Arab American University

Faculty of Graduate Studies

Prediction of Water Demand in North of Palestine Using Artificial Neural Networks

By

Mohammad Tayseer Mahmoud Zaid Alkelani

Supervisor

Prof. Dr. Mohammed Awad

This thesis was submitted in partial fulfillment of the requirements for the Master`s degree in Computer Science

November / 2018

Prediction of Water Demand in North of Palestine Using Artificial Neural Networks

By

Mohammad Tayseer Mahmoud Zaid Alkelani

This thesis was defended successfully on 24/11/2018 and approved by:

Committee Members	Signature
1. Supervisor: Prof. Dr. Mohammed Awad	…................................
2. Internal Examiner: Dr. Mohammed Maree	....
3. External Examiner: Dr. Emad Al-Natsheh	…..……………………

Declaration

I declare that this thesis entitled " Prediction of Water Demand in North of Palestine Using Artificial Neural Networks " is my own work and has been composed solely by myself and does not contain any work from others researcher and has not been submitted for any other degree or scientific qualification except the references is made.

.

Dedication

Proudly, I dedicate my thesis to my parents, as I always feel their prayers in all aspects of my life, I also dedication my thesis to my brothers, sisters, friends, colleagues who are always willing to provide any support.

Finally, a special dedication to my wife and my sons and my daughters who supported me with all they could.

Acknowledgments

First and foremost, I am grateful to my lord, who has given me the strength, enablement, wisdom, knowledge, and understanding required to complete this thesis.

Next, I wish to express my gratitude to my supervisor, Prof. Dr. Mohammed Awad, for his deep help and support and his ideas have made this thesis worth research and effort.

I would like to thank the Arab American University, Faculty of Engineering and Information Technology for providing me a high standard of Education through the duration of a master program in computer science.

I also extend my sincere thanks and appreciation to all the staff of the faculty of engineering and information technology master program.

I appreciate the support rendered by the staff of Jenin municipality, especially the water section. They made the data used in this project available.

I would like to thank all those who have assisted, guided and supported me in my studies leading to this thesis.

Finally, I would like to extend my deepest gratitude to my parents and my wife; they have always given me unremitting support during the preparation of my thesis.

Abstract

Water is very significant to humanity, in all our life fields, such as agriculture, industry, and health. In our daily life, water consumption has been increased in a significant manner. The use of domestic water is considered the most fundamental factor in the chain of water demand and consumption in urban areas. Several models and more computational techniques have been proposed and presented in the last decades to predict the water demand in short or long-term time series, such as linear statistical and mathematical methods for predicting the future quantities of daily, monthly and yearly demand. Other more advanced methods like the regression of time series analysis approaches have been applied. The majority of these methods depends on extrapolating historical trends and linked the demand with socioeconomic variables. It is necessary to evaluate the ability of existing resources to meet future needs and to provide the basis for planning for future systems and improve them to limit the uncertainties in demand predictions.

These uncertainties include the population growth, economic change, changes in consumption habits, and climate change especially in our region, hence the prediction of water demand helps water distribution companies and government knowing and to investigate the expected demand and the impacts on the sustainable development planning. According to a set of challenges that face the water authorities, municipalities and the water sector, in general, the most important limitation is lack tools that can predict the water demand in the future.

The Jenin city municipality does not have computer applications to assist the future needs estimation and the satisfactory distribution of water.

In this thesis, we use linear and nonlinear methods to predict water demand in the city of Jenin, for this purpose, we use different types of Artificial Neural Networks (ANNs) with different learning methods to predict the water demand, and compare our results with a known types of statistical methods. In this context a computational technique which depends on artificial neural networks (ANNs) and a hybrid method of NNs with optimization algorithm is used to predict the short-termwater demand in urban regions. The dataset depends on sets of collected data (extracted from Municipalities Databases) during a specific period of time and hence, we propose a nonlinear model for predictingthe monthly and yearly water demand and finally provide the more accurate prediction model compared with other linear and nonlinear methods. We aim to create a model capable of making an accurate prediction for water demand in the future for the Jenin city. This prediction is made with a time horizon (months or years) depending on the extracted data. This data will be used to feed the neural network model to implement mechanisms and a system that can be employed to predict a short-term for water demand, unlike the current situation where the water authority in Jenin organize the water where supply one day on a week for each region in the city, hence our idea is to organize the water distribution for each region in Jenin city depending on the past consumption. This method is based on the use of K-means clustering algorithm to classify the regions, so the authorities can supply a water for each region with a number of days to ensure a fair distribution.Two applied models of artificial neural networks are used; Multilayer Perceptron (MLP) and Radial Basis Function Neural Networks (RBFNNs) with different learning and optimization algorithms, and one type of linear statistical method called Autoregressive Integrated Moving Average (ARIMA) are applied on the water demand data collected from Jenin city to predict the water demand in the future. The obtained results demonstrate that the MLPNN type is surpassed the RBFNN and ARIMA models in the prediction the water demand values according to lowest error and goodness fit. Furthermore, the experimental result of clustering the regions in the city provides an efficient approach to supply water fairly.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANN	Artificial Neural Network.
NNs	Neural Networks.
KMC	K Means Clustering.
MP	Multilayer Perceptron.
RBF	Radial Basis Function.
ARIMA	Autoregressive Integrated Moving Average.
MLPNN	Multilayer Perceptron Neural Network.
RBFNN	Radial Basis Function Neural Network.
MLPFFNNBP	Multi-Layer Perceptron Feed Forward Neural Network Back Propagation.
AR	Autoregressive.
MA	Moving Average.
ARMA	Autoregressive Moving Average.
GAs	Genetic Algorithms.
MSE	Mean Square Error.
PNA	Palestinian National Authority.
PLO	Palestine Liberation Organization.
PACF	Partial Autocorrelation Function.
ACF	Autocorrelation Function.
SARIMA	Seasonal Autoregressive Integrated Moving Average.
BP	backpropagation.
SSE	Sum of Square Error.
SVD	Singular Value Decomposition.
LA	Learning Algorithm.
LMA	Levenberg–Marquardt Algorithm.
CSV	Comma Separated Values.
AIC	Akaike Information Criterion.
BIC	Bayesian Information Criterion.
KNN	K-Nearest Neighbor.

CHAPTER 1

INTRODUCTION

1 Introduction

The majority of the countries in the world are suffering from problems due to the increasing demand for water in light of the scarcity of resources to obtain sufficient quantities and satisfy the needs of citizens of different needs in different fields [1]. The water used in houses is generally one of the most important parameters in the process of urban water consumption, and the planning of water distribution systems depends on satisfying the consumer demand needs by providing the consumer with adequate volumes of water continually. The absence of regulations or even techniques in terms of organizing the water sector is considered the main problem, also which make it a worse issue is lack of clear and precise strategic planning due to the absence of modern and effective prediction techniques of water demand or even the medium computerized plans in term of smart applications.

The possibility of existing programs and applications that can predict the water demand effectively and reliably in light of the short, medium and long-term is a useful element in the strategic planning and the processes of water scheduling and maintenance; add to that, the distribution and regulating the water demand in a fair and optimal manner [2]. Therefore, the need is required and worth the examining accurately by new and suitable methodologies or techniques. Prediction strategies of water demand are very important to support and help the water authorities and municipalities in identifying the future needs and to develop the necessary plans to find real solutions in the short, medium and long-term plans, so it is considered the essential piece in decision making whether to organizations or communities. Higher efforts from researchers have been made over the past and in recent years to introduce the best and most methods to be a strong alternative and to integrate the numerical paradigms that they employ to make a precise prediction.

The water circumstance in northern Palestine, such as the city of Jenin, is similar to the rest of Palestine cities. But in Jenin city, there is a number of difficulties back to the amount of water leaking in the ground due to weak of networks and interruptions in supply and consumption during periods. The water department in the municipality of Jenin has no programs or applications for estimating future demand for water and calculating and estimating the expected demand is currently dependent on simple statistical calculation methods. Even in the rest of governorates, there are no advanced methods, hence, simple statistical calculations are not enough, neither effective nor reliable to provide predictions and estimates that are appropriate for the nature and characteristics of the different regions.

The collected datasets are available in database records as the consumption of consumers which we obtained and we extracted from Jenin municipality. It is composed of seven years from 2011 until 2017 for 7500 subscribers. We aim to exploit this data and prediction the monthly water demand according to the short-term scenario based on the actual consumption that has been aggregated throughout 84 months. Another type of data was collected which depends on the water consumption in each region (Jenin city divided into 39 regions), we aim to cluster the similar regions to organize the water distribution with the suitable number of days depending on the consumption.

Different methodologies and methods have been used in the field of water demand prediction in urban areas. These methods depend on statistical extrapolation or advanced analytical models, and the choice of appropriate methodology depends on the purpose of prediction for water service companies and also depends on the quality and quantity of data [3]. Although the various models are simple and effective in some linear issues, they fail and show weakness in matters that do not rely only on correlation in the time series such as nonlinear issues [4]. Artificial intelligence technique has begun to emerge strongly, especially artificial neural networks (ANNs) and has been proposing as an effective and powerful tool in prediction and modeling [5].

A majority of researches indicate that the water demand prediction with artificial neural networks (ANNs) is considered an efficient and strong alternative technique in contrary to traditional methods [3]. We aim to make performance evaluation of the different methodologies and adopt the most suitable method for water demand prediction in an urban area (case study Jenin city).

The applied models (MLPNN, RBFNN, and ARIMA) will produce prediction results for the next year.

1.1 Problem Statement

In light of the increase in demand for fresh water for human use, lack of supply at the local level and the scarcity of resources, In addition to the problems related to the quality of the networks, quantities of leakage and in the absence of modern methodologies or even traditional methods able to estimate the future demand of water, with the increase in water consumption, Jenin city suffers from bad distribution process to supply water.

The consumption in 2017 is estimated at about (1160927) cubic meters. This puts us in a major challenge that requires the development of future programs and strategic planning in the medium and long-term goals in order to meet future needs and determine the general trend in the demand series. Where we can exploit the modern scientific computing methods in terms of exploiting the opportunities and improve the sector of water supply. Thus, reflected positively on improving the lives of people and arrive at degrees of greater satisfaction with fair and adequate distribution, Hence, the idea of the study came to contribute to providing scientific solutions to the problem based on the use of advanced and reliable computer models.

1.2 Thesis Objectives

Water demand prediction based on reliable modeling and predictability of future has received considerable attention from researchers at the global and regional levels. Because of the vital role, it plays in the future strategic planning process. In Palestine, the applied methods depend on simple statistical or numerical ways. The primary objective of this thesis is to carry out a systematic investigation to discover and examine the facts so as to establish a more applicable artificial neural networks model or other traditional methods that can be used to predict water demand in short-term period in order to help achieving water resources sustainability in light of the expected increase in the water demands and to improve the distribution of water rations supplied to the regions in a fair manner in Palestine and particularly in the northern regions (Jenin city).

Our proposed system requires the previous time series data of water consumption, hence:

- We aim to prediction the future water demand depends on extracting the patterns of the historical demand data, the data set was collected from Jenin city in the north of Palestine as a case study.
- We aim to applying a general statistical model (ARIMA) to predict water demand value in the short term period (next year).
- We aim to employing and applying more efficient artificial neural networks models with different learning algorithms to predict the water demand in a short-term period.

- We aim to comparing different models depending on their architectures and the prediction result (smallest mean square error) to determine the best model for predicting.
- We aim to applying a smart clustering algorithm (k-means clustering) to group each region in the city with the aim of producing a fair water distribution by controlling the open valves during the time.

1.3 Contribution

The following points summarize the contributions of this work:

- Proposed Artificial Neural Network, to predict the water demand for water authorities and municipalities in north Palestine (Jenin city as a case study).
- Utilized and exploited previously unused sets of raw data for training the ANNs and developing new methods to serving optimal prediction and distribution processes in the water consumption sector and thus to avoid traditional and primitive methods that are not trusted.
- Compared the prediction result produced by the general statistical model, with the proposed neural networks model using different learning algorithms.
- Used a clustering algorithm to group the similar regions in consumption, with the aim of making a fair distribution of water related to the number of water supply days.

1.4 Thesis Outline

This thesis is structured of as follow: Chapter 2 presents the background of the investigated topic, so we introduce the components of water consumption and some factors that affect the water consumption process, and the data used, then we present the time series prediction concept. Then we give a general introduction of stochastic models Autoregressive Integrated Moving Average (ARIMA) that used and carried out in our practical experiences, Artificial Neural Network (ANN) focus on the applied NNs models MLPNN, RBFNN, and the applied clustering method. Chapter 3 covers a general literature review and previously published methods of water demand prediction that used MLPNNBP-LMA, RBFNN-GAs K-means clustering and autoregressive integrated moving average. Chapter 4 discusses and describe the applied models and the methodology of study including the preprocess operations applied on dataset, preprocessing process, data normalization, the detailed operations of performance role MSE, construction the procedural steps in building ANN model, MLPFFNNBP, RBFNN with Genetic Algorithms (GAs), K-means, ARIMA model that used for time series predicting. Results and discussion for the yearly water demand prediction are so demonstrated an in chapter 5 the result of regions clustering to organize the distribution of water in each region. Chapter 6 presents the main conclusions and recommendations and the future works of study.

CHAPTER 2

BACKGROUND

2 Background

Throughout the ages, freshwater and its availability in quantity and quality has been a major concern for all human civilizations for its association and its reflection on the health of society. Demand, supply, service and community health are therefore based on the infrastructure of supply, distribution systems, and future strategic plans, that have the capacity to meet the needs and sustain the success of all elements of Human, industrial and agricultural development [6]. Thus the water sector is an important sector of sustainable development at the national level.

The deliberate neglect of the development of the water sector in Palestine by the Israeli occupation over decades, has contributed, directly, to reduce the chances of real development, despite the launching of the Oslo peace process and the creation of the Palestinian National Authority (PNA) in 1994, the development of the water sector remained constrained by the obstacles imposed by the occupation authorities even within the agreements signed between the Palestine Liberation Organization (PLO) and Israel. The high demand for water and the significant gap between demand and supply in the water sector is one of the major challenges facing the sector over the next few years. The water demand is increasing because of natural population growth and national development requirements, Water and its use due to Israeli obstacles, add to this the large difference and the gap between the per capita Palestinian water compared with the Israeli as shown in Figure 2.1, also the complex licensing requirements from the Israeli side, this is a great challenge, and it is necessary to find creative solutions to supply the necessary quantities of water to different sectors and achieve balance for supply optimal water in Palestine.

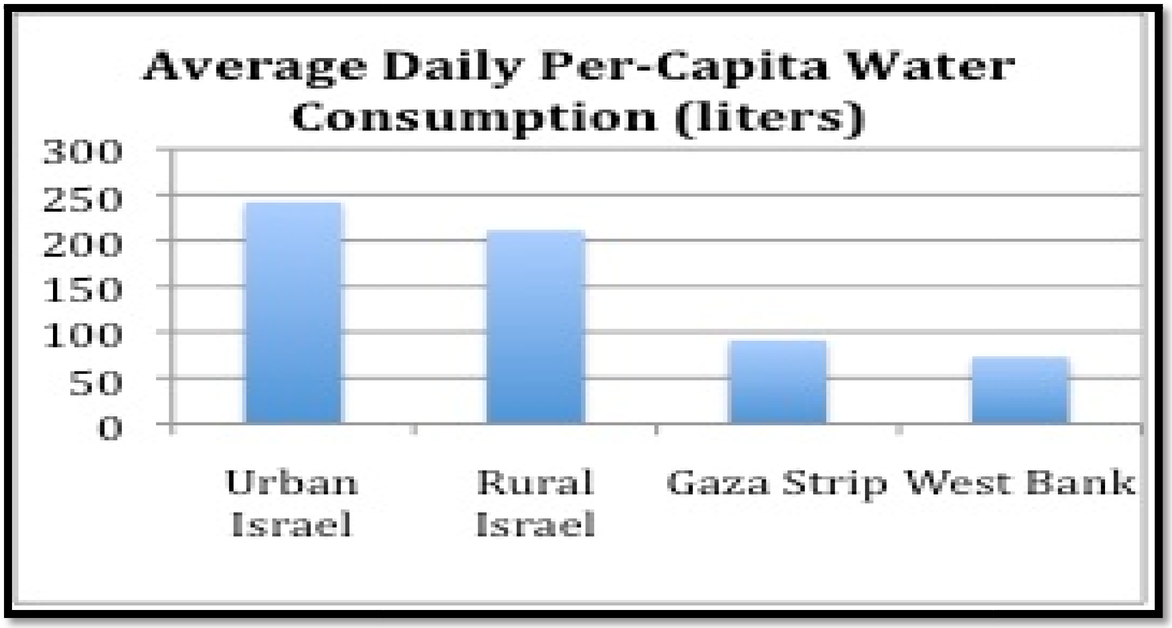

Figure 2-1: Daily per capita share of Palestinian water compared with the Israeli [7]

2.1 Elements of Water Consumption

Water consumption can be classified into different components, so each component contributes significantly to total consumption, home water use (domestic) is water employed for inside and outside household objective, connected with all domestic activity forms. Not only is the population increasing but also the level of consumption doubles for reasons of increasing and the use of modern devices such as dishwashers, automatic sprinklers, has 83 liters/day average per capita consumption of water. This rate ranges between 82.3 liters/day in the West Bank and 84 liters/day in the Gaza Strip during 2016[9].

Taking into consideration that this rate is due to the quantities of water consumed divided by the population as there are some residential communities do not exceed the average per capita consumption of 44 liters per day, while this rate exceeds 100 liters per day in other communities such as Jericho and thus the achieving fairness in the distribution among residential communities is one of the main challenges faced by the State of Palestine [10]. Graphically as shown in figure 2.2, Industrial and commercial consumption, about 20 percent of freshwater and globally used water goes for industrial use, while 10 percent goes for domestic or municipal use and 70 percent for agricultural consumption [11]. The percentage of cultivated land in Palestine is about 15.5% of the total area, of which about 84.5% is Non - irrigated agriculture and of which 15.5% is irrigated, and this irrigated land is consuming about 47% of the water which pumped only from underground wells [12]. Add to that the amounts of water lost and leaking from networks due to the stale of lines and aging over time.

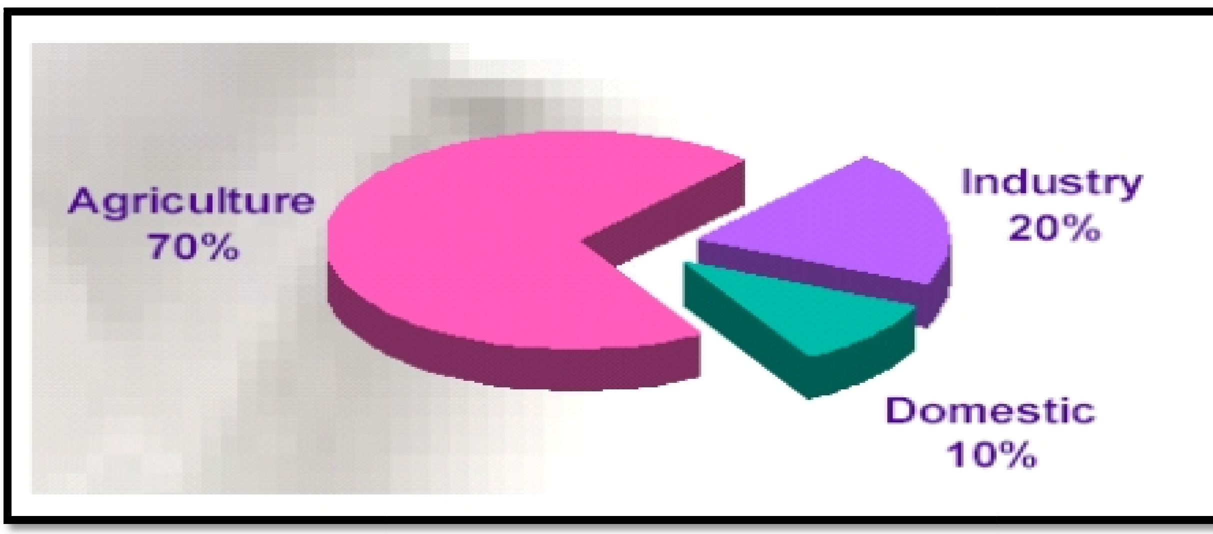

Figure 2-2: Global water consumption by sectors [12]

2.2 Circumstances Influencing the Level of Water Consumption

In the previous section, we reviewed the components of water consumption, knowing that there are some factors affect the level of water consumption, and for each component, there are many implicit factors that have a significant impact. Consequently, patterns and behavior of consumption necessarily depend on the characteristics of the demographical region (Population characteristics of size, distribution, density, growth components, economic and social conditions, income level, etc), and geographic (rural, urban, agricultural, industrial), Figure 2.3 shows the percentage of globally water demand in the domestic, industrial and agricultural sectors [13, 14, 15].

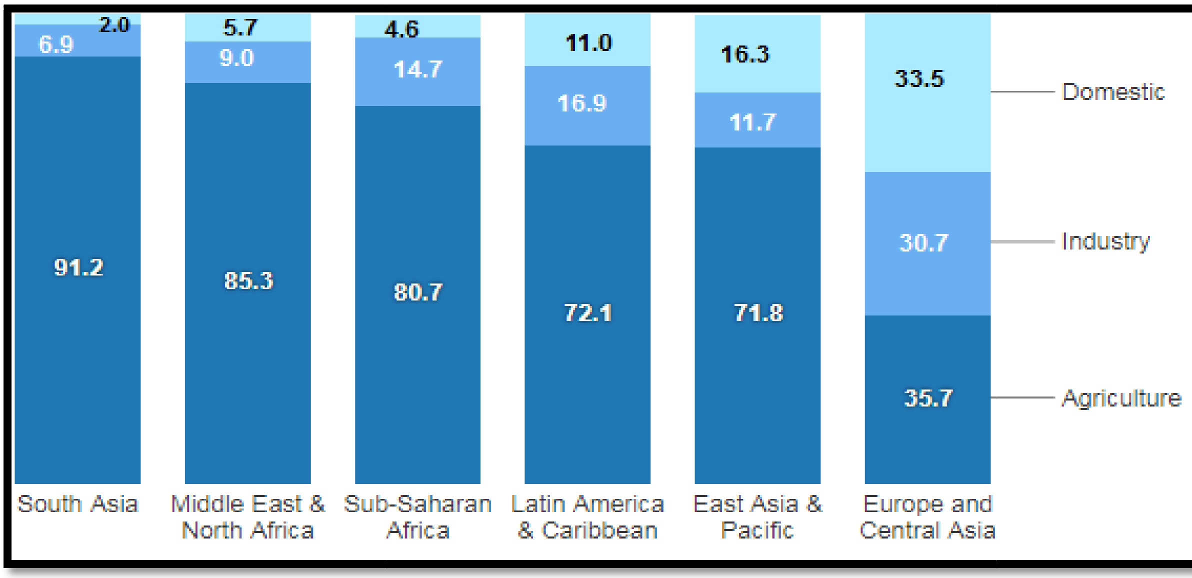

Figure 2-3: A world-wide scheme for water demand [17]

Finally, the losses from water supply networks also have an important aspect and a significant impact on the quantities consumed as a component of water consumption, the losses of water from the water network because of the effects of leakage depends on three basic factors. Firstly, the condition of the pipes covered in the earth and their aging and it depends on the nature and quality of the manufacture of these pipes and so the quality and chemical properties of the water that passes through them. Secondly, the volume of pressure treated in the pipes also has an effect on the aging of the networks and the size of the leakage and losses. Therefore, if the pressure is large, the size of the loss and leakage will be very high in cases of lines explosion. Thirdly, the efficiency and effectiveness of the technical staff, which monitors the leakage of lines and constantly work on maintenance operations, so this illustrate the reasons of developing the efficient leak detection tools and this is strategy has become a fundamental research issue in recent years[16].

