken and other than specified parameter was taken default value in the application. In addition, the number of input variables varied according to table 4 and studied the effects on the model performance.

For each model the regression plot, error histogram and the prediction graph comparing to the actual are presented below.

Table 11 MSE and coefficient of determination results

model	MSE	R			R^2			Adjusted R^2
		Training	test	All	Training	test	All	
1	0.0278	0.9937	0.9636	0.9887	0.9874	0.9285	0.9775	0.9752
2	0.0435	0.9948	0.9338	0.9859	0.9896	0.8720	0.9720	0.9691
3	0.0192	0.9929	0.9735	0.9897	0.9859	0.9477	0.9795	0.9774

4	0.0219	0.9913	0.9701	0.9879	0.9827	0.9411	0.9759	0.9734
5	0.0172	0.9824	0.9749	0.9810	0.9651	0.9504	0.9624	0.9584
6	0.0472	0.9671	0.9173	0.9615	0.9353	0.8414	0.9245	0.9166
7	0.0187	0.9628	0.9824	0.9622	0.9270	0.9651	0.9258	0.9181
8	0.0661	0.9705	0.9482	0.9618	0.9419	0.8991	0.9251	0.9172
9	0.0209	0.9521	0.9718	0.9531	0.9065	0.9444	0.9084	0.8988

As mentioned above mean squared error (MSE) is the average squared difference between model outputs and targets. Lower values are better and zero means no error. Therefore, based on table 9 result model 5 has lowest MSE value but the respective R^2 (0.9624) value was not higher. This was not expectable because a model with higher R^2 (very close to 1) mean the model has captured the overall data trend very well so that the difference between model output and target would be small which is close to zero (lowest MSE). This may arise from model complexity, more complex model with many parameters, which fit very well with training data (low MSE) possibly over fitting which may not generalize well and result in lower R^2.

Therefore, adjusted R^2 was calculated to choose the best performing model and found that model 3 outperforms the other input combinations and selected for prediction purpose. And comparing adjusted R^2 value, from model 3 to model 9 the value reduced at each model this indicate that the removal of variables was affecting model performance or in other word the removed variable was important for the model.

It is observed that the output tracks the targets very well for a total response of adjusted R^2-value =0.9774 and MSE=0.0192 and found that model 3 (7-input combination) outperforms the other input combinations. In this case, the network response is satisfactory, and simulation can be used for entering new inputs.

Therefore, as shown in the figure above the ANN structural model MLP 7-40-1 proved to be sufficiently accurate to predict water supply-demand gap based on the given data. The training (0.9859) and test (0.9477) performance values shows that the model is able to predict values almost equal to the measured values.

The simulation results of water supply –demand gap is presented in figure 14 by plotting the actual and predicted output variables.

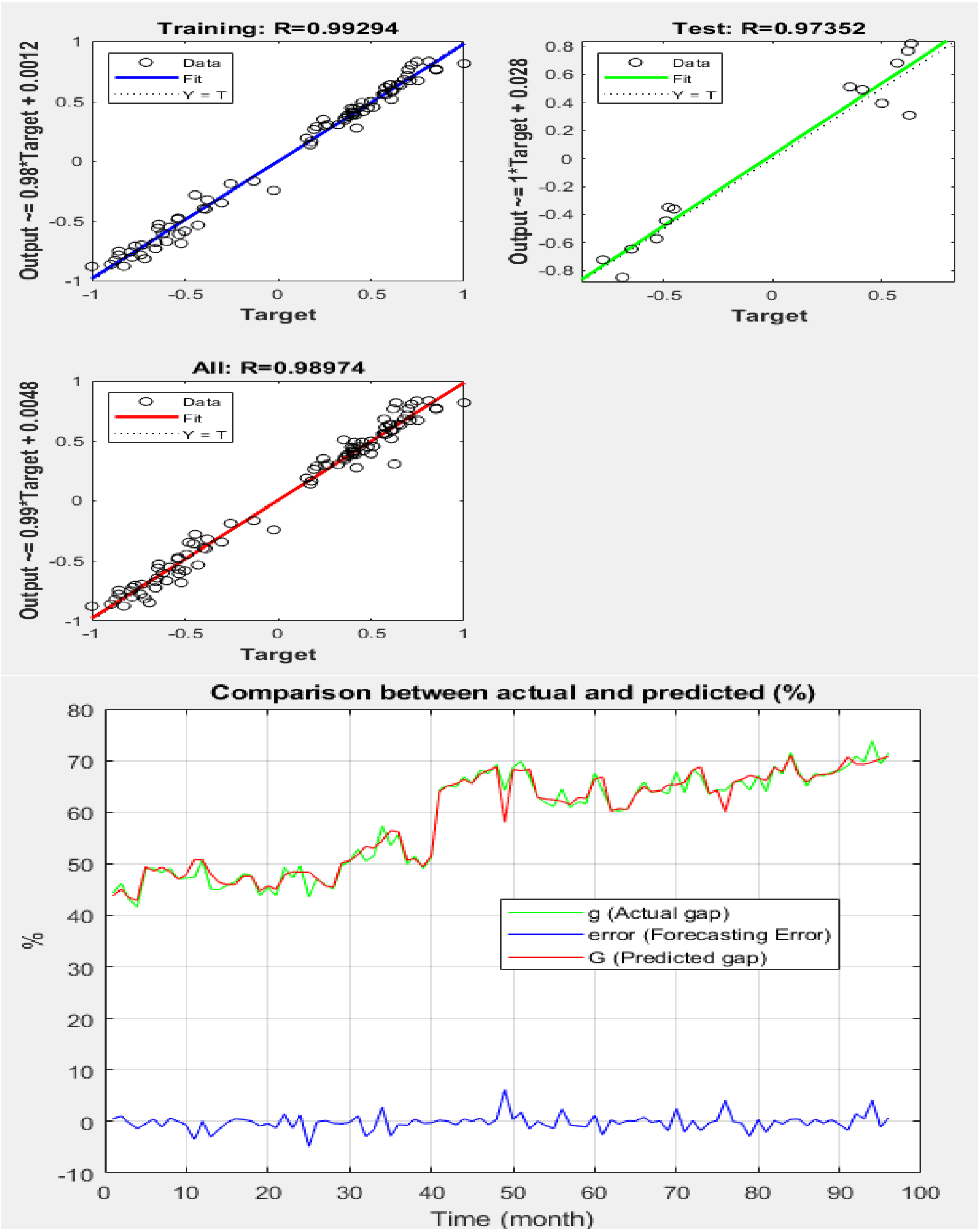


Figure 13 model 3 regression plot and prediction graph

Table 12 Actual gap and simulated/predicted gap for ANN Model 1 from 2022 to 2023 (%)

Date	Actual gap(%)	Predicted gap(%)	Error
31/01/2022	65.77	65.81	-0.04
28/02/2022	66.01	66.30	-0.29
31/03/2022	64.28	67.12	-2.84
30/04/2022	67.09	66.69	0.40
31/05/2022	64.07	66.07	-2.00
30/06/2022	68.96	68.73	0.23
31/07/2022	67.40	67.84	-0.44
31/08/2022	71.46	70.99	0.47
30/09/2022	67.68	67.19	0.49
31/10/2022	65.02	65.80	-0.78
30/11/2022	67.52	67.08	0.44
31/12/2022	67.00	67.32	-0.32
31/01/2023	67.73	67.39	0.34
28/02/2023	67.97	68.46	-0.49
31/03/2023	69.02	70.66	-1.64
30/04/2023	70.81	69.28	1.53
31/05/2023	69.77	69.24	0.53
30/06/2023	73.88	69.68	4.20
31/07/2023	69.32	70.32	-1.00
31/08/2023	71.46	70.74	0.72

As shown above in the table as well as on the figure the prediction errors are more or less moderate or small so the model can be used for prediction purpose.

Therefore, it can be concluded that, to obtain proper estimation of future water supply demand gap the inclusion of past records of water consumption and all-weather parameters and influential factor is necessary.

4.3. Driving Factor Analysis

Here ten water demand or water consumption driving factor were considered and analyzed using LDIM following to kaya identity framework. Even though this method was initially established for decomposing co_2 emission factors it can be adapted for other leading factor which can be decomposed or derived by other factors.

The change in urban water demand is derived by different factors. As mentioned previously for this study only ten factors (Rainfall, sunshine hour, relative humidity, temperature, population, and number of tourists, industrial growth, economic growth livestock number and tariff rate) were considered.

There are other additional water demand driving factor for the considered study area like construction growth rate and ornamental plant plantation but due to resource limitation and data unavailability they became out of consideration for this study. Nowadays in the city the construction development and ornamental plant planting is becoming very high so including these for further study will be important. Consequently, the town water demand will increase.

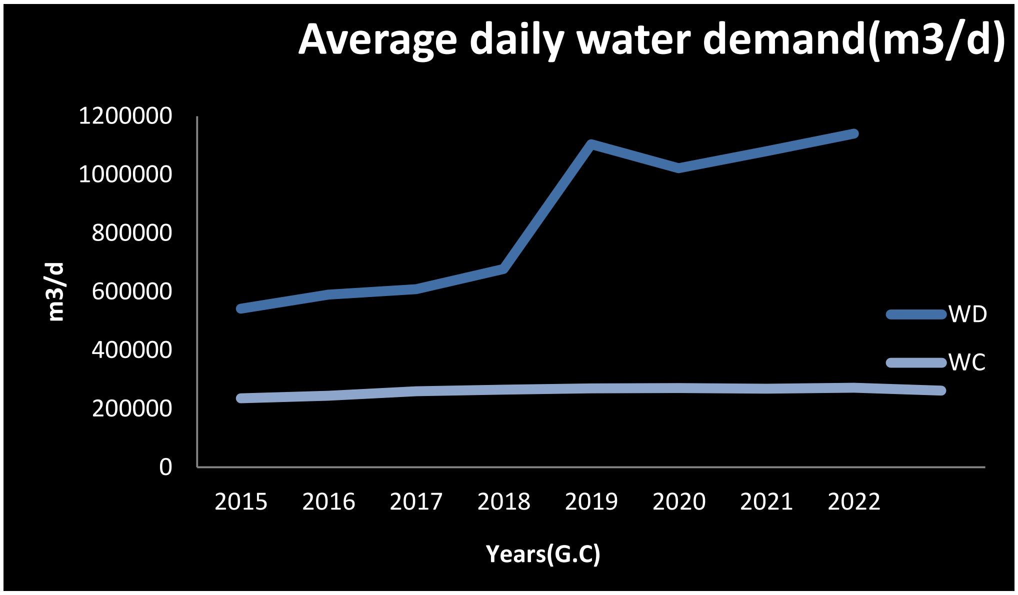


Figure 14 Trend of water demand and consumption resulted from different factors

Figure 25 indicates the trend of water demand for period 2015-2022. The water demand of Addis Ababa increased from 542,278m3/d in 2015 to 1,140,235m3/d in 2022, the percentage of increase in water demand being 54.88 % during the analyzed period. In 2019 there was sharp increase of demand this could be due to covid-19 and the sharp decrease in water demand trend after 2019 could be a situation normalization of post covid-19 era or it could be erroneous data.

Table 13 Data used for decomposition

Year	RF (mm	T (^{0}c)	SS (hr	RH (%)	Pop (no)	To (no)	IG (%)	GR (%)	LN (no)	WT (Br)	WC
2015	3.69	16.2	7.5	52.86	3871000	68079.92	20.5	13.9	563650	9.61	235651
2016	4.64	16.8	5.3	62.71	4040000	66573.58	-6	-45.4	577246	9.61	244834
2017	5.07	16.3	6.8	56.29	4216000	436832.4	0.49	34.5	620864	11.05	259499
2018	5.01	16.2	6.1	59.87	4400000	1155366	9.03	21.4	670139	11.05	265738
2019	5.43	16.7	6.1	57.87	4592000	867033.2	15.8	21.4	719413	12.71	269807
2020	5.25	16.5	6.2	56.53	4794000	385377.7	8.8	-12.0	757276	12.71	271157
2021	4.04	16.2	7	53.34	5006000	453440.6	7.2	18.6	773700	24.37	269040
2022	3.44	16.7	6.6	57.68	5228000	534608.3	10.1	11.3	786486	28.26	271540

4.3.1. Contributing factors

The growth of Water demand which was in turn shows the growth of the gap between water supply and demand is influenced with different factors and these factors are decomposed according to kaya equation and the result was shown below in table 6.

