ADDIS ABABA UNIVERSITY SCHOOL OF GRADUATE STUDIES

ADDIS ABABA INSTITUTE OF TECHNOLOGY (AAiT)

Artificial Neural Network and Fuzzy Logic for Water Demand forecasting (In case of Mekelle city)

BY

G/WAHID ADHANA ABERA

A Thesis Submitted to

The School of Civil and Environmental Engineering

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

Advisor: Dr. Mebrate T.

Addis Ababa University

Addis Ababa, Ethiopia

May 2018

Addis Ababa University
Addis Ababa institute of technology
The School of Civil and Environmental Engineering

This is certify that the thesis prepared by **Gebrewahid Adhana Abera**, entitled: **Artificial Neural Network and Fuzzy Logic for Water Demand forecasting (In case of Mekelle city)** and submitted in partial fulfillment of the requirement for the degree Master of Science (Civil and Environmental Engineering, Major of Water Supply and Environmental Engineering) complies with the regulations of the university and the meets the accepted standards with respect to originality and quality.

Signed by Examining Committee

Internal Examiner____________________ signature_______________ date_________

External Examiner____________________ signature_______________ date_________

Advisor Mebrate Taffese (PhD) signature _______________ date 06/11/2

Chairman________________________ signature_______________ date_________

ACKNOWLEDGEMENTS

Many people and organizations have given a variety of support to me to complete this study. I would like to take this opportunity to express my gratitude to all of them for what they have done for me during the period of studying in Addis Ababa University.

My time at the Addis Ababa University has been extremely rewarding. Thank you to Dr. Mebrate for directing me in the right direction and guided me a lot during I am doing my thesis. His suggestions, comments and encouragements helped me to overcome many problems and open up interesting ideas.

Finally, I would like to thank all of my friends and family who took time out of their lives to assist me through difficult times.

ABSTRACT

The methodology emphasizes the importance of the new predictive model development, and to account for water demand uncertainty, through evaluation of the reliability of model predictive.

In this thesis Artificial Neural Networks (ANN) and fuzzy logic models were used for Mekelle city water consumption predictions analysis. Water consumption data collected from Mekelle city water and sewerage authority and independent monthly climate data were obtained from the National Metrological Agency. Sensitivity analysis was applied to selection most relevant minimum explanatory variables in the forecasting process. These data were used in both ANN and the fuzzy model setting up, testing and validation. To build ANN model the available dataset were divided into 3 subsets: 70% of the data for model development; (15%) of the data are used for training; and (15%) of the data are used validation to determine the optimal number of inputs and optimal number of hidden neurons. Both input variables and the output variable of the water consumption were fuzzified and triangular fuzzy membership functions were created. The Mamdani fuzzy rules in If–Then format with the centroid defuzzification method were employed. Seven ANN model were developed with different weather combination as input variable and Model one were found best with a root mean square error (RMSE) of 30.96, mean absolute percentage error (MAPE) 2.54, correlation coefficient (R^2) of 0.98 and 97.46% of forecasting accuracy. The average absolute percentage error of the fuzzy model was found as 19.2%. Therefore, in this research ANN model is successfully presented for predicting water consumption in Mekelle city with climate inputs, cost of water and population compared with the fuzzy prediction system.

Keywords: *water demand forecasting, ANN, fuzzy logic, climate and sensitivity analysis*

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURE

ABBREVIATIONS

ANN	Artificial Neural Network
ARIM	Autoregressive Integrated Moving Averages
BP	Back Propagation
CSA	Central Statistic Authority
FIS	Fuzzy Inference System
FL	Fuzzy Logic
L/c/day	Liter per capita per day
MAPE	Mean Absolute Percentage Error
MATLAB	Matrix laboratory
MF	Membership Function
MLP	Multi-Layer Perceptron
NGO	None Governmental Organization
NumMFs	Number of Membership function
R^2	coefficient of correlation
RMSE	Root Mean Squared Error
Trimf	Triangular Membership function
UFW	Unaccounted for water

1 INTRODUCTION

Water resources are under huge pressure due to population increase and industrialization. The pressure on water resources is worsened by the fact that water is a finite resource, whose management should be done to ensure that the appropriate usage. The increasing urbanization in the world is one of the major pressures to development today limited resources calls for an approach to ensure sustainability of resource use. One of the major problems in the world today especially in Africa due to urbanization is the scarcity of water resources specifically drinking water supply.

The most important factor in planning and operating a water distribution system is satisfying consumer demand. This means continually providing users with quality water in adequate volumes at reasonable pressure, and so ensuring a reliable water distribution system. Efficiently operating and managing a water supply system requires short term water demand forecasts; and the estimation of future municipal water demand is central to the planning of a water supply system. Water demand forecasting is becoming an essential tool for the design, operation, and management of water supply systems in activities such as planning new developments or system expansion; estimating the size and operation of reservoirs, pumping stations, and pipe capacities; and for urban water management issues.

The problem with drinking water shortage may not be unique to the city of Mekelle in Ethiopia. Due to the vast expansion of cities and towns all over Ethiopia there are many cities and towns in Ethiopia facing sever water shortage. Although one can take such shortages as a side effect of the fast development pace in Ethiopia, the shortage in Mekelle has been an issue for long. According to Mekelle water supply project feasibility study in 2014, water demands in Mekelle have meaningfully increased because of various factors, such as local population growth, migration from other localities, industrial growth and expansion, and consequently, a general rise in the living standards. The water demand of the city expected about 43,000m^3/day and it is current production is only about 28,000m3/day which covers about 65%.

Water demands are highly variable and are affected by factors such as size of city, characteristics of the population, the nature and size of commercial and industrial establishments, climatic conditions, and cost of supply (Altunkaynak, et al., 2004). In addition to water consumption values, we also have information concerning daily values of climate variables: temperature in Celsius, millimeters of rain. All these factors are connected with the water demand behavior. Temperature is the more relevant because it directly influences multiple sources of water consumption such as showers, water for gardens (Herrara, et al., 2010).

This research is interested to conduct a detail study by considering weather parameters with relation to water consumption and to use relatively new methods of water demand prediction models namely Artificial Neural Network and Fuzzy logic. In this research work, influence of climatic variables such as rainfall, maximum and minimum temperature, relative humidity on water consumption will analyze for an urban area. For this past time series of climatic and water consumption data will collect on monthly basis and it is divided into training and testing pattern.

1.1 Problem of the statement

The challenge of sustainability for the urban water system is considerable, because it faces multiple pressures from natural, social, and economic aspects. Most of the decision in urban planning and sustainable development are highly dependent on forecasting of water demand. Research into water consumption in cities has been restricted to the capital or development of new sources. There are lacks of studies relating to the water demand of cities in Ethiopia, similarly in Mekelle city. There appears to be no specific information about water demand forecasting method used. Water scarcity has become a major problem and which affects the economic growth. Hence in order to match supply and demand, proper forecasting of water demand is required. All abovementioned challenges must be resolved specially the development of an adequate model for forecasting water consumption.

Forecasting water demand is no simpler than any other type of demand. Moreover, the complex relationship between water supply and water demand can complicate matters further. The traditional approach is to forecast the water demand based on the population. Water demand is affected by various climatic and socioeconomic factors, government policies and strategy related factors which differ from place to place, thus necessitating the need to develop city specific models to predict water demand.
At present, there are many water demand forecast models which are available in literature but most of them are developed for specific places and circumstances and can only be used for that region. Such studies may be relevant and transferable. Also the problem of domestic use of water has become complex because its dependability on the climate, which is becoming increasingly uncertain and nonlinear relationship to water consumption.

There is a very large set of tools that can be used in this forecasting problem, and so there is an effective need for choosing and understanding the best techniques for a given forecasting application. The technique of artificial neural networks (ANN) and fuzzy logic models were used with algorithm for several civil engineering applications including water demand prediction. This is what this study seeks to identify an appropriate water demand forecasting for Mekelle city.

1.2 Objectives

1.2.1 General objectives

This methodology provide useful information to promote a more efficient model of future water demand forecasting and to ensure the sustainability of urban drinking water supply in the Mekelle city.

1.2.2 Specific objectives

The Specific objectives of this research are particularly focus on the following points.

- To develop a best model for water demand forecasting for Mekelle city
- To demonstrate applicability of the best selected models in the practical issues of the water demand forecasting

1.3 Research Questions

The problems are marked in in Mekelle city due to increase of the urbanization and due to traditional way of demand forecasting. The general issue that leads to a need for the study is to check and analysis the water demand of the city. Based on this the following research questions are proposed and this research tries to answer the following after its completion.

- What is the true water demand of the city?
- What is the appropriate model of water demand forecasting for the city?
- How new demand forecasting method is properly demonstrate to decision makers?

1.4 Scope and limitation

The scope of this research was to develop water demand forecasting model to forecast daily municipal demand for Mekelle city by using weather data, water tariff and population as input parameter. Limitations of this study include the following:

- Only water use on a monthly basis was studied. Other time intervals such as hourly and weekly were not investigated.

- Only residential water use was analyzed. Other categories such as water use from commercial and industrial sectors were not investigated.
- Air temperature, rainfall, relative humidity and sunshine hour were the only fundamental variables included in the implementation of the proposed model.
- In order to compare forecasting performance for different models, water consumption in the forecasting period was assumed known as observed in historical records.

1.5 Organization of the Research

The research is made up by five main Chapters. Chapter one Introduction**:** includes the reasons of this project, the research, objectives and scopes. Chapter two Literature Review: 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 Methodology*:* 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 the water use pattern between Mekelle and other works for other cities. 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

The choice of a forecasting approach for water demand prediction depends on the expected uses for the forecast results, as well as the size and other characteristics of the utility and its service area. Generally, the amount of effort invested in forecasting increases with the size of the utility and with the importance of the decisions that will be influenced by the forecast results (Jones, 2008).

Time series forecasting, the most widely used approach, relies on the direct identification of patterns existing in historical water demand data. Researchers have used regression, exponential smoothing, autoregressive integrated moving averages (ARIMA), artificial neural networks (ANN), and other approaches to model water demand. Some have included economic, demographic, and weather factors in their water demand models. ANN was used for water demand forecasting to identify the relationship between the level of demand and key meteorological variables (Hartley, 1995).

Recently the technique of artificial neural networks (ANN) was used with the back-propagation algorithm for several civil engineering applications (Herrera, 2010). Zhou et al. (2000) developed time series models for daily water consumption in Melbourne, Australia. ANN models have also been used to model weekly peak demand. These demand levels can be formulated as a function of climate variables such as temperature, rainfall and previous water consumption. The model provided 24 hour forecasts using base and seasonal components, by considering the summer and winter months separately exhibited considerable improvement over a single model. In addition to the ANNs, other machine learning methodologies have been applied in forecasting hydraulic time series. Artificial neural networks have recently begun to be used for short term water demand forecasting. A number of studies have compared the use of artificial neural networks for short term urban water demand forecasting with other forecasting methods.

Most studies used gradient descent feed forward and back propagation**.** An ANN, gradient descent feed forward and back propagation, was developed by (Daniel, 1991) in order to estimate water consumption for Canberra. Four input parameters, the monthly rainfall, the number of rainy days in a month, the monthly evaporation and the monthly average temperature were used to predict the average daily per capita water consumption for one month in advance.

Artificial Neural Network ware used to forecast the total cost of municipal water in the districts of Isfahan. The findings reveal that the neural networks are very capable of understanding and simulating patterns of water cost, and thus can be used as a strong tool for forecasting the cost of water. In addition, three parameters of per capita population, per capita area, and per capita green space have the greatest effect on the total cost of water in the municipality of Isfahan (Mohammadzadeha, 2011).

Previously, the most common method of producing demand predictions for the forthcoming day/month was simply to use the profile either of the previous day/month or of the same day the previous week/month. The major drawback with such a method is the lack of consideration of the underlying factors which determine the level of demand on a particular day or month. The set of meteorological and social circumstances that combined to produce the observed consumption on one day may not be applicable to the day for which the prediction is required.

Recently, most of work conducted in the field of water demand forecasting has been based on the application of mathematical algorithms to records of past consumption values and causal meteorological variables. The standard mathematical models achieve acceptable performance in terms of accuracy when the influences of external non-cyclic factors are negligible or absent. Some of the water demands perditions are Artificial intelligence and time series models. Artificial intelligence is separated into two significant concepts; ANNs and fuzzy logic. Both ANNs and fuzzy logic is proved to be useful in modeling and simulation of a system, with one or more variables (Bozokalfa, 2005).

Liu et al. (2002) used an artificial neural network to forecast the monthly usage of municipal water in the city of Weinam, China. They took account of some economical parameters including per capita income, number of family members, and water price. Ghiassi et al. (2008) developed a dynamic artificial neural network model for projecting the urban water demand. They also demonstrated that by using time series water demand data, the dynamic artificial neural network model could provide excellent fit and forecasts without reliance on the explicit inclusion of weather factors.

Adamowski (2008), compared multiple linear regressions, time series analysis, and ANNs as techniques for the forecast modeling of the peak of daily water demand in summer. There are some studies on prediction methods in Iran. In this regard, the use of artificial neural networks in Iran started in 2000 when the municipal water usage in Yazd was forecasted by some economical parameters as well as the volume of precipitation and evaporation.

The literature review shows that ANN techniques have been commonly used to predict the water resources system including water demand modeling, water consumption, and in a few cases forecasting the cost of water. Therefore, there is a need for comparing ANN techniques with fuzzy logic in the field of water demand forecast modeling.

A Fuzzy model was developed to estimate water consumption of Tehran water demand. The effective parameters of the model were: daily average temperature, relative humidity percentage, last day, and last week and last year consumption. The advantages of fuzzy models are their simple structure and easy representation, but in general, these models do not produce good results for forecasting Tehran daily water consumption. Their error values are very high because they cannot learn from input data in the training stage (Dini, 2009).

The application of the fuzzy logic is presented for water consumption values records in Istanbul city, which is the most covered city of Turkey. The predictions are obtained for the last 18 months test data. The overall prediction relative error is less than 10%, which is practically acceptable (Altunkaynak, et al., 2004).

Comparison of constructed models based on two random and non-random approaches showed that the random data input approach produces better results. Neuro fuzzy models with random input produce very good results in comparison with the fuzzy models because these models have a high capability to understand and learn the consumption pattern. In general, in fuzzy and neuro fuzzy models, if the water consumption values are not considered as input parameters, they cannot predict the trend component. In non-linear models such as neuro fuzzy models, the selection of effective parameters with evaluating the correlation coefficient between input and output parameters are necessary. Therefore, the results of models with effective input parameters are better than models with all input parameters (Dini, 2009).

In the neuro fuzzy models, the models which only accept water consumption data such as last day to last week and last year daily water consumption as their input have very good results, so these models are proposed for forecasting daily water consumption in Tehran city (Dini, 2009). Also, these models do not need to forecast the input parameters (temperature and humidity) for the same day that daily water consumption is forecasted. It should be mentioned that all the models based on fuzzy or Neuro Fuzzy logic systems should be updated regularly with more recent data to produce realistic results.

2.1 Artificial Neural Network

The interest generated in neural networks is centered on their ability to learn by producing a mapping between a given input signal and a desired output signal. Through the development of advanced learning algorithms and network architectures, highly complex relationships have successfully been modeled by neural networks where conventional mathematical approaches have failed to provide adequate solutions. This is particularly the case where there exists a complex relationship between a numbers of influencing factors which interact to generate a particular end result; an example of such a relationship is the influence of weather conditions upon the level of water demand. Neural networks allow a 'black box' approach to be applied to such relationships so avoiding the need to explicitly define the exact interrelation between each of the influencing factors and their specific influence on the final result. The process of training a network so that it performs a desired function involves the application of one of the many learning algorithms that have been developed for neural network applications.

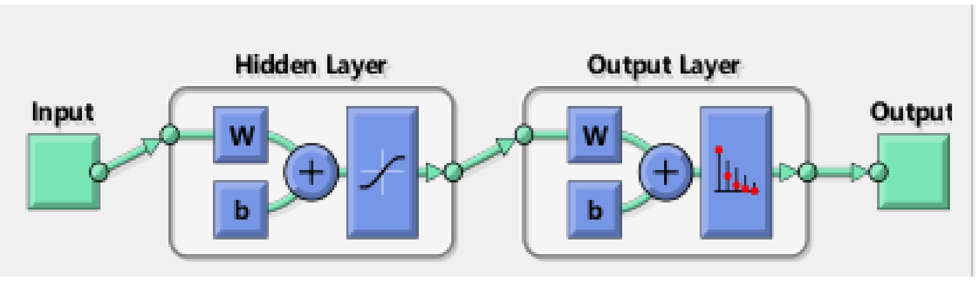


Figure 1 Layers of neural network

2.1.1 Artificial Neural Network Learning Algorithms

There have been a number of attempts to link the study of neural networks directly with the thought processes in operation within the human brain (Lobbrecht, 2002). The general aim of the learning process is to observe the networks performance in response to particular input stimuli and use this information to modify and improve that performance. The predominant method for achieving this improvement in performance is by the adjustment of the weight values associated with the connections between individual processing units within the network. Many different learning algorithms have been developed for neural network applications.

2.1.1.1 Back Propagation (BP) learning algorithm

First, and most known, the Back Propagation (BP) learning algorithm is the most common learning algorithm for Multi-layer perceptron networks (MLP NN) (Michael & Jiří , 2010). The algorithm is based on minimizing the error of neural network output compared to the target value. Basically, the algorithm remembers in a parameter the direction in which the current state in error space was reached. This parameter prevents the algorithm from being stuck in local minima. The momentum is added to the learning rule fitting up to changes in weights which are equal to the sum of the recent changes and new variations calculated using BP. The momentum constant defines the effect of the momentum.

2.1.1.2 Radial Basis Function network

Radial Basis Function neural networks (RBF-NN) belong to the feed-forward models of neural networks. RBF-NN consists of three layers of nodes. The first is the input layer that transports the input vector to each of the nodes in the hidden layer. The third layer consists of one node. It sums up the outputs of the hidden layer of nodes to yield the decision value (Michael & Jiří , 2010). Radial Basis Functions are used for the approximation and interpolation in numerical values. The basis function realizes the transformation of the distance value, calculated from the input vector to its own center, to the output value of the node. The output value is then multiplied by a weighting value or a constant.