2.3 Datasets Description

If we look at recent years, we will find different methodologies and methods that have been used in the field of water demand prediction in urban areas and have been proposed. Some are as simple as simple statistical extrapolation, some are advanced analytical models, and the choice of appropriate methodology depends on the purpose of prediction for water service companies and also depends on the quality and quantity of data [3].

The water circumstance in the northern side of Palestine, such as the city of Jenin, is similar to the rest of Palestine. But in Jenin city, there are more difficulties depending on the amount of water leaking in the ground due to weak of networks and interruptions in supply and consumption during the time, in addition to other factors, such as the scarcity of water resources and they're linked with the Israeli side which controls the valves. The water department in the municipality of Jenin has no programs or applications in terms of estimating the future demand for water and at the most calculating and estimating the quantity foreseeable in simple statistical calculation methods, even in the other governorates also there are no modern methods, thus these simple statistical calculations are not enough and not effective and not reliable to provide predicts and estimates that are appropriate for the nature and characteristics of the different regions.

Studies related to water conditions for supply and demand issues in the West Bank for the years 2000 to 2020 according to the applied research institute Jerusalem indicate that the city of Jenin will require large amounts of water to meet the expected progress in the field of agriculture and industry [19]. Table 2-1show the results of this study.

Table 2-1: Water supply and demand for the Jenin city [19]

Year	Population	Domestic (Mcm)	Irrigation Area(Dunum)	Agricultural (Mcm)	Industrial (Mcm)	Total Water Use (Mcm)
1990	180400	5.20	13100	9.0	0.2	11.70
2000	311340	17.12	108700	74.7	0.6	92.42
2010	426610	31.47	158600	101.4	1.2	134.07
2020	546100	42.60	208500	128.1	1.9	172.60

The dataset used in this thesis consists of recorded water consumption parameters in the Water Department which follow the municipality of Jenin city. This Department measures quantities of water consumed for all subscribers in all neighborhoods of the city. The management of department distributes a group of collectors is reading the active meters and recording readings and quantities of consumption at the end of each month and transferred it to the database.

The dataset was extracted and obtained from Jenin city municipality records in Palestine for the last seven years according to the period between 2011 and 2017, So our data, in the beginning, was composed of the actual monthly consumption for every customer for every year for 7500 customers, thus every customer has 12 reading per year, we aggregated the total consumption for instance in one month for 7500 customers so as to obtain just one value that represents the total consumption for all customers in one month in year, hence we obtained the 84 values represent 84 months and thus represent 7 years, this data Will be used and exploit on all models intended to be built and implemented.

The water supply networks serving the city of Jenin have been established from a very long time. The amount of water leaking from the network was estimated at about 40 percent and may reach about 60 percent before making some improvements in recent years [19]. Figure 2.4 shows the water supply network in the city of Jenin.

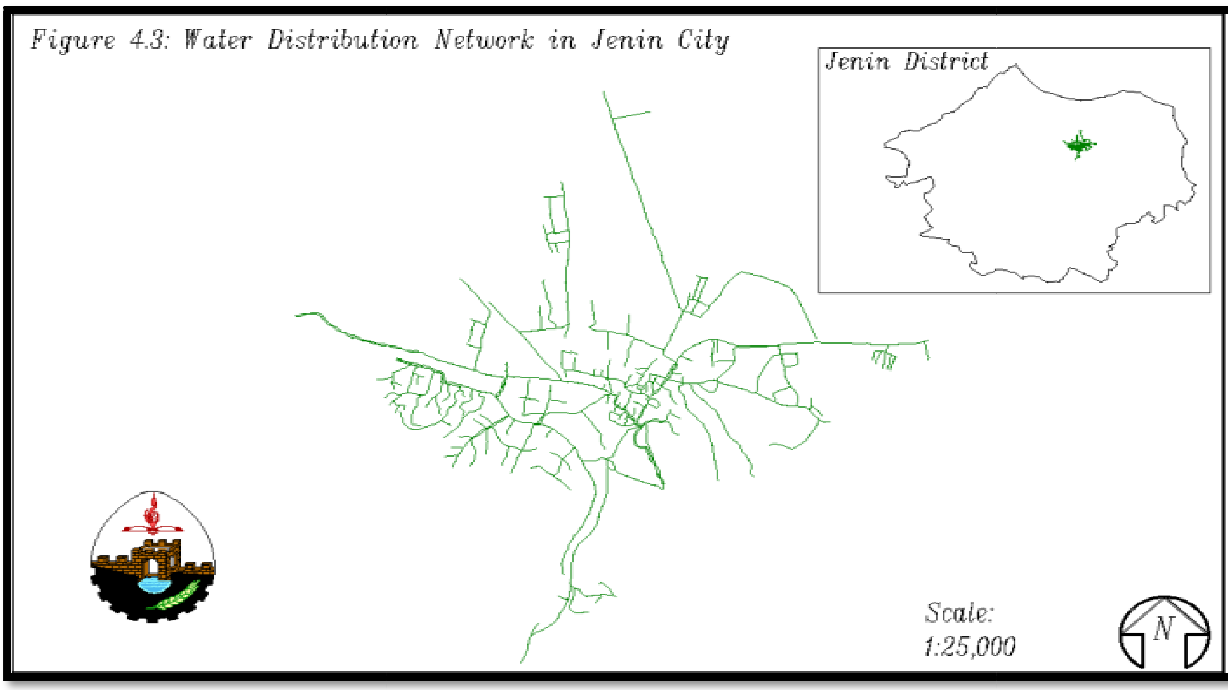

Figure 2-4: Water distribution network in Jenin city [19]

2.4 Time Series

Time series models depend on the hypothesis that any time series have a historical particular recurring statistical that can be manipulated for prediction purposes [20]. The unknown time series function F (X_t) to build a model, that allows obtaining accurate predicts [21]. Time series is a combination that represents the data (X_t) listed and registered over the duration of time for instance, daily, weekly, monthly, yearly, analyzed it to comprehend the past so as to predict the future [22]. Over the years, a wide research effort has been undertaken by researchers to develop effective and powerful models with the aim of improving the possibilities of accurate prediction results. Many models have been developed that predict the time series within the literature in this regard. Stochastic time series models it was used very widely such as ARIMA [18].

Recently, ANNs have increasing attention in the domain of time series prediction, the basic hypothesis to implement this model is that the time series is nonlinear and associated with another parameter known as biologically inspired [23], so it mimics the work of human brain and employing a nonlinear calculation. The more data and notes are closer to each other over time, the more correlated to each other and stronger commonly the chain is taken over separate time $t = 1, 2...n$, $t > 0$.

In general, the time series can be decomposed into four components [24].

First - the general trend, that appear the Smooth Long-Term Direction (increasing or decreasing).

Second - seasonal variation, it is a pattern of change that repeats every year or the period of time (week, month, quarter), so it occurs periodically in a year.

Third - cyclical variation, rise or fall of the time series occurs in a period of time exceeding (one or two, three… years).

Forth - irregular variation, something random does not follow a particular pattern. There are several types of models that are used in time series and usually two types are used mostly, the multiplicative and additive models. The multiplicative model depends on the hypothesis that the 4 parameters of a time series are not independent and thus affect each other; the other model has based on the hypothesis that the 4 parameters are inevitably independent of one another [24].

2.5 Time Series Prediction Using Stochastic Models

A widespread statistical method vastly used to foresee the time series is (ARIMA) model. ARIMA is a term (concept) and expression that stands for Auto Regressive Integrated Moving Average; it is a manner and paradigm that captures a set of different temporal component in time series data, ARIMA is a prediction method that visualizes the future values of a certain series, others call it “Box-Jenkins” [24].

ARIMA is commonly better and more efficient than the exponential smoothing method given that the length of data is moderately and the observations of time series are stationary or stable and the correlation between these observations must exist [24]. In briefly, ARIMA model consists of three component or parameters, non-seasonality *(p, d, and q):*

p - Represent the lags of stationeries series (AR), the order of non-seasonal (AR), we determine it through partial autocorrelation function (PACF).

d - Differencing between the observations (I)

q - Represent the lags of the predict errors (MA), the order of non-seasonal (MA), we determine it through autocorrelation function (ACF).

SARIMA is another type and it similar ARIMA but the first take into account the seasonality issues, thus we can demonstrate the model as ARIMA add to that the part of seasonality *(P, D, Q),* so the adopted form is by employing the following expression:

(p, d, q) X (P, D, Q).

Many applications provide us with the possibility of building ARIMA models such as MINITAB, MATLAB, IBM-SPSS-STATICS, and the most professional is the R-studio, which provides us with ready tools to shorten the long list of procedures, calculations, and representations and suggest to us the best usability. Actually, ARIMA suffering from some limitation, such that the ARIMA method is dealing just with a linearity shape, and it suitable with only for a time series that is stationary, this means that it is variance and mean must be a constant through the time [18], add to that they need a large number of observations to be an efficient manner, So the accuracy of forecasting according to this manner is mostly a variable especially if they represent a linear manner for a nonlinear problems.

2.6 Time Series Prediction Using Artificial Neural Networks

The Neural Network in term of biological status is imitating the human brain. The brain, in general, depends on a small processor of nerve cells, which are called neurons or nodes. The ANNs consisted of a number of simple processors, called neurons, which are approximately similar to the biological neurons in the brain. ANNs have been characterized by many reasons and that makes them very suitable for particular problems, so it has the ability to learn, generalize, does not force any limitation on the input for variables, especially in the complexity observations in predicting, unlike traditional models, it ready to be the powerful alternative technique.

ANNs are basically developed to mimic the human brain, a non-linear calculation executes through neuron according to the input values and perform result values to transfer and fed others neurons, in more clear shape every neuron obtains the input signal or total information from another neuron, the processing is done through activation functions to generate a converted output signal to other nodes and exterior outputs. It should be noted that although each neuron performs its function somewhat slowly but the network in a collective and participatory way progress the huge number of calculations and processes in a manner of high speed and high efficiency [25].

The advantages according to a wide range of parallel distributed processor makes ANN's strong arithmetic tools and enormous in computation, the greatest achievement is the enormous ability to learn according to data situations and initial observations and then produce and generate new cases and examples that haven't seen before, NNs learns depends on the input data and the related output data, hence this is known as the generalization ability of the ANNs [26].

Several and different ANNs models have been proposed especially the multi-layer Perceptron’s MLP, Kohonen's self-organizing networks, and Hopfield-networks, Also there are many types of ANNs, for instance, radial-basis functions, ridge-polynomial, and wavelet networks that have performed very well in the application that is related to approximation and the time series prediction problem is the best example [26].

In general, the past observations of the data series represent the input, as for the future value represents the output and the ANNs calculate this as this function:

$$Xt + 1 = F(Xt, Xt - 1, Xt - 2 \dots, Xt - p) \qquad (2.1)$$

Where X is the observation at time t, figure 2.5 show some explanation, so the ANNs synonymous with the nonlinear autoregressive model that associated time series predicting.

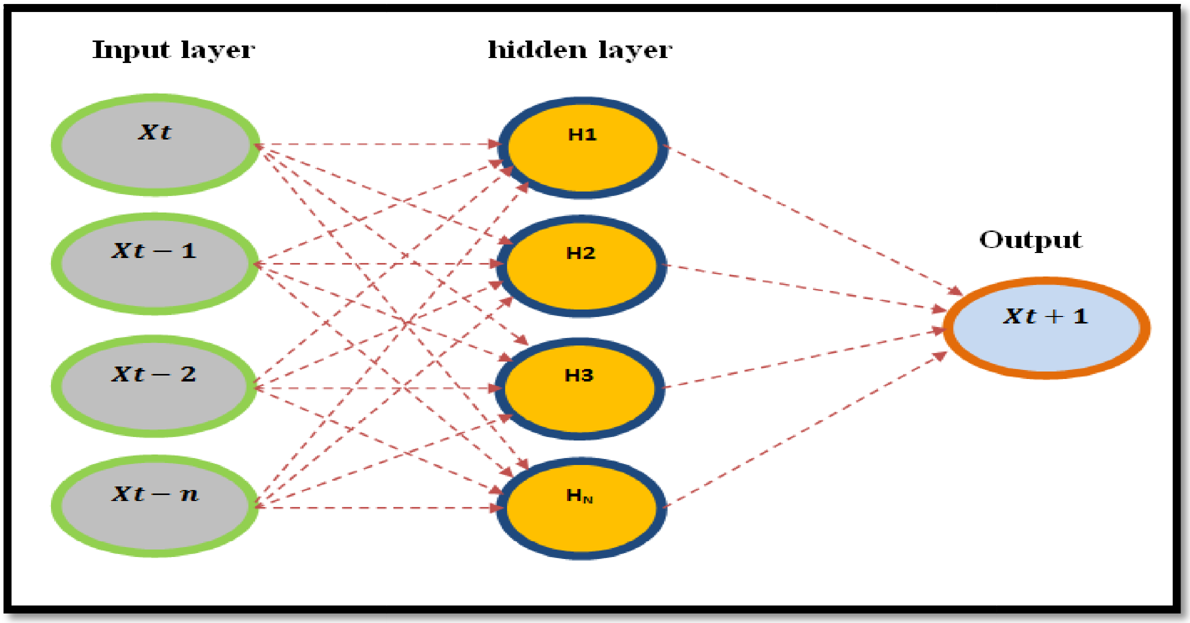

Figure 2-5: General time series prediction using NNs

Neural networks in terms of topology can fall under two categories where there are networks with a single layer and there are networks that can contain multiple layers. It is important to note the important aspect of the flow of data (the feedforward or feedback neural networks), but the central part of the whole process is the subject of learning algorithms, which plays a pivotal role in intracellular communication between neurons to determine and adjust weights. The learning process has two kinds of learning: supervised and unsupervised, we explain and clarify them in different contexts in other sections. The general structure of neural networks can be illustrated in figure 2.6.

Every artificial neural network consists of several compounds: a set of *X* as an input vector that represents the values that are pass-through nodes to the hidden layer. So every node from the input layer to the hidden layer takes a random weight (*W*) value, thus it is an actual value that concise together with an input value to the neuron. At all neuron within the hidden layer, we employing the transfer function that grants propagating of output to the rest of the neurons [27], it is worth mentioning here that the network has the ability to learn, by adjusting its interconnection weights.

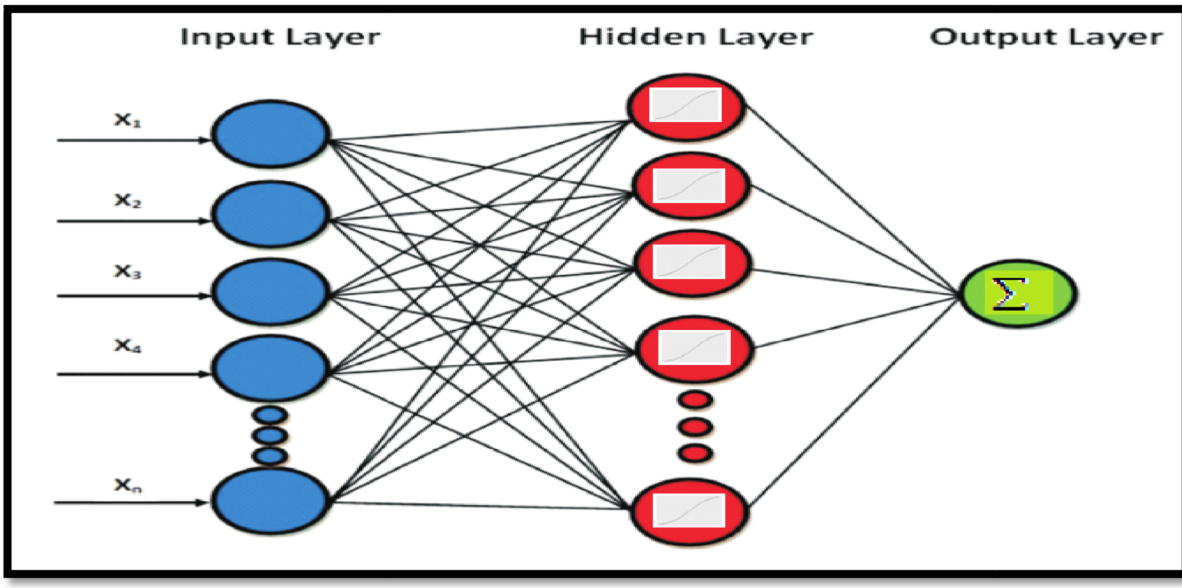

Figure 2-6: Basic Artificial Neural Network

The importance of the hidden layer is that it receives the input through the neurons and analyzes it with the aim of making the appropriate decision about what to do according to the acquired experience and learning [28], Then is passes it to the rest of the neuron cells and nodes through communication after the recruitment and implementation of certain types of activation functions such as the (sigmoid) [29]. This means that neural networks have high capabilities in adaptive learning and training processes. So the network possesses the characteristics of self-organization in order to achieve the desired goals. The output values are calculated and produced and passed to the output neural network layer at the end. Hence we can express the output by employing the subsequent equations.

$$Y_i = F\left(\sum_{j=1}^{n} W_{ij} X_j\right) \qquad (2.2)$$

Where (Y_i) is the network output, (W_{ij}) is the weights, (X_j) is the value of a set of inputs and (n) is the total number of inputs. The most important unit in the neural network is the activation function or transfer function", output a result value called the "unit's activation" [30]. Many activation functions are used in NNs training; the most popular types are the identity function, binary step function, and Uni-Polar Sigmoid Function.

- ***Identity functions:***

  This function is denoted by *f(x) = x* for all *x*, Single layer neural networks make use of a step function while converting a continuously varying input function to a binary output (0 or 1) or a bipolar output (1 or -1).

- ***Binary step function:***

This function makes use of a threshold, so the binary step function with a threshold T is given by:

$$F(x) = \begin{cases} 1 & if x \geq \text{T} \\ 0 & if x < \text{T} \end{cases} \qquad (2.3)$$

- **Uni-Polar Sigmoid Function**:

This is the most popular activation function which specially used in neural networks training by back-propagation algorithms.

$$\text{F(x)} = \frac{1}{1+e^{-x}} \qquad (2.4)$$

Where (x) is the input value, (1) represent the Curve maximum value, (e^{-x}) represents the steepness of the curve. So the sigmoid activation function is used to predict the outcomes such as (1/0).

As we mentioned in earlier, there are many ways of tying the nodes of a neural network together. The modest topologies are the feed-forward network where signals flow in one direction only and hence there is no loop in the signal paths.

Typically, in the input layer, a multiplication of input data by the connection weights value that assigned to a particular input. To clarify this if a neuron has two inputs then two weights also that can be adjusted in an individual manner. Through the learning stage, the neural network can adjust the weights based on the error of the last test result. In the next step is a modified input signal summed up to a single value, so the offset (bias) is also added to the sum, the bias adjusts during the learning phase. The main idea is all the neurons have random weights and biases, hence every learning iteration the weights and biases are gradually changed and hence the next result becomes closer to the desired output. So every time the neural network is gradually shifted towards a new situation where the desired patterns are "learned".

Finally, the values are passed them on to the next layer 'hidden' layer where the result produced by the neurons is used as an output signal. This is performed by feeding the result to the activation/transfer function and the output layer offer the result. We can express the outputs of perception according to the following equation [31].

$$y(x_1, x_2, \dots, x_N) = \sum_{j=1}^{J} w_j \sigma_j \left(\sum_{i=1}^{N} w_{ij} x_i + \theta_j \right) \qquad (2.5)$$

Where j is the number of neurons in the hidden layer, (σ_j)is a transfer function, (w_j) is the synaptic weight between the neuron j and the neuron of the last layer and (w_{ij}) is the synaptic weight between the input i and the neuron j of the hidden layer.

2.6.1 Multilayer Perceptron Neural Networks

Often, single layer Perceptron cannot solve some of the complex problems, so we need more efficient and effective methods. MLPNN have been applied successfully to predict or approximate some difficult and complex problems by training them in a supervised manner with the using of a robust algorithm known as the back-propagation algorithm. In our thesis, the Multilayer Perceptron Feed Forward with Backpropagation process consists of three layers are proposed to optimize the connection weights, so it is generally the mostly applied in ANNs as in figure 2.7. Numerical weight is linked with each of the input node connections, the $W_{[i\text{-}h]}$ exemplify the connections of nodes between the input layer and hidden layer whereas $W_{[h\text{-}o]}$ exemplify the connections of nodes between hidden and an output layer.

The input layer is a group of neurons like (X_i) represents the time of data consumption according to months, so it is the input vector fed forward during the hidden layer (H_j) until output layer $F(X_i)$ that represent the result of calculation by employing a nonlinear function called activation function in back-propagation status.