For the purpose of describing the driving forces of water demand, this subsection analyzed the contribution of RF, T, SH, RH, Po, To, IG, EG, LN, and WT. Table 11 presents the results, calculated by the LMDI technique in formula (24)–(33). Between them, the positive values of factors except rainfall and water tariff indicate the incremental effect on water demand. On the contrary, the negative values indicate the restraining effect.

Table 14 Decomposition - change in % of water demands attributable to each factor, 2015 to 2022

year	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22
ΔWD	9183	14665	6240	4069	1349	-2117	2500
	3.90%	5.99%	2.40%	1.53%	0.50%	-0.78%	0.23%
RF	54839	22589	-3398	21743	-9160	-70496	-43251
	138%	8%	3%	2%	9%	-24%	6%
T	8444	-8658	-1714	8955	-3614	-5411	9017
	4%	-4%	-1%	3%	-1%	-2%	1%
SH	-82815	63279	-31512	2186	5356	14408	-15125
	-36%	26%	-12%	1%	2%	5%	-1%
RH	41028	-27196	16186	-9110	-6325	21126	21141
	17%	-11%	6%	-3%	-2%	8%	8%
Po	10265	10750	11218	11437	11644	11688	11728
	4%	4%	4%	4%	4%	4%	4%
TN	-5375	474252	255416	-76875	-219321	43929	44508
	-2%	194%	98%	-29%	-81%	16%	17%
IG	-725546	981108	765208	149805	-158302	-54201	93347
	-308%	401%	295%	56%	-59%	-20%	35%
GR	-633503	893198	-125485	0	-240866	790407	-133427
	-269%	365%	-48%	0%	-89%	291%	-50%
LN	5726	18363	20056	18998	13874	5795	4430
	2%	8%	8%	7%	5%	2%	2%
WT	0	35199	0	37476	0	175823	39996
	0%	14%	0%	14%	0%	65%	15%

Here as shown in the above table the sum of each driving factor effect is not equal to the total change of water consumption. Theoretically it should become equal as equation 11 but not always; this could be because of interactions and dependencies between the factors. Additionally, measurement errors, data inaccuracies or assumption taken during decomposition process can also contribute to discrepancies between the total change and sum of change in individual factors. One of the assumptions LMDI method assume is that the relation between the factors and the main variable are additive in log space. However, in practice these relationships may be nonlinear, leading to discrepancies between the total change and sum of factor contributions.

Contribution of each factor over consecutive year was calculated and presented below. From 2015 to 2016 water consumption has increased. As shown in figure 26 Industrial growths, economic growth, Rainfall, Relative humidity, tourist number, and sunshine hours were the inhibiting factors in changing water consumption whereas Temperature, population and livestock number were the promoting factors changing water consumption. Meanwhile the major contributing factors were population and temperature and the major inhibiting factor was industrial growth.

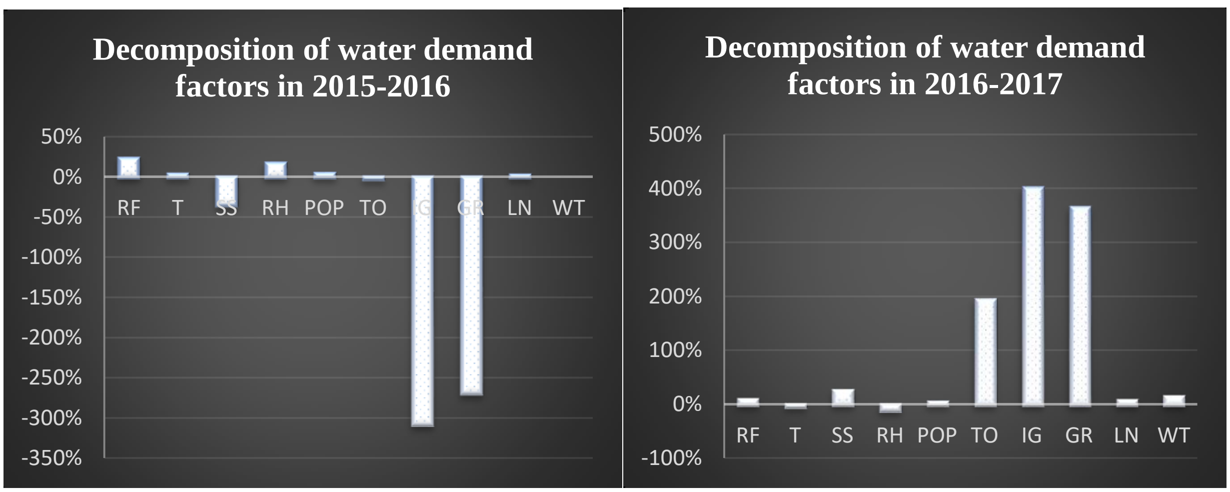


Figure 15 Decomposition of water demand factors in 2015-2016

Climatic effects

ΔWD, which denotes water consumption change, affected positively by temperature during 2015-2016, 2018-2019, and 2021-2022 years, on the other period it was inhibiting factor. Rainfall contribute positively for water usage increment only during period of 2020-2021 and sunshine hours affect negatively during period of 2015-2016, 2017-2018, 2021-2022 and 2022-2023. Relative humidity contributes positively for water usage increment only during period of 2016-2017, 2018-2019 and 2019-2020.

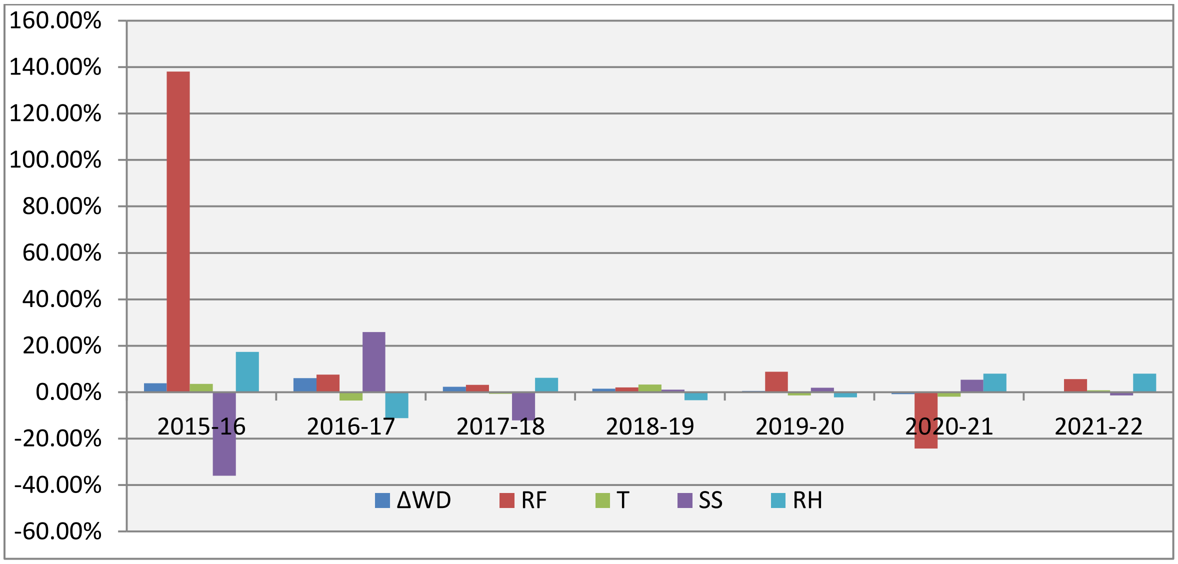


Figure 16 Contribution of meteorological factors in water consumption

Population effect

As shown in figure 28, during all periods from 2015-2022 the contribution of population number was promoting factor. Number of tourist during 2016-2017 showed large increment which was promoting factor for water usage change. In addition during 2017-2018, 2020-2021 and 2021-2022 had promoting factor for water usage. Similar to population, livestock number showed promoting effect throughout all period 2015-2022.

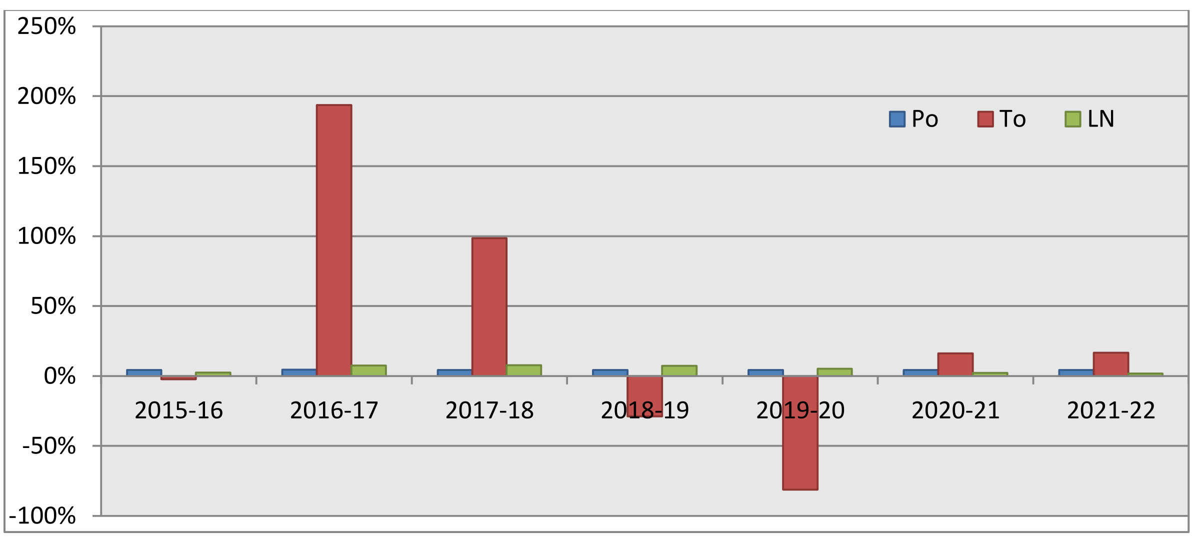


Figure 17 Contribution of population, tourist and livestock factors in water consumption

Economic effect

In figure 29 the contribution of three factors were presented and industrial growth had promoting effect during 2016-2019 and economic growth had promoting effect during 2016-17 and 2020-21. Water tariff shows increment from 2015 to 2022 which inhabited factor for water usage throughout that period.

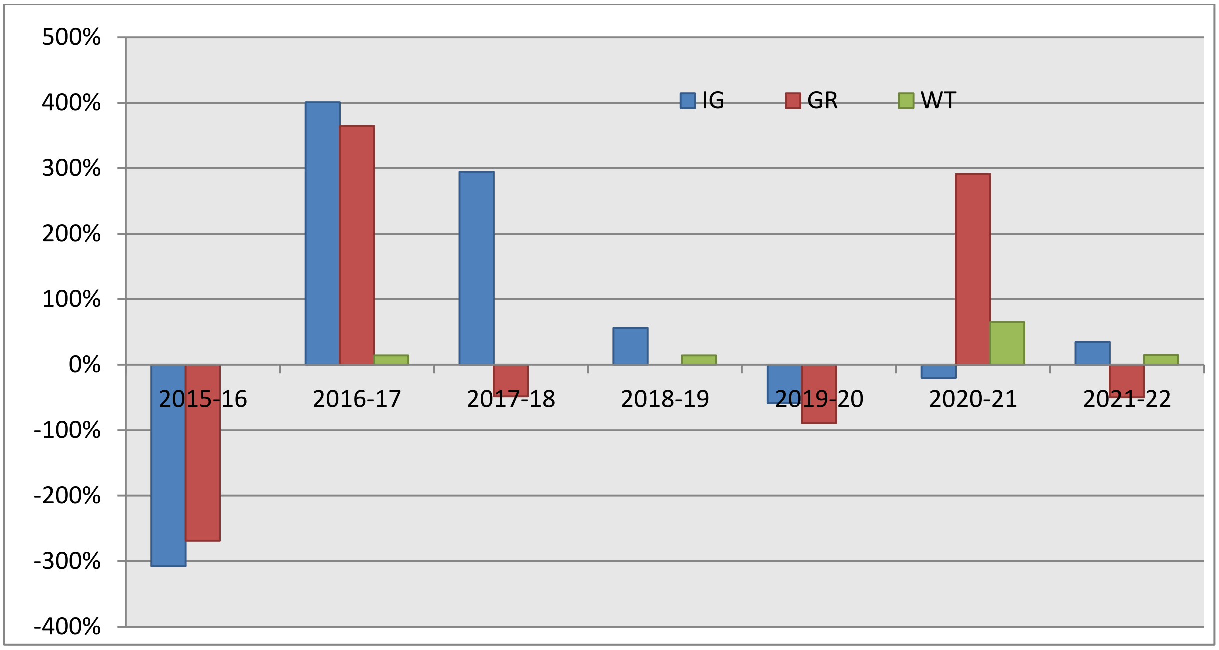


Figure 18 Contribution of Industrial growth, economic growth and water tariff in water consumption

Overall from the analyses it was observed that each factor had different amount of contribution for water usage except population and livestock number on some period had promoting and, on some period, inhibiting effect.