2.2 Fuzzy logic (FL)

FL is a form of multi-valued logic derived from fuzzy set theory that deals with reasoning, which is approximate rather than precise. In contrast to yes/no or 0/1 binary logic, FL provides a set of membership values inclusively between 0 and 1 to indicate the degree of truth (fuzzy). Fuzzy Inference System (FIS) employing fuzzy 'If–Then' rules to quantify human knowledge and reasoning processes without employing precise quantitative analyses (Neeru Gupta et.al, 2012). The fuzzy approach based on linguistic expressions include uncertainty rather than numerical probabilistic, statistical or perturbation approaches. Fuzzy set theory allows the user to capture uncertainties in data. The following steps are necessary for the successful application of fuzzy inference.

i. Fuzzification of the input and output variables by considering convenient language subsets such as high, medium, low, heavy, light, hot, warm, big, small, etc.,

ii. Construction of fuzzy IF–THEN rules based on the expert knowledge and on the basis of available literature in order to model the problem.
iii. The implication part of a fuzzy system is defined as the shaping of the consequent based on the premise (antecedent) part, and finally
iv. The result is a fuzzy set, and therefore, requires defuzzification to arrive at a crisp value, which is required by the researcher.

2.2.1 Types of member ship function

The most common types of MF are triangular MFs, trapezoidal MFs, gaussian MFs, generalized bell MFs, π- shaped membership Function and S- shaped membership function. The simplest membership functions are formed using straight lines. Due to their simple formulas and computational efficiency, both triangular MFs and trapezoidal MFs have been used extensively, especially in real-time implementations. However, since the MFs are composed of straight line segments, they are not smooth at the corner points specified by the parameters (Omar , et al., 2015).

2.2.2 Type Fuzzy Inference

Mamdani's fuzzy inference method is the most commonly seen fuzzy methodology. Mamdani type inference, as defined for the toolbox, expects the output membership functions to be fuzzy sets. After the aggregation process, there is a fuzzy set for each output variable that needs defuzzification.

Takagi-Sugeno-Kang, method of fuzzy is similar to the Mamdani method in many respects. The first two parts of the fuzzy inference process, fuzzifying the inputs and applying the fuzzy operator, are the same. The main difference between Mamdani and Sugeno is that the Sugeno output membership functions are either linear or constant.

2.3 Factors that influence domestic water consumption

The factors that influence domestic water consumption divided in to four categories: economic, socio-demographic, climate and regulations and ordinances. The climate effects should be incorporated in water demand models using associated variables, such as rainfall or temperature (Ferreira, 2014). Ferreira argued that consumption in outdoor uses is very sensitive to climate, mainly with temperature.

Table 1 Factors that influence domestic water consumption

Categories	Sub-categories	Key-variables
Economic	Tariff structure	Marginal block price, fixed component of water and billing period
Socio-demographic	Population	Number of persons in household, number of dependents per household
	Education/Knowledge and awareness	Ratio of the total population to the university students, university graduates, people with 12 years of education
	Income	Average Household income, real income, median household income, per capita income, income per household member, monthly income
	Tourism	Tourism index
Climate	Temperature	Average annual (°C), average temperature in the summer month, average maximum daily temperature, and average monthly temperature
	Rainfall	Annual (mm), total rainfall recorded in the summer months, cumulative monthly rainfall, and average monthly precipitation
	Relative humidity	%
Rules and Regulations	Water regulation and restrictions	Number of daily hours of restrictions

3 MATERIAL AND METHODOLOGY

3.1 Data collection

The study applied mostly secondary data with the use of quantitative research method in the analysis and interpretation of data. A two phase methodology is employed to analyze consumptive water use. The first phase is to conduct the statistical analysis within each major category through Microsoft Excel and XLSTAT Package. The major land use types of residential, commercial, industrial, and open space are combined into three categories. The objective is to explore the regression relationship between the possible influential factors (e.g. population) and total water usage or consumptive water usage. The second phase is to develop the estimation method of urban consumptive water based on major land use categories.

3.2 Study area

According to the census data of the Central Statistic Authority (CSA), the total population of Mekelle in 2007 is 215,914, while the population data of each year thereafter is the forecast data based on census results of relevant department. Therefore, the population forecast of this project is still based on the 2007 census data. Meanwhile by reference with the results of Tahal's population projections and taking into consideration of Mekelle city expansion factors, the 2007-2030 average population growth rate is determined as 3.7% and the 2030 to 2045 rate as 2.7%. Thus, the total population of Mekelle is predicted to reach 497,957 in 2030 and 742,590 in 2045.

Mekelle city infrastructure situation assessment report 2014 shows that geographically Mekelle city is located at about $13^{0}30'$ north and 39^{0} east longitudinal. The altitude of Mekelle varies from 2150 to 2500 m above sea level which makes it to be categorized under Weynadega (mild) type of the agro climatic zone. The city have a large geographic area covering 191.51km^2 and stretches, on average, 14.92 km along the East-West and 14.36 km along the North-South axes, within its administrative boundary (Figure 2).

Mekelle experiences mild climatic condition with annual average maximum temperature of 27.2^{0}C and annual average minimum temperature of 8.4^{0}C. The city has annual average rainfall of 618.3 mm. According to the Ethiopian national metrological agency the months of July and August in combination accounts for some 436.7mm (70.6 %).

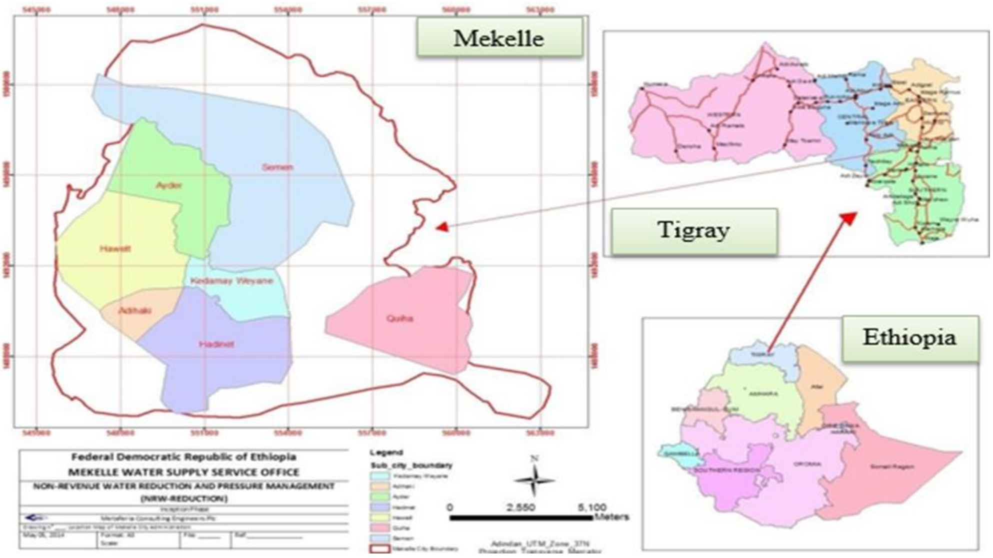


Figure 2 Location map of Mekelle City, Source: Mekelle City Structure Plan Revision

3.2.1 Existing Water Supply, consumption and loss Quantity

Based on Figure 3 below total annual water production of the city is has grown a steady increase from year to year and this implies there is continuous effort to cope up with the corresponding water demand coming from the fast population growth and urbanization of the Mekelle city.

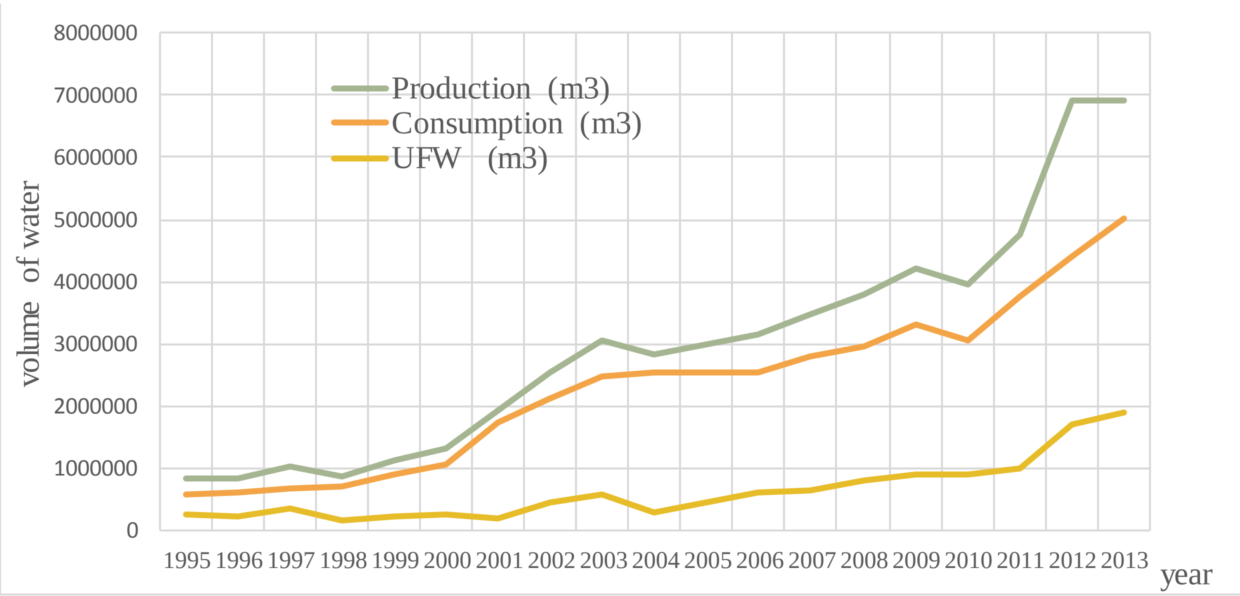


Figure 3 Water Production, Consumption and UFW for Mekelle

3.3 Data collection sources

Information is collected in four major categories: water use, land use, census data, and climate data. All collected secondary data sources and their respective organizations are summarized below in Table 2.

Table 2 List of data source offices

Office	Data
Mekelle city water and sewerage authority	- monthly billed water consumption - water tariff structure
Ethiopian Metrological Agency	Metrological dada - Temperature - Rainfall - Relative humidity - Sunshine hours
Mekelle city	Land use
Other	Influential factors: - Population - population density - housing units, housing density - average income

The water consumption data used in the testing and assessment of the neural network and fuzzy logic predictor was from the Mekelle city water and sewerage authority. This consumption and meteorological data covered the period from Sep-2010 to Dec-2016 and consisted monthly consumption by category. The weather data values extracted, as stored in Appendix 5 in Table 16, were average maximum and minimum temperature in degrees Celsius, rainfall in mm, sunshine hours and humidity.

3.4 Data Quality Analysis

Water consumption data are often imperfect with frequent erroneous time steps, missing data, seasonality, and trends. For this reason effective cleaning and preprocessing of raw data is important to achieve accurate water demand forecasting. In view of this, the main objective of this preprocessing is to assemble a valuable set of input data to be used for building of the models. In the descriptive analysis, the original series of water consumption were presented in Appendix 5 in (Table 16), namely the initial and final time of historical data. The majority of water consumption presents a complete year (September 2010 to December 2016) of historical data with availability is higher than 98%.

Large differences between the consumption values are expected, in terms of average, maximum and minimum flow as well as of the number of domestic clients. The flow values mentioned are affected by the presence of outliers. Most maximum and minimum flow values of the original time series are outliers and their analysis is not necessary at this stage.

The average consumption is more affected by the presence of outliers. Outlier detection analysis is employed for all input data and outlier is detected only on the water consumption. Data shows that water consumption varies from 208,511m^3/month to 10,500,141m^3/month not similar with the average consumption of 2,140,037m^3/month. Concerning the domestic clients in Mekelle city analyzed the number of domestic clients varies between 21096 and 46014. The observation shows that there is a clear difference in the demand profile for the different moths (Figure 4). The R^2 for Figure 4 is 0.4587, it is shows that the raw data have poor correlation coefficient. Therefore proper data cleaning is required.

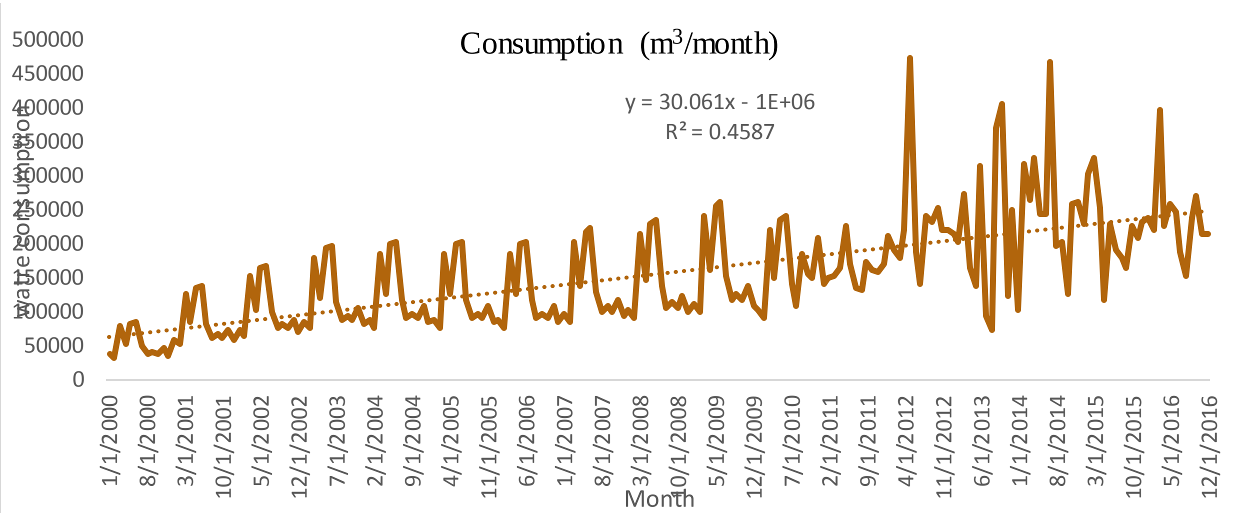


Figure 4 Monthly total billed water consumption profile (trend)

3.4.1 Outlier detection and cleaning

The statistical analyses are conducted in order to check the data quality. Data Quality Checking is the process of reviewing the data to discover inconsistencies and other anomalies and performing data cleaning activities to improve data quality.

A dataset is almost never 100% clean, but the aim is to produce datasets that are as error-free as possible. If the data has mistaken, it is difficult to justifying the results and conclusions. On the other hand, to produce proof that of systematically checked of the data and have eliminated as many errors as possible, then the results will have greater credibility. Two steps data quality analysis was carried out.

- Before data entry: the raw data may has errors are common

- After data manipulation & analysis: values may be truncated when transferring between software or may make a mistake in a calculation

The outlier detection using Dixon test was carried out before data entry and after analysis. This test is used to determine if the largest or smallest value can be considered as being an outlier. This test assumes that the data corresponds to a sample coming from a population that follows a normal distribution. Z-scores are displayed by XLSTAT to identify potential outliers. Z-scores correspond to the standardized sample by Equation 3-1.

$$Z_{score} = \frac{X_i - X_{mean}}{s} \qquad \text{Equation 3-1}$$

Where "s" is standard deviation

Figure 4 and Figure 5 shows the identification of outliers for the billed water consumption of Mekelle city before data entry. The outliers characterized by largest levels of consumption, were identified in the months of November and December 2015 (the two red columns in Figure 4 and Figure 5). In both graphs outliers are detected in the same months. The outliers detected were removed from the flow time series in order to obtain the clean flow time series data.

After removing of the outlier from raw data the data analysis has been carried out. Again the result analyzed was checked against outlier in order to avoid the error during data transferring between software or may make a mistake in a calculation. Figure 7 show that the analyzed data is from outlier.

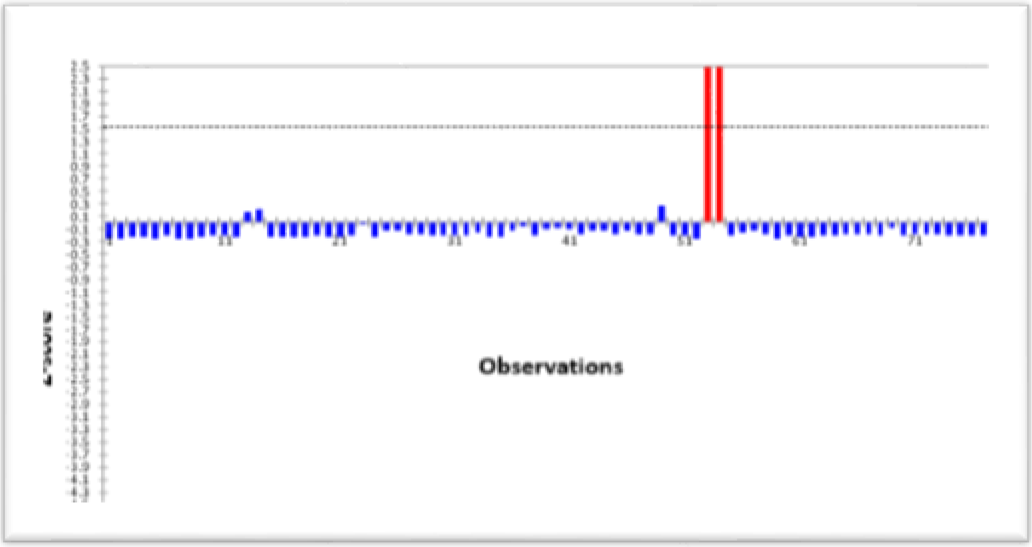


Figure 5 Outlier detection in monthly water consumption compared to the total data series (2010 to 2016)

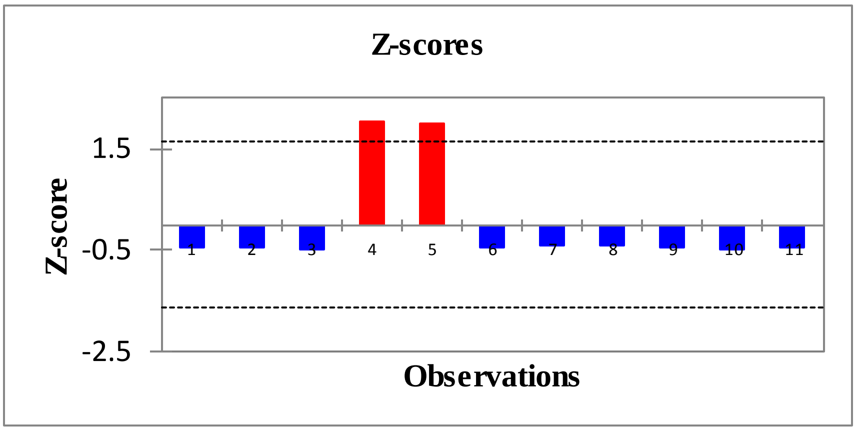


Figure 6 Outlier detection in monthly water consumption compared 2015

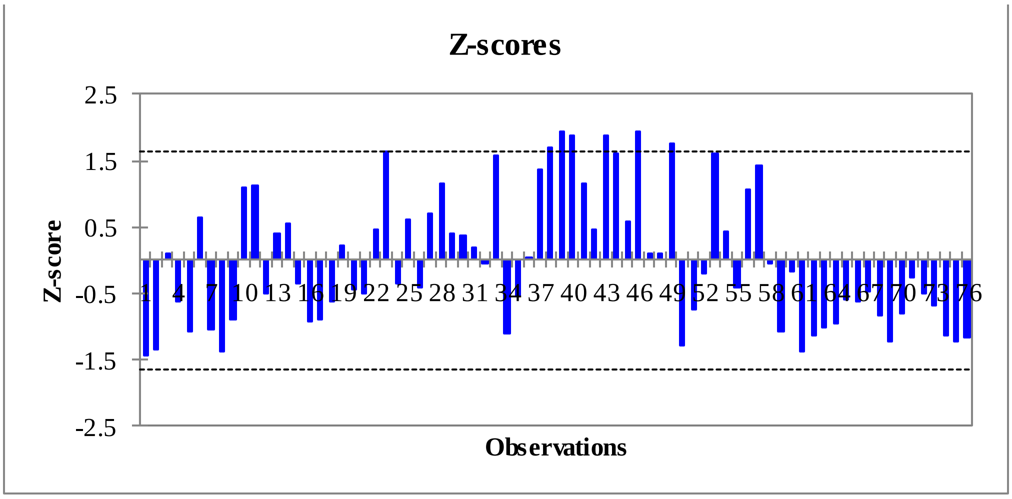


Figure 7 Outlier detection for the analyzed residential water consumption

3.5 Sensitivity Analysis

Sensitivity analysis was performed to arrive at the minimum number of relevant explanatory variables which could be used in the model development with acceptable prediction accuracy. Then, the least sensitive variable was not used in the model development and the remaining variables were used to develop models.