The sigmoid function is employed as the activation function in most cases, so it is represented by the subsequent equation.

$$\mathrm{F(x)} = \frac{1}{1 + e^{-x}} \qquad (2.6)$$

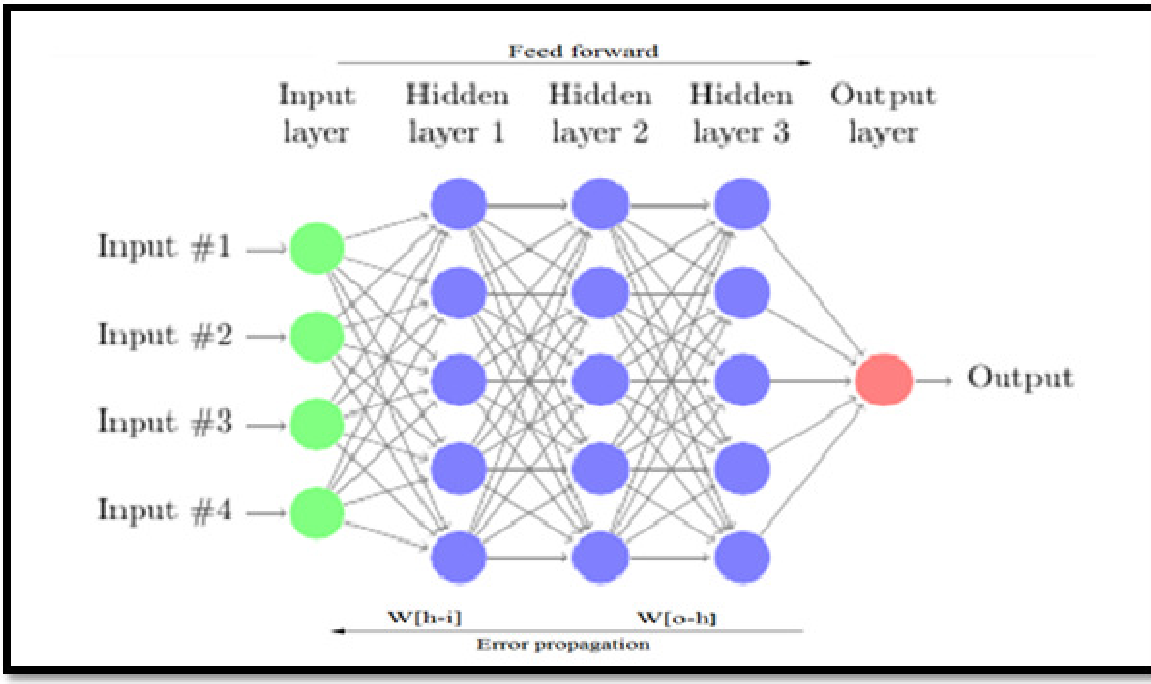

Figure 2-7: Typical three-layer back-propagation neural network

Available data are divided into two parts, a training set, and a test set, the training set is used for assessment and estimate the weights whereas the test set is employed to measuring the generalization efficiency of the network. The training process is like a scheme or diagram between the input data nodes and the output of the network, the data value pushed to the input nodes and the input data nodes are weighted and stacked as accumulate at each node in the hidden layer. The sum is then transformed by the activation function, and finally, the output values are produced, the output of the neural network represented by the subsequent equation.

$$Yi = f\left(\sum_{j=1}^{m} w_{ij} x_j + b_i\right) \qquad (2.7)$$

Where w_{ij} are the weights, x_j is the value of a set of inputs, b_i is the bias, Often enclose as an additional weight, essentially it equals 1. It should be noted here that the extraction of value of the error is a process carried out through back-propagation as we represented through the vector we called it error propagation.

So the network calculated error between the target value and the calculated output value knowing that the weights are updated by employing some training technique and all this process, the network are reiterated until we obtain the satisfy minimize network error, one of these methods is the traditional method Sum of Square Error (SSE) function [32], the error is basically calculated by using the subsequent equation:

$$E_{sse} = \frac{1}{2}\sum_{j=1}^{n}\sum\left(Target_j - Output_j\right)^2 \qquad (2.8)$$

Where Target is the actual value, Output is the predicted value, and (n) is the total number of input data.

We must note that the weights in the first stage are random, according to the error propagation, the weights continue to adjust until the error criteria are satisfied according to Θ, *and* Θ is the threshold value of the prediction process. The weight update in many cases based on the gradient descent algorithm and the following equation represent it:

$$\Delta w_{jk}^{n} = -\mu\frac{dE\left(w_{jk}^{n}\right)}{dw_{jk}} \qquad (2.9)$$

Where (μ) represents the learning rate (take place generally between 0 and 1), (w) is the weight, and (dE) is the derivative, Levenberg Marquardt algorithm for weight updating process was adopted for our thesis.

Typically, in the prediction process, there is a basic rule of stopping training through a prior condition, in our thesis we employing the mean square error value (MSE), the following expression presented it:

$$\text{MSE} = \sum (\text{Actual} - \text{predict})^2 / \text{n} \qquad (2.10)$$

Where (n) is the number of inputs.

2.6.2 Radial Basis Function Neural Network

The network was designed by Darken and Moody (1989) and the network depends on two layers other than the input layer, namely, hidden and output layers. This network converts inputs in a nonlinear way and finds the appropriate curve to give the correct results.

The network is feed-forward and contains two types of learning methods of neural networks, so that the learning between the input layer and the hidden layer is unsupervised and the data is grouped into totals between the input data and the weights of the hidden layer that are initially randomly selected without the need to know the output, In this layer the activation function (Gaussian Radial Basis Functions) is used. The learning between the hidden layer and the output layer is supervised and depends on the degree of error on the outputs. It is called the basis function because the middle layer nodes represent a set of base functions such as the (Function Gaussian); it is a different way which presents the design of neural network as a curve fitting problem in a high dimensional space. The hidden layer of RBFNNs is nonlinear and the output layer is linear, whereas the hidden and output layers of an MLPNN are usually all nonlinear [33].

One feature of the Radial Basis Function Network is that any function of a limited range can be approximated by employing a linear combination of radial basis [34]. The radial basis function network consists of three layers of nodes; these layers are input layer, hidden layer, and an output layer. Each layer in this network is connected with the next layer, meaning that every node in the input layer is connected to all the nodes in the hidden layer, and the hidden layer nodes send their outputs to each node in the output layer. The number of nodes in the input layer depends on the application given to the network, and the number nodes in the hidden layer depend on the complexity of the given application (issue), the following figure 2.8 illustrates the three-layer radial base function network.

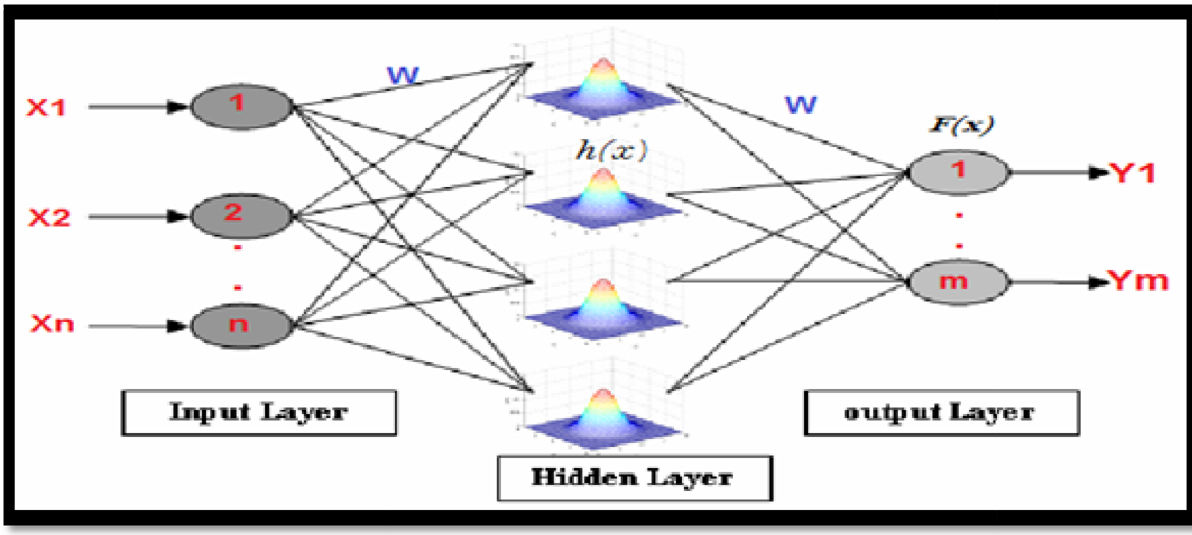

Figure 2-8: Three-layer radial base function network

Note from the previous figure that each node in any layer is connected to all the nodes of another layer by means of weights (*W*), the input to the network is represented by a vector (*X*) and the output of the network is vector (*Y*) and (*n*), (*m*) are the dimensions of the input and output vector. The initial weights of the first processing layer (hidden layer) are initialized as well the output layer as a random number generated and limited between [0-1] and the weight changes in the layers will be as follows:

- Hidden layer: weights are not changed in this layer because the learning is unsupervised.

- Output layer: the weights are changed when the required output is not achieved and it done by mathematical equations explained in the algorithm of the radial basis function.

The output of the neuron in the output layer of an RBF Network, where is given by [35].

$$f(\vec{x}, \phi, w) = \sum_{i=1}^{m} \phi i(\vec{x}).\, w_i \qquad (2.11)$$

Where ϕ is the activation of the radial unit for the input pattern x and w_i is the synaptic weight between the radial unit i and the output neuron [7], the activation of the i^{th} radial unit is dependent on the distance between the input pattern and the hidden unit center, using an Euclidean metric [36], In the hidden layer, the activation function used as a Gaussian, the activation of the i^{th} radial unit can be defined as using the following mathematical form [35]:

$$\phi(\vec{x}, \vec{c}, r) = \exp\left(\frac{||\vec{x} - \vec{c}||}{r}\right) \qquad (2.12)$$

Where $\vec{c}$ is the central point of the function ϕ, r is its radius and $\vec{x}$ is the input vector. In our thesis as an RBFNN, the training procedure is divided into two stages [35, 37].

First, the parameters centers, , and widths of the hidden layer are determined by genetic algorithms (GAs).

Second, the weights connecting the hidden layer with the output layer are determined by Singular Value Decomposition (SVD) algorithms method is adopted and these algorithms explained in special sections.

2.6.3 Learning Algorithms

The neural network learns by giving it a set of examples that must be carefully selected, as this contributes to the speed of network learning and this set of examples is called the training category, so a training algorithm finds a decision function that updates the weights of the network, notes that there are many variations of the training algorithms [38].

It is a difficult task to estimate which training algorithm will produce the best results [39], so typically the algorithms update the network weights and biases to map the best and correctly arbitrary inputs to outputs. Every neural network of all kinds has own learning algorithm, roughly speaking, the basic two learning algorithms strategies:

Supervised learning: All methods of learning or training by a Supervised of artificial neural networks are based on the idea of presenting training data or the network in the form of a pair of forms, namely the input form and the target form, Where learning with a Supervised can be done typically by correct the error.

Unsupervised learning: In this method, the training class is a vector of inputs only, without the presenting of the target on the network and this is called self-learning [40]. Our thesis adopted the supervised learning algorithms according to efficiency in prediction situation like using multilayer perceptron algorithm and also unsupervised learning algorithms in clustering Niberhoods in Jenin city based on the K-means clustering algorithm. Many training algorithms are available to train a neural network such as the Levenberg-Marquardt algorithm, Gradient descent, Quasi-Newton [30]. The Levenberg–Marquardt algorithm is a method to solve non-linear least squares problems. The goal is for minimizing problems which appear mostly in the least squares curve fitting. The Levenberg–Marquardt algorithm like Interlocking mixture, a combination between the Gauss-Newton algorithm (GNA) and the method of gradient descent. Thus, it is an iterative method that determines the minimum of a multivariate function that is defined as the sum of squares of non-linear real-valued functions [41].

Also in the same context, the Gradient Descent is an assistant algorithm that proposes to obtain the optimal solution according to trial and error way. When we have a neural network with a vast amount of parameters Gradient descent is the recommended algorithm. An iterative algorithm to solve nonlinear least squares problems as the Gauss-Newton method can be used to train the neural network. This method exploits a series of calculations as x-values to discover the solution, usually used to find the best fit model.

Radial Basis Function (RBF) also have own training algorithm as Gaussian training algorithm, also Many methods proposed such as employs the K-Means clustering (KMC) algorithm to determine the RBF center location, Also employed the K-nearest neighbor's technique (KNN) to initialization of the radius of each RBF, add to that the Singular value decomposition (SVD) used to optimize the weights, Singular value decomposition algorithm used in RBFNN for optimization of weights matrix of the output layer. Recently genetic algorithm is presented by optimization of its topology includes the parameters (centers, , and weights), so the possibilities of employing GAs to configure an RBFNNs exists and available. In this thesis, we use three different models, the Multi-Layer Perceptron (MLP) with Levenberg Marquardt training algorithm and the RBFNN with genetic algorithm and Auto Regressive Integrated Moving Average (ARIMA) method are carried out to train the obtained dataset.

CHAPTER 3

LITERATURE REVIEW

3 Literature Review

Predicting water demand or the needs of consumers in the future or in the short term is one of the main problems facing the management of water distribution systems. Therefore, it is necessary to estimate the demand for water with high accuracy and thus reduce the cost to the minimum. Given the increasing need and common role in all aspects of water resource planning, there is a need for more efficient planning models and methodologies to estimate water requirements and the need of municipalities and distribution companies in a more reliable and clear manner. The following section describes the research papers work that proposed to deal with water demand prediction and other fields in prediction status.

One of these methods that are based on linear and nonlinear manner was developed by Kofinas, D et al 2014 which provide an urban water demand prediction for the island of Skiathos [3]. The authors used four models to predict water demand (artificial neural network, winters additive exponential smoothing, ARIMA, hybrid). The proposed model results compared with all relative statistic of the prediction methods used and thus they deduced that ANNs is the better fitting model according to the root mean square error (RMSE) and others parameters like high (R square) represent how close the data are to the fitted regression line.

In [4] the authors proposed three different modeling to prediction water demand for the optimal operation of large-scale drinking water, they employed a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model, a Box-Cox transformation, ARMA Errors, Trends and Seasonality (BATS) modeling, and an RBF-based Support Vector Machine model. They established time series modeling methodologies and employed it to capture the dynamics of water demand, these models are trained using the same dataset, and the models are statistically validated of their prediction error. All models proved to be appropriate for the estimating of the demand.

Herrera, Manuel, et al. in [5] proposed a large number of alternative machine learning methods to predict the hourly urban water demand, they employed the Artificial Neural Networks (ANN), Projection Pursuit Regression (PPR), Multivariate Adaptive Regression Splines (MARS), Support Vector Regression (SVR), Random Forests, Weighted pattern-based model for water demand prediction. According to root mean square error (RMSE) and mean absolute error (MAE) adapted in this study and the final results obtained of comparison of all models, they deduced that the support vector regression models (SVR) is the most precise model, followed by (MARS), (PPR) and Random Forests. The experimental results about neural networks have exposed unsatisfied outcome and the author not explained the reason behind it.

A new technique of artificial neural network (ANN) called water demand to predict (WDF) approach in [26] is proposed to modeling and prediction the water demand in urban areas, This model integrates the artificial neural network technique and the techniques of econometrics. It does not require so many input data, It is a familiar interface and anyone with basic experiences in computers can be used it, the study area was chosen in Weinan City in China that consists of 9 years of data. Authors developed a simple ANN model consisting of only one hidden layer with the back-propagation algorithm, the authors have deduced according to analysis and results that the approach of artificial neural network displays the possibility to estimate and formulate water demand for home use in an efficient way. The proposed model shows that the correlation coefficients are more than 90% both for the training data and the testing data.

Ekonomou, L. in [42] presented as a model for long-term energy consumption prediction using artificial neural networks, the author proposed a multilayer perceptron model (MLP) with back-propagation and used for this purpose by testing several possible architectures in order to deduce the one with the best generalization capability, the author adopted 2 hidden layers, with 20 and 17 neurons in each one respectively, Levenberg Marquardt back-propagation learning algorithm and sigmoid transfer function. The network she was capable to minimize the mean square error (MSE) to the final value of (0.010). The results of ANN were compared and evaluated with the results generated by a linear regression method and support vector machine method and with actual energy consumption records and it showing a large reliability and accuracy.

In [18] Zhang, G. Peter, used a hybrid method to establish the effectiveness a new manner of time series prediction by using a hybrid ARIMA and neural network model, the methodology combines the advantages characteristic and unique feature of ARIMA and ANN models in linear and nonlinear modeling, they are used the three well-known data sets (Wolf's sunspot data, Canadian lynx data, and British pound-US dollar exchange rate data). These time series present from variant location and based on very statistical characteristics.

They deduced that the integrated approach can be an efficient way to improve prediction performance, and the empirical results that obtained with three real data sets indicate that the hybrid model is eligible to perform better than each model used in isolation from each other.

Kihoro, J. M. in [23] introduced a comparison between the ANN model and ARIMA model to evaluate models performance, the paper evaluates the performance of Artificial Neural Networks (ANN) and ARIMA models in prediction of seasonal (monthly) Time series by employing the Airline data and two other data sets and make comparison between each models each one in isolation on each other's, each data set an ARIMA model was fitted and the best model was found by check the minimum AIC, AIC, and BIC, also the MLP-Backpropagation with levenberg Marquardt optimization, hyperbolic tangent activation function neurons at the hidden layer and a linear transfer function on the output neuron were used. The training was repeated until mean squared error (MSE) reached and obtained minimum values. The empirical results show that ANN is somewhat better than ARIMA models in prediction ability taken into account the nature of the data may impact the results to a certain degree.

In [43], Ediger, et al. applied the Autoregressive Integrated Moving Average (ARIMA) and seasonal ARIMA (SARIMA) methods to predict the future primary energy demand of Turkey from 2005 to 2020. They studied time series data for a set of items (hard coal, lignite, asphaltite, petrocoke, wood, animal and plant remains, oil, natural gas, hydropower, geothermal heat electricity, and solar) in order to predict the future total primary energy. The mean-square error (MSE) is adopted for the reliability of the proposed model, according to the obtained results, the authors concluded that the ARIMA approach for prediction of the total primary energy demand seems to be reliable and hence the ARIMA and SARIMA models can effective to use for prediction of energy demand.

Chen, Ching-Fu, et al. in [44], they evaluated the efficiency of different methods that are used for prediction inbound air travel arrivals to Taiwan. Several models are studied by comparing their performance functionality. They used the Holt-Winters method, the seasonal ARIMA (SARIMA) model, and the GM (1, 1) grey prediction model, to check the accuracy of predict measurement the adopted the mean absolute percent error (MAPE) in empirical results, they compared the performance between all methods. All fitted models have shown a good prediction performance in terms of the MAPE criterion, the SARIMA model offers the best one for prediction inbound air travel arrivals to Taiwan.

The authors recommended that must be more empirical proof to support this finding as a good fit for a particular period because it maybe does not guarantee good predicts in the future.

Valipour, Mohammad, et al. in [45], investigated the accuracy of predicting the Dez reservoir inflow at the TalehZang station by employing and applied the ARMA, ARIMA, and dynamic autoregressive ANN with sigmoid activity function models. The dataset where used include the Monthly flow for 42 years and used for training models. The accuracy of prediction models was investigated by recent 5 years' data. The root mean square error (RMSE) and the mean bias error (MBE) were employing to nominate the best framework between ARMA, ARIMA, and artificial neural network models; because the time series feature is stationary the authors deduced that the ARIMA model has a better performance than ARMA model based on training and prediction phases. On the other hand, the dynamic autoregressive artificial neural network model with sigmoid activity function was excellent when compared with ARMA and ARIMA models for the desired prediction.

MemarianHadi, Siva Kumar Balasundram in [46], introduced the predictive performance of the Radial Basis Function (RBF) and Multi-Layer Perceptron (MLP) to predict the sediment load in a tropical watershed. Time series data of daily sediment discharge and water discharge at the Langat River, Malaysia obtained from 1997 through 2008 recorded were used for training and testing the networks, The MLP was trained by a backpropagation algorithm, for transfer activation function the logistic function and the hyperbolic tangent are employed and each weight in the network was adapting by updating the present value by gradient descent learning algorithm, the authors applied the same datasets for network training and testing in both types network, so they divided data by employing 54% for training and 14% for cross-validation and the rest of the data equal 32% is employed for Network testing. Experimental result indicated according to the minimum MSE obtained of the RBF network was larger than the MLP network and the MLP network achieved the best-fitted output to the cross-validation data set than the RBF network. Also, MLP network in testing data achieved the lowest MSE and NMSE when compared with RBF network, add to that the MLP network showed additionally qualified in term of depiction the fluctuations in daily sediment load than the RBF network.

In [35] Awad, introduced a new efficient method of optimizing Radial Basis Function Neural Networks Parameters using Genetic Algorithms, the author integrated and exploited the power of genetic algorithms to improve and get better the centers and radius of RBFNs, also he applied the singular value decomposition (SVD) to optimize the weights of network. The proposed algorithm was applied to cases of one and two dimension. The outcomes show that the approach achieves better-normalized root mean square error than traditional algorithms. Also, it achieves theminimal complexity of calculation to optimize each parameter alone and produce the additional accurate prediction.

Francesca, et al. in [47], the author used two models one of them depends on the artificial neural network to represent water distribution systems characterized by a very different number of users, and Data-driven models based on linear and non-linear data-driven techniques, consequently Both the models provide high-medium forecasting accuracy which produce good prediction.

JAIN, et al. in [48], The artificial neural networks (ANNs), regression, time series analysis technique, have been examining to model the short-term water demand predicts, precisely two-time series models, five regression models, and six ANN models, where progressing and developing. All the above models were carried out to select the appropriate of each technique investigated to model water demand predicts. According to the results gained in this study and the comparative analysis, it has been shown that the models of ANN technique have performed better than the models using classic and traditional techniques of time series analysis and regression, so the complex ANN model performed the best among all the models developed in this study.

In [49] Ishmael S, et al, Two machine learning techniques have been used and tested in this paper, the artificial neural networks and the support vector machines are used to predict short-term and long-term water demands, two type of neural networks architectures is used, MLPNNs and RBFNNs, SVM experiment was including of many models, the models were compared to each other in order to locate the Support Vector Genius (SVG). The result produces that artificial neural networks perform prediction result better than SVM.

Xing, Ying, Zhenwei You, et al. in [50], improve the new way of back-propagation (BP) by means of two heuristic algorithms, they developed the back-propagation by combining it with genetic algorithm (GA) and particle swarm optimization (PSO), in order to do real-life water demand prediction of Beijing city, they tried to increasing the accuracy of prediction process by the testing and verification of the three algorithms (back-propagation, back-propagation with genetic algorithm , back-propagation with particle swarm optimization).although the execution time consumed is take longer time but both back-propagation with genetic algorithm , back-propagation with particle swarm optimization performed with higher accuracy and less errors than back-propagation .

González, et al. in [51], the authors presented a new approach to predicting the daily demand for water in the short term in the irrigation sector. In this paper the researchers developed a Bayesian framework approach and a multi-objective genetic algorithm to find the optimal neural network with the best values for coefficient (R^2) and Standard Error Prediction (SEP), two best models were created in order to achieve the best coefficient and Standard Error Prediction. One of the models is sophisticated and improved on the models developed by others in term of measurements are higher accuracy and the better. The authors used just two subsets (training and testing sets) rather than three subsets (training, validation and testing sets), add to that the joint optimization of the ANN architecture and the fundamental parameters of the Bayesian algorithm, thus the predictive model which developed made more accurately.