From table 11 for single period it could be drawn which factor has high contribution and which has less for instance the decomposed driving factor contribution for water consumption during 2018-2019 periods is shown below and tourist number contribution rate was leading factor for water usage increment, following to that economic growth, livestock number and population number contribute for water consumption change.

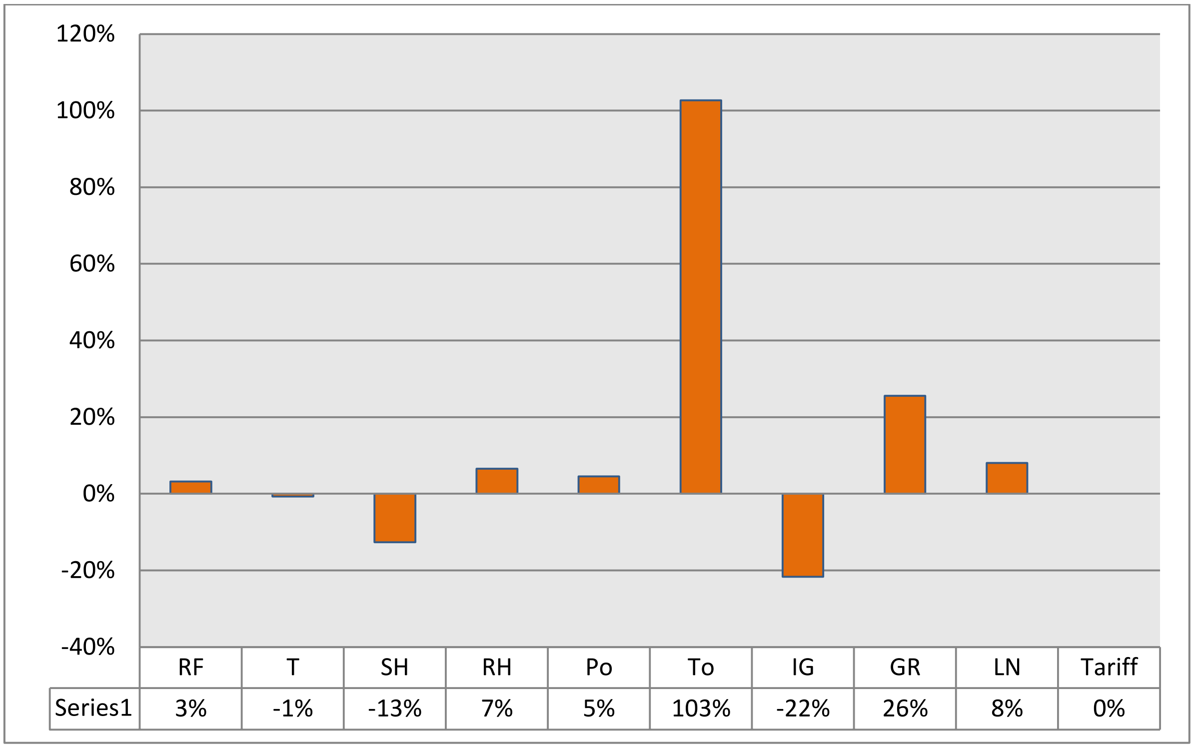


Figure 19 Decomposition of water demand from 2018-2019

5. Conclusion and recommendation

5.1. Conclusion

After the application of Artificial Neural Networks as a forecasting technique to develop a model that forecast urban water supply demand gap, where dependent and independent variables are analyzed and processed in building the neural network through the processes of training, validation and testing, obtaining results that behave with greater similarity to actual demand supply gap the following conclusion are drawn for this study

- Water demand in Addis Ababa is increasing, requiring more water production for supply. Despite many improvements, the current system provides only 58 liter/capita/day. Unmet demand is expected and the city need improvement to meet the projected demands
- As a result of correlation analysis, the correlation coefficient between livestock number and population become 0.935 and this showed that there was multi collinearity. Consequently, livestock number removed from independent variables.
- The applied automatic hyper parameter tuning (grid search) for activation function, training algorithm, number of neurons in hidden layer and epoch number resulted with 40 number of neurons in hidden layer, Bayesian regularization back propagation training algorithm, hyperbolic tangent sigmoid transfer function (tansig) activation function and 200 epoch number found to be best for neural network creation.
- All models' coefficient of determination (R^2-value) between the actual and predicted output variables is greater than 0.9, from which we could interpret that there exists a strong positive relationship among the variables. Therefore, the model developed in this work has an acceptable generalization capability and accuracy. As a result, the neural network modeling could effectively simulate and predict the water supply –demand gap of Addis Ababa city for the given variables.
- Out of different models, model-3 with 7 number of independent variables showed to have better adjusted R^2 value, this prove that artificial neural network is data driven model because it learns from the input data provided to it. The more data provided the better learning will be.

- Generally, it is concluded that, ANN provides an effective analyzing and diagnosing tool to understand and simulate the non-linear behavior of the water supply-demand gap and can be used as a valuable performance assessment tool water provider and decision makers.
- The ten components of the change in water demand analyzed in this study were rainfall (ΔRF), temperature (ΔT), sunshine hours (ΔSS), relative humidity (ΔRH), population (ΔPop), industrial growth(ΔIG), economic growth(ΔEG), tourist number(ΔTN), livestock number(ΔLN) and water tariff(ΔWT)). Among them two positive factors (ΔPop and ΔLN) have the expanding effect on water consumption throughout all periods while the other factors had the inhibiting and expanding effect in different period of the considered years.
- From 2015–2016, population and temperature were the major factors impacting the increase of total water consumption while industrial growth had inhibiting effect.
- From 2021-2022, industrial growth, tourist number, population and temperature have expanding effect while other have inhibiting effect.
- Generally, for driving factor analysis LMDI method could be a good factor decomposer provided that data interaction and independency is absent.

5.2. Recommendation

As a result of this study the following recommendation are drawn.

- Severe weather and population explosion will be most common in the coming future, so it is very essential to conduct more studies related to different data analytic techniques and ANN techniques that could ease the simulation of urban water demand predictive models.
- Sufficient data unavailability and accuracy is the major challenge faced during this research work. So, government agencies that are responsible in water supply and water resource management should necessarily provide free access to reliable data through online resources to the public so that the development of these kinds of models will become easier.
- Based on my literature review, I also recommend further studies on data pre-processing and use of hybrid models like as BSA-ANN and CSA-ANN for water supply- demand prediction.
- One of the major constraints of the study was the lack of systematic data availability. If longer dataset were available, the result of the model could have been better. Therefore, it is

recommended that AWSA should also emphasize on the collection and management of data which can aid in analysis and development of water demand in the future.

- This study provides a first water demand decomposition analysis using the LDMI method in Addis Ababa. Due to the difficulties in data acquisition and a number of presumption analyses, the results may not be conclusive. Therefore, future studies must focus on a more comprehensive analysis of all environmental and economic factors.

6. Reference

1. Leon, L.P., B. Chaplot, and A. Solomon, Water consumption forecasting using soft computing–a case study, Trinidad and Tobago. Water Supply, 2020. **20**(8): p. 3576-3584.
2. Davis, M.L., Water and wastewater engineering: design principles and practice. 2010: McGraw-Hill Education.
3. Muhammad, A.U., X. Li, and J. Feng. Artificial intelligence approaches for urban water demand forecasting: a review. in Machine Learning and Intelligent Communications: 4th International Conference, MLICOM 2019, Nanjing, China, August 24–25, 2019, Proceedings 4. 2019. Springer.
4. Joshua, I.D., et al., ANALYSIS OF COMMUNITY-BASED PATTERN OF WATER DEMAND AND SUPPLY. Journal of Environmental Science and Sustainable Development. **6**(2): p. 3.
5. Arampatzis, G., et al. A water demand forecasting methodology for supporting day-to-day management of water distribution systems. in Proceedings of the 12th International Conference “Protection & Restoration of the Environment”, Skiathos, Greece. 2014.
6. An, M., et al., The gap of water supply—Demand and its driving factors: From water footprint view in Huaihe River Basin. Plos one, 2021. **16**(3): p. e0247604.
7. Adamowski, J., et al., Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resources Research, 2012. **48**(1).
8. Wang, Z., et al., Analysis of the water demand-supply gap and scarcity index in lower Amu Darya River Basin, Central Asia. International Journal of Environmental Research and Public Health, 2022. **19**(2): p. 743.
9. Billings, R.B. and C.V. Jones, Forecasting urban water demand. 2011: American Water Works Association.
10. Boryczko, K. Forecasting water use in selected water supply system. in E3S Web of Conferences. 2017. EDP Sciences.
11. Station, C.I.A.s.O.C.o.C., Water Demand Forecasting. October 2016, City Internal Auditor’s Office.
12. Truneh, D. and N. Getachew, Ethiopia - Addis AbabaWater and Sewage Authority (AAWSA), A. Ababa, W.a. Sewage, and A. (AAWSA), Editors.: Addis Ababa.
13. JICA, <THE PROJECT FOR STRENGTHENING ADDIS ABABA WATER AND SEWERAGE AUTHORITY’S MANAGEMENT CAPACITY OF NON-REVENUE WATER REDUCTION IN FEDERAL DEMOCRATIC REPUBLIC OF ETHIOPIA>. February 2023, JICA.
14. Alemu, Z.A. and M.O. Dioha, Modelling scenarios for sustainable water supply and demand in Addis Ababa city, Ethiopia. Environmental Systems Research, 2020. **9**: p. 1-14.
15. Moreda, T.A.a.F., <Addis Ababas Chronic Urban Water Supply The Ticking Time Bomb.pdf>. ASPIRE, 2023.
16. Moreda, D.T.A.a.D.F., <Addis Ababas Chronic Urban Water Supply The Ticking Time Bomb.pdf>. we aspire, 2023.

17. Guidelines, W.W.P.F., CHAPTER 5 WATER DEMAND FORECASTING., in WATER DEMAND FORECASTING.February 2002.

18. Haque, M.M., et al., Assessing the significance of climate and community factors on urban water demand. International Journal of Sustainable Built Environment, 2015. **4**(2): p. 222-230.
19. Li, Y., S. Wang, and B. Chen, Driving force analysis of the consumption of water and energy in China based on LMDI method. Energy Procedia, 2019. **158**: p. 4318-4322.
20. Lobbrecht, A., Y. Dibike, and D. Solomatine, Applications of neural networks and fuzzy logic to integrated water management project report. IHE, Delft, 2002.
21. basin, I.C.o.t.P.R., Water Demand forcasting.pdf.
22. Msiza, I.S., F.V. Nelwamondo, and T. Marwala, Water demand prediction using artificial neural networks and support vector regression. 2008.
23. Kassahun, G.S., Predictive modelling of Kaliti wastewater treatment plant performance using Artificial Neural Networks. Chemical Engineering with Specialization in Environmental Engineering, 2012.
24. Janssen, T., Hyperparameter tuning for Artificial Neural Networks, in Department of Mechanical Engineering. 2022, Eindhoven university of technology: Eindhoven,.