The climate variables were entered as independent and water consumption as the dependent variable. Regarding the statistical model, the certainty of a reasonable number of dominant variables guided the use of multiple regression analysis.

Numerous studies have employed linear and nonlinear regression to establish water consumption models while applying ANN. Some of linear regression have included rainfall, air temperature, family income and the cost of water as independent variables (An-Chi Huang, et al., 2017). Amaury de Souza, et al., (2015) develop a multiple linear regression statistical modelling framework of urban water consumption forecast for the city of Aquidauana, Brazil from year 2005 to 2014, monthly data, using multiple linear regression.

Based on literatures a typical polynomial first order model of water consumption is expressed is developed to determine the water consumption of Mekelle city (Equation 3-2). All available explanatory variables were combined together to create a regression model.

$$WD=\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta n Xn \qquad \text{Equation 3-2}$$

Where

β_i = coefficients for the input variables

β_0 = constant of the regression model

Xi =Input variable

$$WD = 1219.2 + 0.95 * R + 6.31 * TMX - 12.1 * TMN - 0.71 * SH - 2.80 * HU - 8.33E - 07 * TARIFF - 8.2E - 04 * POP$$

Where WD: water demand (L/c/month)

R : Rainfall

TMX : Maximum Temperature

TMN : Minimum Temperature

SH : Sunshine Hour

HU : Humidity

TARIFF: Water Tariff

POP: Population

The Sensitivity index (αi) to estimate the contribution of the variable of xi to the uncertainty of water demand (WD).

$$\boldsymbol{\alpha_i = \frac{\partial WD}{\partial X_i} \frac{X_{i_{mean}}}{WD_{mean}}} \qquad (\mathbf{i = 1, 2, 3, \ldots, n}) \qquad \text{Equation 3-3}$$

$$U_i = \alpha_i (COV)_i \qquad (i = 1, 2, 3, \ldots, n) \qquad \text{Equation 3-4}$$

Where, α_i = sensitivity factor of random variable i

Ui = uncertainty of random variable i

WD_{mean} = mean of WD

Xi = random variable i

$X_{(mean)i}$ = mean of xi

COVi = Coefficient of variation of random variable i

Table 3 The Uncertainty analysis for input variable selection

Variable	**Sensitivity(U)**
Rainfall	0.56
Temperature(max)	1.65
Temperature(min)	2.15
Sunshine Hour	0.04
Humidity	-0.89
TARIFF	-0.0000004

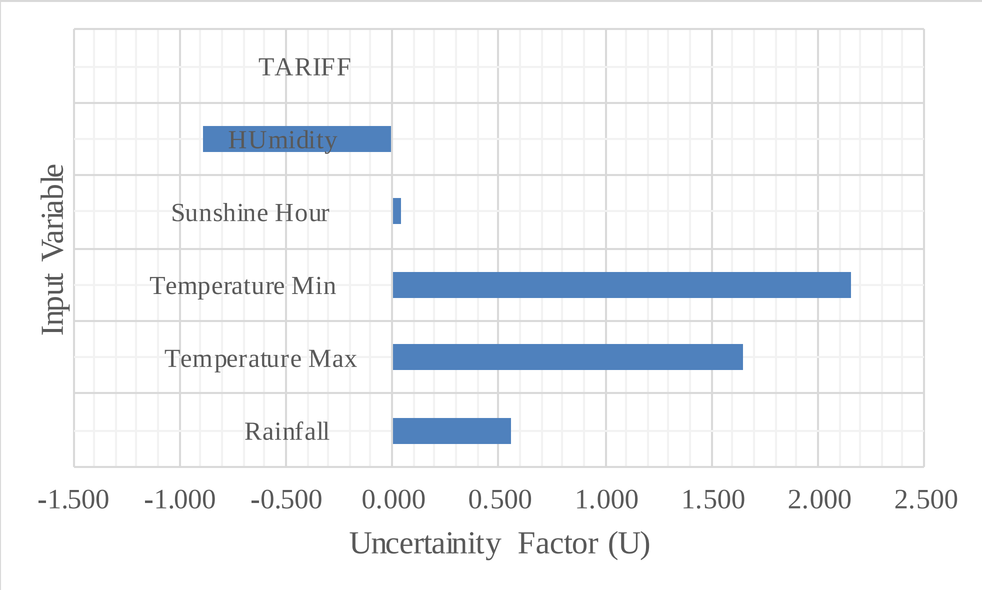


Figure 8 Uncertainty analysis graph

The meteorological variables; rainfall, relative humidity, maximum temperature and minimum temperature show great contribution to the water consumption forecasting in Mekelle city sunshine hour and water tariff show less contribution as shown in Figure 8.

3.6 Artificial Neural Network (ANN) for water demand modeling

The architecture of an artificial neural network defines how its several neurons are arranged, or placed, in relation to each other. These arrangements are structured essentially by directing the synaptic connections of the neurons. The topology of a given neural network, within a particular architecture, can be defined as the different structural compositions it can assume. A particular type of ANN is the called feed forward neural network, which consists of neurons organized into layers where outputs from one layer are used as inputs into the following layer. There are no cycles or loops in the network, no feed-back connections. Most frequently used example is the multi-layer perceptron (MLP) with a sigmoid transfer function and a gradient descent method of training called the back-propagation learning algorithm. In practical usage, the MLPs are known for their ability to approximate non-linear relations and therefore, when speaking about an ANN, the MLP is considered in the following text.

Moreover, one can consist of neurons with logistic activation function, while the other one can consist of neurons with the hyperbolic tangent as the activation function. On the other hand, training a particular architecture involves applying a set of ordinated steps to adjust the weights and initiations of its neurons. Hence, such adjustment process, also known as learning algorithm, aims to tune the network so that its outputs are close to the desired values.

The synaptic weights are parameters of an ANN to be determined during the training process. The type of the activation function is usually chosen in accordance with the type of a function to be approximated. In the case of continuous problems, the sigmoid activation function is the most common choice.

3.6.1 Main Architectures of Artificial Neural Networks

In general, an artificial neural network can be divided into three parts, named layers, which are known as input layer, hidden and output layer. The main architectures of artificial neural networks, considering the neuron disposition, as well as how they are interconnected and how its layers are composed can be divided as single layer feed forward network, multilayer feed forward networks, and recurrent networks and mesh networks.

Several issues have to be considered in order to build ANN models that show good water demand forecasting for different demand forecast horizons. These issues include the choice of the ANN structure, the transfer function, the training algorithm, the ANN parameters and the ANN input structure.

With regard to the ANN structure different scholars used feed forward in their research works. The investigated ANN structures included: the Feed Forward ANNs (Bishop, 1995) with one and two hidden layers, the Jordan ANN and the Elman ANN (Romano, 2012). These included the logistic and the hyperbolic tangent transfer functions for the neurons in the hidden layers and the logistic, hyperbolic tangent and linear transfer functions for the neuron in the output layer.

The Feed Forward ANNs with a hyperbolic tangent transfer function for the neurons in the single hidden layer and a linear transfer function for the neuron in the output layer, trained using the Back Propagation method were identified as the most suitable candidates for faster training and better predictive accuracy (Romano, 2012).

In this research Multilayer perceptron (MLPs) of feed-forward ANN model architecture with hidden layer is used for modeling the demand for Mekelle city. Feed forward back propagation method with Levenberg-Marquardt technique is used as the ANN training procedure because of its advantages in relation to the other gradient-descent techniques.

3.7 Residential Water demand forecasting using ANN

Since residential housing is the dominant water consumer in Mekelle city, this study made an extra analysis of climate in relation to residential water use they are mostly dependent for specific climate condition. Based on this idea, seven combinations of weather variables were considered in order to investigate their effect on water consumption. The MLP neural network method are applied for predicting the water demand in Mekelle city based on the combinations shown in (Table 4). The combination of climate variable examines the degree correlation of relationship between potential determinant factors and dependent variable of total water usage (i.e. Residential consumptive water usage correlates with climate variables).

Table 4 weather variables combinations for ANN model building

Model number	Water consumption	Rainfall	Maximum temperature	Minimum temperature	Relative humidity	Sunshine hour	Cost	Population
1	✓	✓	✓	✓	✓	✓	✓	✓
2	✓		✓	✓	✓	✓	✓	✓
3	✓			✓	✓	✓	✓	✓
4	✓				✓	✓	✓	✓
5	✓					✓	✓	✓
6	✓	✓					✓	✓
7	✓		✓				✓	✓

3.7.1 Input variables preparation for ANN

Regarding accessibility of data, input data are selected from 76 months (Sep-2010 to Dec-2016) water consumption time series, maximum and minimum air temperature, population, sunshine hour and humidity. To change the dataset to the dimensionless units, it is necessary to normalize the data matrix where the variables have been measured in different units. The inputs and output parameters of the prediction were normalized between -1 and 1(Equation 3-6).

$$z_i = \frac{x_i - x_{mean}}{s_i} \qquad \text{Equation 3-5}$$

$$z_{norm} = \left(\frac{z_i - z_{min}}{z_{max} - z_{min}}\right) * 2 - 1 \qquad \text{Equation 3-6}$$

Where x_i is raw Dataset, x_{mean} is mean value, s_i standard deviation of the raw data, z_i is the standardized dataset (Z-score), Z_{min} and Z_{max} are the minimum and maximum values of the i^{th} dataset and z_{norm} is the normalized dataset.

Table 5 the parameters used in ANN and Fuzzy model building

Parameter	Short notation for parameters	Parameters used in ANN model	Parameters used in Fuzzy model
Time (month)	T	x1	x1
Water consumption	Wc	x2	x2
Rainfall	Rf	x3	x3
Humidity	Hu	x4	x4
Temperature	Tem	x5	x5
Sunshine Hour	Hu	X6	X6
Water tariff	Cost	X7	X7
Water demand (output)	WD	y1	y1

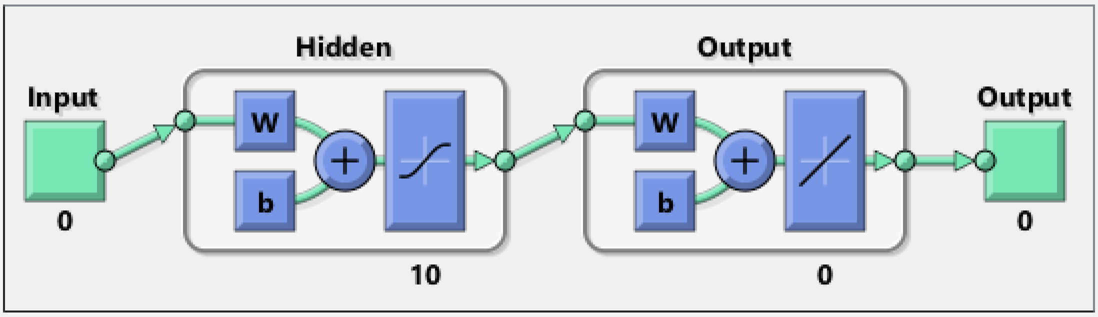
Data for function fitting problems are set up for a neural network by organizing the data into two matrices, the input matrix ‘x’ and the target matrix ‘co’ (water consumption). Each i^{th} column of the input matrix will have 7 elements representing the input data value is already known. Each corresponding column of the target matrix will have one element, representing the billed water consumption as shown in the MatLab program below.

```
%input data
load MCR.txt    % Mekelle city residential water consumption
time=MCR(:,1);              % September 2010 to  December 2016
inputvariable=MCR(:,2:8); % input variables used for water
demand modelling
consumption=MCR(:,9);      % Target variables which going to be
modeled
x=inputvariable';
co=consumption';
```

3.7.2 Fitting neural network

Two 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. Layers which are not output layers are called hidden layers.

```
setdemorandstream(15)    % set the training and testing
randomly
net = fitnet(10);          % first trial
view(net)
```



Two layer are used for model-1 with 30 neurons in the input layer and one neurons for output layer. In general, more difficult problems require more neurons, and perhaps more layers.

3.7.3 Testing the Neural Network

The mean squared error of the trained neural network can now be measured with respect to the testing samples. This will give us a sense of how well the network will do when applied to data from the real world. The input layer has 7 input variables and output have sizes of one variables, because the network has not yet been configured to match our input and target data. This will happen when the network is trained.

Root Mean Squared Error (RMSE) = $\sqrt{\text{MSE}}$

```
[net,tr] = train(net,x,co);
testX =(:,tr.testInd);% separates input into three sets:
training, validation, and testing
testT = co(:,tr.testInd);% separates targets into three sets:
training, validation, and testing
testco x = net(testX);
perf = mse(net,testT,testco);
```

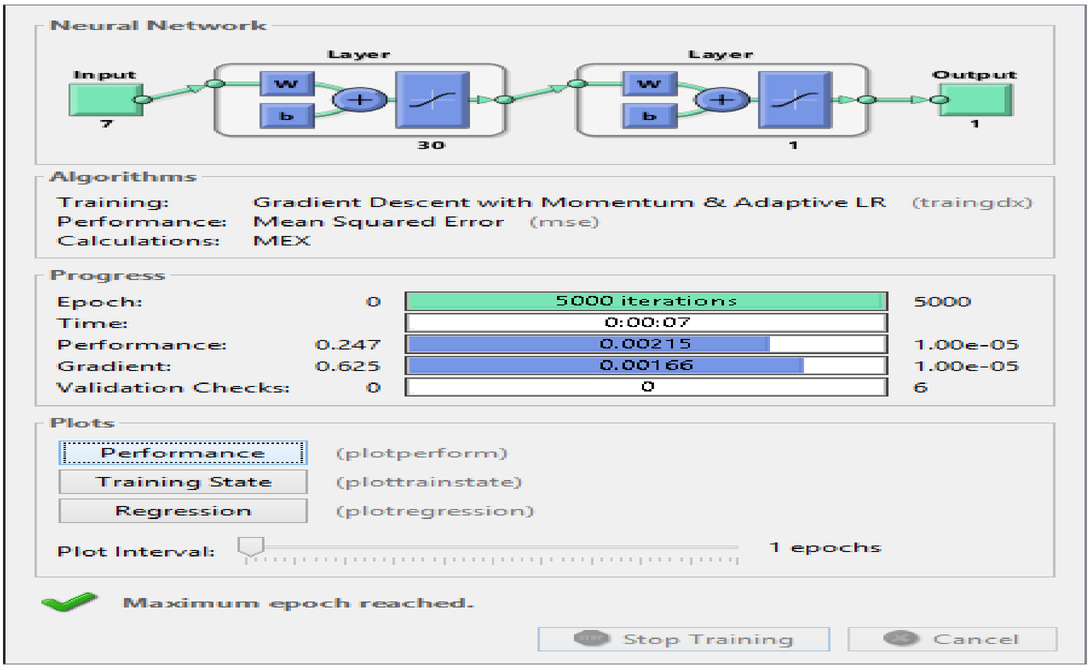
Figure 9 Artificial Neural Network for monthly Water Demand Forecasting

3.7.4 Training the Neural Network

After preparation the ANN number of layer and testing parameters, the network is ready to be trained. The samples are automatically divided into training, validation and test sets. The training set is used to teach the network. Training continues as long as the network continues improving on the validation set. The test set provides a completely independent measure of network accuracy.

One of the most relevant features of artificial neural networks is their capability of learning from the presentation of samples, which expresses the system behavior. Hence, after the network has learned the relationship between inputs and outputs, it can generalize solutions, meaning that the network can produce an output which is close to the expected output of any given input values. Therefore, the training process of a neural network consists of applying the required ordinated steps for tuning the synaptic weights and thresholds of its neurons, in order to generalize the solutions produced by its outputs. The set of ordinated steps used for training the network is called learning algorithm. During its execution, the network will thus be able to extract discriminant features about the system being mapped from samples acquired from the system. The NN Training Tool shows the network being trained and the algorithms used to train it. It also displays the training state during training and the criteria which stopped training will be highlighted in green. The buttons at the bottom open useful plots which can be opened during and after training. The overall training is done as the scripts below.

```
[xn,minx,maxx,con,minco,maxco]=premnmx(x,co); % Normalization
of Data(scale
% inputs and targets so that they fall in the range [-1,1]).
NodeNum1=30; % number of nodes of the first layer in hidden
layer
TypeNum = 1; % output vector dimensions
TF1 = 'tansig'; %transfer function for the 1st layer
TF2 = 'tansig'; %transfer function for the 2st layer
net=newff(minmax(xn),[NodeNum1,TypeNum],{TF1 TF2},'traingdx');
nntraintool
% creat network 'traingdm'
net.trainParam.show=50;
net.trainParam.epochs=5000;        % Number of training times
net.trainParam.goal=1e-5;          % accuracy of training
net.trainParam.lr=0.01;             % learning rate
net=train(net,xn,con);
% use the trained network to simulate the test data
p2n=tramnmx(x,minx,maxx);% Normalization of test data
an=sim(net,p2n);
[w]= postmnmx(an,minco,maxco);% Reverse normalization of Data,
also the desired result
nntraintool
```



3.7.5 Check the Neural Network Training Performance

Performance is measured in terms of mean squared error, and shown in log scale. The mean squared error rapidly decreased as the network was trained. Performance is shown for each of the training, validation and test sets. The version of the network that did best on the validation set is was after training.