Xiao, et al. in [52], the authors introduce a new approach namely a hybrid EEMD-Elman neural network model to predicting the hourly campus water demand, they are combined the Elman neural network (ENN) and Empirical mode decomposition (EMD) method and carried out the prediction of water consumption hourly for 31 days on the campus of Hebei University of Engineering. The actual water consumption and the predicted results were carried out over the EEMD-improved ENN model and compared with the single Backpropagation (BP) and Elman neural network (ENN), the results indicate that the combined model has a good performance in prediction hourly water demand and achieved minimum error and more accuracy but the complexity was increased.

CHAPTER 4
THE PROPOSED METHODOLGIES

4 The Proposed Methodologies

This chapter presents the utilized methods in our work. We have used three applied techniques were used to perform the prediction process, with the aim of modeling the future water demand in Jenin city, it is necessary to gather data for training purposes. Thus we will concentrate and take consideration on the process of data gathering which will have used for neural networks and Stochastic Models that are used to predict future water demand based on the pattern in previous time series values which obtained from the municipality of Jenin city in the north of West Bank, Palestine. Hence the collected data will be exploited to train in different neural networks and also others models with different learning algorithms in order to predict the future demand relating to actual demand. All models will be discussed in details.

4.1 Data Collection and Preparation

The water demand prediction in this work depends on the historical data, thus required data were extracted mainly from the municipality of Jenin city database management system because it is the only source usually has the best information about the amounts of consumption. As known, the water distribution networks consist of water meter for each user.

The assumed values were collected, and then preprocessing operations have been applied, validated and rearranged to generate a training dataset which contain many of the periods. Each period of them contains values for input and output data.

In this thesis, it is important to deal with regular time series data to construct data training matrix, since that the records have not been obtained for years prior to 2011, the period of the data which include the water consumption starts from 2011 to 2017. In the first stage, we were provided with raw data through different format, some of it CSV file format and another it Excel files and separated from each other and contained monthly consumption quantities per subscriber for a full year for a total of approximately 7700 subscribers for each excel sheet as figure 4.1 illustrate that.

NO	customer_id	customer name	total consumption	2011/1	2011/2	2011/3	2011/4	2011/5	2011/6	2011/7	2011/8	2011/9	2011/10	2011/11	2011/12
1	W0000001	[redacted]	32	9		0	0	1	5	2	0	9	3	2	1
2	W0000003	[redacted]	239	20	12	40	33	21	19	4	14	4	28	40	4
3	W0000004	[redacted]	97	1	4	39	1	2	5	18	6	18	1	0	2
4	W0000005	[redacted]	91	10	6	2	1	1	9	8	10	10	20	10	4
5	W0000006	[redacted]	192	10	6	24	8	15	7	7	8	10	35	49	13
6	W0000007	[redacted]	71	3	4	10	6	4	4	5	6	6	5	8	10
7	W0000008	[redacted]	74	8	4	1	4	7	4	6	8	7	5	10	10
8	W0000009	[redacted]	80	11	6	2	5	4	6	7	9	15	4	5	6
9	W0000010	[redacted]	191	7	4	23	15	13	20	16	15	29	11	27	11
10	W0000011	[redacted]	121	13	7	15	19	13	4	10	5	11	7	8	9
11	W0000012	[redacted]	135	4	4	2	2	18	9	12	4	7	5	30	38
.	.	.	285	0	4	0	37	50	50	48	10	29	20	24	13
.	.	.	53	2	4	4	0	4	2	9	4	8	6	10	0
.	.	.	79	3	4	2	0	1	11	14	8	6	6	17	7
7700	.	.	159	14	8	52	5	9	11	15	8	14	9	4	10

Figure 4-1: Consumption quantities per subscriber for a full year

In the second stage, we transformed data to uniform format and aggregated all the consumption of subscribers for all months and years within a unified database to facilitate sorting and maintaining a precise data copy, taking into consideration the adoption of all the subscribers and the active meters only and dropping the inactive meters to obtain correct and reliable data for future processing. The dataset values then aggregate to compute the total consumption vector for each month individually in order to normalize according to the following equation:

$$y_i = \frac{(x_i - \min(x))}{(\max(x) - \min(x))} \qquad (4.1)$$

Where (x_i) is the actual consumption and (y_i) is the normalized value, *min* and *max* are the maximum and minimum values for actual consumption [53].

Figure 4-2 shows the aggregated and normalized consumption values.

No	period	consumption	Normalize	No	period	consumption	Normalize	No	period	consumption	Normalize	No	period	consumption	Normalize
1	1/1/2011	87922	0.3469	22	1/10/2012	100315	0.5185	43	1/7/2014	88144	0.3499	64	1/4/2016	101247	0.5314
2	1/2/2011	62872	0.0000	23	1/11/2012	102506	0.5488	44	1/8/2014	135093	1.0000	65	1/5/2016	107857	0.6229
3	1/3/2011	92815	0.4146	24	1/12/2012	90849	0.3874	45	1/9/2014	132464	0.9636	66	1/6/2016	91098	0.3908
4	1/4/2011	86630	0.3290	25	1/1/2013	96753	0.4691	46	1/10/2014	126675	0.8834	67	1/7/2016	101135	0.5298
5	1/5/2011	80095	0.2385	26	1/2/2013	85356	0.3113	47	1/11/2014	85396	0.3119	68	1/8/2016	122636	0.8275
6	1/6/2011	78519	0.2167	27	1/3/2013	80811	0.2484	48	1/12/2014	99532	0.5076	69	1/9/2016	101980	0.5415
7	1/7/2011	82955	0.2781	28	1/4/2013	98963	0.4997	49	1/1/2015	107301	0.6152	70	1/10/2016	91279	0.3933
8	1/8/2011	74941	0.1671	29	1/5/2013	104256	0.5730	50	1/2/2015	79610	0.2318	71	1/11/2016	88208	0.3508
9	1/9/2011	127914	0.9006	30	1/6/2013	96515	0.4658	51	1/3/2015	82943	0.2779	72	1/12/2016	127649	0.8969
10	1/10/2011	106002	0.5972	31	1/7/2013	98566	0.4942	52	1/4/2015	91379	0.3947	73	1/1/2017	108796	0.6359
11	1/11/2011	115280	0.7257	32	1/8/2013	107047	0.6117	53	1/5/2015	95155	0.4470	74	1/2/2017	94845	0.4427
12	1/12/2011	84630	0.3013	33	1/9/2013	124322	0.8509	54	1/6/2015	108161	0.6271	75	1/3/2017	94269	0.4347
13	1/1/2012	90324	0.3801	34	1/10/2013	111649	0.6754	55	1/7/2015	83022	0.2790	76	1/4/2017	93726	0.4272
14	1/2/2012	84518	0.2997	35	1/11/2013	101258	0.5315	56	1/8/2015	130066	0.9304	77	1/5/2017	97876	0.4847
15	1/3/2012	83559	0.2864	36	1/12/2013	88596	0.3562	57	1/9/2015	118142	0.7653	78	1/6/2017	92142	0.4053
16	1/4/2012	86828	0.3317	37	1/1/2014	94194	0.4337	58	1/10/2015	117532	0.7568	79	1/7/2017	110706	0.6623
17	1/5/2012	104071	0.5705	38	1/2/2014	77720	0.2056	59	1/11/2015	106424	0.6030	80	1/8/2017	109670	0.6480
18	1/6/2012	105533	0.5907	39	1/3/2014	85769	0.3170	60	1/12/2015	95267	0.4486	81	1/9/2017	110216	0.6555
19	1/7/2012	104202	0.5723	40	1/4/2014	101406	0.5336	61	1/1/2016	97948	0.4857	82	1/10/2017	88352	0.3528
20	1/8/2012	99820	0.5116	41	1/5/2014	113229	0.6973	62	1/2/2016	90744	0.3859	83	1/11/2017	96403	0.4643
21	1/9/2012	113595	0.7023	42	1/6/2014	103408	0.5613	63	1/3/2016	91198	0.3922	84	1/12/2017	97753	0.4830

Figure 4-2: aggregated and normalized consumption values

Our Input data vector includes 84 entries that represent the number of months in the all years from 2011 to 2017, Also the target data vector includes 84 entries that represent the actual consumption values for each month from (January - 2011 to December - 2017). The total consumption of water subscribers over the seven years was collected from all neighborhoods of Jenin city which is 39 districts according to the map as shown in figure 4.3 which was provided to us by the engineers in the municipality committed to the organization and planed the water department and technical issues.

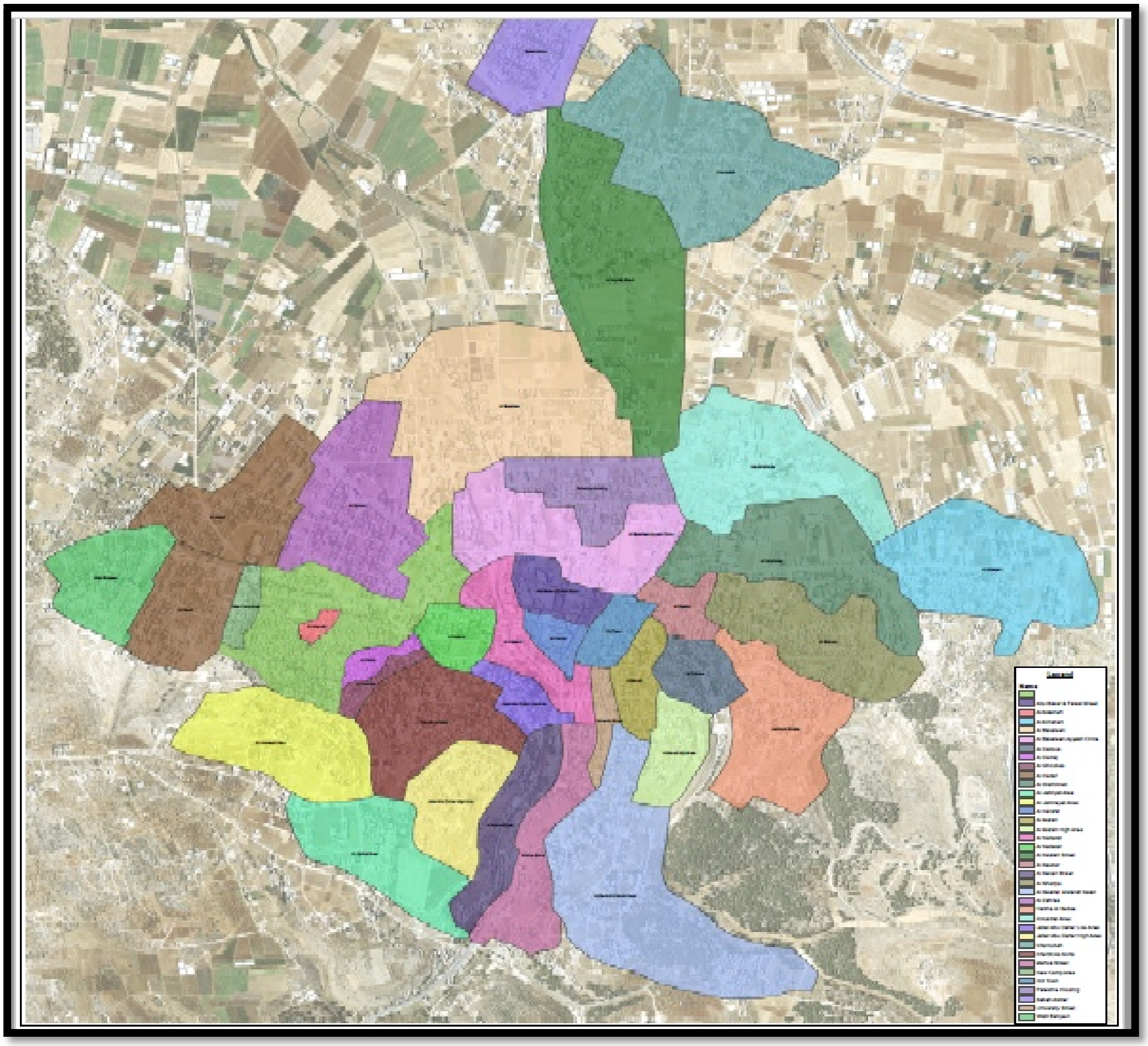

Figure 4-3: Neighborhoods of Jenin city

4.2 Design the Proposed Applied Models

The goal of this study is to develop a method capable of predicting water consumption. In an attempt to build models, systems, and applications in order to estimate the future needs of something, it is necessary to take into account the amount of accuracy that will be achieved. Therefore, the decision to build models for the future prediction of water needs to be as accurate and reliable as possible so as to produce and choose the best one.

Specifically, with regard to time series prediction, it is known that there is a relationship between past observations and what will be or will occur in the future. This also applies to our case under consideration regarding the predictability and estimation of future demand for water. Regardless of the models used or methodologies adopted and applied remains the main objective of a time series is to construct a model to infer future unknown data from existing data by reducing the error between actual and desired values. Many methodologies and methods have been proposed over the past years to assess future needs and have been investigated in several areas, domains, and fields. In the following sections, we will review in detail the methods that have been investigated and applied in this thesis. In our work, we used two learning algorithms depending on two types of neural networks also we used the multi-linear regression (MLR) ARIMA models as illustrated in Figure 4.4, The next sections present in detail these algorithms.

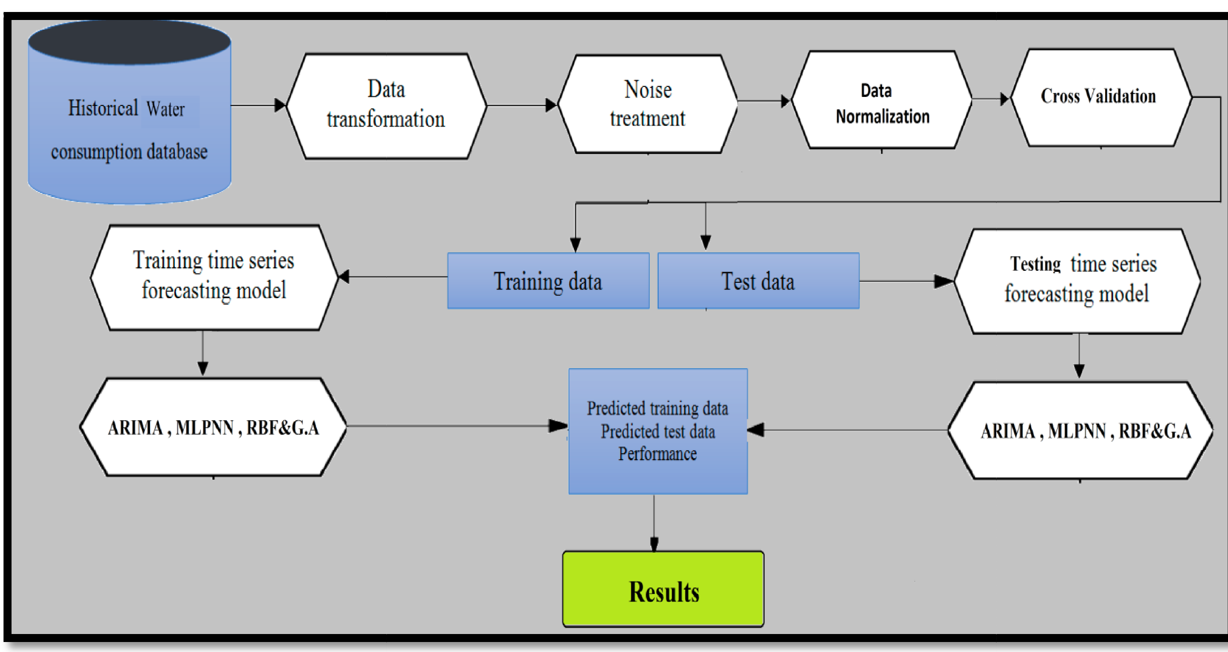

Figure 4-4: General method procedure

From the above figure we find that architecture passes and consists of several stages: -

- Initially, all dataset is stored in the municipality database.
- This data is stored in several different formats where it must pass by process of converting data or information from one format to another.
- Drop and delete Noisy data that meaningless or corrupt data. Hence, its include any data that cannot be understood and interpreted correctly by machines
- Data normalized to organizing data in a database belonging to us, including the creating new tables and redesign relationships between those tables and then to convert all water consumption values to range between zero and one to facilitate computation process in MATLAB.
- Cross-validation by separation the dataset into two-part, training and test set, so as to use the test data to evaluate the quality of model trained based on training data and to measure the performance on the unseen data.
- Training and testing the time series of all models (ARIMA, MLPNN, RBFNN).
- Executing the prediction according to all models (ARIMA, MLPNN, and RBFNN) and finally obtain the results.

4.3 Cross-Validation

Cross-validation is a practical and reliable way for testing the predicting capability for used methods. Hence it’s necessary for any non-linear models or neural networks techniques to have a training set and test set in order to verify and to check its validity and estimate the quality of a neural network. Also, Cross-validation is the best choice to avoid minimum overfitting or underfitting. Many styles and varied types of cross-validation can use [54], but a core at the end is a similar and in the same meaning.

The simplest kind of cross-validation is the holdout method or sometimes named test sample estimation, the dataset distributes or divided into two mutually independent subsets one of them called a training set and the other called a test set. It is widespread to allocate about (0.70) of the data as a partition of the training set and about (0.30) as a partition of the test set. Approximator employs a function using just the training dataset. Next, the Approximator is carried out to predict the output values for the testing dataset, knowing that this test partition Not seen before.

Another way can be used is K-fold cross-validation, where the data set is partitioned into k subsets, and test sample estimation method is repeated k times individually [55]. Every time, we select one of the subsets to be the test set and the other subsets are placed unitedly to be a training set. Thus each data point captures in a test set precisely. We always look at less expensive computing processes and fewer consumption times more in running so we start to apply the easiest and fastest way to get the desired results, maybe the emergence of problems with regard to (over/under fitting) and therefore we back to other ways such as K-fold technique to be used. In our thesis, we divided data points into two sets randomly, and we selected the portion of each set as a Rate between (70%) for the training set and (30%) for testing set.

4.4 ARIMA Time Series Predictor

(ARIMA) Autoregressive Integrated Moving Average, (Box-Jenkins) and multi-linear regression models have been broadly implemented in various predictions in many fields, such as financial, industrial, commercial, climatic and demographic fields, but accuracy is variable because they represent a linear manner of nonlinear systems [56]. ARIMA method based on short-term forecasting was employed in many fields as we mentioned in previous, also including water demand forecasting applications with many services [57].

As such, the use of ARIMA model is proven to be effective in demand forecasting technology, thus the outcome produced by this model produces precise results compared with the results generated by the modern techniques presented in this thesis. This section presents a classification in detailed of the ARIMA approach and presents the algorithm applied during the work reported in this thesis to give corresponding results for water demand data. Numerous mathematical time series forecasting approaches, like exponential smoothing, assume that every data point in the time series values will consist of the deterministic mean and a random error. Most often time series characterized by a large grade of dependency between sequential observations. So the time series of water demand giving such dependency between observations because of a great daily, weekly or monthly cycles inside the consumption patterns, hence the ARIMA forecasting manner is intended to benefit from this dependency in order to create the prediction.

The technique of ARIMA must achieve many essential components, so the components are:

1- Autoregressive (AR), it is a time series paradigm use the past observations as input to fed a regression equation to predict the value in future vision, based on a linear combination [58]. For instance, we can predict the values for later time ($t+1$) by given two times of past observation (*t-1, t-2*) as a regression model by employing the following equation:

$$X(t+1) = b0 + b1 * x(t-1) + b2 * x(t-2) \qquad (4.2)$$

Where [b0, b1, b2] are coefficients obtained through optimizing process on training data, It represents the value and amount of the regression slope of the dependent variable on the independent variable and X is representing the input.

2- Integrated (I): it finds the difference between the observations that represent the actual and raw data, In other words, we subtract the current observation from previous observation [43]. Differencing is a very good way of change or converts a non-stationary time series to another stationary. Basically, this approach can remove the trend if we saw that the rising of time series is rising at a fixed rate. In this way, we can apply only once, this is called “first differenced”, mathematically as an equation.

$$\acute{\mathrm{Y}}\mathrm{t} = \mathrm{Yt} - \mathrm{Y}(\mathrm{t}-1) \qquad (4.3)$$

If achieved purpose of making the curve of time series is stationary and the trend was removed then the first-differenced will be enough, if not we can repeat the procedure again and this called “second differenced”, mathematically as an equation.

$$Y *= Yt - 2Y(t-1) + Y(t-2) \qquad (4.4)$$

3- Moving Average (*MA*): determine that the output value relies in a linear way on the current value and previous values, this model uses the dependency among the observation and between lags of the predicted error in moving average approaches

[59, 44]. In other meaning, each value of *X (t)* is immediately linked just to the random error in the past *E(t-1)*, and to the current error *E(t),* so the subsequent equation represents it:

$$Xt = \mu + \varepsilon t + \theta 1\varepsilon(t-1) + \ldots + \Theta q\varepsilon(t-q) \qquad (4.5)$$

Where μ: mean of the time series, Θ: parameters model, ($\varepsilon t,\ \varepsilon_{t-1},\ \varepsilon_{t-q}$): white noise error represent residual errors series x_t and represent the difference between an observed value and a predicted value from a time series model at a particular time t.

In general, select the best model of ARIMA and the model order is not easy, and the user must have a good experience and high level of skills because of ARIMA model includes a large number of details and the statistical equations. In briefly, ARIMA model consists of three component or parameters, non-seasonality *(p, d, q):*

p- Represent the lags of stationeries series (AR), the order of non-seasonal (AR), we determine it through partial autocorrelation function (PACF).

d- Differencing between the observations (I)

q- Represent the lags of the predict errors (MA), the order of non-seasonal (MA), we determine it through autocorrelation function (ACF).

SARIMA is another type and it similar ARIMA but the first take into account the seasonality issues, thus we can demonstrate the model as ARIMA add to that the part of seasonality *(P, D, Q)*, so the adopted form is by employing the following expression:

(p, d, q)X (P, D, Q).

Many applications provide us with the possibility of building ARIMA models such as MINITAB, MATLAB, IBM-SPSS-STATICS, and the most beautiful is the R-studio, which provides us with a ready tool to shorten the long list of procedures, calculations and representations and suggest to us the best model, that are represented in the tool (AUTOARIMA).

4.4.1 Box-Jenkins Methodology

Despite the diversity and different methodologies that can be exploited and applied to solve the problem of adopting a certain methodology, which must lead to the goal we seek, our greatest concern remains is how to nominate and select a suitable paradigm that achieve to accurate predict according to an observation of historical pattern in the data and how to discover the optimal performance model. The expert statisticians Box and Jenkins [60], in (1970) establish a practical approach to create the ARIMA model, which given the best match to time series. This approach has an essential significance in the domain of time series analysis and forecast [61].

The Box-Jenkins technique is based on the assumption that there is no any particular pattern in the historical data of the time series to be predicted. But the alternative is to use a three-step iterative approach of model identification, parameter estimation, and diagnostic checking to determine the best model from a certain number of ARIMA models [62]. The process is repeated continuously many times until an accurate model is finally discovered, and then this model can be used for predicting future values of the time series, the general Box-Jenkins predict method is shown in figure 4.5.