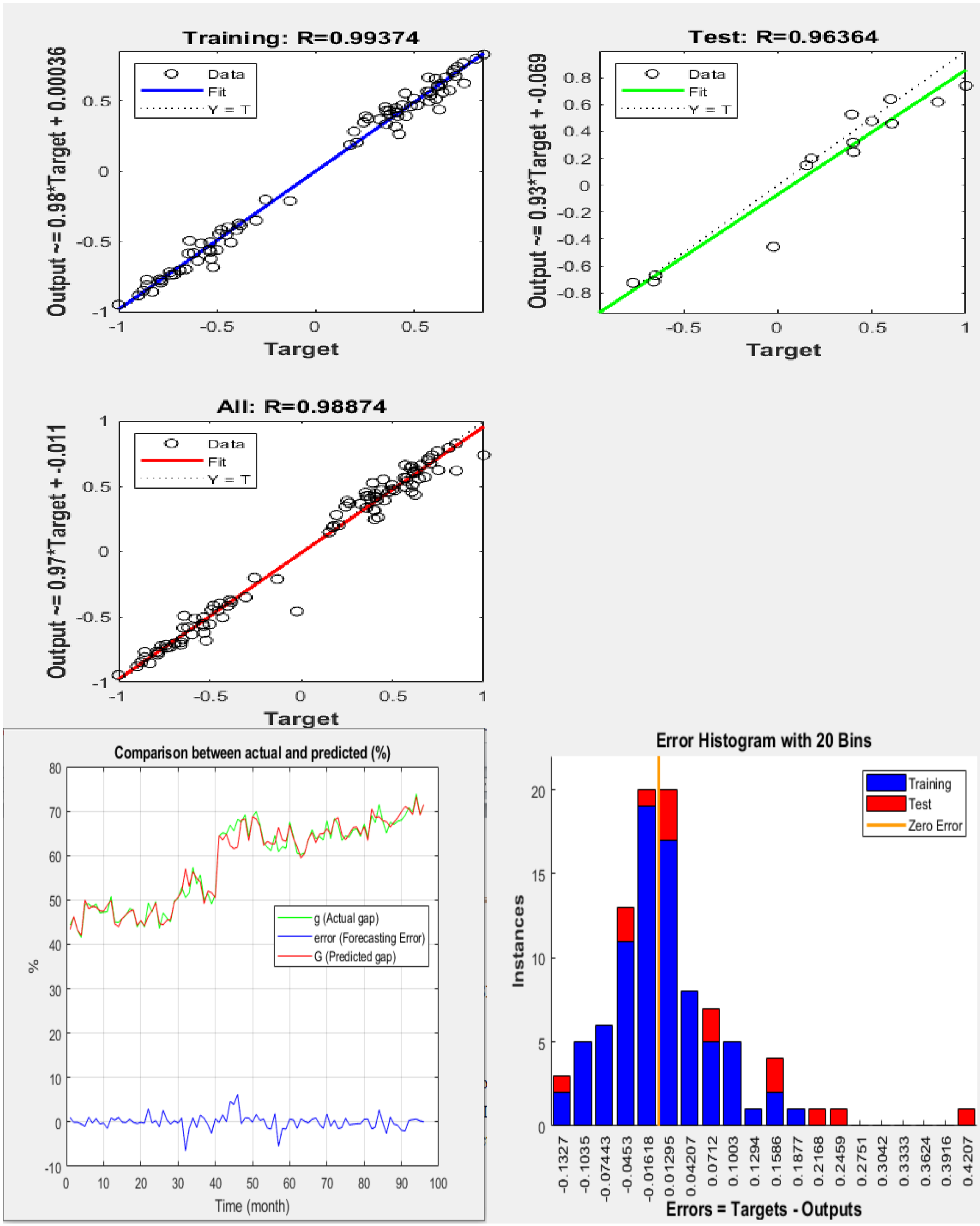
25. O'Reilly, G., C. Bezuidenhout, and J. Bezuidenhout, Artificial neural networks: applications in the drinking water sector. Water Science and Technology: Water Supply, 2018. **18**(6): p. 1869-1887.
26. Hristev, R., The ANN book. 1998, Edition.
27. Bishop, C.M. and N.M. Nasrabadi, Pattern recognition and machine learning. Vol. 4. 2006: Springer.
28. Basheer, I.A. and M. Hajmeer, Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods, 2000. **43**(1): p. 3-31.
29. Chen, Y., et al., A review of the artificial neural network models for water quality prediction. Applied Sciences, 2020. **10**(17): p. 5776.
30. Shirkoohi, M.G., M. Doghri, and S. Duchesne, Short-term water demand predictions coupling an artificial neural network model and a genetic algorithm. Water Supply, 2021. **21**(5): p. 2374-2386.
31. de Souza Groppo, G., M.A. Costa, and M. Libânio, Predicting water demand: A review of the methods employed and future possibilities. Water Supply, 2019. **19**(8): p. 2179-2198.
32. Awad, M. and M. Zaid-Alkelani, Prediction of water demand using artificial neural networks models and statistical model. International Journal of Intelligent Systems and Applications, 2019. **11**(9): p. 40.
33. Lorente-Leyva, L.L., et al. Artificial neural networks for urban water demand forecasting: a case study. in Journal of Physics: Conference Series. 2019. IOP Publishing.
34. Alnajjar, H.Y. and O. Üçüncü, Performance prediction and control for wastewater treatment plants using artificial neural network modeling of mechanical and biological treatment. Archives of Environmental Protection, 2023. **49**(2).
35. AAWSA, ADDIS ABABA WATER AND SEWERAGE AUTHORITY BUSINESS PLAN 2011 - 2020 FINAL REPORT. 2011, AAWSA: Addis Ababa
36. Malato, G., Hyperparameter tuning. Grid search and random search. Your Data Teacher, 2021. **19**.

37. Zhao, T. and B. Shao, Domestic water consumption and its influencing factors in the yellow river basin based on logarithmic mean divisia index and decoupling theory. Sustainability, 2022. **14**(24): p. 16866.
38. Tang, M., Driving Forces of Industrial Water Pollutant Emission from Spatial-Dynamic Perspective in China: Analysis Based on Kaya Equation and LMDI Decomposition. Open Journal of Social Sciences, 2019. **7**(10): p. 132.
39. Bektas, A., Log- Mean Divisia Index Method. Ministry of Energy and Natural Resources: Turkey.
40. He, H. and R.J. Myers, Log Mean Divisia Index decomposition analysis of the demand for building materials: Application to concrete, dwellings, and the UK. Environmental Science & Technology, 2021. **55**(5): p. 2767-2778.
41. Agency, N.G.i., The Geographic Names Server (GNS) 2012.
42. Machine, W., Census 2007 Tables: Addis Ababa. 2010.
43. Enterprise, O.w.w.d.a.s., <Akaki Draft Design Report - October 23 2014_FINAL.pdf>, in draft. 2014, ADDIS ABABA WATER AND SEWERAGE AUTHORITY.
44. Sime, I., <ADDIS ABABA WATER SUPPLY STAGE III.pdf>. 1998, AAWSA: Addis Ababa.
45. Fan, L., et al., Factors affecting domestic water consumption in rural households upon access to improved water supply: Insights from the Wei River Basin, China. PloS one, 2013. **8**(8): p. e71977.
46. Magar, V., et al., Innovative Inter Quartile Range-based Outlier Detection and Removal Technique for Teaching Staff Performance Feedback Analysis. Journal of Engineering Education Transformations, 2024. **37**(3).
47. Al-Ghamdi, A.-B., S. Kamel, and M. Khayyat, A Hybrid Neural Network-based Approach for Forecasting Water Demand. Computers, Materials & Continua, 2022. **73**(1).
48. Al Shamisi, M.H., A.H. Assi, and H.A. Hejase, Using MATLAB to develop artificial neural network models for predicting global solar radiation in Al Ain City–UAE, in Engineering education and research using MATLAB. 2011, Citeseer.