3.7.6 Regression plotted for all the trained samples and Fitting Error

Another measure of how well the neural network has fit the data is the regression plot. The linear correlation coefficient is a commonly adopted measure of dependence between variables. Here the regression is plotted across all samples. The regression plot shows the actual network outputs plotted in terms of the associated target values.

```
plotregression(co,w) % 'w' is the reverse normalization of
Data, also the desired result
```

Another third measure of how well the neural network has fit data is the error (e) histogram. This shows how the error sizes are distributed. Typically most errors are near zero, with very few errors far from that.

```
error = co - w;
ploterrhist(error)
```

3.7.7 Simulate a time-series model

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

```
init_sys = idpoly[1 -1 0.2]; % initialize the system
e = iddata(randn(76,1));
data = sim(init_sys,e);
% Estimate model for the simulated data.
sys = armax(data(1:76),
% Obtain a 1 step-ahead prediction for the estimated model.
K = 1;
WD = predict(sys,data,K); % 'WD'is water demand
% Analyze the prediction.
figure;
plot(co,'g')
hold on
grid on
plot(error,'b')
plot(w,'r')
ylabel('litre/Month');
xlabel('Time (month)');
legend('co (Actual Consumption)','error (Forecasting
Error)','w (Predicted data)')
grid on
% Create title
title({'Comparison between actual and predicted
(litre/month/capita)'});
```

3.8 Fuzzy Logic Residential Water Demand Forecasting Model Building

Fuzzy Logic, the theory of approximate reasoning by means of which a powerful technique for reasoning of imprecise and uncertain information was proposed Zadeh 1975. The general structure of the fuzzy logic modeling is presented in Figure 10. The model basically consists of four components: Fuzzification, fuzzy rule base, fuzzy output engine, and defuzzification. Fuzzification converts each piece of input data to degrees of membership by lookup in one or more several membership functions.

The concept of fuzzy rule base is the If-Then rule, which is a mathematical interpretation of the linguistic If-Then rule but there are no mathematical equations and model parameters. All the uncertainties, nonlinear relationships, and model complications are included in the descriptive fuzzy inference procedure in the form of If-Then format. Defuzzification is the process of retranslate the fuzzy output into a crisp value using Membership functions.

The key idea in fuzzy logic is allowance of partial belongings of any object to different subsets of a universal set instead of complete membership to a single set. The membership function (MF) helps the partial belongings numerically which have values between 0 and 1. Fuzzy membership functions may take many forms; in fact in practical applications, simple linear functions, like triangular, trapezoidal ones, are preferable (Bozokalfa, 2005).

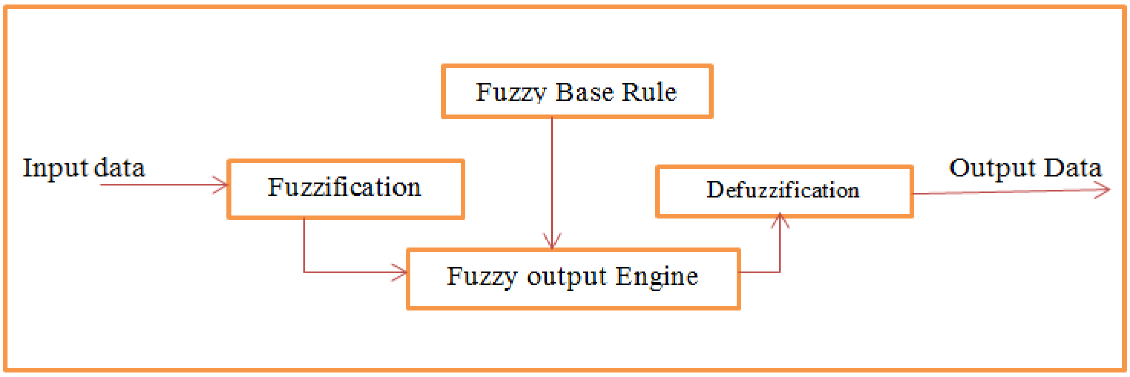


Figure 10 the basic structure of the fuzzy logic modeling

3.8.1 Fuzzy Rule Construction

The application of the fuzzy logic for water consumption forecasting in Mekelle city is starting with defining the input and output variables. The total input variables used in this research are totally six including four climate parameters (temperature, humidity, rainfall and sunshine hours), cost of water (water tariff) and population served (Figure 11). The overall input-output defining process expressed as the following scripts.

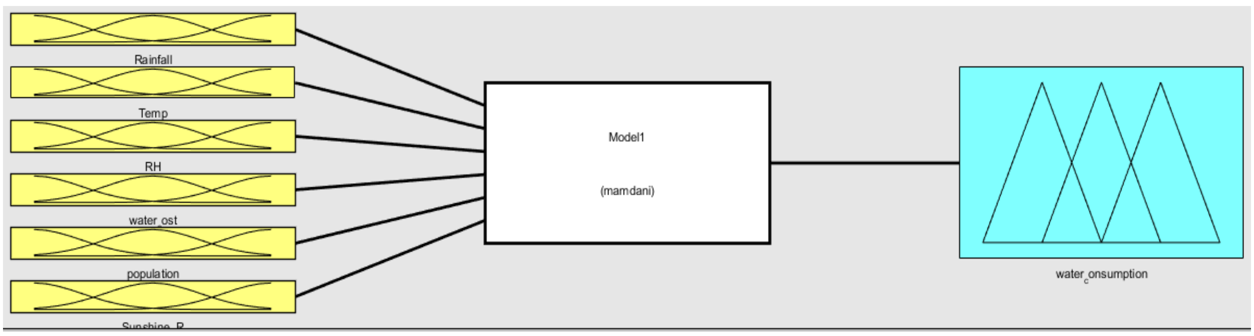


Figure 11 Defining input and output for fuzzy logic designer

The fuzzy inference technique is used Mamdani method which was proposed by Mamdani and Assailant. In Mamdani's model the fuzzy implication is modeled by Mamdani's minimum operator, the conjunction is min and the aggregation of the rules the max operator is used.
The overall fuzzy rules depend upon the number of input data and selection of the membership function. In this research triangular membership function is used to construct fuzzy rules. A fuzzy set is called triangular fuzzy number with peak (or center) *a*, left width *a* > 0 and right width *c* > 0 if its membership function is as (Equation 3.6) form (Zadeh, 1965). The triangular curve is a function of a vector, x (input paramater), and depends on three scalar parameters depend upon the magitud of input parameter like low avarage and maximamum. Example if low=*a*, average =*b* and maximum=*c* , therefore *a*, *b*, and *c*, as given by Equation 3-7

$$\boldsymbol{f(x{:}\,a,b,c)} = \left\{ \begin{matrix} \mathbf{0},\ \boldsymbol{x \le a} \\ \frac{x-a}{b-a},\ \boldsymbol{a \le x \le b} \\ \frac{c-x}{c-b},\ \boldsymbol{b \le x \le c} \\ \mathbf{0},\ \boldsymbol{c \le x} \end{matrix} \right\} \qquad \text{Equation 3-7}$$

The parameters *a* and *c* locate the "feet" of the triangle and the parameter *b* locates the peak. Base on Equation 4-6 the following triangular membership function were created in MatLab Tool box of Fuzzy logic designer depending upon the input parameters as the following program.

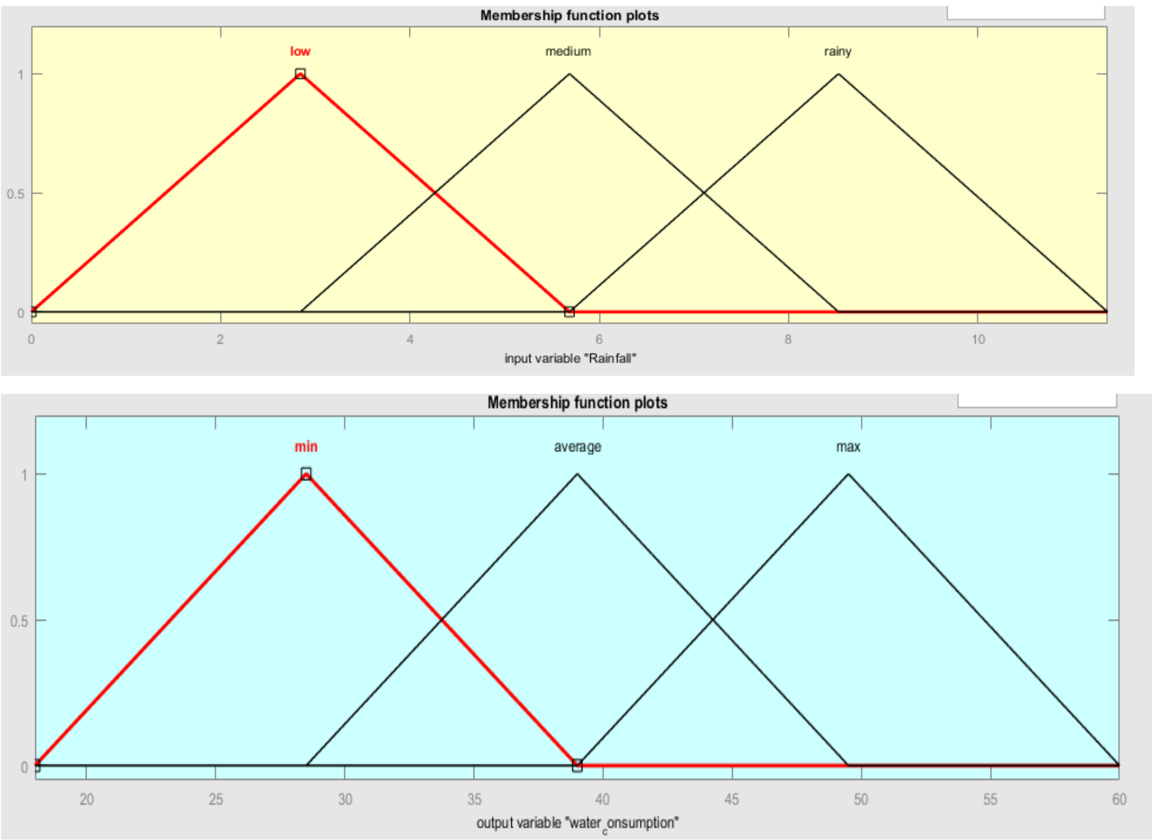


Figure 12 Fuzzy logic triangular membership function

The here there are 6 input paramaters and the resulting number of rules will be 64 totally ($number\ of\ fuzzy\ rule = 2^n$ where n is number input paramater). Some rules may be conflicting to each other. To resolve these conflicting rules assign a degree is necessary to each generated rule and keep only one rule from a conflicting group that has the maximum degree. In this way not only is the conflict problem resolved, but also the number of rules is greatly reduced to 25 rules. The degree of a rule is defined as follows: the construction of fuzzy rules.

$$\boldsymbol{D(rule) = V_{min} * V_{average} * V_{max} * y} \qquad \text{Equation 3-8}$$

Where V_{min}, $V_{average}$ and V_{max} are triangular membership 1, 2 and 3 conflicting rules respectively and y is the output.

4 RESULTS AND DISCUSSION

This section summarizes the analysis done and the results obtained when forecasting demands with a monthly forecast periodicity. This corresponds to a situation where the observed consumption data are collected to the Mekelle city water supply and sewerage authority. In this chapter specially is discuss about the results found earlier in the methodology and answer the research questions outlined in the research questions section. Finally the results of ANN were compared with fuzzy logic and answer the last research question.

4.1 Artificial neural network prediction results

To obtain the optimum structure of model ANN the model input data are divided into different combination with different hidden layers. This needs several models to be constructed are presented in Table 4. A comparative analysis was performed to evaluate the relative performance of each modeling technique investigated in this study. After trying several configurations, the optimum network architecture developed was based on one input layer and one hidden layer with 30 neurons using the training algorithm (Figure 9). For all of the seven models data were divided as 70% training, 15% validation and 15% testing. It is also observed that with increasing the hidden layers, the model error indicators are smaller.

Table 6 Correlation coefficient (R^2) of the seven combinations of weather variables and water consumption

Model Number	**Number of hidden neurons**										
	6	10	15	20	25	30	35	40	45	50	55
	Correlation coefficient(R^2)										
1	**0.83**	**0.94**	**0.98**	**0.99**	**0.98**	**0.99**	**0.97**	**0.97**	**0.98**	**0.97**	**0.98**
2	0.67	0.94	0.95	0.98	0.97	**0.99**	0.96	0.97	0.95	0.98	0.96
3	0.66	0.86	0.93	0.92	0.95	**0.91**	0.95	0.96	0.94	0.95	0.95
4	0.52	0.55	0.76	0.80	0.90	**0.91**	0.87	0.81	0.85	0.90	0.90
5	0.45	0.52	0.60	0.60	0.58	**0.62**	0.67	0.64	0.64	0.69	0.63
6	0.23	0.23	0.23	0.22	0.25	**0.32**	0.32	0.38	0.35	0.19	0.38
7	0.01	0.01	0.01	0.01	0.01	**0.01**	0.03	0.01	0.03	0.04	0.03

The training was stopped when the validation error increases for some specified and or default number of iterations. The weights and biases at the minimum of the validation error are returned finally. Model 1 for residential water consumption converged after 9 epochs with a root mean square error (RMSE) of 30.96 and mean absolute percentage error (MAPE) 2.4 for training the subset and all the input variables were retained at 0.001confidence level. The output obtained though simulation by the developed network was found to have a very good correlation (R^2) 0.98. The other models means square error (MSE), root means square error (RMSE), mean absolute percentage error (MAPE) and correlation (R^2) was summarized as shown in Table 8.

Similarly model two through model seven have the same process with model one but the only difference is each model has one less input weather parameter respectively. As the input parameter for each model decreases the correlation coefficient (R^2 value) also decreases rapidly and the mean squared error increase.

Table 7 Actual inputs and simulated outputs of residential water consumption for ANN Model 1 from 2015 to 2016 (litre/month)

	Consumption(L/c/month)					Consumption(L/c/month)			
Year	Actual	Forecasted	Error	Percentage Error	Year	Actual	Forecasted	Error	Percentage Error
1/1/2015	1060	1060.7	-0.7	0.1	1/1/2016	712	707.7	4.3	0.6
2/1/2015	1245	1227.8	17.2	1.4	2/1/2016	712	702.5	9.5	1.3
3/1/2015	738	737.7	0.3	0.0	3/1/2016	727	729.3	-2.3	0.3
4/1/2015	971	971.1	-0.1	0.0	4/1/2016	675	668.7	6.3	0.9
5/1/2015	1030	1029.2	0.8	0.1	5/1/2016	1201	1197.9	3.1	0.3
6/1/2015	798	797.1	0.9	0.1	6/1/2016	674	675.0	-1.0	0.2
7/1/2015	636	631.2	4.8	0.8	7/1/2016	764	764.0	0.0	0.0
8/1/2015	779	776.8	2.2	0.3	8/1/2016	726	724.5	1.5	0.2
9/1/2015	576	586.7	-10.7	1.8	9/1/2016	696	695.5	0.5	0.1
10/1/2015	545	623.3	-78.3	12.6	10/1/2016	612	623.6	-11.6	1.9
11/1/2015	709	641.5	67.5	10.5	11/1/2016	612	609.3	2.7	0.4
12/1/2015	642	653.6	-11.6	1.8	12/1/2016	612	619.0	-7.0	1.1

Table 8 the means square error (MSE), root means square error (RMSE), mean absolute percentage error (MAPE) and correlation (R^2) of residential water consumption for the developed ANN models

Model	MAPE	MSE	RMSE	R^2
1	2.54	958.35	30.96	0.9801
2	6.4	14289	119.54	0.9801
3	7.07	6131.8	78.31	0.8281
4	7.46	5906.6	76.85	0.8281
5	12.75	25180	158.68	0.3844
6	17.04	44403	210.72	0.1024
7	22.41	64848	254.65	0.0001

Figure 13 to 19 show that the seven models results graphically plotted to show R^2, the error distribution by error histogram and the actual and forecasted water consumption. It can be seen that error indices for model 1 is less than all the remaining models because in model 1 water consumption, population, water tariff and weather data such as maximum and minimum temperature, rainfall, sunshine hour and relative humidity are considered as the input parameter, having considerable effects on the results. Therefore, it can be concluded that, to obtain proper estimation of future water demand the inclusion of past records of water consumption and all weather parameters is necessary.

The results of the ANN models are represented in Table 6. The comparison of the results with different hidden layer was analyzed. The ANN network architecture with 30 hidden layer shows that better results with R^2 is 0.98. One reason is that the effects of selecting the optimal number of hidden layer is more realistically in order to avoid both under and over fitting condition.

1.1.1 Model-1

The input data selected for model 1 were from 76 months (Sep-2010 to Dec-2016) water consumption, population, water tariff and weather data such as maximum and minimum temperature, rainfall, sunshine hour and relative humidity was used. This model shows that the error is -20 to 36 liter/capital/month as shown in the error histogram in Figure 13 which is accumulated near zero error line.

Performance is measured in terms of mean squared error, and shown in log scale. The mean squared error rapidly decreased as the network was trained. Performance is shown for each of the training, validation and test sets. The version of the network that did best on the validation set was after training.