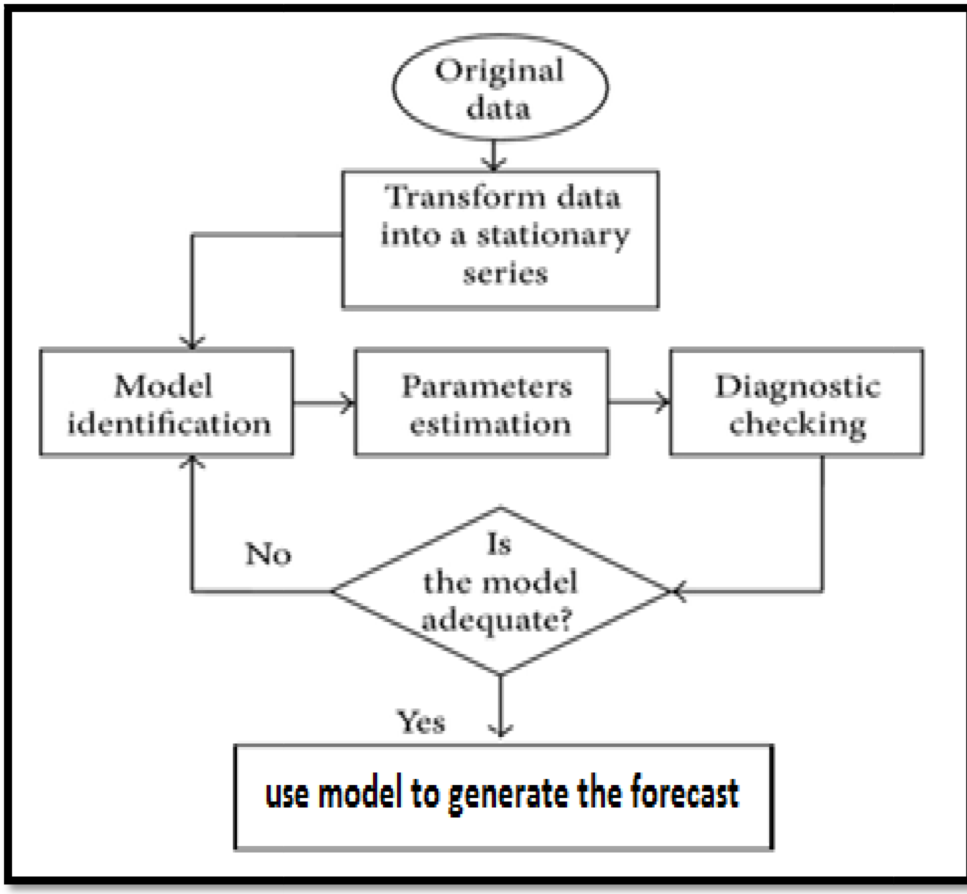

Figure 4-5: Box-Jenkins manner for determination best model

To provide the practical approach of Box and Jenkins in order to build the models of ARIMA we must go through multiple stages starting from fetching data and ending with the adoption of a predictive model that achieves the goal.

First stages are called identification, we aim at this stage to determine the order of the model wanted (*p, d, q*) so as to take and know the data properties and the major features of series, such that capture the trend both (increasing/decreasing), also the series is (stationary-nonstationary), add to that to determine seasonality or not for the time series according to use graphical procedures (plotting the series, ACF/ autocorrelation function and PACF/

Partial autocorrelation function). So a time series must be stationary and it is “stable”, that means the mean is constant over time (there is no trend) and the correlation structure between every spot in the time series in observed value of series remains constant over time, so we always demanding to remove the trend and seasonality in order to move on to the next stage.

Hence the ACF plots can help us to determine the order of the moving average (MA) (q) model, add to that the PACF plots areas very significant role for determining the order of the Autoregressive (AR) (p) model.

The next stage is called the estimation and diagnostic checking to determine and selection the best model, which includes the estimation of the parameters of the several models using previous stage and produce the first best selection of models based on (whether using information criteria the Akaike Information Criterion AIC or the Bayesian Information Criterion BIC) [63][64], it is widely used to measure the goodness fit in any estimated statistical model, and which are defined below according the equations [63]:

$$AIC(p) = n \ln\left(\frac{\hat{\sigma}_e^2}{n}\right) + 2p \qquad (4.6)$$

$$BIC(p) = n \ln\left(\frac{\hat{\sigma}_e^2}{n}\right) + p + p\ln(\mathrm{n}) \qquad (4.7)$$

Where (n) is the number of observations to fit the model, (p) is the number of parameters in the model and (σ) is the sum of sample squared residuals. The best model order is picked by the number of model parameters, which minimizes either AIC or BIC [24].

Model 1: The general procedure that was used for Box-Jenkins Algorithm

Input : data points

Step -1: Plot series

Step -2: is variance stable

  2a. No, apply transformation, go to 1

  2b. yes, continue

Step -3: obtain ACF and BCF

Step -4: is mean stationary?

  4a. No, apply regular and seasonal differencing

  4b. yes, continue

Step -5: Model selection

Step -6: Estimate parameter value

Step -7: Are residuals uncorrelated?

  7a. No, modify model, go to 5

  7b. yes, continue

Step -8: are parameters significant and uncorrelated?

  8a. No, modify model, go to 5

  8b. yes, continue

Step -9: forecast

Figure 4-6: ARIMA Box-Jenkins Model

Note that already we must have more models to select the best one and has the lowest value, once we have adopted one of the models according to the previous step, this model must be subjected to several diagnoses to determine if this model is sufficient and achieves the goal.

The final stage is called the diagnostic. In this stage, we are concerned about three issues; first is (P value for Ljung-box statistic) and this means that all values of (P-value) must be higher than 0.05 and therefore we do not reject the initial hypothesis that indicating that the model meets the purpose. The second issue is ACF of residual, we concern about the violation if exist, else that our assumption for adopted model stay meets our need. The final issue is standardized residual, and must care about the randomization that does not indicate to the pattern, so we don't need for any pattern, add to that there is additive diagnose include the observance of normality, so we will review all of the above and explain it in a practical way in the experimental results chapter. Based on all stages above, we now carry out the predicting, it is applied step to produce the prediction based on the model you decide is best.

4.5 Time Series Prediction Using Artificial Neural Networks Models

Artificial neural networks model is a computational model derived from the biological neural networks. They include a huge number of processing component which performs their work in simultaneous and parallel ways. ANNs are used in broad areas mostly used in classification (pattern recognition [85], Image Processing [84], Character recognition [86]) and prediction problems, hence they dealing with very complex data in order to be understood by humans at the end.

ANNs can deal easily with large amounts of inputs under the noise and inaccurate information. According to these features, they will be idealistic and perfect to address the problem of water demand under imprecise circumstances of data and thus to offer an accurate future prediction.

The ANNs is a methodology for solving our problems, this manner must accomplish through training, validating and testing phase before used in any application. As soon as a network has been built to serving or solving a certain problem, the network must be trained and begin the process with initial weights which picked according to the random approach [65]. Then, training/learning begins based on different methods like (Supervised / Unsupervised) and adjusting parameters that represent the data set, so the error of this phase is minimized. The major advantages of neural networks are their ability to generalize, this means that a trained network could classify new data which have the same characteristics of the trained data category, knowing that it had never seen her before [66]. In order to determine the performance and ability of network on new examples that are not trained during learning, the validation set is used also to avoid over-fitting and stop the training at an appropriate time [67]. Testing is the process of measuring and evaluating the matching degree of the network between their output and target data.

Nowadays the ANNs are the most widely implemented methodologies in prediction large sectors like (energy consumption, water demand, weather conditions, and more domains). Since the complexity of these sectors is very high because of many factors, the ability of the ANN in performing the non-linear analysis is perfect for carrying out the prediction [68]. The evolution of the methodology used in prediction processes in artificial neural networks, in general, can be measured by developing and reducing error value and minimizing them. Therefore, the network during the training process continuously calculates and evaluates the value of the error between the target data and the desired results so that the error value is as low as possible and this is the main principle in the work and performance of these networks. In our work, we used the mean square error (MSE) to measure the progress grade in error reduction. The following equation is used to the MSE [69].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(T - Y)^2 \qquad (4.8)$$

Where n is the input data, T is the target data and (Y) is the network output. In our work, we used two learning algorithms depending on two types of neural networks also we used the multi-linear regression (MLR) ARIMA models. The next sections present these algorithms in detail.

4.6 Multi-Layer Perceptron Feed Forward Network Back Propagation Algorithm

The basic rule of Multi-Layer Feed Forward Network is that all connections flow in one direction, meaning that the data is streaming from the input layer to the output layer. On the other hand, the Back Propagation Network is a type of gradient descent mechanism with backward error propagation [70].

Comparisons continue between the initial outputs of the network and between the real values by modifying the network parameters repeatedly and continuously to obtain the lowest error value, the greater the amount of data available, the higher the performance of BPN. In recent times, according to the results of experiments and practical practices and the use of various statistical and traditional algorithms, artificial neural networks, especially multilayer Perceptron, are very effective and powerful and alternative to other methods, unlike traditional statistical methods, multilayer Perceptron does not care about any prior assumptions about data distribution or relationships between them.

A major issue and feature are that it has non-linear functions and can be trained to accurate generalization when presented with new, unseen data [70]. A multi-layer perceptron network is capable of creating an approximation of any function and smooth, it is a measurable function between input and output vectors depending on the appropriate choice of weights and connections [70]. Multi-layer Perceptron depend on the supervised way to learn, through the training may be the results not equal the desired output, so the error is calculated and this error belonging to variation between the desired and actual output. Thus the training uses the size of this error signal to decide the degree of weights must be adjusted or modified, hence we obtain the lowest error of the Multilayer Perceptron. Optimizing and adjusting the weights is the core of the training process in multilayer Perceptron which determines through it the feasibility and ability of network to give us the optimal solutions about certain problem in an accurate manner, according to many techniques used in this situation like (gradient descent, Gauss-Newton, and the Levenberg – Marquardt) algorithm, we will explain the adopted technique in our work in a particular section.

Predicting the future demand for water consumption in many areas was performed according to the numerical models, statistical methods, and time series prediction based on the applications of statistical linear and nonlinear techniques. In our work, the Multilayer feed-forward with back-propagation neural networks model is adapted to predict the future demand of the water consumption. Figure 4.7 illustrates the process that emulates the same way of prediction using Multi-Layer Perceptron Feed Forward Network with Back Propagation Algorithm models.

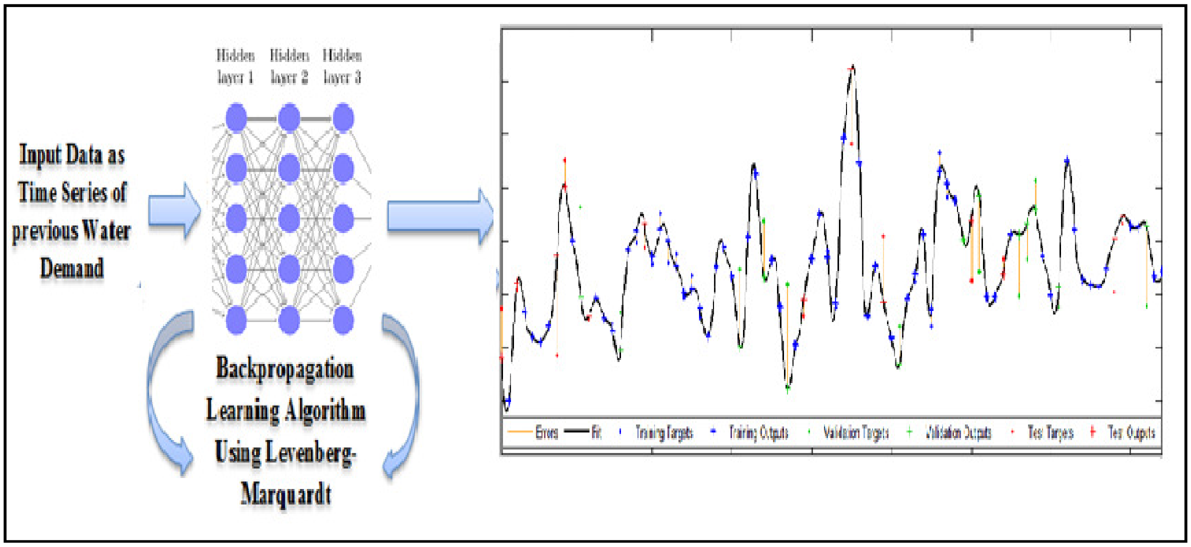

Figure 4-7: Time series prediction process using MLPFFBPNNs models

Multilayer Perceptron feed-forward neural networks with back-propagation is an MLPNN that forwards the inputs and the weights from one layer to the next layer then backward to update the weights from layer to layer in order to update the weights in each layer [71].

4.6.1 Training a Multi-Layer- Perceptron-Backpropagation Algorithm

The applied procedure to execute the learning operation in NNs is called training Algorithm [72]. So there are several different training algorithms with different Characteristics and performance. The most important training algorithms that are known for Neural Networks are Gradient descent [87], Newton's method, Quasi-Newton [88], Levenberg Marquardt algorithm [89]. In general, the objective of training is to minimize the error which defines as the following equation [72]:

$$E = \frac{1}{p}\sum_{p=1}^{p} Ep \qquad (4.9)$$

Where (p) represents the total numbers of training patterns, and (E_p) represent the error of

Training patterns, (Ep) is calculated according to the following formula:

$$\mathrm{Ep} = \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{Oi} - \mathrm{Ti})^2 \qquad (4.10)$$

Where (n) is the total number of output nodes, (O_i) is the network output at the (i^{th}) output node and (T_i) is the target output at the (i^{th}) output node.

So in each training algorithm, we seek to reduce and decrease the error by adjusting the weights and biases.

In our work, we used a Multi-Layer Perceptron Back-propagation algorithm based on a Levenberg-Marquardt algorithm to optimize the weights and we will explain this approach in the following section.

4.6.2 Levenberg – Marquardt Algorithm (LMA)

Levenberg Marquardt Algorithm is applied to solve nonlinear least squares problems. This algorithm is based on the preferred manner and this curve-fitting method is a combination and mixture of two methods: the gradient descent and the Gauss-Newton. [41], in the other meaning, the iterative update is dependent on the value of an algorithmic parameter; (λ) is a non-negative damping factor which facilitates the graph. The update is Gauss-Newton if (λ) is small and close to the optimal value whiles a gradient descent if (λ) is large.

The hessian matrix of quadratic error can be expressed as the following equation:

$$H^{ij} = \frac{\partial^2 E}{\partial w^i \partial w^j} = \sum_{l=1}^{N}\left[\frac{\partial y_l'}{\partial w^i}\frac{\partial y_l'}{\partial w^j} + \left(y_l' - y_l\right)\frac{\partial^2 y_l'}{\partial w^i \partial w^j}\right] \qquad (4.11)$$

Where (E) is the error function, (w^i) is an i^{th} element of input layer weight, (w^j) is a j^{th} element of output layer weight,($y_l^{'}$) is the derivative output of 1^{th} the neuron, (y_l) is output of l^{th} the neuron.

The second edition of this expression is neglected when considering that the errors are random and de-correlated with the second derivative of the output, so it can be considered as a white noise. From the first edition of the Hessian the following matrix is constructed [73]:

$$J^{ij} = \sum_{l=1}^{N} \left[\frac{\partial y_l^{'}}{\partial w^i} \frac{\partial y_l^{'}}{\partial w^j} \right] + \lambda I_M \qquad (4.12)$$

Where I_M is the identity matrix of order M, and λ is a parameter that makes a function similar to learning in the back-propagation algorithm. The modification of the weights with the Levenberg-Marquardt method in the μ-i_{th} learning cycle is given by the following expression:

$$W(\mu) = W(\mu - 1) - J^{-1}(\mu - 1)\nabla E(\mu - 1) \qquad (4.13)$$

Or

$$\Delta W = (\mathrm{JJT} + \lambda l_M).j^T E \qquad (4.14)$$

The matrix J can be calculated from the following auxiliary matrix:

$$\mathrm{Ja}_l^i = \frac{\partial y_l^{'}}{\partial w^i} \qquad (4.15)$$

Getting:

$$\mathrm{J} = \mathrm{Ja}^T \mathrm{Ja} + \lambda \mathrm{I_M} \qquad (4.16)$$

As we saw we didn't need more complicated calculations except for the back-propagation algorithm. So the auxiliary matrix (J_a) also gives the error gradient [74]:

$$\nabla E^{i} = \frac{\partial E}{\partial w^{i}} = \sum_{l=1}^{N} \mathrm{Ja}_{l}^{i} \left(y_{l}^{'} - y_{l} \right) \qquad (4.17)$$

The Initialization process can be done either by fixed weights or random, in our case we used random initialization.

Firstly the predicted output of the first hidden layer must be calculated by employing the sigmoid activation function, and this output will be the input of the second hidden layer. Then we calculate the predicted output for the second hidden layer also.

This stage called the forward phase and at the end of this stage, we just have two values, target data that represents the actual data and also the output of the network.

Between target data and the output of the network, there are some errors which we aim to reduce it to as little as possible.

In the backward phase, we calculate the mean squared error between the target and the calculated output by employing the Levenberg Marquardt algorithms to train the network and update the weights, this process runs and Several cycles resume of forward and backward and repeat until getting the threshold and obtain the lowest value of mean squared error.

Figure 4.8 shows Pseudocode of the general procedures that were used in building our prediction model and explains the steps of the algorithm.

Algorithm: The General Procedure That Was Used in Building the Prediction Model

Input: Datasets consists of time series of **84** points
Output: prediction the future water demand result;
Data Preprocessing:
  Normalize (input time series, target)
  Initialize the weights wji of the MLPNN with random values.
  Determine input pattern Xt: (Xt_1, Xt_2, ….X tn).
  Determine target output from the collected data.

Forward Phase: Feed Forward
For each iteration in the training process, do

Calculate the prediction output of the hidden layer L1 of the MLPNN $out_{L1} = f^1(\sum_{i=1}^{n} X_i . w_{iL1})$

Calculate the prediction output of the second layer L2 use output of L1 as input $Y = f^2(\sum_{j=1}^{n} out_{L1} . w_{jL2})$

% Where f^1 and f^2 are the activation functions for output layer and hidden layer calculated using the following expressions:

$$f^1 = \frac{1}{1 + e^{-X}} \quad \text{and} \quad f^2 = X$$

Backsword Phase: Backpropagation
Calculate the error that using the following expression $\nabla E^i = \frac{\partial E}{\partial w^i} = \sum_{l=1}^{N} \text{Ja}_l^i(y_l' - y_l)$
Where E is error function, **J** is the Hessian matrix of a function f of n variables calculated as: $J^{ij} = \sum_{l=1}^{N}\left[\frac{\partial y_l'}{\partial w^i}\frac{\partial y_l'}{\partial w^j}\right] + \lambda I_M$, I_M is the identity matrix of order M, and λ is a parameter that makes a function similar to learning in the backpropagation algorithm., w_i is i^{th} element of input layer weight, w_j is j^{th} element of output layer weight, y_l' is the derivative output of l^{th} neuron. y_l is output of l^{th} neuron.

Update the weights using the recursive algorithm, starting with the output neurons and backward until reaching the input layer, and tuning the weights in the following way:
$W(\mu) = W(\mu - 1) - J^{-1}(\mu - 1)\nabla E(\mu - 1)$ Where μ-ith learning cycle.

Repeat the training process until getting the threshold
Return the output Result.

End
  Test ← dataset.
  Present the next year prediction result.

Figure 4-8: A general procedure used MLPFFNNBP

4.7 Radial Basis Function Neural Network Model

The main idea of the RBF Networks originates from the theory of function approximation. In the previous parts, we have presented in details the Multi-Layer Perceptron MLP networks and how can learn. Radial Basis Function RBF Networks employ a different way to some extent such as they are two-layer feed-forward networks and the hidden nodes implement a set of radial basis functions (Gaussian functions), add to that the (training/learning) is very fast.

Typically, the first step in training the RBF Network is presented by the choice of the number of radial units and the parameters of each one of them. Many methods proposed such as employs the K-Means clustering (KMC) algorithm to determine the RBF center location [75]. In recent years, there is a big interest in optimizing the radial unit's parameters of Radial Basis Function Networks by employing the concept of "Evolutionary Algorithms "[76], so the principle of integrating genetic algorithms (GA) with neural networks emerged in the late 1980s.

4.7.1 Learning RBFNN with Genetic Algorithms Approach

There is a problem of neural networks in determining the prior parameters before starting any training, thus there is no fine rule so as to optimize these parameters, note that these parameters are essential and most important to decide the success of the training process [35]. In general, the learning method for (RBFNNs) are represented by optimization of its topology includes the parameters (centers, radii, and weights) and a number of neurons in hidden layer, so the possibilities of employing a (GA) to configure an RBFNNs exists and available.

In [35], a common approach presents several ways that depend on optimizing the topology of RBFNNs and parameters centers c, r using GAs, weights w calculated by means of methods of resolution of linear equations, in this algorithm the author used the singular values decomposition (SVD) to solve this system of linear equations and assign the weight w for RBFNNs to calculate the output. In our thesis we adopted the same way and approach to optimizing the topology of RBFNNs and parameters centers c, r using GAs, weights w calculated by singular values decomposition (SVD). The goal of combining RBFNNs and GAs to enhance the RBFNNs parameters (centers, radios), thus the need is to select a suitable place in the input data space for the nominated neurons.

Figure 4.9 illustrates the process that emulates the same way of prediction using RBFNN with Genetic Algorithms models.