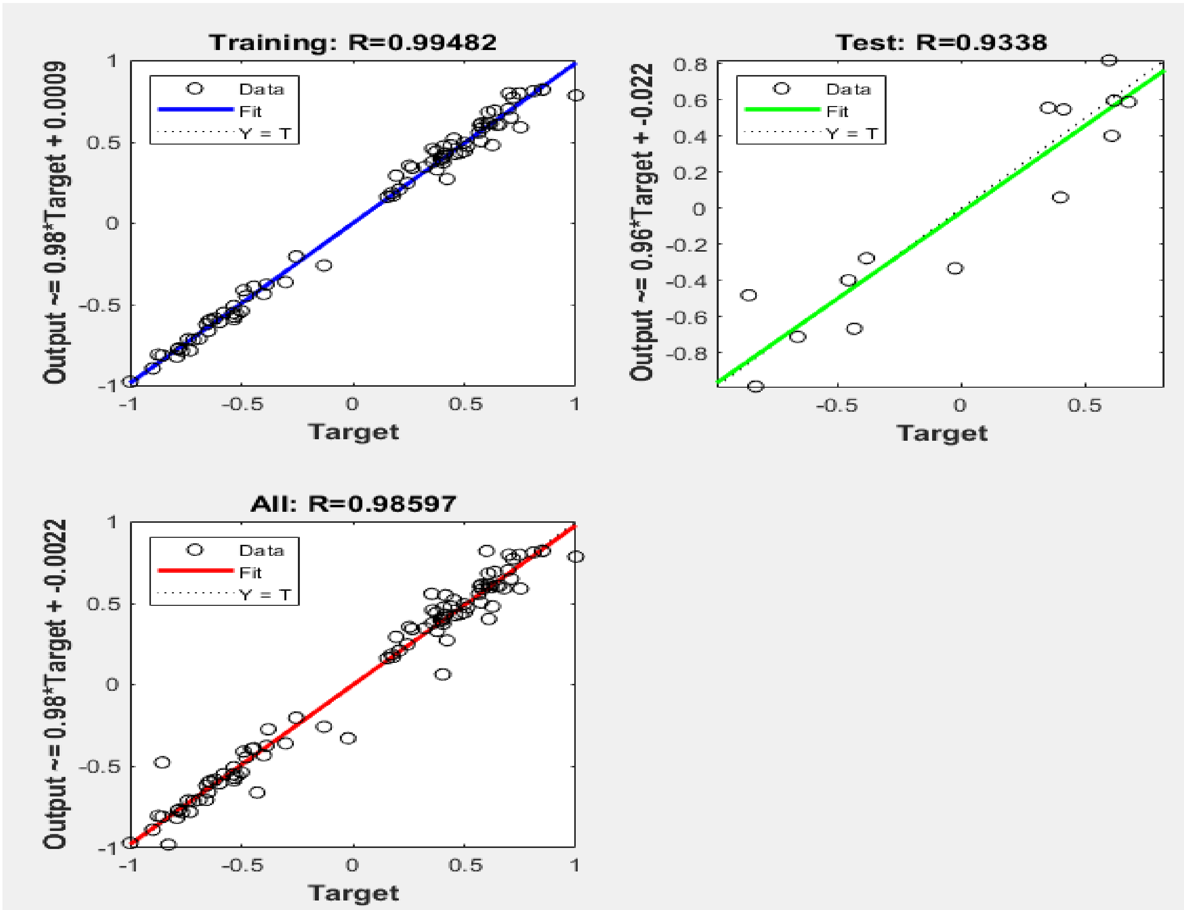
7. Annex

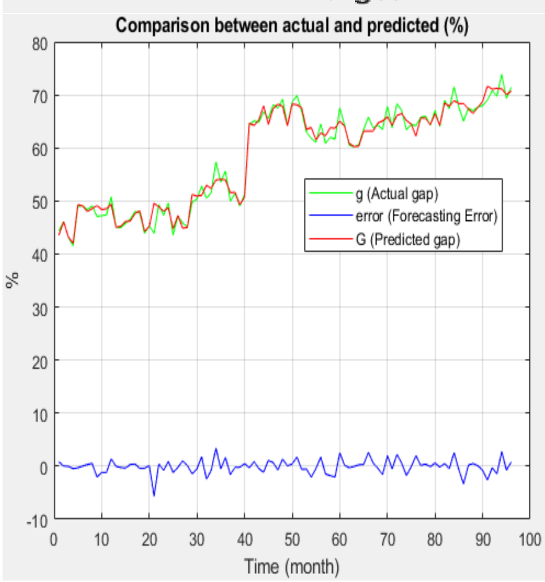
1. All model regression plot

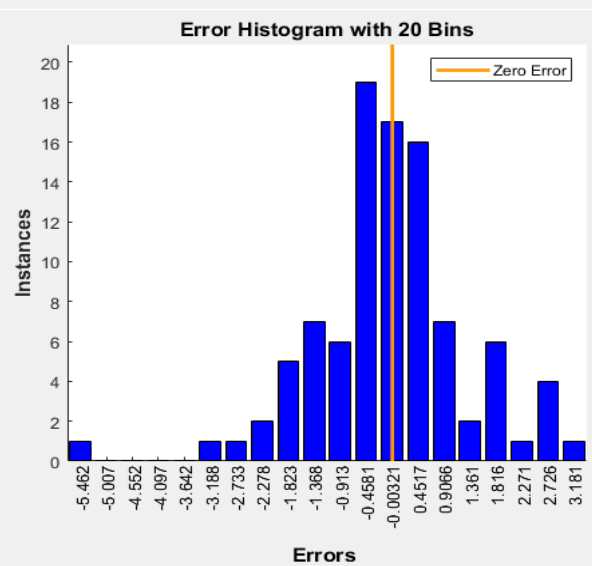
Model-1:-9-40-1



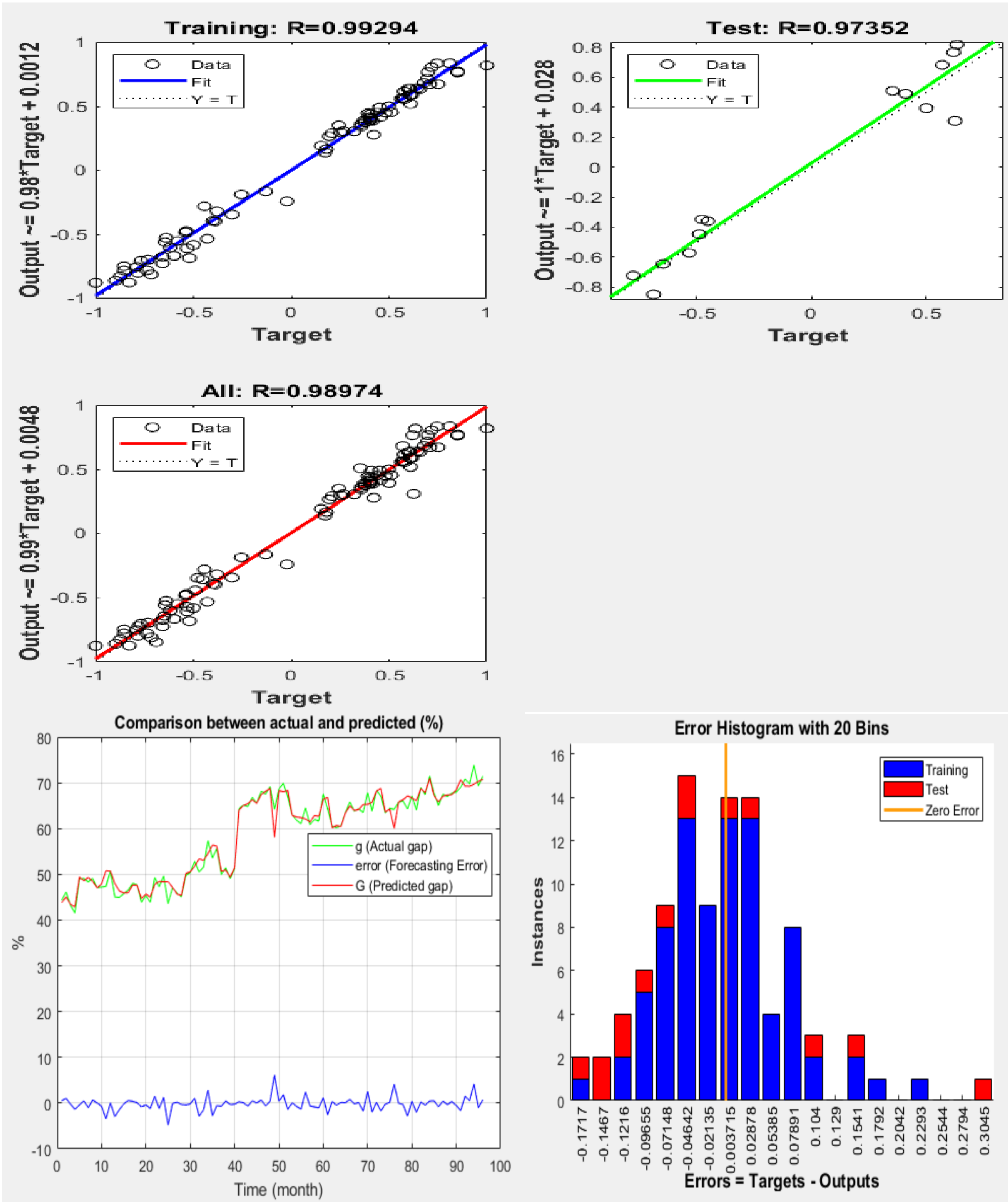
Model-2:-8-40-1


Training: R=0.99482
Output ~= 0.98*Target + 0.0009
Target
Data
Fit
Y = T
Test: R=0.9338
Output ~= 0.96*Target + -0.022
Target
Data
Fit
Y = T
All: R=0.98597
Output ~= 0.98*Target + -0.0022
Target
Data
Fit
Y = T

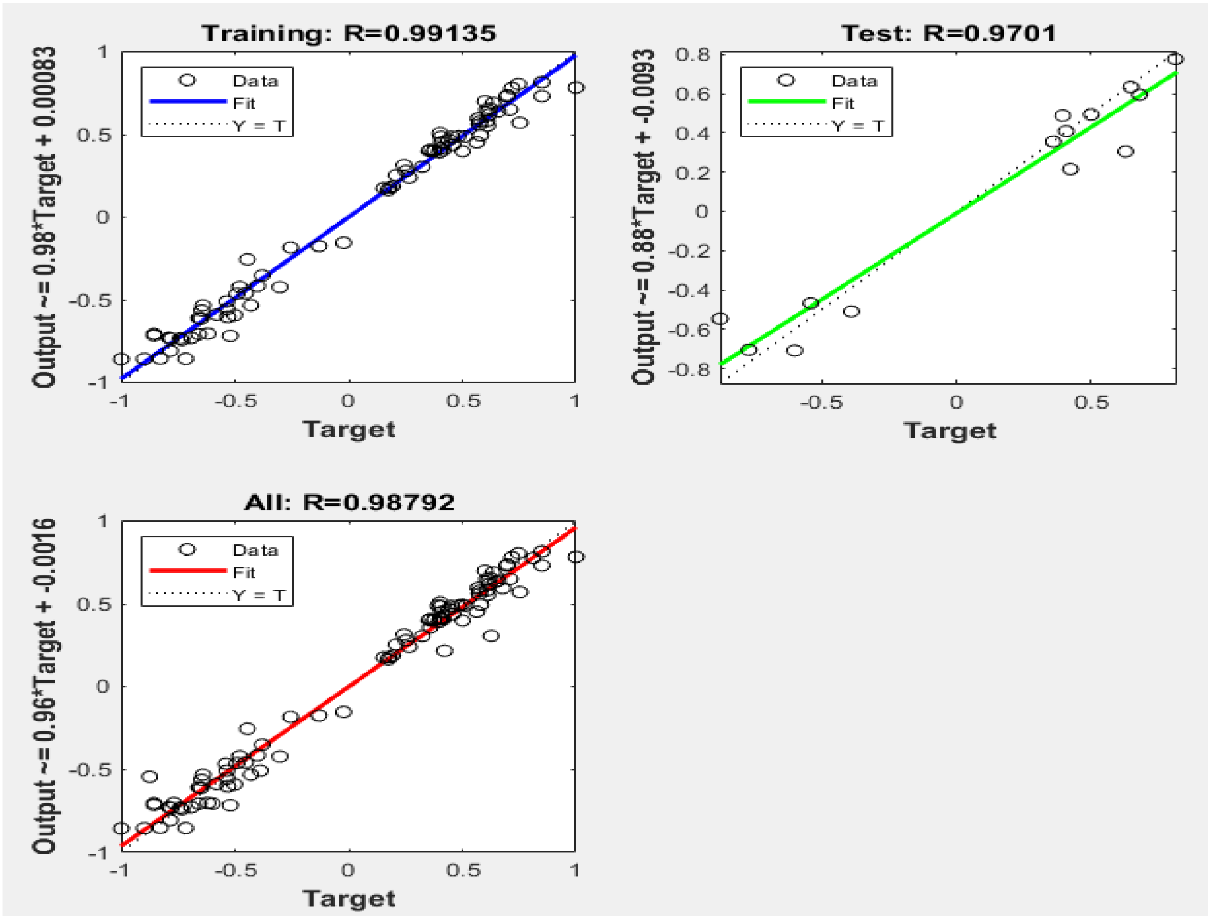

Comparison between actual and predicted (%)
g (Actual gap)
error (Forecasting Error)
G (Predicted gap)
%
Time (month)


Error Histogram with 20 Bins
Zero Error
Instances
Errors
-5.462
-5.007
-4.552
-4.097
-3.642
-3.188
-2.733
-2.278
-1.823
-1.368
-0.913
-0.4581
-0.00321
0.4517
0.9066
1.361
1.816
2.271
2.726
3.181

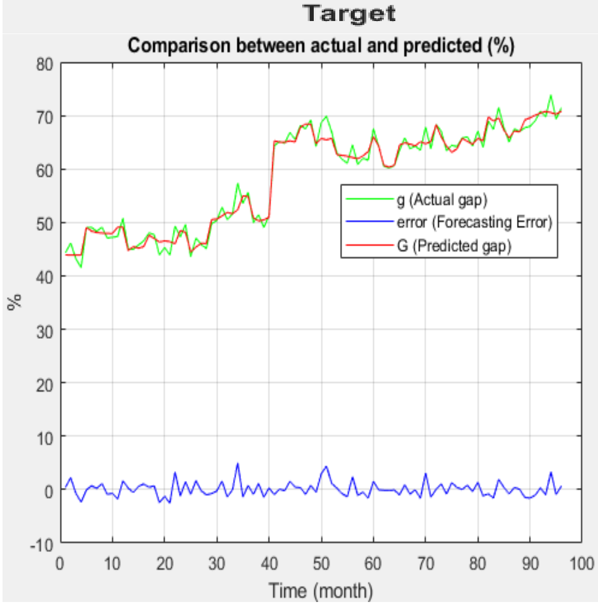
Model-3:-7-40-1


Training: R=0.99294
Output ~= 0.98*Target + 0.0012
Target
Data
Fit
Y = T
Test: R=0.97352
Output ~=1*Target + 0.028
All: R=0.98974
Output ~= 0.99*Target + 0.0048
Comparison between actual and predicted (%)
g (Actual gap)
error (Forecasting Error)
G (Predicted gap)
%
Time (month)
Error Histogram with 20 Bins
Training
Test
Zero Error
Instances
Errors = Targets - Outputs