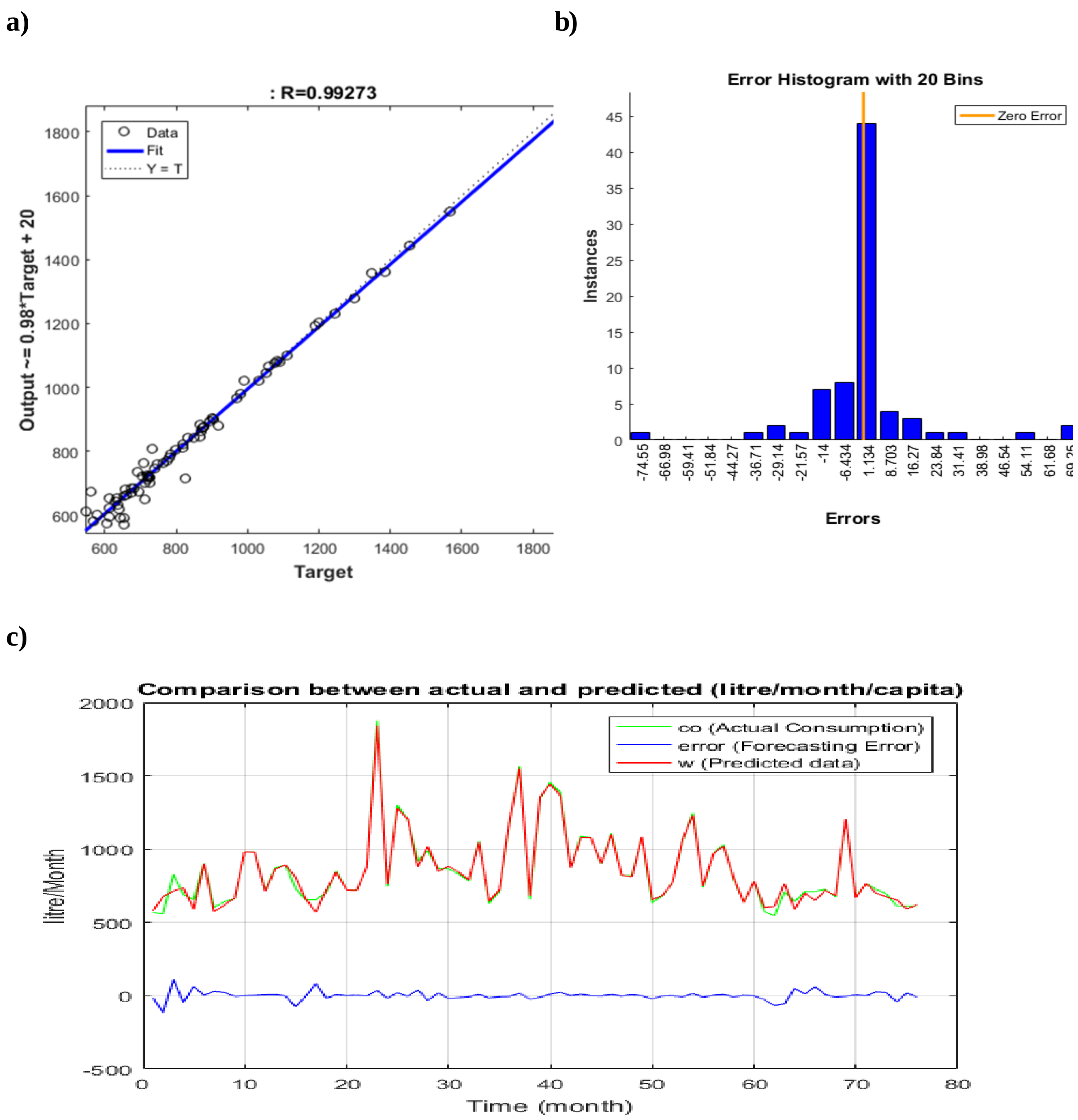


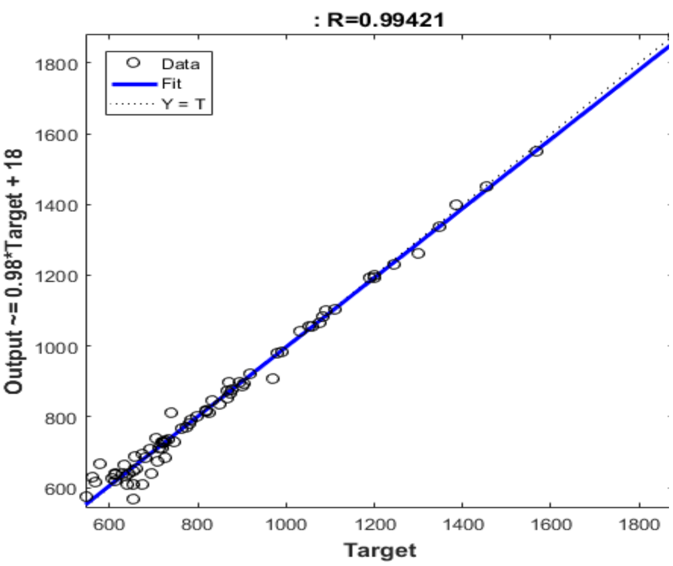
Figure 13 Comparison of the actual and forecasted residential water consumption of Model-1

4.1.1 Model-2

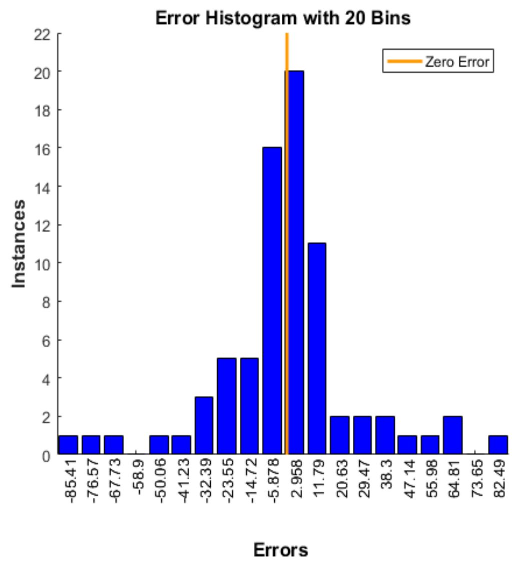
Input data are selected from 76 months (September-2010 to December-2016) water consumption time series and all weather data used in model 1 was used except rainfall (rainfall is not considered as input parameter as in Table 5). The training, testing, checking and simulation of the time series are similar to model 1. In this model the error histogram shows that the error is -50 to 64 litre/capital/month which is start to increase the error relative to model

1. Therefore from this analysis the error histogram shows that rainfall is one of the effective which affects water demand forecasting.

a)



b)



c)

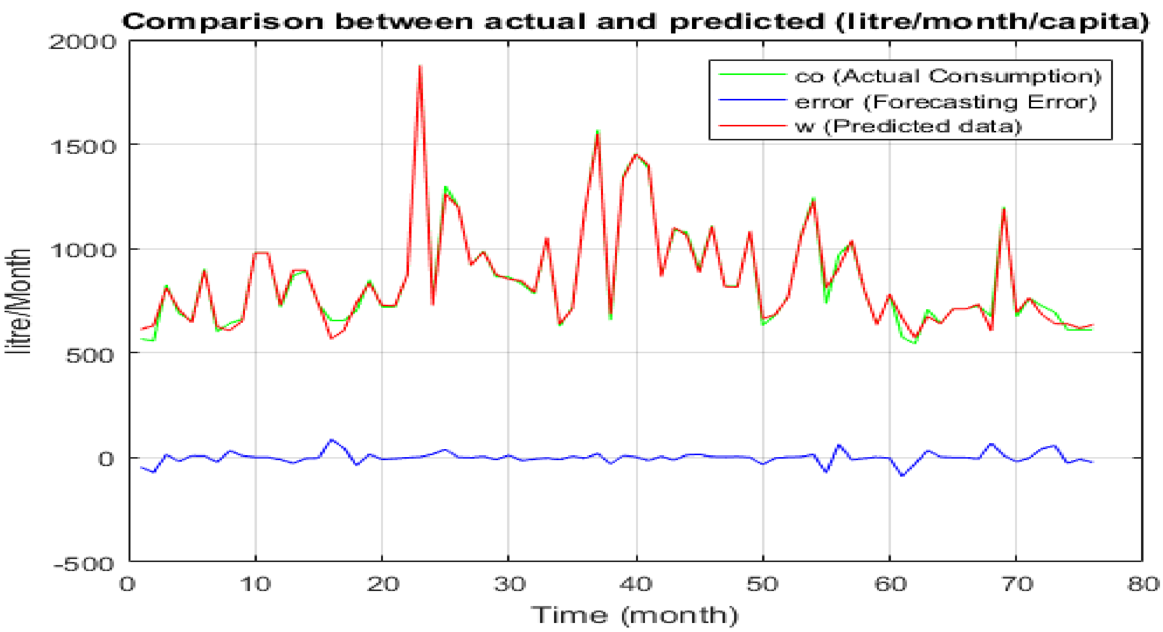


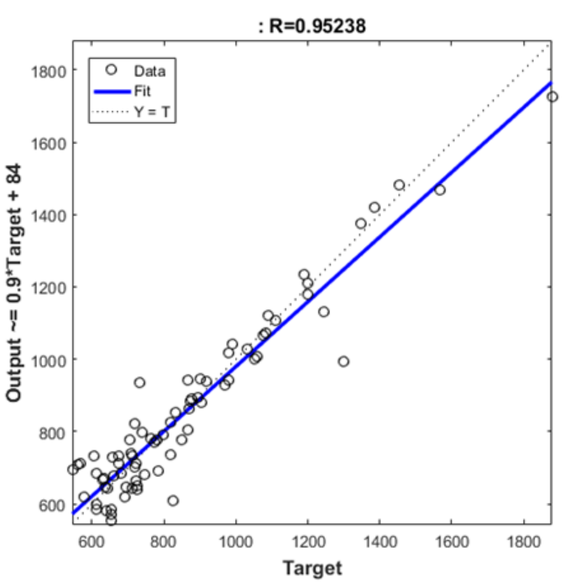
Figure 14Comparison of the actual and forecasted residential water consumption of Model-2

4.1.2 Model-3

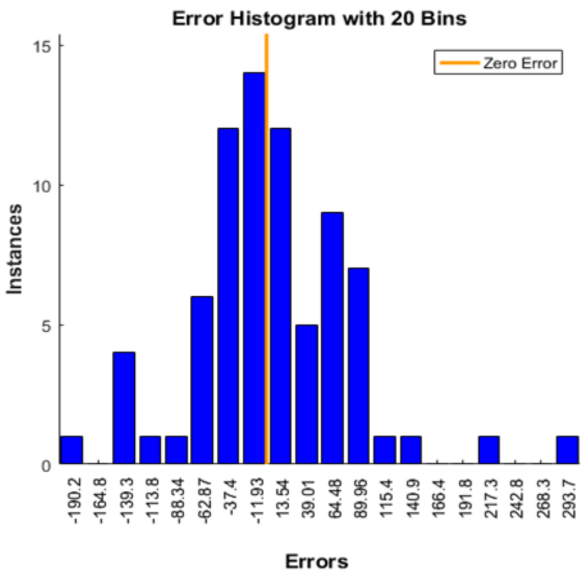
For this model also, input data are selected from 76 months (Sep-2010 to Dec-2016) water consumption time series and all weather data used in model 2 was used except the maximum temperature (rainfall and maximum temperature are not considered as input parameter as in Table 4). The training, testing, checking and simulation of the time series are similar to model

1. In this model the error histogram shows that the error is -164 to 140 litre/capital/month which is far from the error line.

a)



b)



c)

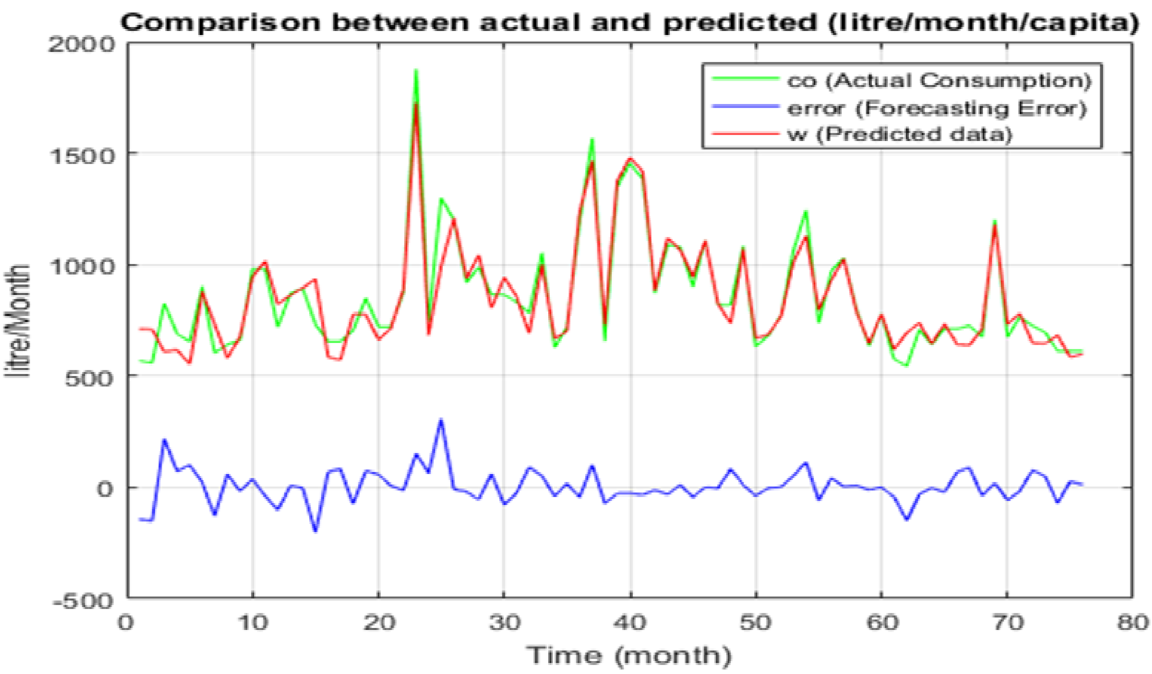
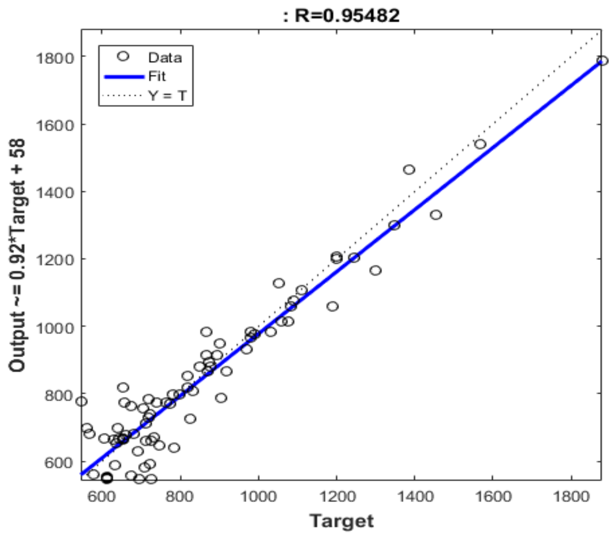


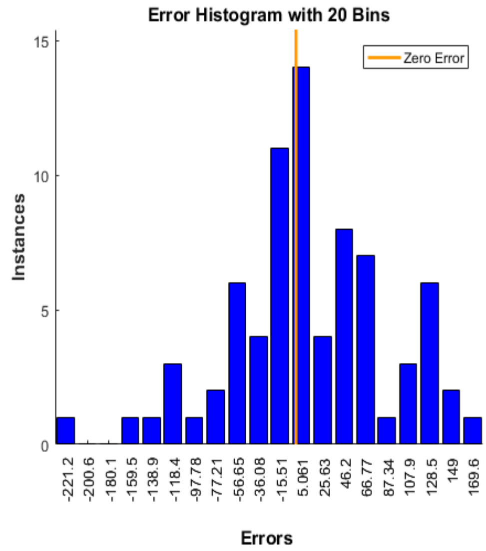
Figure 15 Comparison of the actual and forecasted residential water consumption of Model-3

4.1.3 Model-4

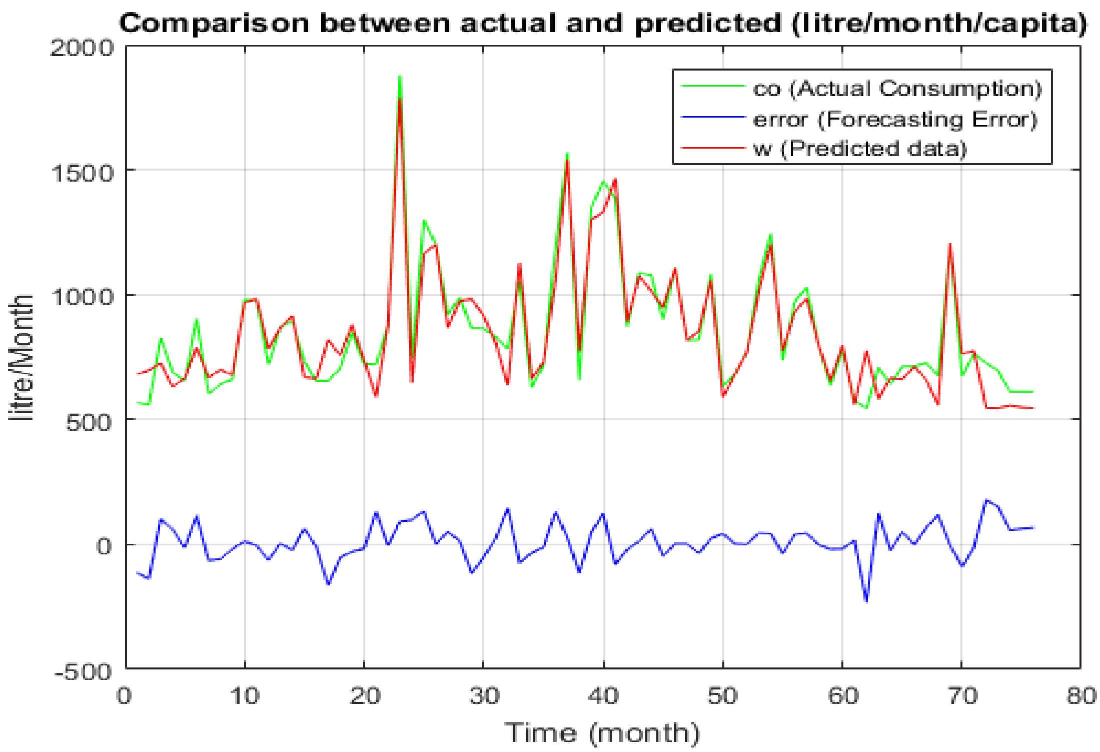
For this model also, input data are selected from 76 months (Sep-2010 to Dec-2016) water consumption time series and all weather data used in model 3 was used except minimum temperature (rainfall, maximum temperature and minimum temperature are not considered as input parameter as in Table 5).



a)



b)



c)

Figure 16 Comparison of the actual and forecasted residential water consumption of Model-4

4.1.4 Model-5

For this model also, input data are selected from 76 months (Sep-2010 to Dec-2016) water consumption time series and all weather data used in model 4 was used except Sunshine hour (rainfall, maximum temperature, minimum temperature and Sunshine hour are not considered as input parameter as in Table 5). The training, testing, checking and simulation of the time series are similar to model 4.