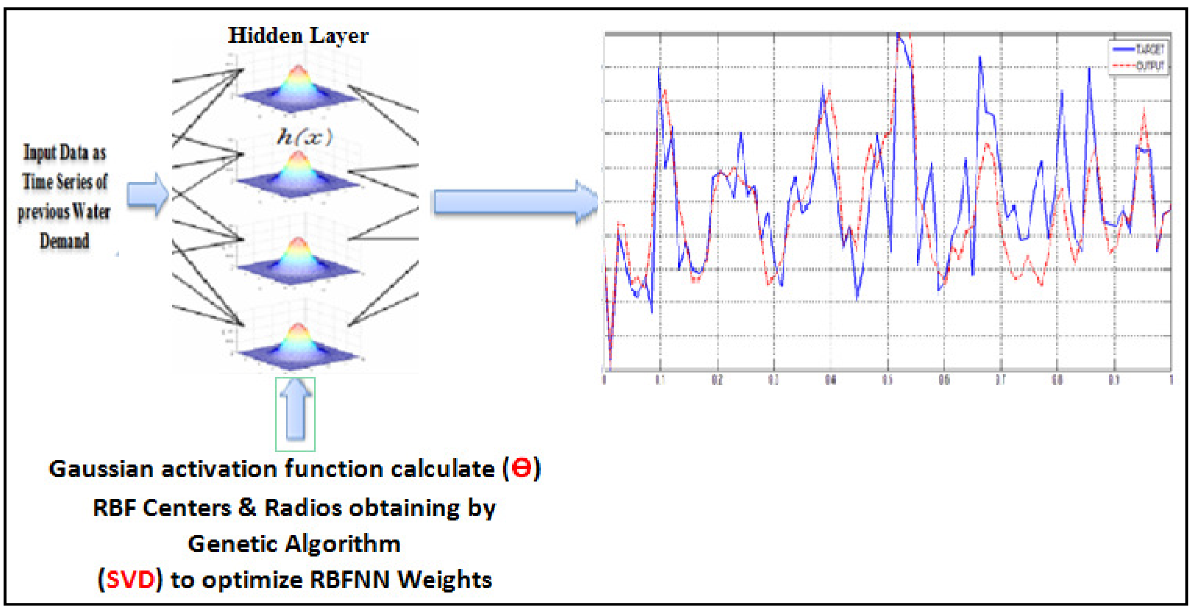

Figure 4-9: Time series prediction process using RBFNNs models

4.7.2 Genetic Algorithms (GAs)

We know that the algorithm is a set of instructions that aim to solve a problem or accomplish a certain goal. Genetic algorithm is a method of randomized research used in computing to identify the correct solution and closest to the optimal solution from set of possible solutions that constitute the so-called search space through a series of steps that rely on comparison and find the "distance" between the solutions and then the selection of Appropriate solutions and dependent on the formation of other solutions more appropriate and closer to the optimal solution. GeneticAlgorithms (GAs) were developed by John Holland in 1975; It is based on the characterization of the problem by representing the chromosomes representing the solutions through one of coding method.

Next, a series of mathematical processes derived from biological processes such as (Crossover), (Selection) and (Mutation) are applied to a group of chromosomes representing the last generation. Hence, by selecting the best chromosome, the best solution is to find the desired solution, so the search algorithm starts with a set of solutions Randomization and ending with choosing just one of the groups of solutions. Through the integration of genetic algorithms with RBF algorithm, we aim to exploit the strength of genetic algorithms in the selection of best and appropriate parameters and to employ them in the work of RBF algorithm to carry out our prediction at the end.

4.7.3 RBFNN with Genetic Algorithm Process

The principle of the work according to this integration algorithm as follows; In the beginning, we must note that each individual of the population is composed of two real vector parts representing the center and its real random values also represent the radius, so we can visualize the individual (chromosome) contains variable lengths as this series

$[c_1|r_1, c_2|r_2\ldots, c_n\ |r_n]$ [35].

We get the initial population of (c: center, r: radios) randomly by using a number of certain individuals, note that each individual length depends on the number of RBF (number of a neuron). The Singular value decomposition (SVD) is used immediately to optimize the weights.

Now the c: center, r: radios and also the w: weight become ready, and we can find and calculation the output of RBF network, for the evaluation function, it is the tool that measures the value of the fitness in each chromosome, so we consider the mean square error between the actual output and the predicted output is the fitness function [35]. In order to stop the algorithm, we used the criterion of the upper limit of the generation. So when the genetic algorithm moves from generation to generation and before approaching the upper limit of the generation without significant changes in fitness values, we end the process. Many methods can be used to perform individual's selection, in this methodology the Geometric Ranking method was adopted, so the following equation explains it [35]:

$$p[individual1\ selection - i] = d^{+}(1-d)^{r-1} \tag{4.18}$$

$$d^{+} = \frac{d}{1-(1-d)^{s}} \tag{4.19}$$

Where (d) is the probability of selecting the best individual, (r) is the line of the individual, where 1 is the best, the (s) is the size of the population [35]. The crossover using to produces two new individuals as in the following equations:

$$\vec{x} = r\vec{x} + \ (1-r)\vec{y} \tag{4.20}$$

$$\vec{y} = (1 - r)\vec{x} + r\vec{y} \tag{4.21}$$

Where ($\overline{X}$ *and* $\overline{Y}$) are two vectors belonging to individuals (parents) of the population,(*r*) is the probability of crossover between (*0, 1*) [35]. The uniform mutation is employed to select one element randomly according to the following equation [35]:

$$\acute{X}\imath = \begin{Bmatrix} U(a\imath, \acute{b\imath}) & \imath f \quad \imath = \jmath \\ X\imath & otherw\imath se \end{Bmatrix} \tag{4.22}$$

Where a_i and b_i are down and top level for every variable *i*. Every step we increase the number of [RBF] and continue to repeat the process until the number of iteration less than generation number or less than the threshold (error- rate).

4.7.4 Singular Value Decomposition (SVD)

Reducing a huge dimensional data set from a set of data points to a moderate dimensional space that which presents the base of the original data in such a way as to allow easy and accurate perception or interpretation and to order it from maximum variation to the least variation is an optimal methodology. The singular value decomposition (SVD) according to a linear algebra is a factorization of real data or complex, square or no square matrix. So, assume matrix A with m rows and n columns with rank *r*, then A can be factorized into three matrices [77]:

$$A = U\Sigma V^T \tag{4.23}$$

Where:

- *U* is an m x m orthogonal matrix.
- V^T is the conjugate transpose of an $n*n$ orthogonal matrix.
- $\sum$ is the $m*n$ diagonal matrix with nonnegative real numbers on the diagonal which are known as the singular values of *A*.

- The m columns of U and n columns of V called the left-singular and right-singular vectors of U and V respectively.
- The singular values of Σ are arranged as $\sigma 1 \geq \sigma 2 \ldots \geq \sigma r \geq 0$, where the largest singular values precede the smallest and they appear on the main diagonal of Σ.
- The numbers $\sigma_1^2 \geq \sigma_2^2 \geq \sigma_r^2$ are the Eigenvalues of AA^T and A^TA.

The radial basis function, we can use the Singular value decomposition algorithm in order to optimize the weights matrix of the output layer.

SVD style can calculate the weights by the optimal way and the following equation to express that.

$$\vec{Z} = \vec{w}\Phi \tag{4.24}$$

Where Z is the RBFNN output, w is the weighted vector; Φ is the matrix of the Gaussian activation function.

By using the two following equations the singular value decomposition is a strong option to find more solutions.

$$A = U\Sigma\left(\frac{1}{\sigma}\right)V^T \tag{4.25}$$

$$\vec{W} = \left[U\Sigma\left(\frac{1}{\sigma}\right)V^T\right]\vec{Z} \tag{4.26}$$

- Input: Dataset consist of time series of 84 points
- Output: prediction the future water demand result
- Data processing:
  - Normalize (input time series, target)
  - Initialize the weight W of RBF-nn using SVD
  - Determine input pattern X_t: ($x_{t1}, x_{t2}, \ldots, x_{tn}$)
  - Determine target output from collected data
- Get initial population of (c: center, r: radios) randomly by using # of certain individuals
- Each individual length depends on the number of RBF [c1|r1, c2|r2... c_n |r $_n$]
- Calculate the output of (RBF-nn) using

$$f(\vec{x}, \phi, w) = \sum_{i=1}^{m} \phi i(\vec{x}).w_i$$

- Calculate the MSE of (RBF-nn) using

$$\mathbf{MSE} = \sum (Target - Forecast)^2 \,/n$$

- Apply selection using Geometric Ranking method using

$$p[individual1\ selection - i] = d^{+}(1-d)^{r-1} \quad , \quad d^{+} = \frac{d}{1-(1-d)^{s}}$$

Where (d) is the probability of selecting the best individual, (r) is the line of the individual, where 1 is the best, The (s) is the size of the population

- Apply crossover to produces two new individuals using

$$\vec{x} = r\vec{x} + (1-r)\vec{y}$$

$$\vec{y} = (1-r)\vec{x} + r\vec{y}$$

Where ($\overline{X}$ and $\overline{Y}$) are two vectors belonging to individuals (parents) of the population, (r) is the probability Of crossover between (0, 1)

- Apply mutation to select one element randomly

$$\acute{X}\, i = \begin{cases} U(ai, \acute{b}i) & if\ i = j \\ Xi & otherwise \end{cases}$$

- Increase the number of [RBF]
- Repeat
- Stop if the number of iteration less than generation number or less than the threshold

Figure 4-10: A general procedure used RBFNN-GAs

4.8 Clustering of Regions Depending on Water Demand and Consumption

Based on the preliminary results of population, housing and establishment's census 2017, the population of the Jenin city and its camp is about 60325 people distributed in many different neighborhoods within the borders of the municipality [80], these neighborhoods vary in degree of overcrowding with respect to the number of buildings, population and monthly water consumption extracted from municipal databases. During the period of preparation of this thesis, engineers from the Municipality of Jenin cooperated to help clarify and explain these neighborhoods and their details regarding the demarcation of their boundaries and locations on the city map based on a logical and real distribution of the supply zones. According to the map in Figure 4.11, we note that the city of Jenin is divided into 39 neighborhoods where the municipality of the city is responsible for providing and pumping the required quantities of water. These 39 neighborhoods, as shown in the previous figure, consume about 1160927 cubic meters according to the statistics of 2017. The quantities of water supplied to these sectors are one day a week and the equivalent of four days per month for each sector, regardless of the area of each neighborhood, regardless of the number of subscribers and meters active, and therefore every neighborhood, whether the size of users little or large, he has four days' supply per month.

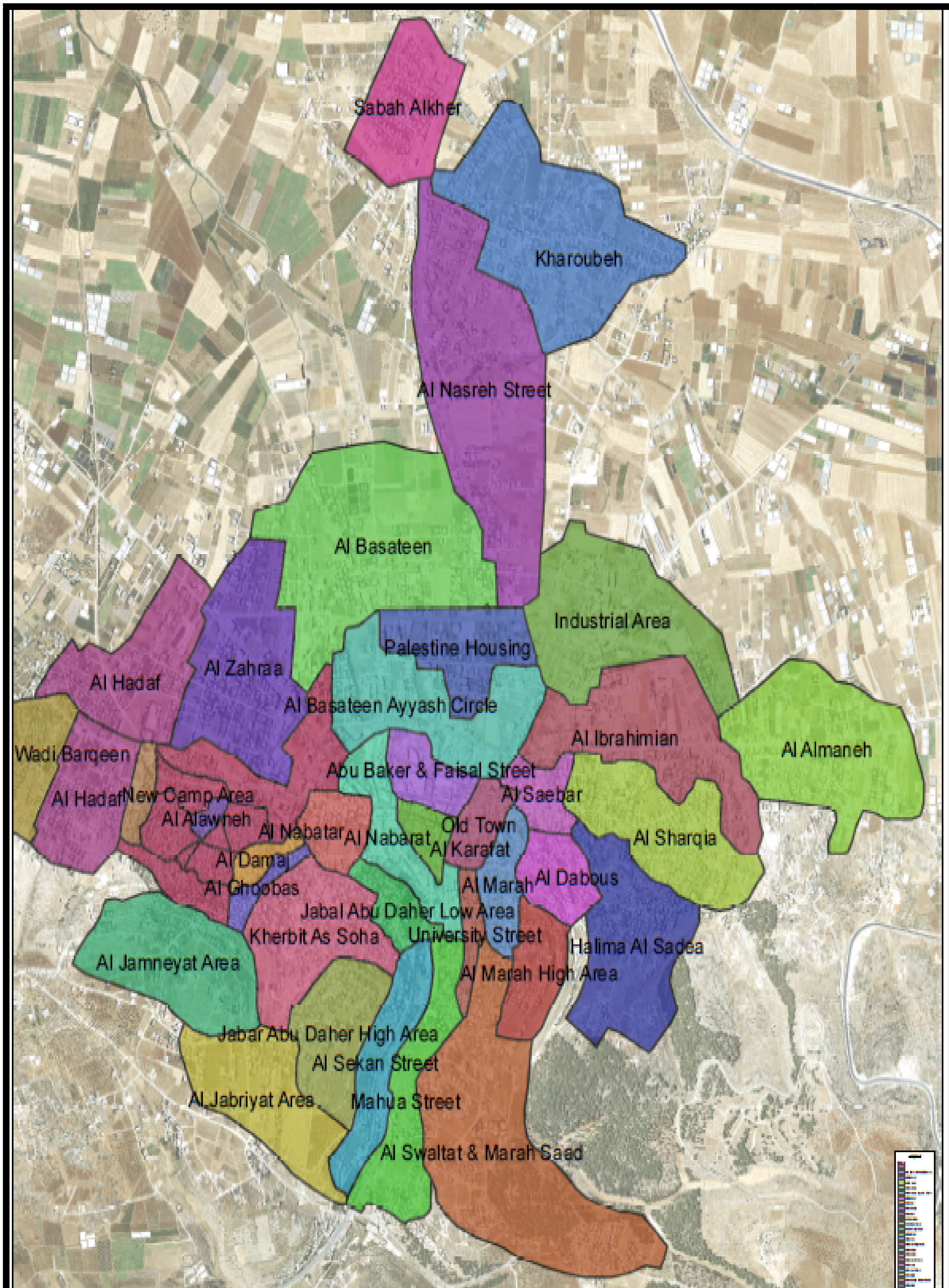

Figure 4-11: Jenin city supply zones

In light of the method adopted in the water supply, we find that it is unfair and lacks an optimal distribution, Therefore, there are small areas that provided with a number of days of pumping water more than the real needs and there are large areas that get a number of days of pumping water far below the need according to the numbers and quantities of water consumed and located in our hands for all regions. In figure 4.12 we can note the distribution of quantities of water demand for the regions and how much it differs between them and however they get an equal number of supply days.

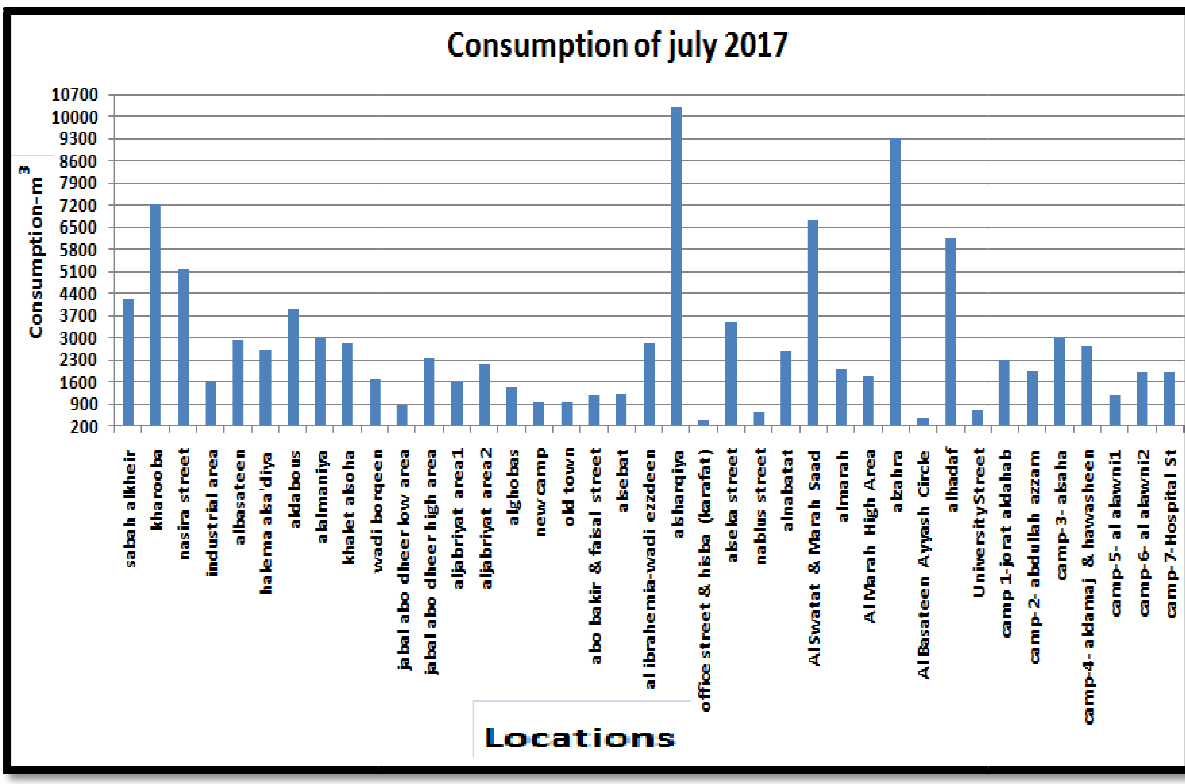

Figure 4-12: distribution of quantities of water demand

In fact, we face a real problem and we seek effective and fair solutions in terms of equitable and optimal distribution, which achieves higher levels of consumer satisfaction than the traditional method used in Jenin city.

4.8.1 Data Clustering with K-Means Clustering Algorithm

Data clustering is the operation of putting data points in similar clusters; it is a section of data mining. The clustering algorithm separates a data set into varied clusters since the similarity between points within a particular cluster is greater than the similarity between two points within two different clusters [81]. The idea, in general, is simple in it is natural and very close to human in its way of thinking, so whenever dealing with a large amount of data we tend to summarize the vast amount of data into a few groups or categories in order to facilitate the analysis.

Clustering algorithms are widely used in data classifications, data compression, and data model construction, since if we can find clusters of data, a model of the problem can be built on the basis of those clusters. There are various algorithms applied in the data clustering process, and we will examine the simplest algorithm called the K-means clustering algorithm. K-means clustering is a kind or class of unsupervised learning, which is applied with unorganized or unlabeled data with defined categories or groups [82]. The purpose of this algorithm is to find groups or clusters of data, with the number of sets, declare by the variable *K*. The algorithm runs frequently to allocate each data node to one of the *K* sets based on characteristic similarity. In other words, this algorithm is used to collect several data points depending on their properties to the K clusters, and the clustering process is done by reducing the distances between the data and the cluster center [83]. The general steps of the K-means clustering algorithm are shown in figure 4.13.

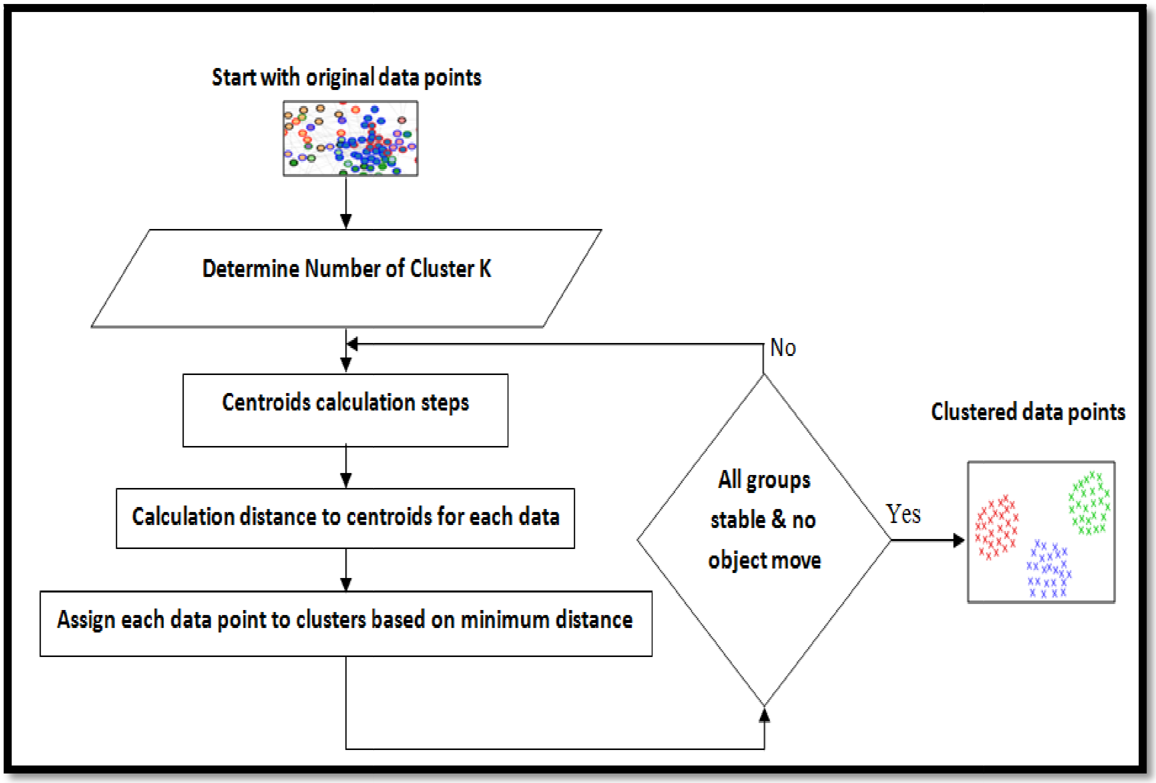

Figure 4-13: General steps of K-means clustering algorithm

The performance and effectiveness of this algorithm depend on the initial positions of the clusters centers, and it is recommended that this algorithm is run several times with different positions each time from previous times. So the (K-means) aims to minimize total cluster variance or the squared error function and this given by:

$$J(c) = \sum_{i=1}^{k} \sum_{j=1}^{n} (\| Xj - Ci \|)^2 \qquad (4.27)$$

Where k is the number of clusters, n is the number of data points in the i^{th} cluster, X_j: data point, C_i: centroid for cluster I, $\|X_j - C_i\|$: is the Euclidean distance between X_j and C_i. The general process for K-means clustering is illustrated as shown in figure 4.14.

Model 1: The general procedure that was used for k-means clustering Algorithmic

Input : data points (X_i) into 'k' sets, number of 'k' (cluster) are formerly determined.

Step -1: Pick out 'k' points at randomly as cluster centers.

Step -2: Allocate data points or nodes (X_i) to their closest and nearest cluster center according to the Euclidean distance function.

Step -3: Recalculate the centroid or new cluster center (position) of all objects in each cluster by using, $Ci = (1/ni)\sum_{j=1}^{n} Xi$, 'ni' represents the number of data points in i^{th} cluster.

Step -4: Recalculate the distance between each data point and new obtained cluster centers as the same steps in (2 and 3) until no data point reassigned and all data points in clusters are stable.

Output: data point with cluster memberships

Figure 4-14: General procedure of k-means clustering Model

K-Means is comparatively an efficient method, therefore we need to designate the number of clusters, and hence the final results often end at a local optimum. Unluckily there is no uniform notional method to find the best number of clusters, but the alternative approach is to compare the results of many rounds with various 'k' and adopt the best one according to a predefined condition. In the next chapter, we present the practical and scientific results produced by all the models proposed in this thesis. There will be a presentation of the results by graphical methods, charts, and tables which will be closer to understanding and transfer the idea of the practical experiments of each model.