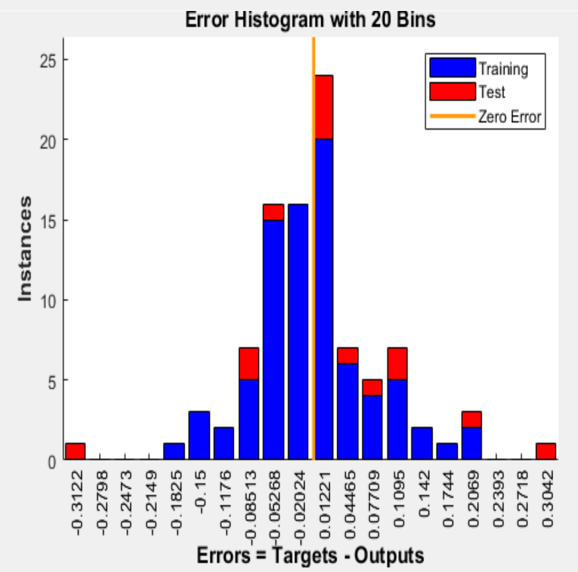
Model-4:-6-40-1



Training: R=0.99135
Output ~= 0.98*Target + 0.00083
Test: R=0.9701
Output ~= 0.88*Target + -0.0093
All: R=0.98792
Output ~= 0.96*Target + -0.0016
Data
Fit
Y = T
Target

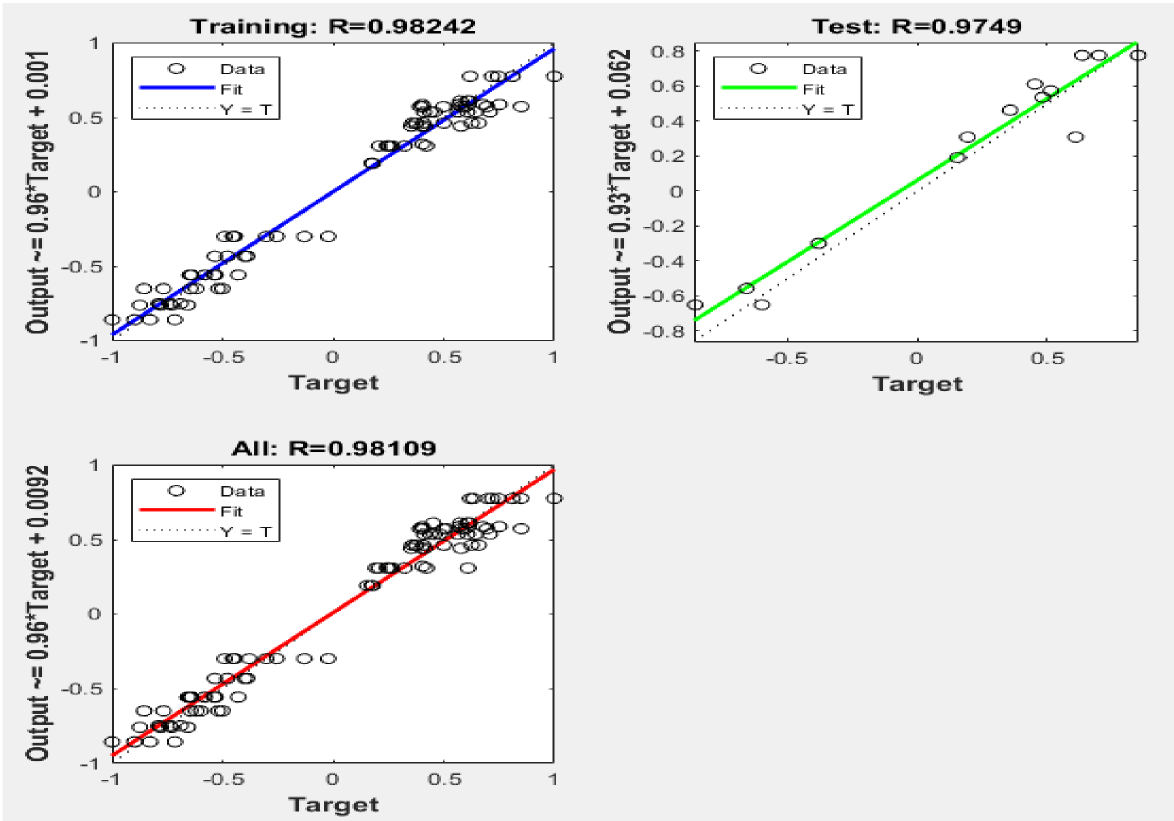


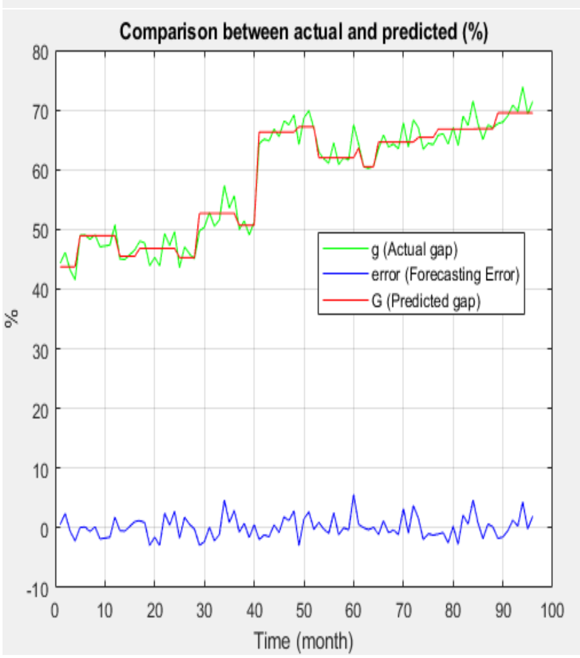
Comparison between actual and predicted (%)
g (Actual gap)
error (Forecasting Error)
G (Predicted gap)
%
Time (month)

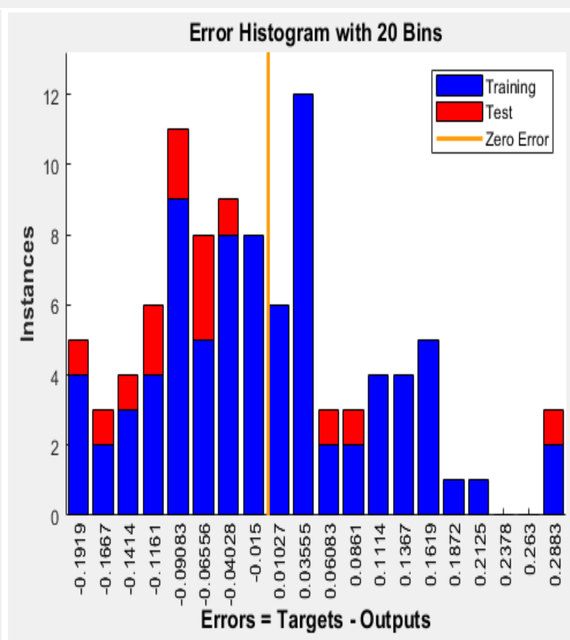


Error Histogram with 20 Bins
Training
Test
Zero Error
Instances
Errors = Targets - Outputs

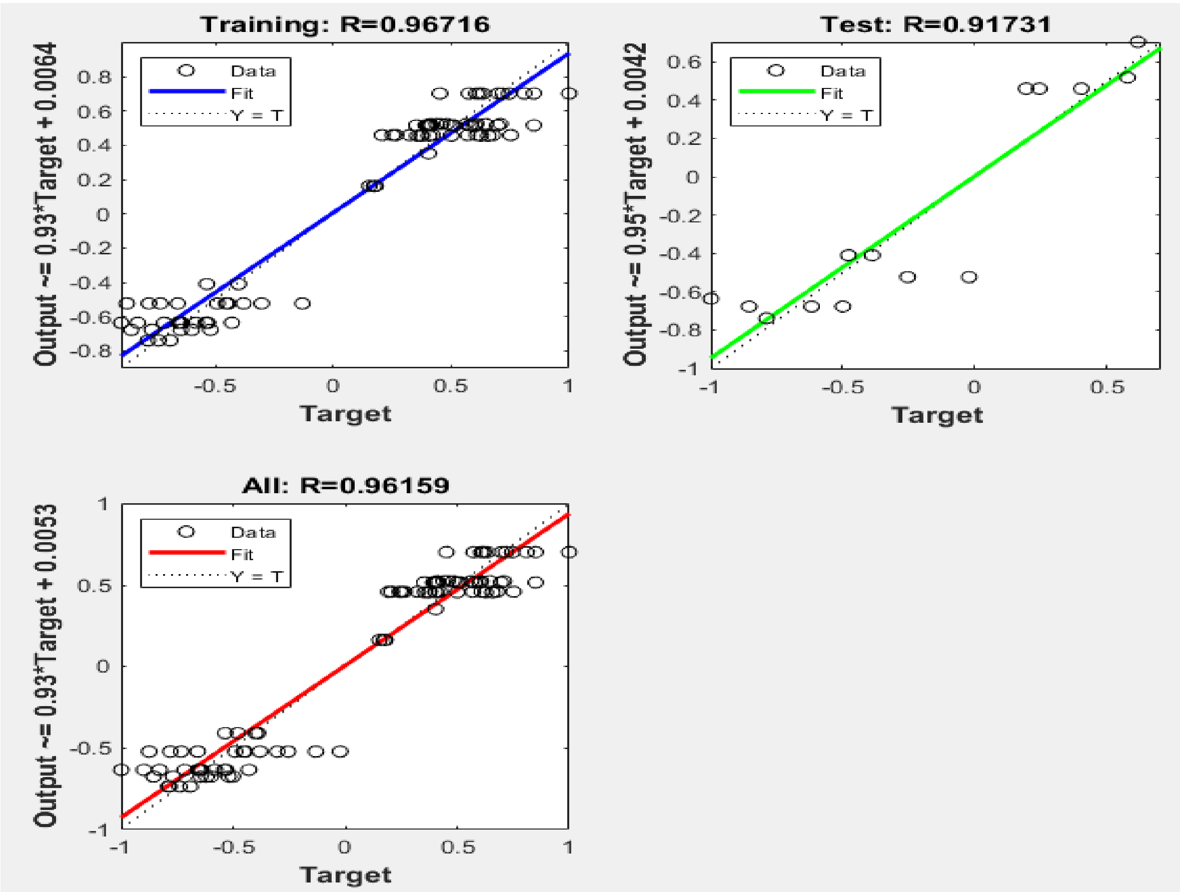
Model-5:-5-40-1


Training: R=0.98242
Output ~= 0.96*Target + 0.001
Target
Data
Fit
Y = T
Test: R=0.9749
Output ~= 0.93*Target + 0.062
Target
Data
Fit
Y = T
All: R=0.98109
Output ~= 0.96*Target + 0.0092
Target
Data
Fit
Y = T


Comparison between actual and predicted (%)
%
Time (month)
g (Actual gap)
error (Forecasting Error)
G (Predicted gap)

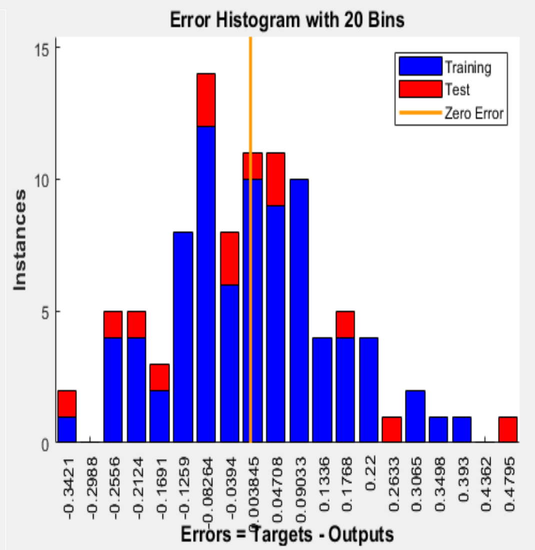

Error Histogram with 20 Bins
Instances
Errors = Targets - Outputs
Training
Test
Zero Error