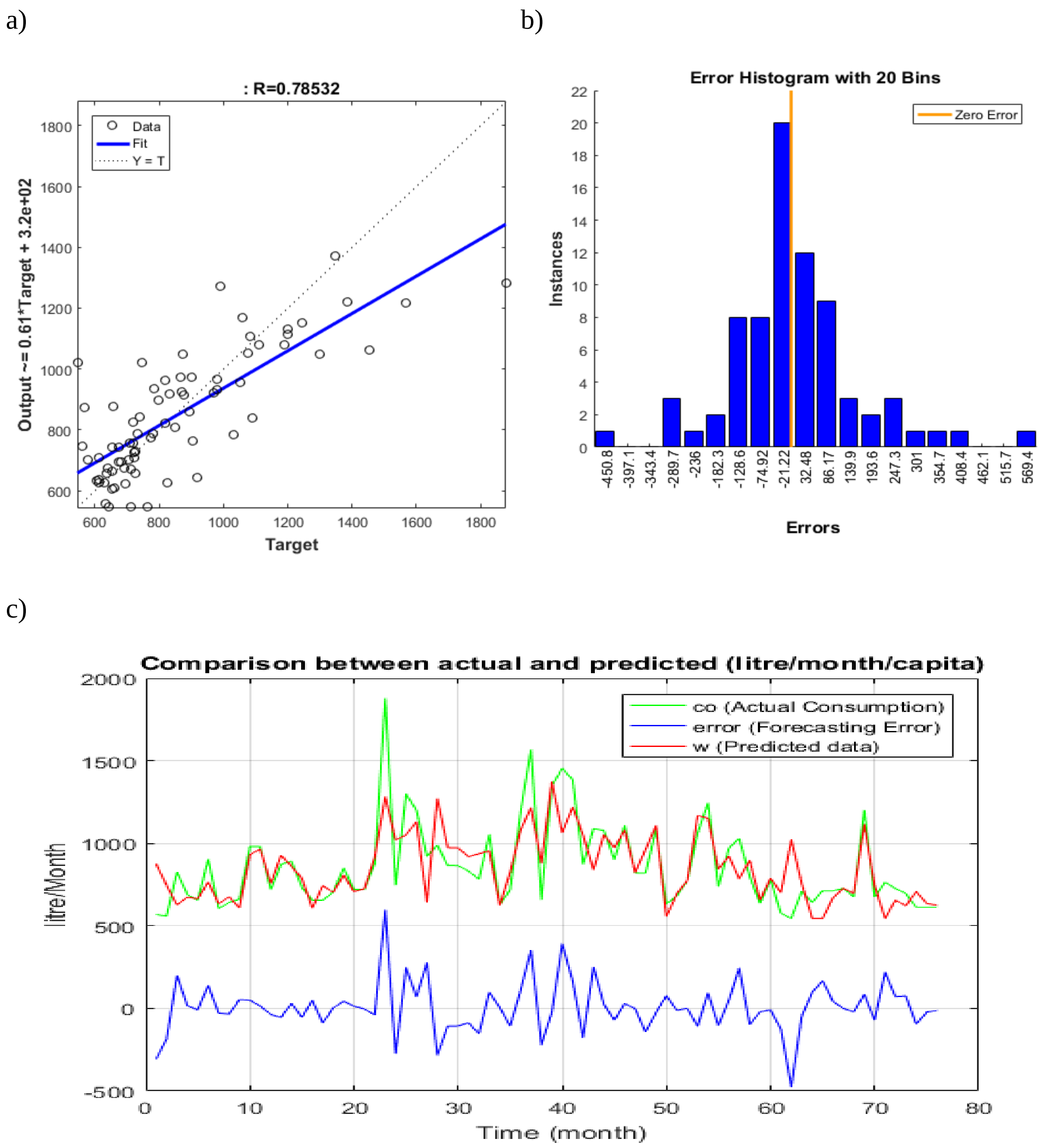
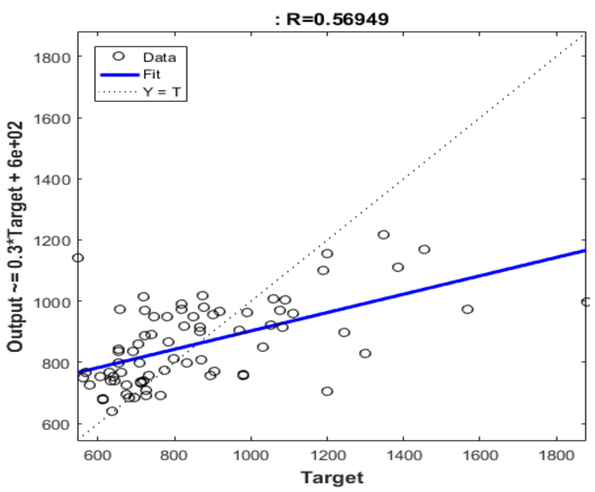


Figure 17 Comparison of the actual and forecasted residential water consumption of Model-5

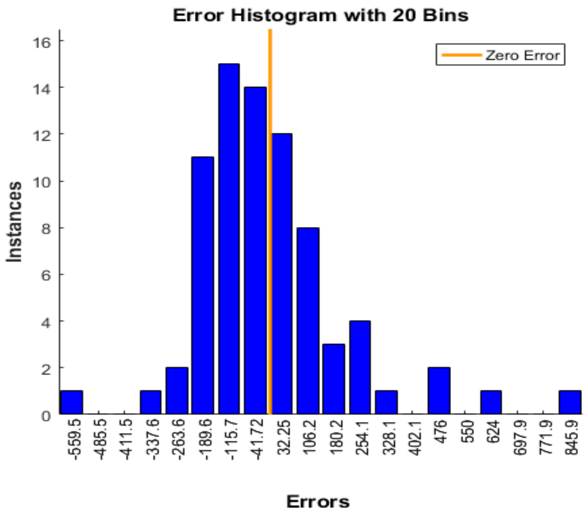
4.1.5 Model-6

For this model also, input data are selected from 76months (Sep-2010 to Dec-2016) water consumption time series and all weather data used in model 5 was used except relative humidity (rainfall, maximum temperature, minimum temperature, Sunshine hour are not considered as input parameter and relative humidity as in Table 4). The training, testing, checking and simulation of the time series are similar to model 5.

a)



b)



c)

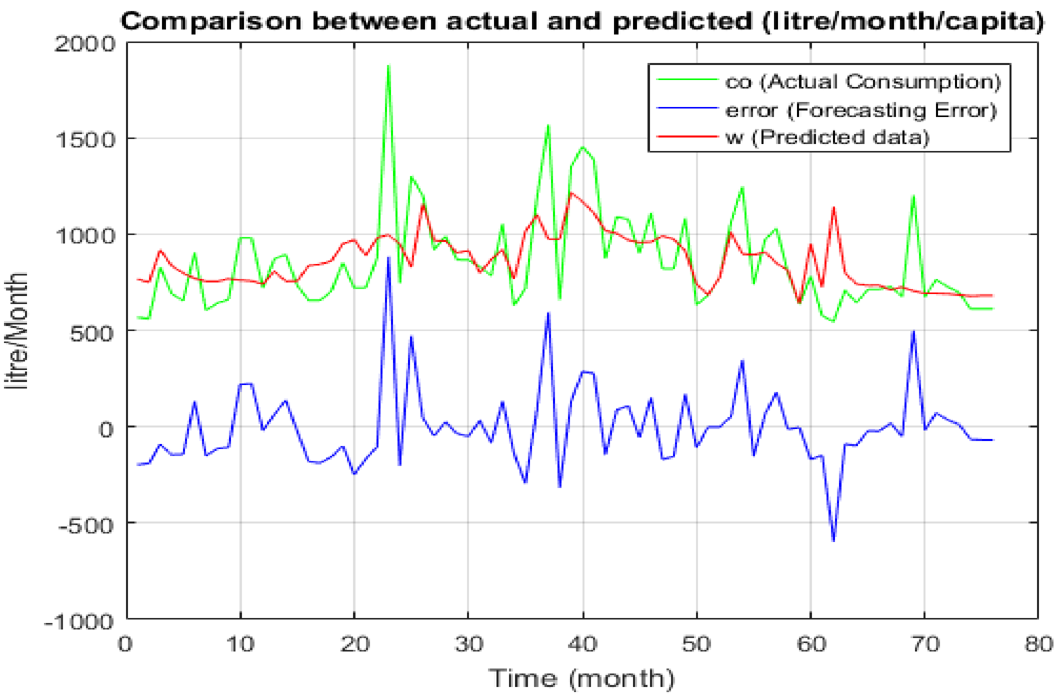
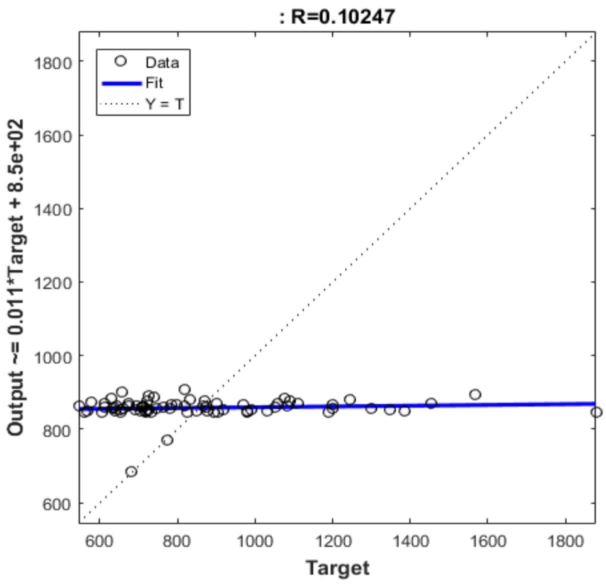


Figure 18 Comparison of the actual and forecasted residential water consumption of Model-6

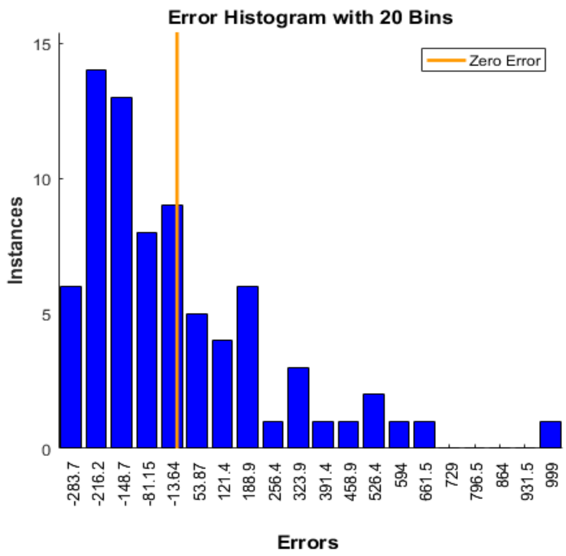
4.1.6 Model-7

Input data are selected from 76 months (Sep-2010 to Dec-2016) water consumption time series and all weather data used in model 6 was used except the cost of water (rainfall, maximum temperature, minimum temperature, Sunshine hour parameter and relative humidity are not considered as input as in Table 4). The training, testing, checking and simulation of the time series are similar to model 6.

a)



b)



c)

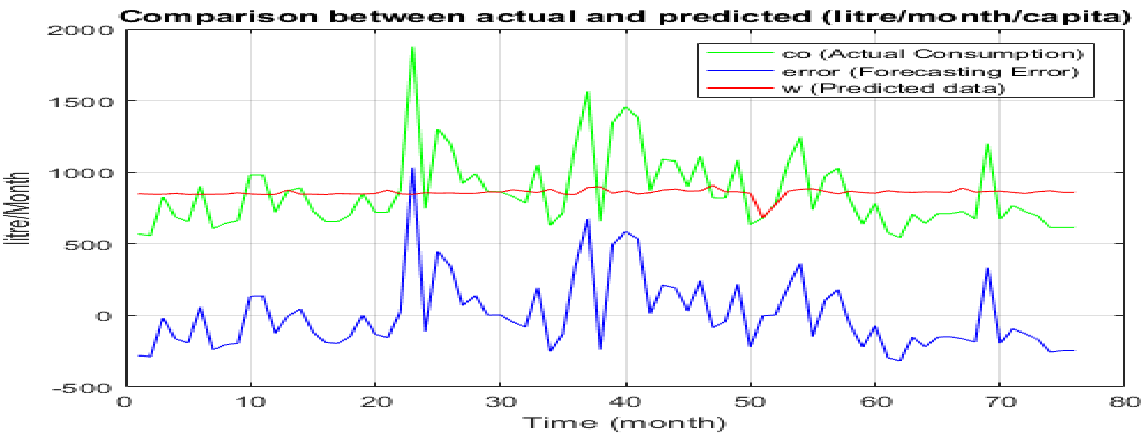


Figure 19 Comparison of the actual and forecasted residential water consumption of Model-7

4.2 Fuzzy logic prediction results

The following effective parameters were selected after a comprehensive investigation by regression between the meteorological parameters and water consumption:

- Daily average temperature, because of its high correlation with water consumption and possibility of future forecast by the meteorological organization with high accuracy.
- Rainfall because of its high correlation with water consumption.
- Relative humidity, which showed a high negative correlation.
- Monthly water consumption

After determining effective parameters has determined, all made from this general rule: "if daily average temperature is hot and relative humidity percent is extremely low, then daily consumption is high". It should be noted that by using the last year water consumption, the trend component is considered automatically and effects of most of the parameters which may influence the water consumption.

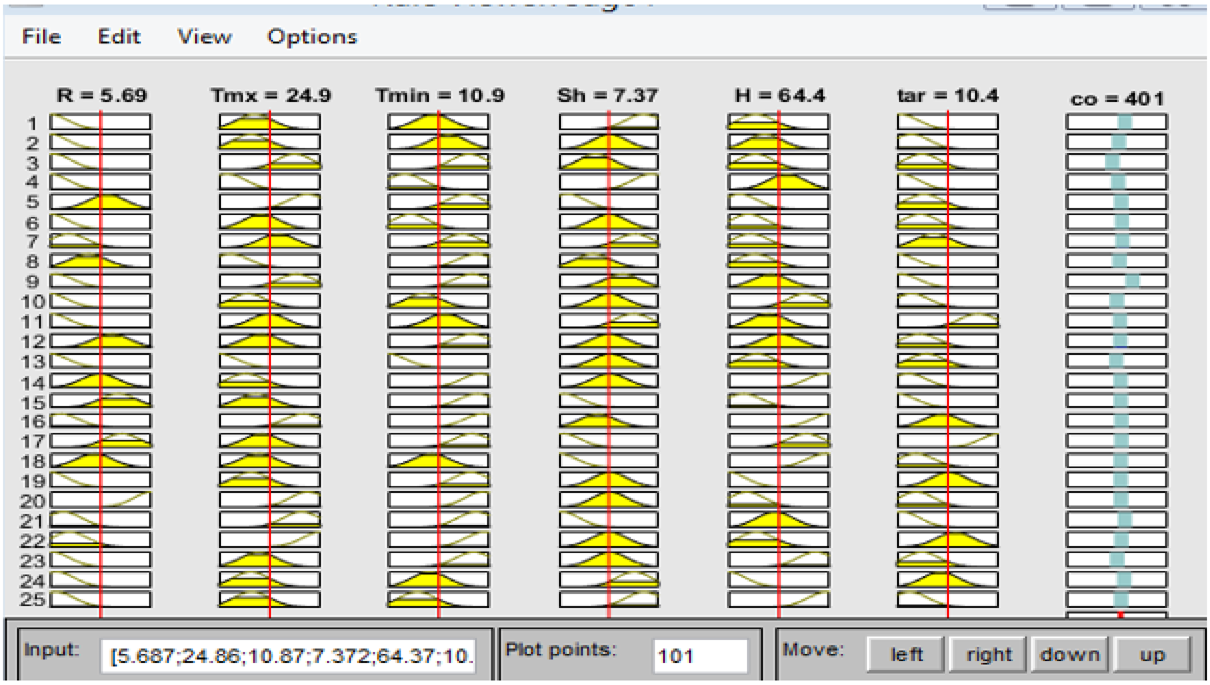


Figure 20 Rule viewer that simulates the entire using Mamdani fuzzy inference

In fuzzy models, the relations between meteorological parameters and water consumption were used to make fuzzy rules. The number and the type of memberships function assigned to input variable was chosen with trial and error method. After iterating the model, triangular and trapezoidal membership function were developed to compere the forecasting capability of fuzzy logic system with Mamdani and Segno fuzzy inference system. Outputs demonstrate that the results do not show high accuracy as shown in **Table 9** and Table 10.

Table 9 Forecasted water consumption using Mamdani and Segno fuzzy inference system with triangular Membership function

Actual (litre/c/month)	Mamdani fuzzy inference system		Segno fuzzy inference system	
	Forecasted (litre/c/month)	Percentage Error (%)	Forecasted (litre/c/month)	Percentage Error (%)
501.16	911	-81.78	1460	-191.324
806.04	401	50.25	353	56.206
1843.42	2020	-9.58	719	60.996

Table 10 Forecasted water consumption using Mamdani and Segno fuzzy inference system with trapezoidal Membership function

Actual (litre/c/month)	Mamdani fuzzy inference system		Segno fuzzy inference system	
	Forecasted (litre/c/month)	Percentage Error (%)	Forecasted (litre/c/month)	Percentage Error (%)
501.2	616.6	-23.0	560.7	-11.9
806.0	1012.0	-25.6	956.4	-18.7
1843.4	1674.5	9.2	1345.2	27.0

Mamdani approach have clear procedures, i.e. fuzzification, logic decision and defuzzificiation procedure. Sugeno approach, however, does not have an explicit defuzzification procedure. For the Mamdani approach, the output of each IF–THEN rule will be a fuzzy set for the output variable, so that the step of defuzzification is necessary so as to obtain crisp value of the output variable. However, in Sugeno method, the conclusion of each IF–THEN inference rule is a scalar rather than a fuzzy set for the output variable.

Fuzzy rule-based systems can be used as suitable representations of simple and complex physical systems. A small number of fuzzy sets leads to unreliable predictions whereas a large number leads to many calculations but better prediction. Therefore, to increase the forecasting accuracy of fuzzy logic system it is necessary to increase the number of input parameters.

4.3 Evaluating Forecast Accuracy (calibration)

The accuracy of forecasts is evaluated by comparing them with observed demand. This evaluation provides insights in recommending changes to existing models in order to reduce deviations in future forecasts. Prior to observing future forecasts, however, such evaluations can form the basis for model selection. The general approach is to consider competing forecasting models in a sequence of steps:

(i) divide the data set into an estimation period and a hold-out period;

(ii) use the estimation period to model demand;

(iii) evaluate the accuracy of the models by comparing the forecasts with the observed values for both the estimation period and the hold-out period and

(iv) select the best model based on its performance, as measured by any of the loss functions specified in (Equation 4.1 to 4.5)

4.3.1 ANN model calibration and Validation

The validation process of our algorithm consists of selecting the best training algorithm according to the regression parameter and average error obtained during the global process. The available dataset for training the ANN is divided into 3 subsets: model development data (70%) are used for calculating the gradient and updating the model parameters; training data (15%) are used in the stopping criterion of the training process (cross-validation); and validation data (15%) are used to determine the optimal number of inputs and optimal number of hidden neurons.