CHAPTER 5
EXPERIMENTAL RESULT

5 Experimental Results

In this chapter, the procedural steps in building ARIMA and ANN model was applied in order to create a new ARIMA, ANN model capable to make an accurate prediction for water demand using the input variables which we have reviewed and mentioned in the previous Chapters.

5.1 Experimental Procedure

In all experiments procedures and applied models which we have designed, we tested them by employing the MATLAB R2013a and R-studio (R language) under Windows 7 with Core i3-M 380 CPU 2.53GHz, 4GB RAM memory. MATLAB is a high-performance programming platform designed to perform calculations, visualization and programming in an environment easy to use where problems and solutions are expressed in a mathematical notation. MATLAB is an interactive system whose basic element is an array that does not require dimensioning. This allows solving many computational problems, specifically those that involve vectors and matrices, in a much shorter time. R is an environment and programming language with a focus on statistical analysis. It is one of the languages most used in research by the statistical community, being also very popular in the field of data mining, biomedical research, bioinformatics, and financial mathematics. [79].

We intend to offer the outcomes that were produced in the practical experiments of our predictions by providing numerical and graphical results according to the monthly prediction based on all models that we developed and constructed. After that, we intend also to offer the comparison and the results of all models and experiments between different neural networks and ARIMA model. We provide the monthly prediction results for Multi-

Layer Perceptron Feed Forward Network with Back Propagation Algorithm, Radial Basis Function Neural Network Models with Genetic algorithms to optimize the centers and ARIMA Time Series Predictor using Box-Jenkins Methodology. Also, we provide the new way in order to clustering all regions of Jenin city so as to give a fair and optimal distribution water supply according to K-means clustering algorithm results. Table 5.1and 5.2 show all applied models used in our practical experiences.

Table 5-1: Water demand prediction applied models

No	Models	Prediction Results
1	ARIMA- Box-Jenkins Methodology	Monthly
2	MLPFFNN with Back Propagation Algorithm	Monthly
3	RBF with Genetic Algorithm	Monthly

The K-means clustering Algorithm used to organize the water distribution dates, regarding the consumption in each region.

Table 5-2: Applied model - K means algorithm

4	K-means clustering Algorithm	Monthly

5.1.1 Monthly Prediction (ARIMA- Box-Jenkins) Methodology

To prediction the water demand for the future according to Box-Jenkins Methodology, we must pass through many stages, so at the beginning we convert the actual consumption values over seven years starting from (Jan-2011) to (Dec-2017) to normalized data format, so that all values will take place between the value of [0,1]. We obtained 84 value represent 84 months to prediction a new 12 months that represent the year of 2018. After we explore the major features of series such that captured the trend both (increasing/decreasing), also the series is (stationary-nonstationary), add to that determined seasonality for the time series according to used graphical procedures (plotting the series, ACF/ autocorrelation function and PACF/ Partial autocorrelation function). So we made a time series is stationary and it is “stable”, hence we removed the trend and seasonality in order to move on to the next stage, figure 5.1 depicts this procedure.

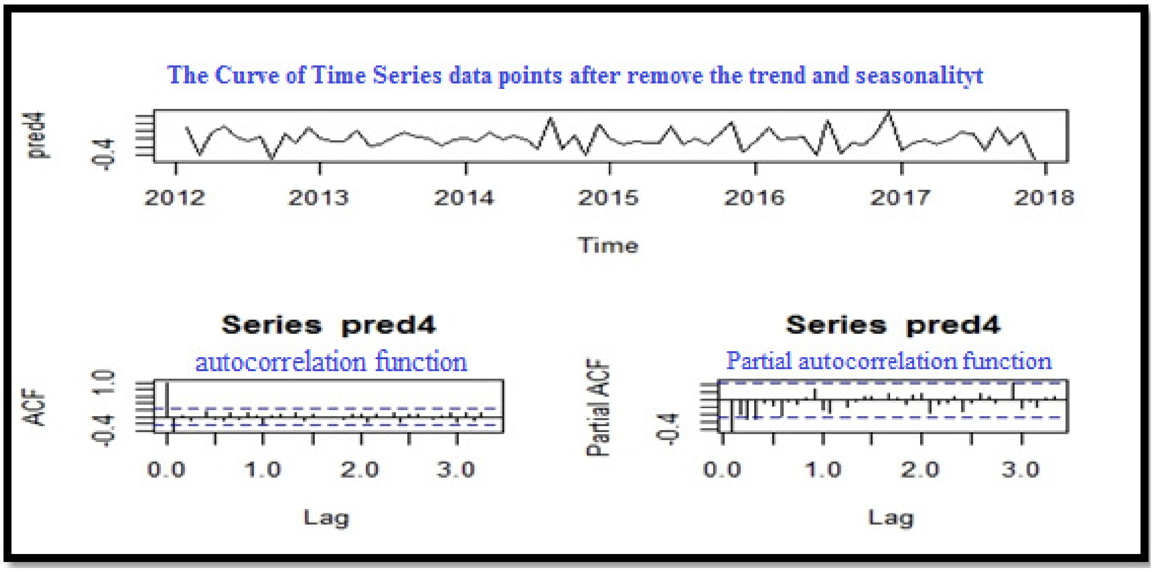

Figure 5-1: graphical procedures for stable time Series after removing the trend and seasonality

According to the ACF plots we tried to determine the order of the moving average (MA) (q) model, and according to the PACF plots we tried to determine the order of the Autoregressive (AR) (p) model. To estimate and diagnostic checking stage we have determined and selected the best model including the estimation of the parameters of the several 10 models using the previous stage and produce the first best selection of models based on the Akaike Information Criterion AIC and the Bayesian Information Criterion BIC. Since we repeated and tested more than one model, we were careful about three issues; the first is *(P value for Ljung-box statistic)* and we sure that all values of (P-value) is higher than 0.05 and therefore we don’t reject the initial hypothesis that indicating that the model meets our purpose. The second issue is ACF of residual; we sure about the violation not exist, so our adopted model stay meets our need. The final thing is standardized residual and we also sure that randomization does not indicate to the pattern, figure 5.2 depict that.

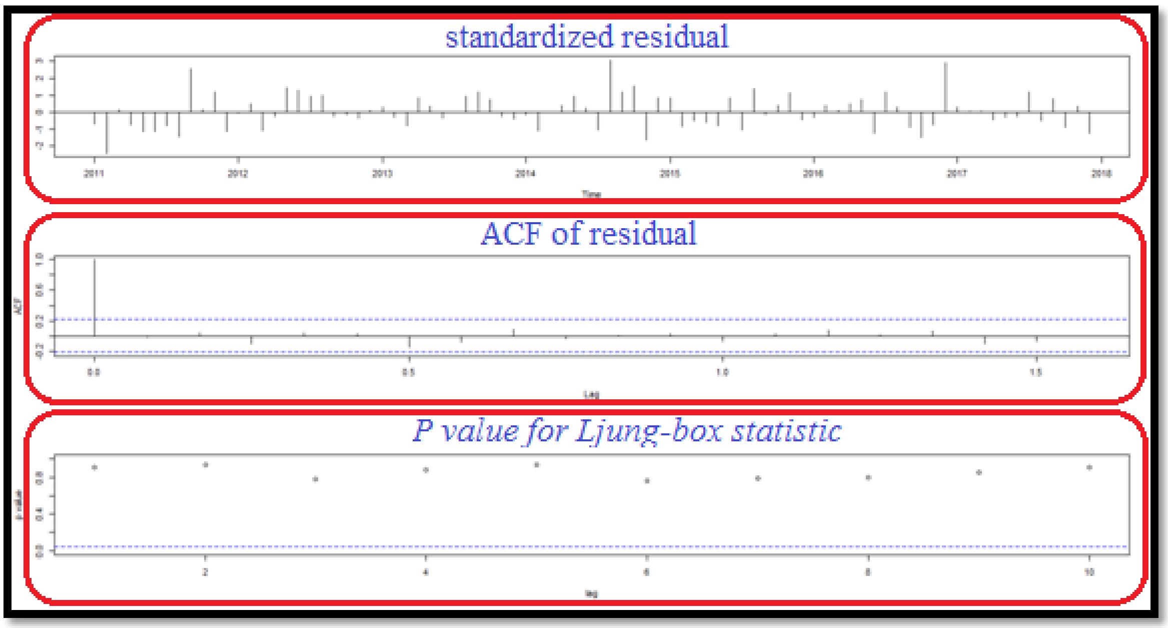

Figure 5-2: Estimate and diagnostic checking stage

Cross-validation model is carried out by partition the data into two subsets; the training set will use the majority of the data. The larger amount of data is used to set up a model in training. For our dataset is divided into two sets, 70% for the training set and 30% for testing set. So figure 5.3 depict this.

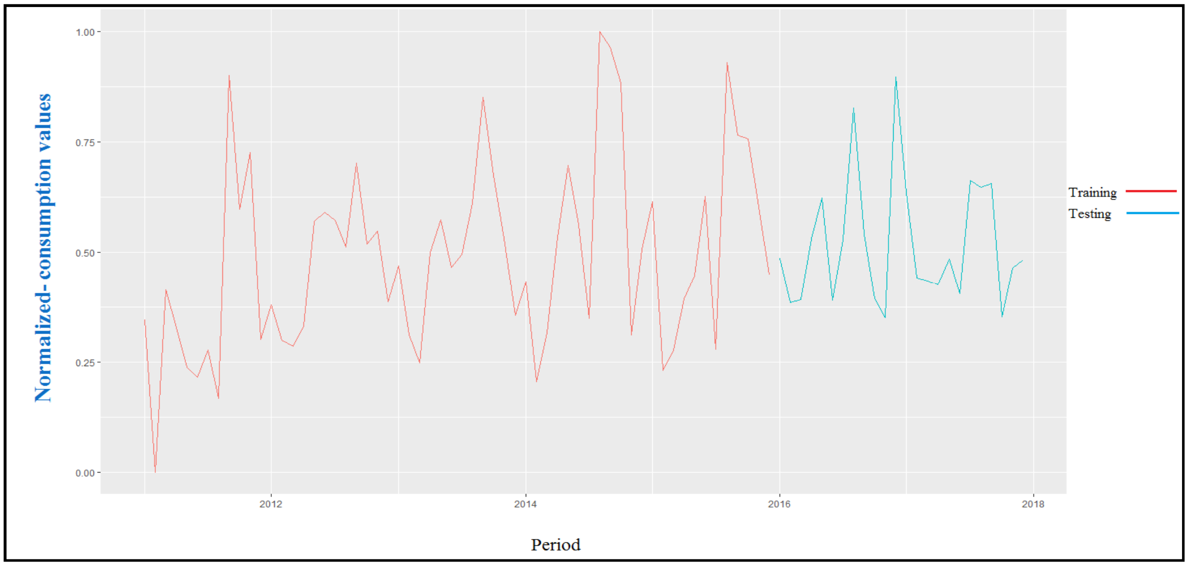

Figure 5-3: ARIMA Cross-validation model

Based on all the above stages, we carried out the prediction, this step is applied to produce the prediction according to the model that we are decided is the best one. Here we intend to generate a new 12 months that represent adding one year (2018) and then present the results that the ARIMA model returned to be compared later with the rest of the methodologies used in this thesis, so figure 5.4 illustrate the prediction process.

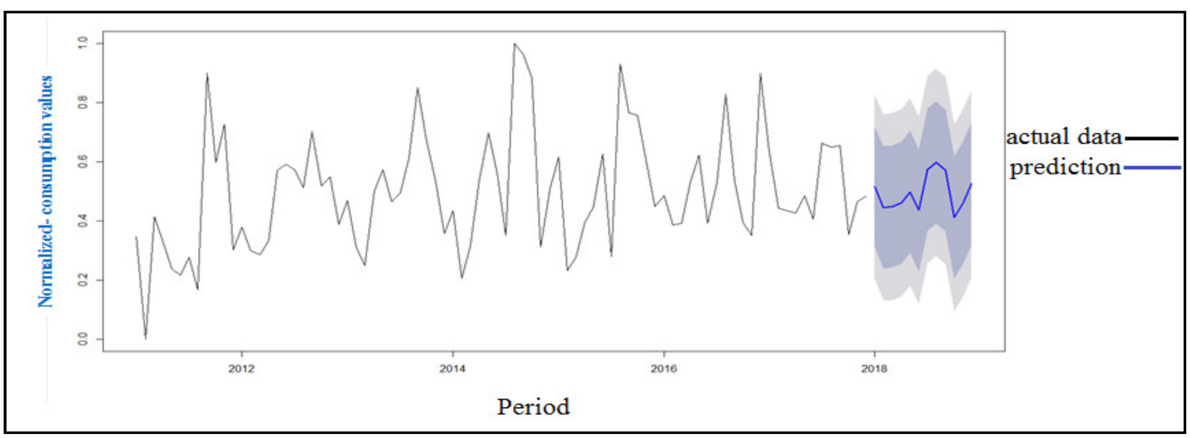

Figure 5-4: ARIMA best prediction result for 2018

According to what we have the dataset that we mentioned in previous places and other section in this thesis. As we have seen in the figure we carried out the prediction based on the R-studio, we built a special and specific model for our problem, ARIMA tried to produce the best results and do the best effort based on the error in training, testing values, and the fitting. The best root mean squared error (RMSE) was obtained is depicted in Table 5.3.

Table 5-3: best root mean squared error (RMSE)

Training set (RMSE)	Test set (RMSE)
0.1696096	0.2257397

That means that the mean squared error for training is (MSE_{train} = 0.028) and the mean squared error for testing is (MSE_{test} = 0.051), Knowing that MSE = $(RMSE)^2$. In a more graphic way, we can see the comparison produced by the ARIMA model between the actual data representing the test data that came from reality and the prediction values according to the model adopted, so figure 5.5 illustrates this.

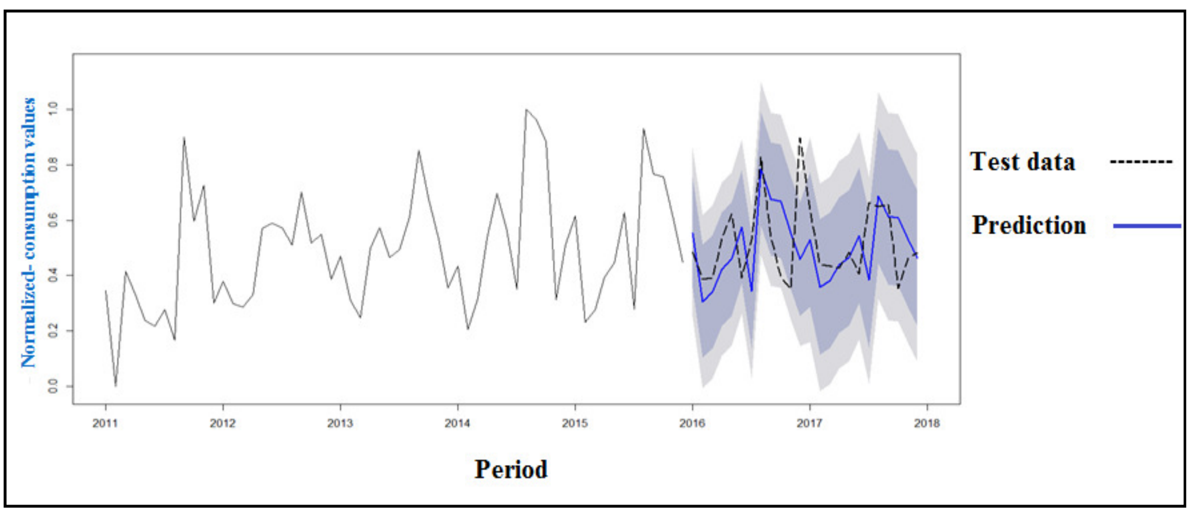

Figure 5-5: Comparison between reality and the prediction values

From the results of month's prediction as in table 5-3, and figures 5.4 and 5.5, it's clear that the proposed ARIMA model produces some good results in some extent based on the mean squared error of training and testing sets that obtained for the future prediction of water consumption demand.

5.1.2 Monthly Prediction (MLPFFNN with Back Propagation) Model

To predict the water demand in monthly style, we compute the total consumption, for instance in one month for 7500 customers so as to obtain just one value that represents the total consumption for all customers in one month in the year, hence we obtained the 84 values that represent 84 months and thus represent 7 years. We use the MATLAB R2013a to produce the predicting, the outcomes of values are given a number of neurons that represent the series of neurons used in every Multilayer Perceptron Feed Forward Back-propagation, number of Iterations that represent the number of the execution cycle of every Multilayer Perceptron Feed Forward Back-propagation, mean squared error for training (MSE_{train}), mean squared error for testing (MSE_{test}). Cross-validation model is carried out by dividing the data into two sets as we do the same thing for this Methodology, so each set is a percentage value as 70% for the training set and 30% for testing set.

Table 5.4 shows the mean square error values and iterations number of training and testing incrementally starting from (5) neurons to (80) neurons by increasing (5) neuron in every cycle.

Table 5-4: MLPFFNNBP (monthly water demand Prediction Result)

Number Of Neurons	MSE_{train}	MSE_{test}	Number OF Iteration
5	3.22535E-02	4.38334E-02	8
10	2.81714E-02	4.49053E-02	13
15	1.71894E-02	3.98097E-02	8
20	1.87715E-02	3.33202E-02	8
25	1.66460E-02	2.95369E-02	8
30	1.26527E-02	4.82637E-02	7
35	1.35805E-02	3.09674E-02	8
40	1.29510E-02	2.03809E-02	7
45	5.43525E-03	3.88319E-02	8
50	7.37251E-03	8.79243E-02	8
55	1.28419E-02	3.82644E-02	7
60	5.91304E-04	3.35725E-02	8
65	3.19636E-03	6.79492E-02	7
70	1.86360E-03	1.00880E-01	8
75	8.28180E-03	1.29360E-01	5
80	3.81429E-05	1.46403E-01	6

According to table 5.4 and figure 5.6, it is clear that the Multilayer Perceptron Feed Forward Back-propagation approach made a very good result for the prediction according to an appropriate quantity of neurons in the hidden layer, thus our prediction mean square error (MSE) on 60 neurons achieves perfect result for the future of monthly water demand with mean square error amount for training did not exceed the (0.0005) and mean square error amount for testing did not exceed the (0.033) also.

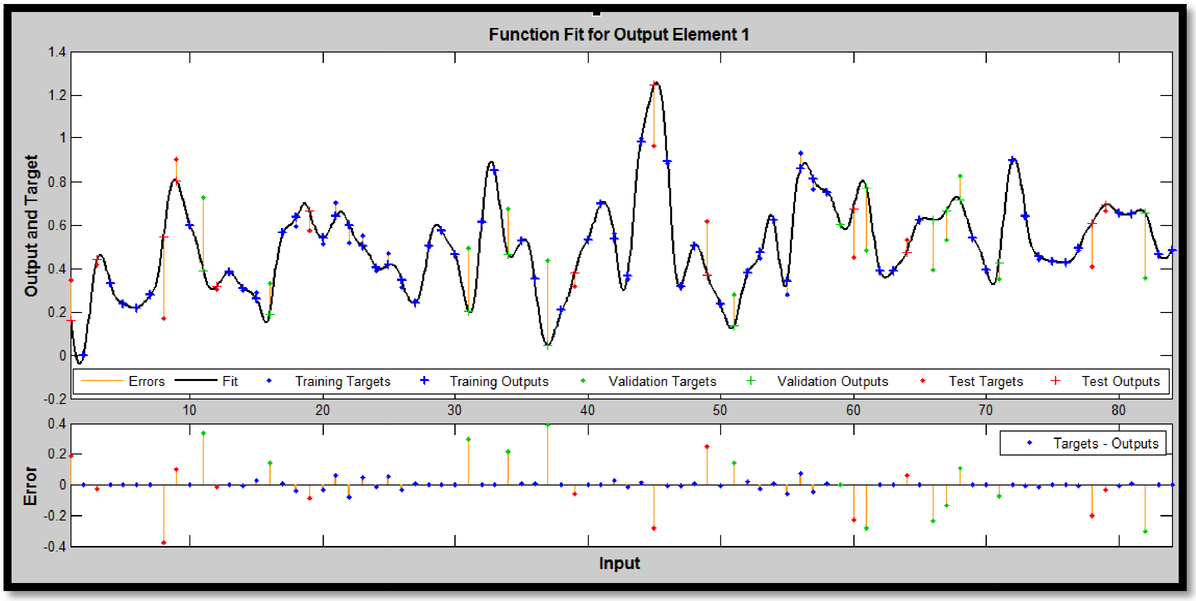

Figure 5-6: MLPFFNNBP, Best monthly water demand prediction Result

In a more graphic way, we can see the comparison produced by the MLPFFNNBP model between the actual data representing the test data that came from reality and the prediction values according to the model adopted, so figure 5.7 illustrates this.

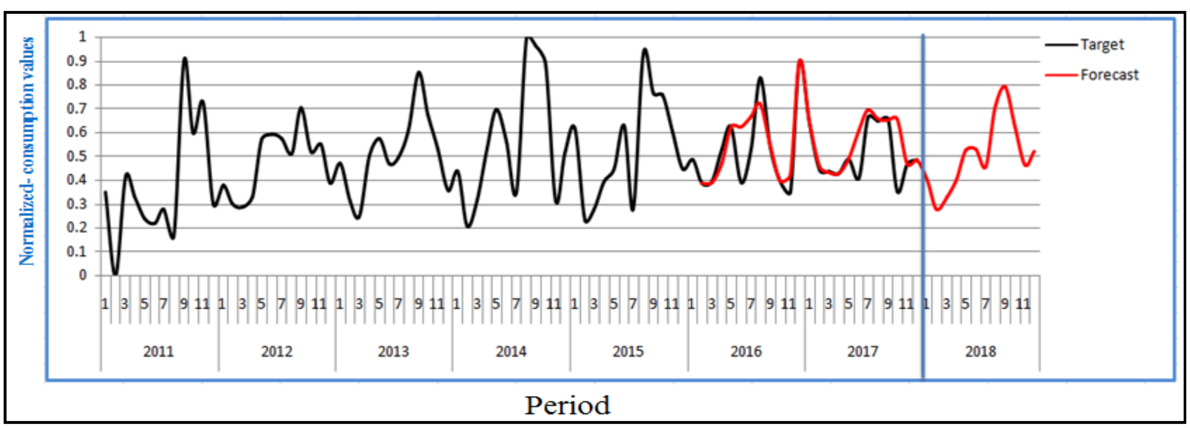

Figure 5-7: Comparison between reality and the prediction values

When the PB algorithm with LMA is used to train MLPNN it is important to let it is executed until its parameters have converged, which decrease the effect of the initial randomness of the values of the weights. As shown the figures 5.6, 5.7, the MLPNN with BP can predict the next year values with a good accuracy. From table 5.4, the values of the MSE during the test period in accepted were it's not very well, this because the pattern of the training process is not ordered during the tested data of the last 7 years. This problem in the patterns will appear in the next applied model which use RBFNN optimized with GAs.