Model-6:-4-40-1



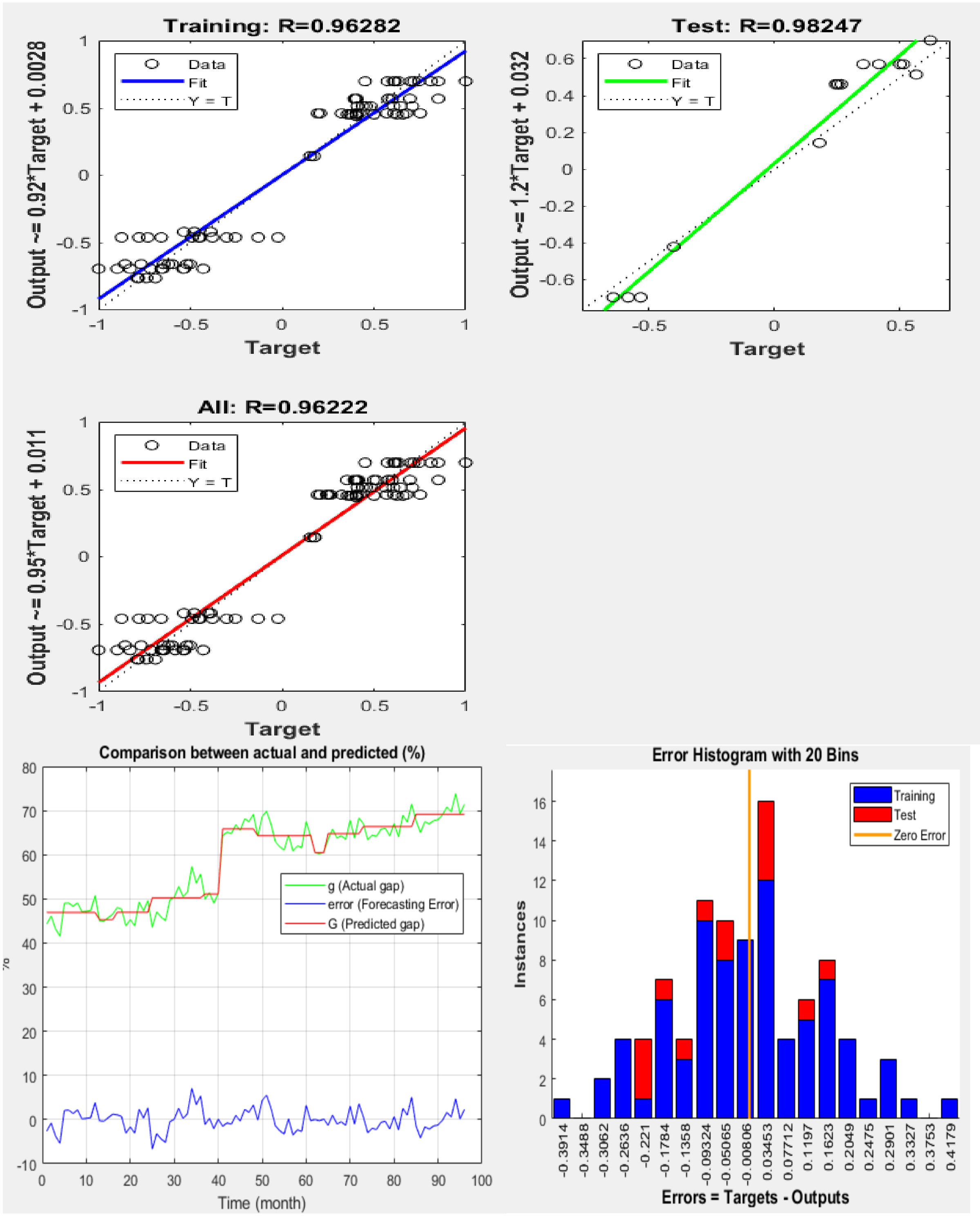
Training: R=0.96716
Output ~= 0.93*Target + 0.0064
Target
Data
Fit
Y = T
Test: R=0.91731
Output ~= 0.95*Target + 0.0042
Target
All: R=0.96159
Output ~= 0.93*Target + 0.0053

Comparison between actual and predicted (%)
g (Actual gap)
error (Forecasting Error)
G (Predicted gap)
%
Time (month)

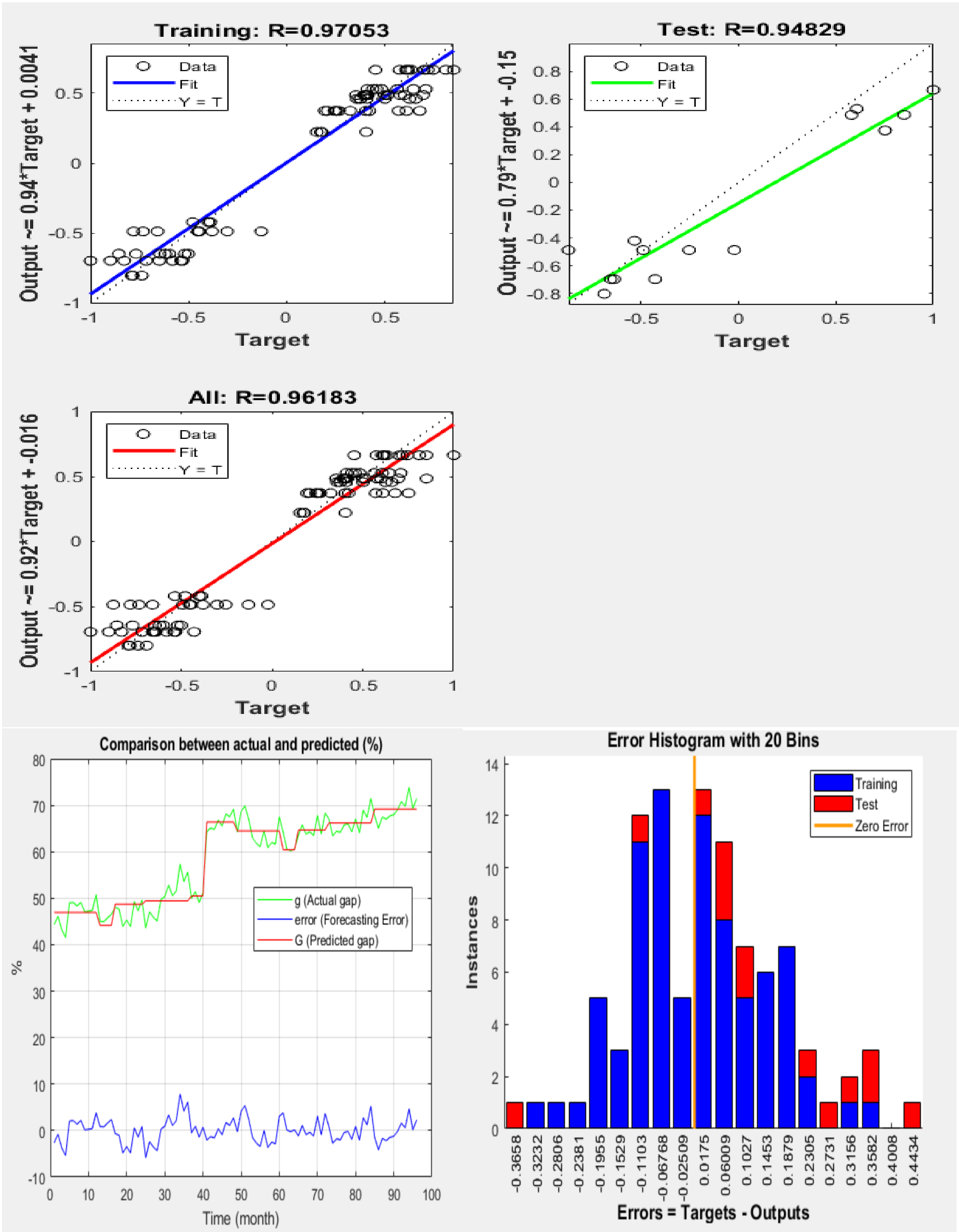


Error Histogram with 20 Bins
Training
Test
Zero Error
Instances
Errors = Targets - Outputs

Model-7:-3-40-1

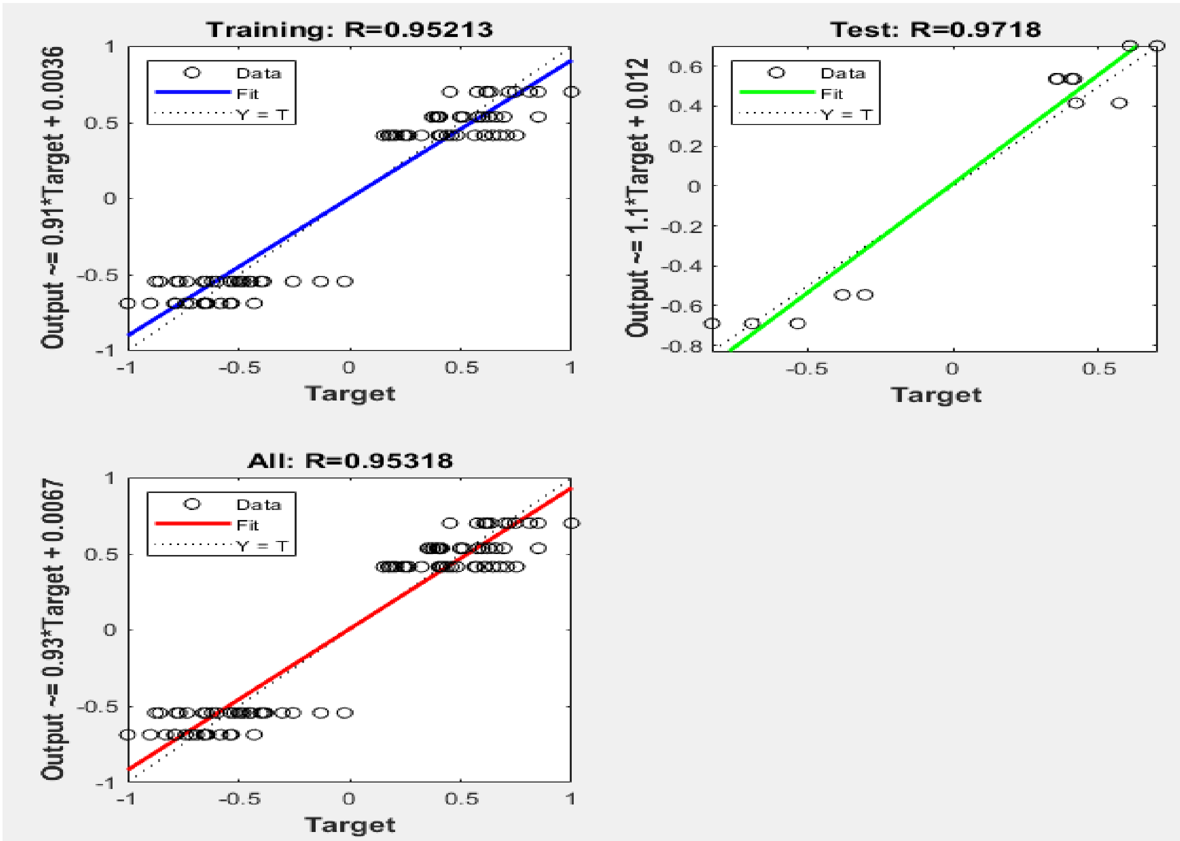

Training: R=0.96282
Output ~= 0.92*Target + 0.0028
Target
Data
Fit
Y = T
Test: R=0.98247
Output ~= 1.2*Target + 0.032
All: R=0.96222
Output ~= 0.95*Target + 0.011
Comparison between actual and predicted (%)
g (Actual gap)
error (Forecasting Error)
G (Predicted gap)
Time (month)
Error Histogram with 20 Bins
Training
Test
Zero Error
Instances
Errors = Targets - Outputs