The way that available data are divided into training, testing, and validation subsets can have a significant influence on the performance of an artificial neural network (ANN). Despite numerous studies, no systematic approach has been developed for the optimal division of data for ANN models. Two methodologies were developed for dividing data into representative subsets, namely, a genetic algorithm (GA) and a self-organizing map (SOM) and these were compared with the conventional approach commonly used in the literature, which involves an arbitrary division of the data (Gavin J. Bowden, et al., 2002).
Thus the data splitting methods in this paper based on simple random sampling (conventional approach) technique. It is a semi-deterministic method, in which every k^{th} sample from a random starting point is selected to form the training, test and validation datasets. In implementing systematic sampling in this study, the data are first ordered in increasing values along the output variable dimension. Then the sampling interval is determined based on the training and test data proportions specified by the user. Thereafter, a starting point is randomly selected and training samples are drawn first, followed by the test samples. Finally, sampled data are allocated into the training, testing, and validation sets.

A set of ANNs with different number of inputs and different number of hidden neurons is trained. The optimal ANN is chosen as one with the lowest mean squared error (MSE) on the validation data. MSE is defined as follows (Equation 4.1 to 4.5)

Where N = number of data samples

Y_t = actual data of the i^{th} sample

$\hat{Y}t$ = model output for the i^{th} sample

$$\boldsymbol{Mean\ Absolute\ Deviation\ (MAD)} = \frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{t}=\mathbf{1}}^{\boldsymbol{N}}\left|\boldsymbol{Y_t} - \hat{Y}_{\boldsymbol{t}}\right| \qquad \text{Equation 4-1}$$

$$\boldsymbol{Mean\ Absolute\ Percentage\ Error\ (MAPE) = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|}$$ Equation 4-2

$$\boldsymbol{Mean\ squared\ Error\ (MSE) = \frac{1}{N}\sum_{t=1}^{N}\left(Y_t - \hat{Y}_t\right)^2}$$ Equation 4-3

$$\boldsymbol{Root\ Mean\ Squared\ Error\ (RMSE) = \sqrt{MSE}}$$ Equation 4-4

$$\boldsymbol{Accuracy\ = 100 - \mathrm{MAPE}}$$ Equation 4-5

In each of these functions, the forecast error is measured by the difference between the observed and forecast values, represented by $Y_t - \hat{Y}_t$. When comparing forecast performance for a given sample data, the model with the least value of MAPE a chosen loss function is supposed the most accurate. The accuracy of developed artificial neural network is listed in Table 11

Table 11 Accuracy of the calibrated artificial neural network for water consumption forecasting model

Model	Accuracy (%)
1	97.46
2	93.6
3	92.93
4	92.54
5	87.25
6	82.96
7	77.59

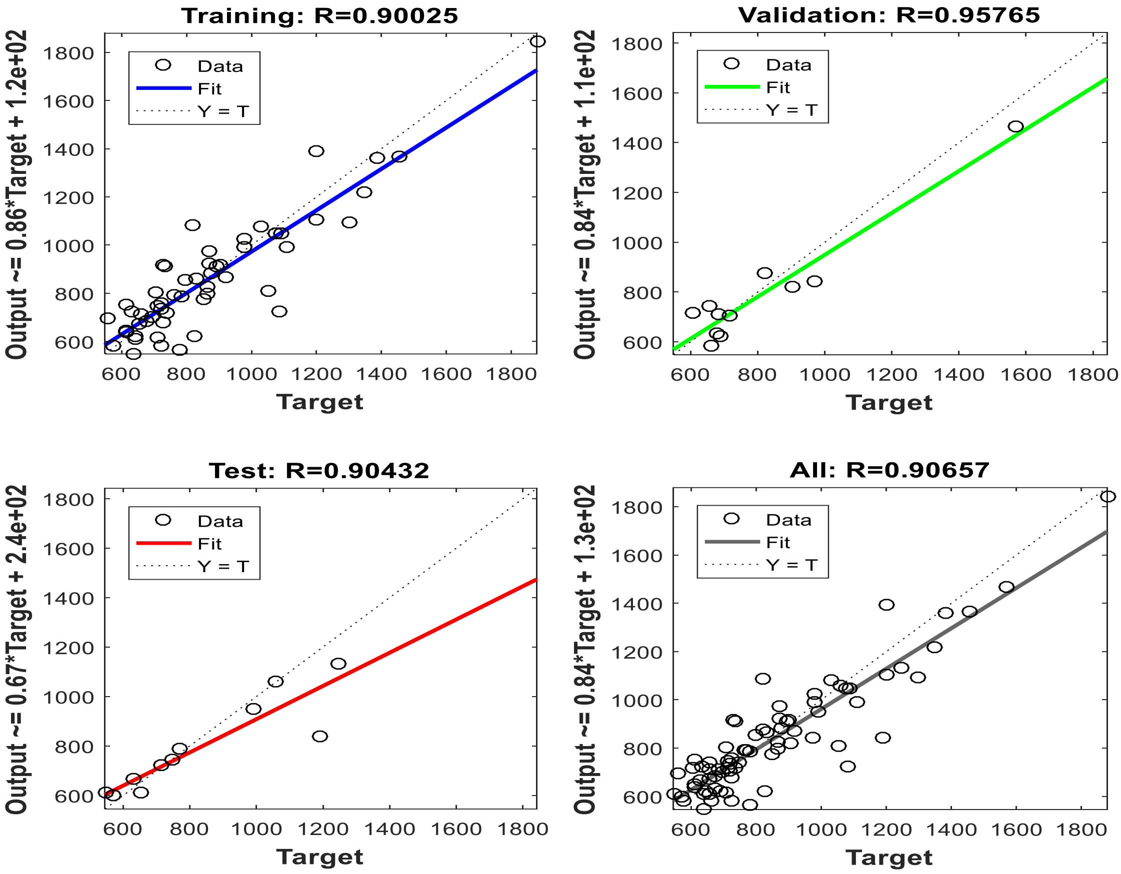


Figure 21 ANN model testing and validation for model 1

The validation of the model was performed by percentage error method as indicated in Figure 22 and Table 12. In Figure 22 the errors histogram of monthly water consumption calibration, indicating that 52.6% of model estimates fall within the percentage error of -3.7% to 0.5% and 30.3% within the range 4.8% to 9.1%. The modeled monthly total water consumption is compared with the observed monthly water consumption (Figure 13c). To quantitatively evaluate the performance of the statistical model was analyzed the correlations between observed and modeled water use series (Figure 13a). Thus the statistical models developed in this study have reasonably good performance in describing the observed monthly urban water consumption variations.

Table 12 Percentage error for ANN mode calibration

Bin (%)	Frequency	%	Cumulative (%)
-20.8	1	1.3	1.3
-16.5	0	0.0	1.3
-12.3	0	0.0	1.3
-8.0	2	2.6	3.9
-3.7	4	5.3	9.2
0.5	40	52.6	61.8
4.8	23	30.3	92.1
9.1	3	3.9	96.1
More	3	3.95	100

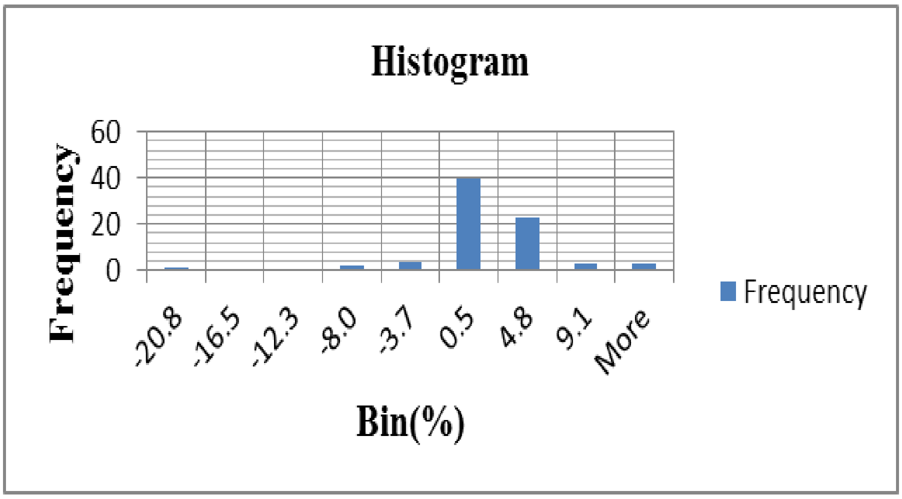


Figure 22 Percentage Error histogram for calibration

4.4 Comparison to other water demand forecasting by ANN an Fuzzy logic

When comparing the performance of ANN model, which shows best result in this research, with MATLAB other models. It is observed that comparable results with the models of (Alvisi et al., 2007) and (Ghiassi et al., 2008) develop ANN model and get R^2 of 99% accuracy, and seemingly better performance than the model of (Herrera et al., 2010), (Zoe, et al., 2011) develop a comparative evaluation of the stepwise regression method and the ANN model results show that the ANN obtains produced R^2 = 84%.

A research work of water consumption is predicted using fuzzy model using different membership function and rules criteria by Surendra, et al. (2014). From the result triangular membership with twelve rules criteria performed better compared, to the result got in this paper, with accuracy of 89.637%.

However, it is very difficult to compare forecast accuracies of different models, because the forecast accuracy highly depends on the variability of the water consumption and availability of influential factors in the study area of the research to be conduct. This indicates that the accuracy of forecasting models obtained by simulations on different datasets cannot be compared objectively. Therefore it is difficult to draw conclusions about the performance of different models found in literature of different places.

4.5 Water Demand forecasting

According to the census data of the central statistic authority, the total population of Mekelle in 2007 is 215,914 and average population growth rate 2007 to 2030 is determined as 3.7% and the 2030 to 2045 rate as 2.7%. Mekelle water supply project feasibility study report shows that the total population of Mekelle is predicted to reach 497,957 in 2030 and 742,590 in 2045. The average water consumption is computed based the artificial neural network result shows that 84Letre/cap/d and 101Litre/cap/d for 2030 and 2045 respectively as shown in Table 13

Table 13 Forecasted water demand from 2017 to 2045

Year	Population	Per Capita(litre/c/day)	Domestic Demand(m3/day)	Total Domestic Demand(m3/day)
2017	349283	76	26425.3	44592.7
2020	420263	79	32941.19	55143.55
2025	469600	81	38134.75	61778.29
2030	497957	**84**	41894.78	67869.55
2035	569931	87	49617.95	80381.08
2040	652306	90	58764.86	91232.44
2045	746589	**101**	75355.63	116989.6

Non domestic water consumption (including industrial, commercial, institutional, greening, firefighting and other water consumption) is determined based on prediction of Tahal. Non-domestic water use of Mekelle city is about 35% of the domestic water demand for 2030 and 2045. The water loss of the city is estimated to be 20% for 2030 and 15 for 2045 of the domestic plus non-domestic water demand of the city. Taken all into account, the total urban water demand will be **67869.55**m^3/day and **116989.6** m^3/day for 2030 and 2045 respectively.

5 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

One of the primary tasks of water utility is to accurately predict the demand requirements at future times. Results obtained from demand forecasting process are used in planning and operation. A feed-forward ANN model using back propagation and if-then rule based fuzzy logic (FL) were developed to forecast monthly water consumption for Mekelle city. The results show that an appropriate (97.46%) simulation accuracy can be achieved using artificial neural network and fuzzy logic shows poor forecasting accuracy with 19.2% average absolute percentage error. It is therefore concluded that the neural network approach for modelling demand is capable of yielding good results and can be considered as a substitution to traditional demand forecasting approaches. Furthermore, the developed methodology can be used to forecast water in the water utilities with good precision.

The developed methodology is compact, adaptable and simple to apply. Neural Network can learn to approximate any function just by using example data that is representative of the desired task. They are model free estimators, which are capable of solving complex problem based on the presentation of a large number of training data. Neural Networks estimate a function without mathematical description of how the outputs functionally depend on the inputs. They represent a good approach that is potentially robust and fault tolerant. In this work, water demand forecasting method based on neural network was implemented using MATLAB. The system performs better results than some other systems. Finally the water demand of Mekelle city was forecasted for which is **67869.55** m^3/day and **116989.6** m^3/day for 2030 and 2045 respectively.

5.2 Recommendation

The following four recommendations can be made with regard to future research based on the limitations of the present research.

1. The first is the need to improve data availability because the artificial neural network forecasting a very hungry of data. Data on urban water use should be recorded according to sectors in the city. Outdoor such as garden irrigation and parks water use should also be included in the statistics. If possible, a city-wide, or even nation-wide data base on water use should be established in order to provide detailed and systematic information about water use in the cities.

2. Further research should be undertaken on the relationship between water use and factors affecting along with the improvement of data availability. In Mekelle city, the relationship between water demand and water price has not been studied at all. Therefore attention should be paid to this field, especially from the perspective of the development of urban infrastructure.
3. The accuracy of the ANN model is 97.46% and it can further be improved if more influential factors are used as input, in addition to the factor used in this research. This would be possible based on further research on the relationships between water use and the explanatory factors. It would also be desirable to develop the model towards a complete system dynamic model, which takes the feedback effects of water supply into account.
4. The forecasting model developed deliberately serves as a city wide approach to the management of water resources in Mekelle city. The parameters and initial variables depends greatly on the situation in each city under investigation, and these can be quite different from city to city. Therefore, it is recommended that this model, or the structure and the general idea behind it, be used to forecast water demand for Mekelle city by those who are intimately familiar with the local water use situation.

REFERENCE

1. A.H. Lobbrecht, Y. D. D. S., October, 2002. *Applications of Neural Networks and Fuzzy Logic to Integrated Water Management*, Delft: s.n.
2. Adamowski, J., 2010. Comparison of Multivariate Regression and Artificial Neural Networks for Peak Urban Water-Demand Forecasting: Evaluation of Different ANN Learning Algorithms. *Journal of hydrologic engineering © asce / October 2010 / 729.*
3. Adamowski, J. F., 2008. Peak Daily Water Demand Forecast Modeling Using Artificial Neural Networks. *Journal of water resources planning and management*.
4. Altunkaynak, A., Cakmakci, M. & Zoger, M., 2004. Water Consumption Prediction of Istanbul City by Using Fuzzy Logic Approach. *Water Resources Management*.
5. Alvisi, S., 2007. A short-term, pattern-based model for water-demandforecasting. *Journal of Hydroinformatics.*
6. Amaury de Souza, et al., 2015 . Climatic Variations and Consumption of Urban Water. *Atmospheric and Climate Sciences,* pp. 292-301.
7. An-Chi Huang, Tzong-Yeang Lee, Yu-Chen Lin & Chung-Fu Huang, 2017. Factor Analysis and Estimation Model of Water Consumption of Government Institutions in Taiwan.
8. Bishop, 1995. *Neural networks for pattern recognition*. New York: Oxford University Press.
9. Bozokalfa, G., 2005. Artificial neural networks and fuzzy logic models for cooling load prediction.
10. Cabral, M. S. F., 2014. *Water Demand Projection in Distribution Systems using a Novel Scenario Planning Approach*. s.l.:s.n.
11. Chen, W. & Jin, R., 2002. *Analytical Variance-Based Global Sensitivity Analysis in Simulation-Based Design under Uncertainty,* s.l.: University of Michigan.
12. Daniel, 1991. Neural networks applications in hydrology and water resource engineering.
13. Daniell, 1991. *Neural networks applications in hydrology and water resource engineering*. s.l.:s.n.
14. Dini, M. T. a. M., 2009. Fuzzy and neuro-Fuzzy forshort term water demand forecasting in Tehran. *Iranian Journal of Science & Technology.*
15. Edara, P. K., October 2003. *Mode Choice Modeling Using Artificial Neural Networks.* Falls Church, Virginia: s.n.

16. Feasiblity study, 2014. Ethiopia Mekelle Water Supply Project Feasibility study.
17. Ferreira, M. S., 2014. *Water Demand Projection in Distribution Systems using a Novel Scenario Planning Approach.* s.l.:s.n.
18. Gavin J. Bowden, Holger R. Maier & Graeme C. Dandy, 2002. Optimal division of data for neural network models in water resources applications. *Water resource research,* Volume VOL. 38.
19. Ghiassi, M., 2008. Urban Water Demand Forecasting with a Dynamic Artificial Neural Network Model.
20. Hartley, J. A., 1995. *A neural network and rule based System application in water demand forecasting.* Brunei University: s.n.
21. Herrara, M., Luís , T. & Joaquín , I., 2010. Predictive models for forecasting hourly urban water demand. *Journal of Hydrology,* Volume 387, pp. 213-234.
22. Herrera, 2010. Predictive models for forecasting hourly urban water demand. *Journal of Hydrology.*
23. Jones, 2008. *Forecasting urban water demand, American Water Works Association.* s.l.:s.n.
24. Kame'enui, A. E., 2003. *Water Demand Forecasting in the Puget Sound Region: Short and Long-term Models.* University of Washington: s.n.
25. Kerachian, F. F. &. R., 2016. Municipal water demand forecasting under peculiar fluctuations in population: a case study of Mashhad, a tourist city. *Hydrological Sciences Journal, 61:8, 1524-1534, DOI: 10.1080/02626667.2015.1027208.*
26. Lobbrecht, A., 2002. Applications of Neural Networks and Fuzzy Logic to Integrated Water Management.
27. M. TABESH AND M. DINI, 2009. FUZZY AND NEURO-FUZZY MODELS FOR SHORT-TERM WATER DEMAND FORECASTING IN TEHRAN. *Iranian Journal of Science & Technology, Transaction B, Engineering.*
28. Mahmoud S. Nasr et.al,, 2012. Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT. *Alexandria Engineering Journal.*
29. Michael , Š. & Jiří , Š., 2010. Artificial Neural Networks Numerical Forecasting of Economic Time Series. *Mendel University in Brno, Faculty of Business and Economics, Dept. of Informatics Czech Republic.*
30. Mohammadzadeha, A., 2011. Forecasting the Cost of Water Using a Neural Network Method in the Municipality of Isfahan.

31. Neeru Gupta et.al, 2012. Applications of Neural Networks and Fuzzy Logic for Integrated Water Management: Review of Theory and Applications. *International Journal of Engineering, Science and Metallurgy.*
32. Omar , A. M. A., Aous, Y. A. & Balasem , S. S., 2015. Comparison between the Effects of Different Types of Membership Functions on Fuzzy Logic Controller Performance. *International Journal of Emerging Engineering Research and Technology,* Volume Volume 3, pp. PP 76-83.
33. Ozeger, M., 2005. Water Consumption Prediction of Istanbul City by Using Fuzzy Logic Approach. *Water Resources Management.*
34. Reitermanov´, Z., 2010. Data Splitting. Volume Part I, p. 31–36.
35. Romano, M., 2012. *ADAPTIVE WATER DEMAND FORECASTING FOR NEAR REALTIME MANAGEMENT OF SMART WATER DISTRIBUTION SYSTEMS.* s.l.:s.n.
36. S.L.Zhou, 2000. Frequency analysis of water consumption for metropolitan area of Melbourne. *Journal of Hydrology.*
37. Surendra H. J and Paresh Chandra Deka , 2014. Development of a Fuzzy Logic Based Model using different Membership and Rules criteria for predicting water consumption using Climatic variables. *International Journal of Scientific & Engineering Research.*
38. Viumdal, H., 2013. *Developing membership functions for fuzzy-neural applications.* Tel-Tek: s.n.
39. Zadeh, 1965. *Fuzzy Sets, Information and Control.* s.l.:s.n.
40. Z., M., M. & Z., 2011. Prediction of Water Demands in a Water Treatment Plant Using an Artificial Neural Network Model. *Journal of Water Management Modeling,* pp. 241-16.
41. Zoe J. Y. Zhu, W. G. B. M. a. E. M., 2011. Prediction of Water Demands in a Water Treatment Plant Using an Artificial Neural Network Model. *Journal of Water Management Modeling R241-16.*
42. Zoe J. Y. Zhu, W. G. B. M. a. E. M., 2011. Prediction of Water Demands in a Water Treatment Plant Using an Artificial Neural Network Model.