Model-8:-2-40-1

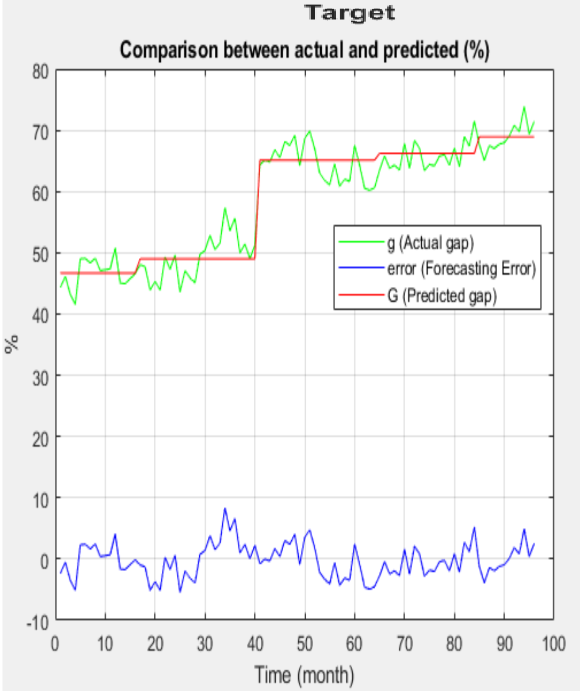


Training: R=0.97053
Output ~= 0.94*Target + 0.0041
Target
Data
Fit
Y = T
Test: R=0.94829
Output ~= 0.79*Target + -0.15
Target
All: R=0.96183
Output ~= 0.92*Target + -0.016
Target
Comparison between actual and predicted (%)
g (Actual gap)
error (Forecasting Error)
G (Predicted gap)
%
Time (month)
Error Histogram with 20 Bins
Training
Test
Zero Error
Instances
Errors = Targets - Outputs

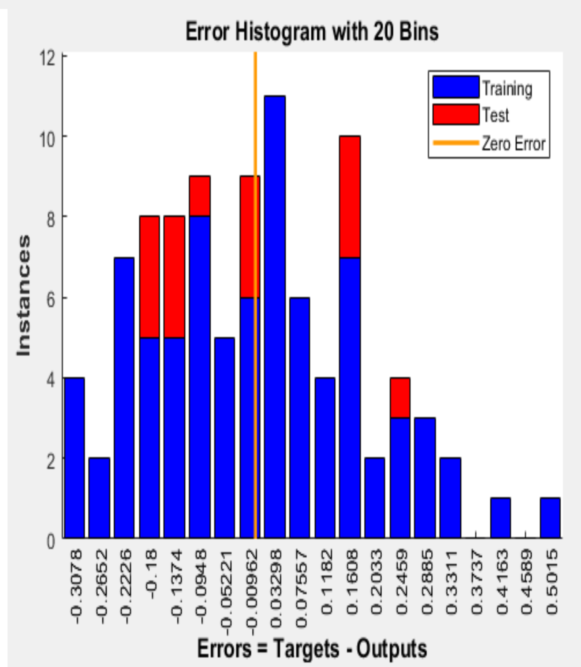
Model-9:-1-40-1



Training: R=0.95213
Output ~= 0.91*Target + 0.0036
Target
Data
Fit
Y = T
Test: R=0.9718
Output ~= 1.1*Target + 0.012
All: R=0.95318
Output ~= 0.93*Target + 0.0067



Comparison between actual and predicted (%)
g (Actual gap)
error (Forecasting Error)
G (Predicted gap)
%
Time (month)



Error Histogram with 20 Bins
Training
Test
Zero Error
Instances
Errors = Targets - Outputs

2. Actual gap calculation 2015-2021

Date	Pop	WC(m3/day)	per Capita Consumption (/c/d)	Pop	Net Average Deamand(m3/d)	per Capita Demand (l/c/d)	NRW (m3/d)	Demand surplus or Deficit (m3/d)	NRW %	Gross Average Daily Demand(m3/day)	Factor by which demand met (%)	Gap(%)
30/09/2015	3,871,000	233,437	60		419,320	83	122958	(185,882.73)	31.4	542,278	0.56	44.33
31/10/2015	3,871,000	225,856	58		419,320	83	122958	(193,463.90)	31.4	542,278	0.54	46.14
30/11/2015	3,871,000	238,234	62		419,320	83	122958	(181,085.83)	31.4	542,278	0.57	43.19
31/12/2015	3,871,000	245,077	63		419,320	83	122958	(174,243.30)	31.4	542,278	0.58	41.55
31/01/2016	4,040,000	230,351	57		451,687	91	124045	(221,336.27)	30.1	575,732	0.51	49.00
29/02/2016	4,040,000	238,233	59		467,871	91	124045	(229,637.37)	30.1	591,916	0.51	49.08
31/03/2016	4,040,000	241,798	60		467,871	91	124045	(226,072.80)	30.1	591,916	0.52	48.32
30/04/2016	4,040,000	238,025	59		467,871	91	124045	(229,845.30)	30.1	591,916	0.51	49.13
31/05/2016	4,040,000	247,585	61		467,871	91	124045	(220,285.97)	30.1	591,916	0.53	47.08
30/06/2016	4,040,000	246,929	61		467,871	91	124045	(220,941.43)	30.1	591,916	0.53	47.22
31/07/2016	4,040,000	246,274	61		467,871	91	124045	(221,596.90)	30.1	591,916	0.53	47.36
31/08/2016	4,040,000	230,314	57		467,871	91	124045	(237,556.97)	30.1	591,916	0.49	50.77
30/09/2016	4,040,000	257,133	64		467,871	91	124045	(210,737.40)	30.1	591,916	0.55	45.04
31/10/2016	4,040,000	257,567	64		467,871	91	124045	(210,303.47)	30.1	591,916	0.55	44.95
30/11/2016	4,040,000	253,708	63		467,871	91	124045	(214,162.47)	30.1	591,916	0.54	45.77
31/12/2016	4,040,000	250,088	62		467,871	91	124045	(217,782.53)	30.1	591,916	0.53	46.55
31/01/2017	4,216,000	251,508	60		484,054	94	124616	(232,546.30)	28.8	608,670	0.52	48.04
28/02/2017	4,216,000	253,088	60		484,054	94	124616	(230,966.13)	28.8	608,670	0.52	47.71
31/03/2017	4,216,000	271,586	64		484,054	94	124616	(212,467.60)	28.8	608,670	0.56	43.89
30/04/2017	4,216,000	264,757	63		484,054	94	124616	(219,296.57)	28.8	608,670	0.55	45.30
31/05/2017	4,216,000	271,672	64		484,054	94	124616	(212,382.23)	28.8	608,670	0.56	43.88
30/06/2017	4,216,000	245,370	58		484,054	94	124616	(238,684.43)	28.8	608,670	0.51	49.31
31/07/2017	4,216,000	255,204	61		484,054	94	124616	(228,850.00)	28.8	608,670	0.53	47.28
31/08/2017	4,216,000	243,847	58		484,054	94	124616	(240,206.97)	28.8	608,670	0.50	49.62
30/09/2017	4,216,000	273,082	65		484,054	94	124616	(210,971.67)	28.8	608,670	0.56	43.58
31/10/2017	4,216,000	256,192	61		484,054	94	124616	(227,861.53)	28.8	608,670	0.53	47.07
30/11/2017	4,216,000	261,947	62		484,054	94	124616	(222,106.57)	28.8	608,670	0.54	45.88
31/12/2017	4,216,000	265,728	63		484,054	94	124616	(218,325.60)	28.8	608,670	0.55	45.10
31/01/2018	4,400,000	277,687	63		552978	105	124677	(275,290.87)	27.6	677,655	0.50	49.78
28/02/2018	4,400,000	274,372	62		552978	105	124677	(278,605.60)	27.6	677,655	0.50	50.38
31/03/2018	4,400,000	260,899	59		552978	105	124677	(292,079.00)	27.6	677,655	0.47	52.82
30/04/2018	4,400,000	273,607	62		552978	105	124677	(279,370.53)	27.6	677,655	0.49	50.52
31/05/2018	4,400,000	267,789	61		552978	105	124677	(285,189.47)	27.6	677,655	0.48	51.57
30/06/2018	4,400,000	235,871	54		552978	105	124677	(317,106.70)	27.6	677,655	0.43	57.35
31/07/2018	4,400,000	256,619	58		552978	105	124677	(296,359.20)	27.6	677,655	0.46	53.59
31/08/2018	4,400,000	245,493	56		552978	105	124677	(307,485.00)	27.6	677,655	0.44	55.61
30/09/2018	4,400,000	276,671	63		552978	105	124677	(276,306.60)	27.6	677,655	0.50	49.97
31/10/2018	4,400,000	268,573	61		552978	105	124677	(284,405.03)	27.6	677,655	0.49	51.43
30/11/2018	4,400,000	281,669	64		552978	105	124677	(271,308.87)	27.6	677,655	0.51	49.06
31/12/2018	4,400,000	269,606	61		552978	105	124677	(283,372.37)	27.6	677,655	0.49	51.24
31/01/2019	4,592,000	289,393	63	4,235,773	811,680	158	292,205	-522,288	36	1,103,885	0.36	64.35
28/02/2019	4,592,000	283,220	62	4,235,773	811,680	158	292,205	-528,460	36	1,103,885	0.35	65.11
31/03/2019	4,592,000	285,644	62	4,235,773	811,680	158	292,205	-526,036	36	1,103,885	0.35	64.81
30/04/2019	4,592,000	269,291	59	4,235,773	811,680	158	292,205	-542,389	36	1,103,885	0.33	66.82
31/05/2019	4,592,000	279,919	61	4,235,773	811,680	158	292,205	-531,761	36	1,103,885	0.34	65.51
30/06/2019	4,592,000	258,622	56	4,235,773	811,680	158	292,205	-553,058	36	1,103,885	0.32	68.14
31/07/2019	4,592,000	263,667	57	4,235,773	811,680	158	292,205	-548,013	36	1,103,885	0.32	67.52
31/08/2019	4,592,000	250,235	54	4,235,773	811,680	158	292,205	-561,445	36	1,103,885	0.31	69.17
30/09/2019	4,592,000	290,409	63	4,235,773	811,680	158	292,205	-521,271	36	1,103,885	0.36	64.22
31/10/2019	4,592,000	254,311	55	4,235,773	811,680	158	292,205	-557,369	36	1,103,885	0.31	68.67
30/11/2019	4,592,000	244,584	53	4,235,773	811,680	158	292,205	-567,096	36	1,103,885	0.30	69.87
31/12/2019	4,592,000	268,394	58	4,235,773	811,680	158	292,205	-543,286	36	1,103,885	0.33	66.93
31/01/2020	4,794,000	266,795	56	4,362,846	719,870	165	302,417	-453,075	42.01	1,022,287	0.37	62.94
29/02/2020	4,794,000	275,019	57	4,362,846	719,870	165	302,417	-444,851	42.01	1,022,287	0.38	61.80
31/03/2020	4,794,000	280,273	58	4,362,846	719,870	165	302,417	-439,597	42.01	1,022,287	0.39	61.07
30/04/2020	4,794,000	255,344	53	4,362,846	719,870	165	302,417	-464,526	42.01	1,022,287	0.35	64.53
31/05/2020	4,794,000	281,901	59	4,362,846	719,870	165	302,417	-437,969	42.01	1,022,287	0.39	60.84
30/06/2020	4,794,000	273,455	57	4,362,846	719,870	165	302,417	-446,415	42.01	1,022,287	0.38	62.01
31/07/2020	4,794,000	276,037	58	4,362,846	719,870	165	302,417	-443,833	42.01	1,022,287	0.38	61.65
31/08/2020	4,794,000	233,483	49	4,362,846	719,870	165	302,417	-486,387	42.01	1,022,287	0.32	67.57
30/09/2020	4,794,000	257,319	54	4,362,846	719,870	165	302,417	-462,551	42.01	1,022,287	0.36	64.25
31/10/2020	4,794,000	284,204	59	4,362,846	719,870	165	302,417	-435,666	42.01	1,022,287	0.39	60.52
30/11/2020	4,794,000	286,516	60	4,362,846	719,870	165	302,417	-433,355	42.01	1,022,287	0.40	60.20
31/12/2020	4,794,000	283,536	59	4,362,846	719,870	165	302,417	-436,334	42.01	1,022,287	0.39	60.61
31/01/2021	5,006,000	280,556	56	4,493,732	768,428	171	311,943	-487,872	40.6	1,080,371	0.37	63.49
28/02/2021	5,006,000	262,822	53	4,493,732	768,428	171	311,943	-505,606	40.6	1,080,371	0.34	65.80
31/03/2021	5,006,000	278,177	56	4,493,732	768,428	171	311,943	-490,251	40.6	1,080,371	0.36	63.80
30/04/2021	5,006,000	274,429	55	4,493,732	768,428	171	311,943	-493,999	40.6	1,080,371	0.36	64.29