APPENDICES

Appendix 1: MATLAB code for the BP neural network method

```
%% MATLAB code for the BP neural network method.
This demonstration illustrates how a function fitting neural
network can estimate to forecast water demand of Mekelle city.
The Problem: Predict water demand.To build a neural network that
can estimate the water demand the following input data were used
    1. Water consumption data (Litre/c/month)
    2. Maximum and Minimum temperature (^{o}C)
    3. Rainfall (mm)
    4. Relative Humidity (%)
    5. Sunshine Hour (Hr)
    6. Number of users
%input data
load MCR.txt % Mekelle city residential water consumption
time=MCR(:,1);              % September 2010 to  December 2016
inputvariable=MCR(:,2:8);  % input variables used for water
demand modelling
consumption=MCR(:,9);% Target variables which going to be
modeled
x=inputvariable';
co=consumption';
 %% Fitting neural network
setdemorandstream(15) % set the training and testing randomly
 net = fitnet(10);
view(net)
%% 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,x,co);
testX  = x(:,tr.testInd);% separates input into three sets:
training, validation, and testing
```

```
testT = co(:,tr.testInd);% separates targets into three sets:
training, validation, and testing
testco = net(testX);
perf = mse(net,testT,testco);
 %% Training the Neural Network
[xn,minx,maxx,con,minco,maxco]=premnmx(x,co);%Normalization of
Data(scale inputs and targets so that they fall in the range [-
1,1]).
NodeNum1=30; % number of nodes of the first layer in hidden layer
TypeNum = 1; % output vector dimensions
TF1 = 'tansig'; %transfer function for the 1st layer
TF2 = 'tansig'; %transfer function for the 2st layer
 net=newff(minmax(xn),[NodeNum1,TypeNum],{TF1 TF2},'traingdx');
nntraintool
 %creat network 'traingdm'
net.trainParam.show=50;
net.trainParam.epochs=5000; %Number of training times
net.trainParam.goal=1e-5; %accuracy of training
net.trainParam.lr=0.01; %learning rate
net=train(net,xn,con);
 % use the trained network to simulate the test data
p2n=tramnmx(x,minx,maxx);%Normalization of test data
an=sim(net,p2n);
[w]= postmnmx(an,minco,maxco);%Reverse normalization of Data,
also the desired result
nntraintool
%% Check the Neural Network Training Performance
% Performance is measured in terms of mean squared error, and
shown in log scale.
 plotperform(tr)
%% Testing the Neural Network
A measure of how well the neural network has fit the data is the
regression plot.
plotregression(co,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 = co - w; % 'W' is Forecasted water consumption
ploterrhist(error)
%% 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(co,'g')
hold on
grid on
plot(error,'b')
plot(w,'r')
ylabel('litre/Month');
xlabel('Time (month)');
legend('co (Actual Consumption)','error (Forecasting Error)','w
(Predicted data)')
grid on
% Create title
title({'Comparison between actual and predicted
(litre/month/capita)'});
```

Appendix 2: Actual inputs and simulated outputs of ANN Model

Table 14 actual inputs and simulated outputs of ANN Model 1 from Sept 2010 to Dec 2016

	Consumption(L/c/month)					Consumption(L/c/month)			
Year	Actual	Forecasted	Error	Percentage Error	Year	Actual	Forecasted	Error	Percentage error
9/1/2010	568	573.57	-5.57	0.97	11/1/2013	1117	1113.01	3.89	0.35
10/1/2010	559	588.50	-29.50	5.01	12/1/2013	1166	1163.57	2.23	0.19
11/1/2010	826	819.06	6.94	0.85	1/1/2014	1215	1232.60	-17.90	1.45
12/1/2010	690	702.64	-12.64	1.80	2/1/2014	872	876.18	-4.18	0.48
1/1/2011	654	632.55	21.45	3.39	3/1/2014	1089	1119.96	-30.96	2.76
2/1/2011	903	907.47	-4.47	0.49	4/1/2014	1077	1058.08	18.92	1.79
3/1/2011	604	638.56	-34.56	5.41	5/1/2014	901	898.43	2.57	0.29
4/1/2011	641	586.72	54.28	9.25	6/1/2014	1109	1110.28	-1.28	0.11
5/1/2011	661	661.83	-0.83	0.13	7/1/2014	820	819.68	0.32	0.04
6/1/2011	980	978.33	1.67	0.17	8/1/2014	819	818.96	0.04	0.01
7/1/2011	980	980.57	-0.57	0.06	9/1/2014	1083	1084.11	-1.11	0.10
8/1/2011	720	722.17	-2.17	0.30	10/1/2014	632	599.95	32.05	5.34
9/1/2011	871	868.87	2.13	0.24	11/1/2014	682	683.29	-1.29	0.19
10/1/2011	893	892.27	0.73	0.08	12/1/2014	772	771.51	0.49	0.06
11/1/2011	733	746.54	-13.54	1.81	1/1/2015	1060	1060.66	-0.66	0.06
12/1/2011	655	658.61	-3.61	0.55	2/1/2015	1245	1227.82	17.18	1.40
1/1/2012	655	661.56	-6.56	0.99	3/1/2015	738	737.74	0.26	0.03
2/1/2012	704	705.42	-1.42	0.20	4/1/2015	971	971.11	-0.11	0.01
3/1/2012	850	836.68	13.32	1.59	5/1/2015	1030	1029.19	0.81	0.08
4/1/2012	721	732.09	-11.09	1.51	6/1/2015	798	797.05	0.95	0.12
5/1/2012	721	723.60	-2.60	0.36	7/1/2015	636	631.20	4.80	0.76
6/1/2012	876	877.54	-1.54	0.18	8/1/2015	779	776.83	2.17	0.28
7/1/2012	1064	1061.78	2.22	0.21	9/1/2015	576	586.70	-10.70	1.82
8/1/2012	745	746.53	-1.53	0.21	10/1/2015	545	623.33	-78.33	12.57
9/1/2012	1299	1225.97	73.03	5.96	11/1/2015	709	641.49	67.51	10.52
10/1/2012	1201	1209.11	-8.11	0.67	12/1/2015	642	653.56	-11.56	1.77
11/1/2012	919	913.79	5.21	0.57	1/1/2016	712	707.70	4.30	0.61

12/1/2012	989	985.71	3.29	0.33	2/1/2016	712	702.54	9.46	1.35
1/1/2013	866	867.08	-1.08	0.12	3/1/2016	727	729.29	-2.29	0.31
2/1/2013	865	864.82	0.18	0.02	4/1/2016	675	668.69	6.31	0.94
3/1/2013	831	832.67	-1.67	0.20	5/1/2016	1201	1197.87	3.13	0.26
4/1/2013	783	795.08	-12.08	1.52	6/1/2016	674	675.02	-1.02	0.15
5/1/2013	1053	1053.54	-0.54	0.05	7/1/2016	764	764.03	-0.03	0.00
6/1/2013	629	628.50	0.50	0.08	8/1/2016	726	724.46	1.54	0.21
7/1/2013	718	717.27	0.73	0.10	9/1/2016	696	695.50	0.50	0.07
8/1/2013	1189	1188.08	0.92	0.08	10/1/2016	612	623.65	-11.65	1.87
9/1/2013	1018	1018.40	-0.90	0.09	11/1/2016	612	609.34	2.66	0.44
10/1/2013	1068	1071.26	-3.26	0.30	12/1/2016	612	619.04	-7.04	1.14

Appendix 3: Input Range Membership function for fuzzy logic rule construction

Input Range

```
[Input1]
Name='Rainfall'
Range=[0 11.37]
[Input2]
Name='Temp'
Range=[7.7968 28.1484]
[Input3]
Name='RH'
Range=[46.4233 82.326]
NumMFs=3
[Input4]
Name='water cost'
Range=[4 15]
[Input5]
Name='population'
Range=[236650 350211]
[Input6]
Name='Sunshine_HR'
```

```
Range=[4.1419 10.6019]
[Output1]
Name='water consumption' Range=[18 60]
```

Membership function

```
[Input1]%Rainfall (mm)
NumMFs=3
MF1='low':'trimf',[0 2.8425 5.685]
MF2='medium':'trimf',[2.8425 5.685 8.5275]
MF3='rainy':'trimf',[5.685 8.5275 11.37
[Input2]% Temperature (0C)
NumMFs=3
MF1='cold':'trimf',[7.7968 12.8847 17.97286]
MF2='average':'trimf',[12.8847 17.97286 23.0605]
MF3='hot':'trimf',[17.97286 23.0605 28.1484]
[Input3]%Humidity (%)
NumMFs=3
MF1='low':'trimf',[46.46 55.437 64.41]
MF2='average':'trimf',[55.399 64.375 73.351]
MF3='humid':'trimf',[64.375 73.351 82.3226]
[Input4]% Cost of water (ETB)
NumMFs=3
MF1='low_cost':'trimf',[443048 38450613.25 76458178.5]
MF2='average_cost':'trimf',[38450613.25 76458178.5
114465743.75]
MF3='high_cost':'trimf',[76458178.5 114465743.75 152473309]
[Input5] %Population
NumMFs=3
MF1='small':'trimf',[4 6.75 9.5]
MF2='average':'trimf',[6.75 9.5 12.5]
MF3='large':'trimf',[9.5 12.25 15]
[Input6] % Sunshine Hour (hours)
NumMFs=3
MF1='low':'trimf',[4.1419 5.7569 7.3719]
MF2='average':'trimf',[5.7569 7.3719 8.9869]
```

```
          MF3='shiny':'trimf',[7.3719 8.9869 10.6019]
[Output] % Per capita consumption (L/c/day)
NumMFs=3
               MF1='min':'trimf',[18 28.5 39]
               MF2='average':'trimf',[28.5 39 49.5]
               MF3='max':'trimf',[39 49.5 60]
```

Appendix 4: Dixon test for outliers for the forecasted water consumption from ANN model

XLSTAT 2014.5.03 - Dixon test for outliers

Significance level (%): 5

Iterations: Unlimited

Number of simulations: 1000000

Summary statistics:

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Value	76	0	76	573.570	1113.013	802.98	159.247

Dixon test for outliers / Two-tailed test:

R10 (Observed value)	0.024
R10 (Critical value)	0.229
p-value (Two-tailed)	0.477
Alpha	0.05

The p-value has been computed using 1000000 Monte Carlo simulations. 99% confidence interval on the p-value:] 0.476, 0.479 [

Test interpretation:

H0: There is no outlier in the data

Ha: The minimum or maximum value is an outlier

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0. The risk to reject the null hypothesis H0 while it is true is 47.74%.

Table 15 Z-scores for outlier detection

No	Value	Z-score	No	Value	Z-score	No	Value	Z-score	No	Value	Z-score
1	573.6	-1.4	20	732.1	-0.4	39	1113.0	1.9	58	797.1	0.0
2	588.5	-1.3	21	723.6	-0.5	40	1100.8	1.9	59	631.2	-1.1
3	819.1	0.1	22	877.5	0.5	41	986.9	1.2	60	776.8	-0.2
4	702.6	-0.6	23	1061.8	1.6	42	876.2	0.5	61	586.7	-1.4
5	632.6	-1.1	24	746.5	-0.4	43	1101.0	1.9	62	623.3	-1.1

6	907.5	0.7	25	898.9	0.6	44	1058.1	1.6	63	641.5	-1.0
7	638.6	-1.0	26	739.6	-0.4	45	898.4	0.6	64	653.6	-0.9
8	586.7	-1.4	27	913.8	0.7	46	1110.3	1.9	65	707.7	-0.6
9	661.8	-0.9	28	985.7	1.1	47	819.7	0.1	66	702.5	-0.6
10	978.3	1.1	29	867.1	0.4	48	819.0	0.1	67	729.3	-0.5
11	980.6	1.1	30	864.8	0.4	49	1084.1	1.8	68	668.7	-0.8
12	722.2	-0.5	31	832.7	0.2	50	600.0	-1.3	69	608.1	-1.2
13	868.9	0.4	32	795.1	0.0	51	683.3	-0.8	70	675.0	-0.8
14	892.3	0.6	33	1053.5	1.6	52	771.5	-0.2	71	764.0	-0.2
15	746.5	-0.4	34	628.5	-1.1	53	1060.7	1.6	72	724.5	-0.5
16	658.6	-0.9	35	717.3	-0.5	54	871.8	0.4	73	695.5	-0.7
17	661.6	-0.9	36	806.0	0.0	55	737.7	-0.4	74	623.6	-1.1
18	705.4	-0.6	37	1018.4	1.4	56	971.1	1.1	75	609.3	-1.2
19	836.7	0.2	38	1071.3	1.7	57	1029.2	1.4	76	619.0	-1.2

Appendix 5: The collected weather and water consumption time series data

Table 16 Mekelle city weather data

	year	Jan	Feb	Mar	April	May	Jun	July	Aug	Sept	Oct	Nov	Dec
	2010									2.3	0.0	0.0	0.0
	2011	0.1	0.3	0.7	0.6	2.0	1.5	7.7	4.3	0.7	0.0	1.1	0.0
	2012	0.0	0.1	0.1	0.6	1.2	1.5	5.4	5.9	1.8	0.7	0.2	0.0
	2013	0.0	0.1	1.0	0.5	0.4	6.1	5.0	0.6	0.0	0.2	0.0	0.0
	2014	0.2	1.4	2.1	0.9	0.4	6.4	8.0	4.2	4.2	0.3	4.2	0.0
Average	2015	0.0	0.8	0.0	0.9	0.4	6.4	6.9	0.2	4.2	0.0	0.6	0.0
Rainfall	2016	0.0	0.8	2.2	0.7	1.6	11.4	5.6	0.6	0.0	0.0	0.0	2.2
	2010									23.7	23.9	22.3	22.0
	2011	22.1	24.9	23.6	26.6	26.0	26.7	23.9	22.3	23.9	23.9	22.7	22.1
	2012	23.5	24.6	25.2	26.0	27.0	26.9	22.4	22.7	24.3	23.7	23.4	23.2
	2013	24.3	25.0	26.3	26.9	23.6	27.0	24.1	22.3	24.8	24.1	23.2	21.6
Average	2014	23.3	24.7	25.4	25.7	26.5	27.3	24.3	22.9	23.6	23.7	22.8	23.7
Temp(Max)	2015	23.6	23.6	23.7	23.6	24.9	27.2	24.3	24.1	25.4	23.6	23.9	22.6
	2016	23.5	25.0	28.1	26.0	26.9	27.3	23.3	23.6	25.4	24.8	23.0	22.2
	2010									11.8	10.6	9.9	8.7
	2011	9.3	9.1	11.2	12.9	13.3	13.2	13.2	13.1	10.8	9.7	10.7	8.8
	2012	8.9	9.5	11.4	12.8	13.5	13.1	13.4	13.0	11.1	9.6	10.5	9.2
	2013	9.2	10.6	12.2	12.9	11.4	13.3	10.4	12.9	11.3	10.4	10.3	7.8
	2014	9.6	9.9	12.3	12.8	12.7	12.5	12.9	12.9	11.5	11.5	10.3	11.4
Average	2015	8.5	10.3	12.8	12.9	13.6	12.5	12.9	13.0	12.3	11.5	10.9	10.9
Temp(Min)	2016	9.9	10.9	13.6	13.9	13.4	13.5	13.4	11.4	12.0	11.6	9.4	9.1
	2010									9.2	9.6	9.1	8.9
	2011	10.6	7.8	9.6	8.2	6.5	5.4	4.4	6.3	9.8	7.4	10.4	10.1
	2012	10.5	9.2	8.8	10.0	6.7	4.3	5.0	6.4	10.1	9.6	10.1	7.5
	2013	7.4	9.0	9.4	7.4	6.2	4.7	4.1	6.7	9.0	9.6	10.1	7.5
	2014	9.8	9.6	9.1	9.1	6.5	4.7	4.4	7.4	7.4	9.1	7.4	10.1
Sunshine	2015	9.9	9.0	7.4	7.4	6.6	4.8	7.4	7.4	7.4	7.4	7.4	7.4
hour	2016	7.4	7.4	7.4	7.4	7.5	7.4	7.4	7.4	7.4	7.4	7.4	7.4
Relative	2010									55.4	56.3	58.1	63.8
humidity	2011	63.8	54.6	57.7	61.5	73.1	53.1	51.9	55.9	48.7	81.8	49.3	67.1
average	2012	50.9	49.4	58.9	48.6	82.3	63.0	55.9	53.6	53.3	81.1	50.0	56.1

2013	53.7	47.0	59.0	59.4	59.6	51.1	80.7	61.7	60.7	53.1	63.9	56.6
2014	63.9	56.6	54.4	50.6	56.1	50.5	73.9	65.9	80.0	65.9	46.5	58.5
2015	58.5	46.4	46.4	46.8	53.0	52.1	64.9	77.3	61.9	59.4	62.0	73.6
2016	73.7	56.0	56.5	61.8	50.3	53.0	80.6	59.4	62.8	53.5	54.6	58.7

Table 17 Total monthly billed water consumption in m^3/month

Year	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sept	Oct	Nov	Dec
2010									547281	284663	254346	356988
2011	303623	235013	250166	232695	274911	216963	222808	197378	475929	320851	509428	518282
2012	365370	422606	335736	397787	676320	300521	258603	277402	321377	323723	299619	326490
2013	377705	363085	432314	308708	339491	437005	401325	416329	384968	407580	404067	392712
2014	465903	453973	432321	505743	406796	384302	554350	308831	479168	514585	498674	435395
2015	349680	515852	557841	426231	258514	383540	1395909	343455	363343	208511	0500141	10377422
2016	406024	392241	566015	387153	419683	393651	271908	285854	433019	405039	377396	415097

Table 18 Water Supply Tariff for Mekelle City birr/m^3

Range of Consumption in m^3	Residents	Commercial	Governmental and NGO's	Public Taps and Hydrants	Water Truck
0-5	4	6	6	4	15
5-10	5	8	8		
10-20	8	10	10		
20-50	10	12	12		
50-100	12	15	15		
>100	15	15	15		

Source: Mekelle City Water Supply Office