5.1.3 Monthly Prediction RBFNN with Genetic Algorithm Model

To prediction the water demand in monthly style, we also compute the total consumption as we do the same steps in MLPFFNN with Back Propagation model. Using RBFNN with a Genetic algorithm to improve and optimize the selection of parameters centers c, radios *r*. The results of applying the proposed model for data is depicted in figure 5.8 and table 5-5 for errors values. The table 5-5 shows the mean square error values of training and testing incrementally starting from (5) neurons to (60) neurons by increasing (5) neuron in every cycle.

Table 5-5: RBF with Genetic Algorithm (monthly water demand Prediction Result)

Number Of Neurons	MSE_{train}	MSE_{test}
5	2.77E-02	1.92E-01
10	2.69E-02	1.30E-01
15	2.10E-02	1.75E-01
20	1.77E-02	1.59E-01
25	1.60E-02	2.20E-01
30	1.54E-02	2.40E-01
35	1.41E-02	2.20E-01
40	1.25E-02	2.30E-01
45	1.30E-02	2.40E-01
50	1.60E-02	2.40E-01
55	1.93E-02	2.50E-01
60	2.59E-02	2.50E-01

According to table 5-5 and figure 5.8, it is clear that the RBF with a genetic algorithm approach made a weak result for the prediction according to an appropriate quantity of neurons in the hidden layer, thus our prediction mean square error (MSE) on 60 neurons appear low result for the future of monthly water demand with mean square error amount for training about (0.025) and mean square error amount for testing about (0.250) also.

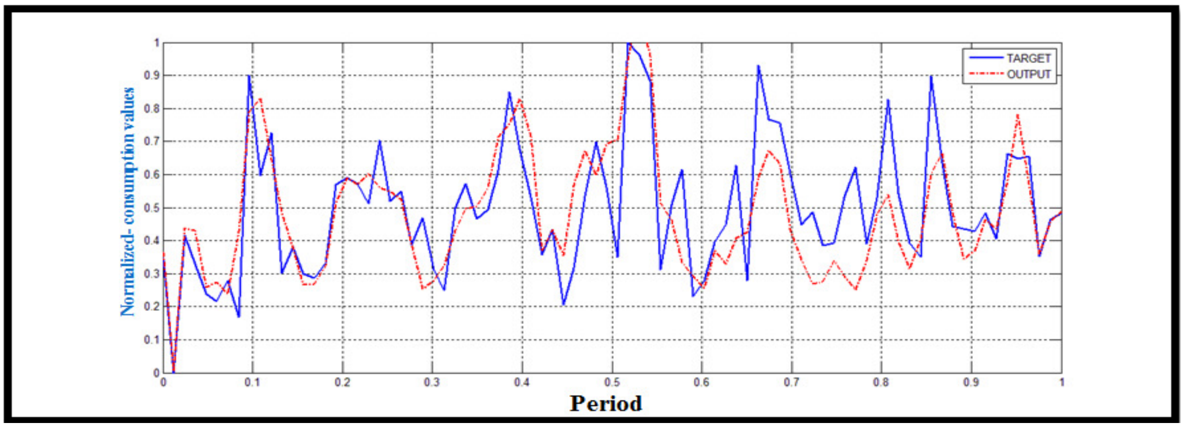

Figure 5-8: RBFNN-GAs, Best monthly water demands Prediction Result

In more graphic way, we can see the comparison produced by the RBF with Genetic Algorithm model between the actual data representing the test data that came from reality and the prediction values according to the model adopted, so figure 5.9 illustrates this.

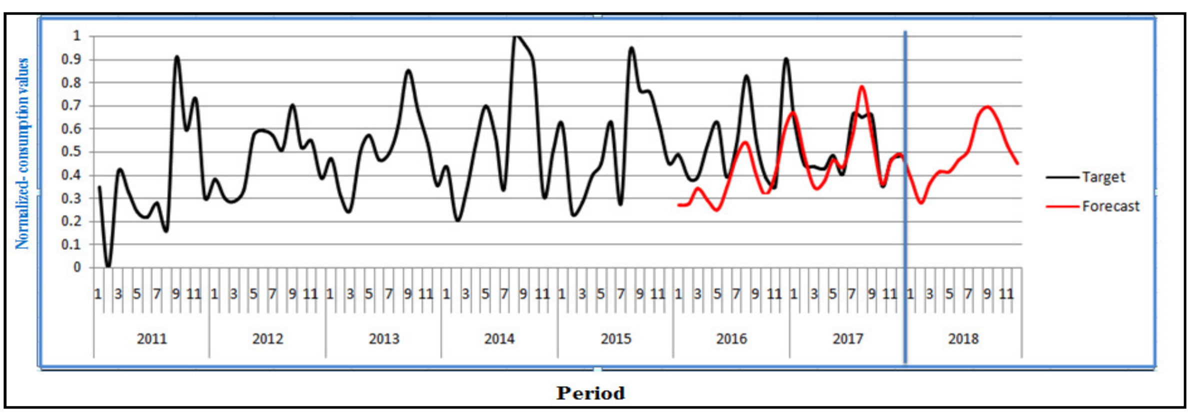

Figure 5-9: Comparison between reality and the prediction values

5.2 Comparison and Discussion

we made a comparison according to the performance from the perspective of the error value versus the number of neurons employed, in figure 5.10 the values of errors for MLPFFNNBP and RBFNN with Genetic Algorithms is represented versus a number of neurons to display the trend of error for predicting. We can see the excellence of MLP with back-propagation from the first prediction process until the end by employing the increased number of neurons. In the other hand we can see the debility for RBF with Genetic Algorithms; hence there is a clear and visible superiority in favor of MLPFFNNBP.

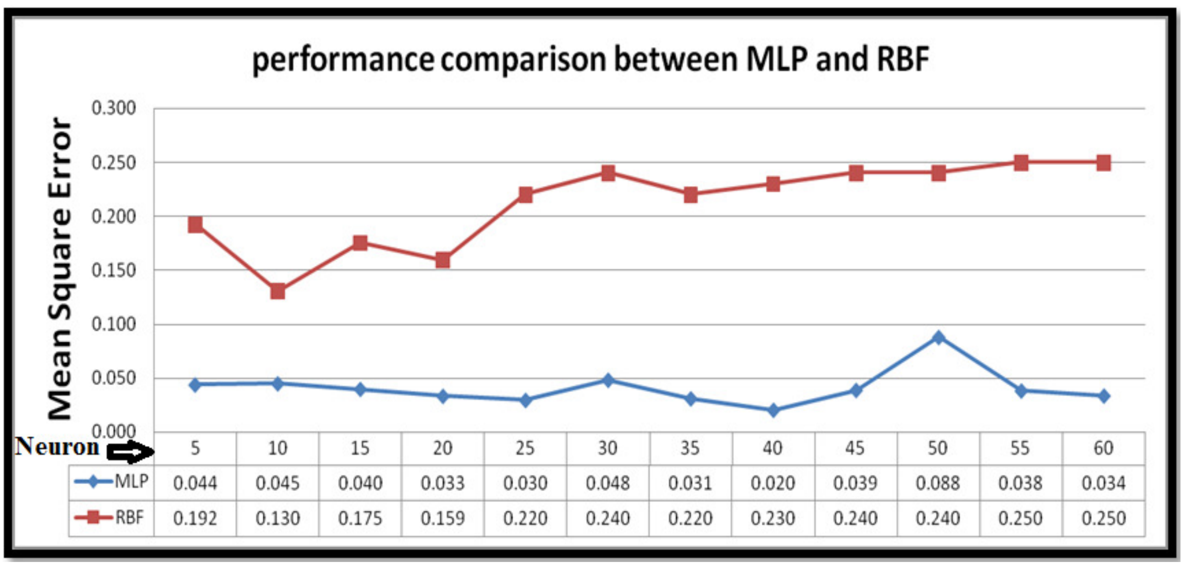

Figure 5-10: MSE Errors trend of the two neural networks applied models

We present a comparison between all results that obtained from applied models both the neural network and Stochastic Models in this thesis. The following table offers the best values of mean square error (MSE) that generated by MLPFFNNBP, RBFNN and ARIMA training and testing stage that utilized to predict water demand consumption. The selected (ANNs) models and ARIMA has been used in order to predict the municipality of Jenin short-term water consumption. Predictions have been made from the year 2011 until 2017. Table 5-6 present the results produced by the developed ANNs models add to that the ARIMA model.

Table 5-6: Mean square error (MSE) Value for Three Model Comparison

Models	Number of Neurons	MSE_{train}	MSE_{test}
MLPBP with LMA	60	0.0005	0.033
RBF with G.A	60	0.025	0.250
ARIMA	Not support it	0.028	0.051

This experimental result shows the potential of the applied neural networks models especially the MLPNN with BP in the control in real time of monthly prediction of water demand since they are met during the generalization correlations of 98% and MSE magnitudes of 2%.

According to the previous table, we can notice clearly that the prediction result based on (MSE) which generated by the MLPBP model is preeminent among the other models and it achieves the best result and accuracy, the minimum MSE obtained during the training process of the ARIMA model and RBFNN is worse than that in the MLPNN. Thus, the MLPNN produced a more accurate output than the RBFNN and ARIMA models. The testing showed that the RBF with genetic algorithm in particular generated unsatisfied value even if compared with ARIMA model and ARIMA had resulted not too far away from MLP with BP in water demand simulation but the MLPFFNNBP stayed the best and superb in results.

5.3 Clustering Regions Using K-Means Clustering Algorithm

In order to carry out an optimal and equitable distribution of water supply, we initially extracted and calculated the total consumption of each area separately during a given month by selecting individual months from 1 to 11 of 2017 for 39 districts representing all districts of Jenin City. Table 5.7 represents one of the three cases for the other months 1, 7, 11.

After all, months was extracted and calculated as the same previous table, we obtained three tables contain all districts and their consumption of water in certain months. To start our work by using the K-means clustering algorithm we already adopted a number of clusters (K=15) represent the maximum number of water supply days in a month, a number of data points (X_i=39) represent the total consumption of water for all districts in the city. We adopted the K=15 because we obtained the best results after we tried different values according to try and error, another reason we thought that this value represents adequate days and satisfy the customers in the large and overcrowded area.

Table 5-7: Total consumption water of January-2017

No	Location	Consumption of January(m^3)
1	Sabah Alkheir	3394
2	Kharooba	6368
3	NasiraStreet	3379
4	Industrial Area	2210
5	Albasateen	3305
6	HalemaAlsa'diya	2070
7	Aldabous	3255
8	Almanya	3453
9	KhaletAlsoha	2266
10	WadiBorqeen	1330
11	Jabal Abo Dheer Low Area	927
12	Jabal Abo Dheer High Area	2102
13	Aljabriyat Area1	1356
14	Aljabriyat Area2	1102
15	Alghobas	1614
16	New Camp	917

17	Old Town	748
18	Abo Bakir&Faisal Street	1313
19	Alsebat	1387
20	Al Ibrahemia-WadiEzzdeen	2380
21	Alsharqiya	8846
22	Office Street &Hisba (Karafat)	1083
23	Alseka Street	4020
24	Nablus Street	721
25	Alnabatat	2229
26	Al Swatat& Marah Saad	6624
27	Almirah	1966
28	Al Marah High Area	2480
29	Alzahra	10479
30	Al BasateenAyyash Circle	780
31	Alhadeff	5184
32	University Street	605
33	Camp 1-JoratAldahab	2194
34	Camp-2- AbdullahAzzam	2462
35	Camp-3- Alsaha	3026
36	Camp-4- Aldamaj&Hawasheen	3346
37	Camp-5- Al Alawni1	1130
38	Camp-6- Al Alawni2	1261
39	Camp-7-Hospital St	1862

5.3.1 Monthly Clustering with K-Means Model

To perform our clustering process according to the monthly style we calculated the total monthly consumption per neighborhood separately until we finally get a 39 value to be input to the algorithm. Figure (5.11, 5.12, and 5.13) shows the regional distribution that represents the actual water consumption in all Jenin neighborhoods for January, July, and November in 2017, respectively, pointing to the large differences in consumption between the neighborhoods.

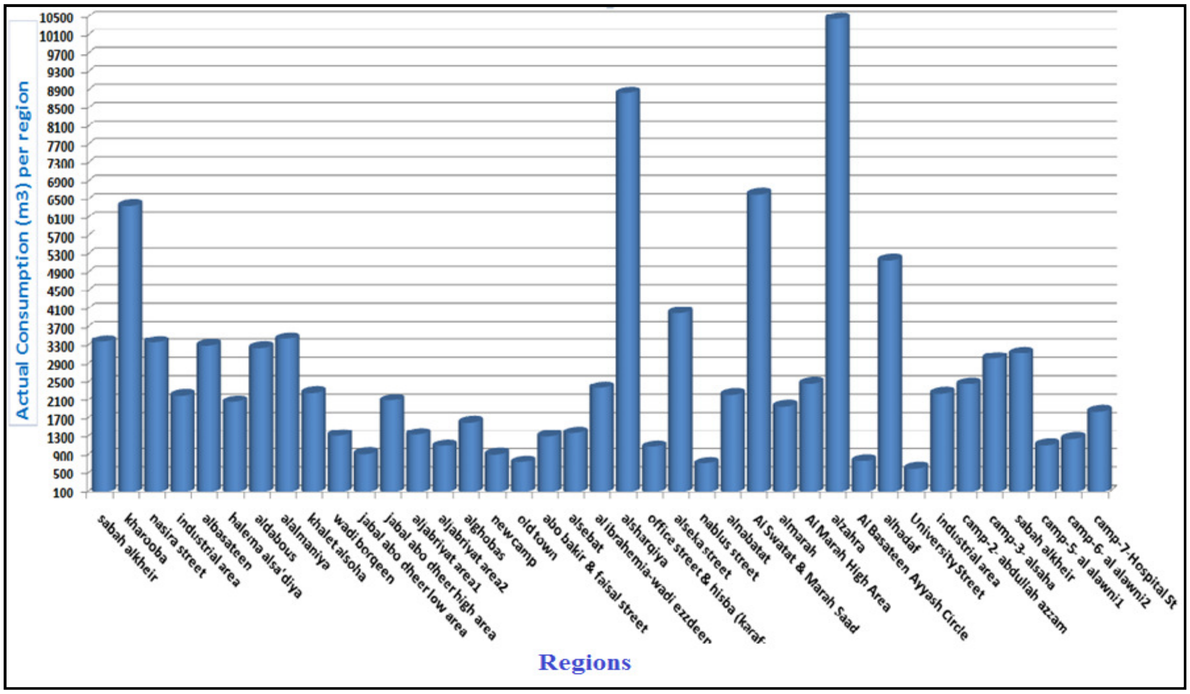

Figure 5-11: January-2017- actual water consumption according to regional distribution

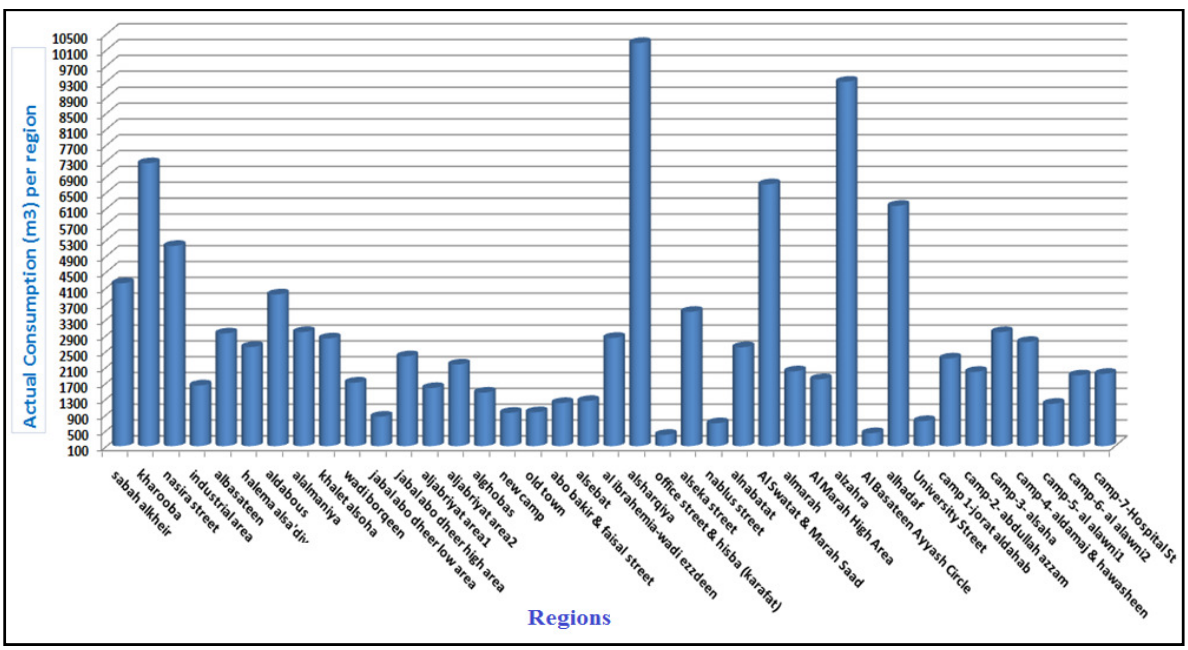

Figure 5-12: July – 2017- actual water consumption according to regional distribution

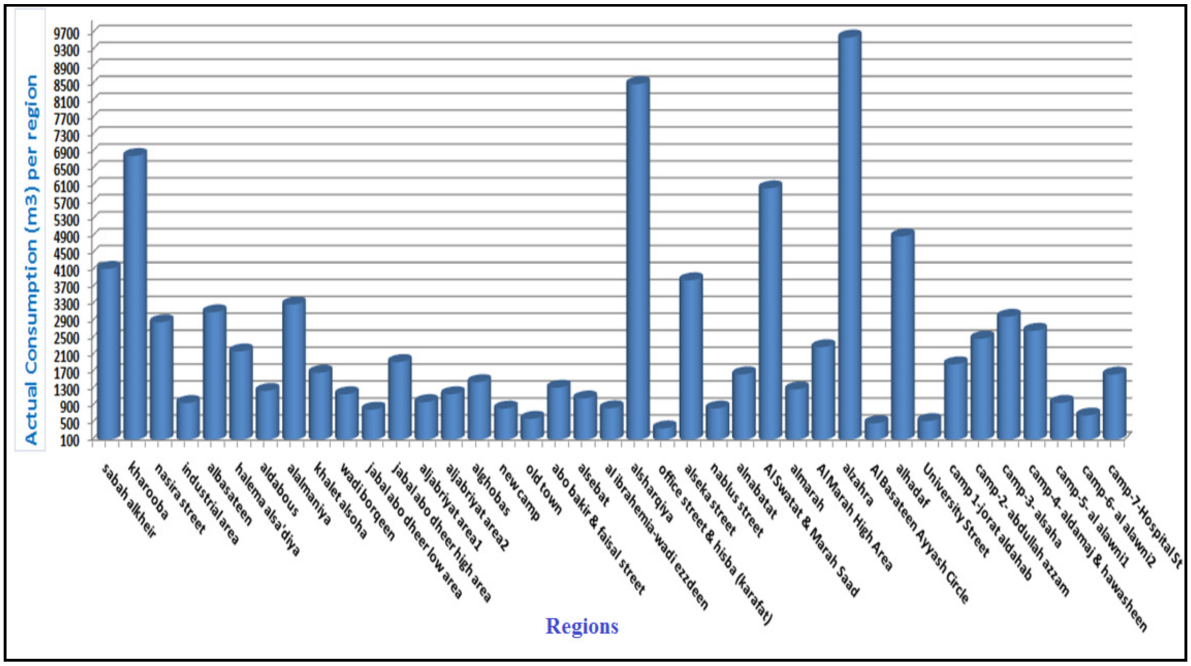

Figure 5-13: November -2017- actual water consumption according to regional distribution

We carried out our proposed way by using a 39 value as input to the algorithm to get 39 outputs to represent the index of clusters for each value. Table 5-8, 5-9 and 5-10 shows the final result was obtained according to the K-means clustering algorithm based on 39 neighborhoods for January, July, and November in 2017, respectively.

Table 5-8: K-means clustering algorithm index result for January-2017

Original Data Set			Outputs Generated by K-Means	Proposed Number of Supply
No	**Location**	**Con-Jan-M^3**	**Index / Cluster No**	**Days**
1	Sabah Alkheir	3394	2	5
2	Kharooba	6368	4	9
3	Nasira Street	3379	2	5
4	Industrial Area	2210	14	3
5	Albasateen	3305	2	5
6	HalemaAlsa'diya	2070	15	3
7	Aldabous	3255	2	5
8	Alalmaniya	3453	2	5
9	KhaletAlsoha	2266	14	3
10	WadiBorqeen	1330	5	2
11	Jabal Abo Dheer Low Area	927	11	1
12	Jabal Abo Dheer High Area	2102	15	3
13	Aljabriyat Area1	1356	5	2
14	Aljabriyat Area2	1102	11	2
15	Alghobas	1614	8	2
16	New Camp	917	11	1
17	Old Town	748	13	1
18	Abo Bakir& Faisal Street	1313	5	2
19	Alsebat	1387	5	2
20	Al Ibrahemia-WadiEzzdeen	2380	6	4
21	Alsharqiya	8846	3	13
22	Office Street &Hisba (Kar	1083	11	2
23	Alseka Street	4020	9	6
24	Nablus Street	721	13	1
25	Alnabatat	2229	14	3
26	Al Swatat& Marah Saad	6624	4	10
27	Almarah	1966	1	3
28	Al Marah High Area	2480	6	4
29	Alzahra	10479	3	15
30	Al BasateenAyyash Circle	780	13	1
31	Alhadaf	5184	7	7
32	University Street	605	12	1
33	Camp 1-Jorat Aldahab	2254	14	3
34	Camp-2- Abdullah Azzam	2462	6	4
35	Camp-3- Alsaha	3026	2	4
36	Camp-4- Aldamaj&Hawa	3135.2	2	5
37	Camp-5- Al Alawni1	1130	11	2
38	Camp-6- Al Alawni2	1261	10	2
39	Camp-7-Hospital St	1862	1	3

Table 5-9: K-means clustering algorithm index result for July -2017

Original Data Set			Outputs Generated by K-Means	Proposed Number of Supply
No	**Location**	**Con-Jul-M^3**	**Index / Cluster No**	**Days**
1	Sabah Alkheir	4208	8	6
2	Kharooba	7236	4	11
3	Nasira Street	5149	13	7
4	Industrial Area	1630	14	2
5	Albasateen	2942	7	4
6	HalemaAlsa'diya	2600	9	4
7	Aldabous	3926	8	6
8	Alalmaniya	2976	12	4
9	KhaletAlsoha	2817	9	4
10	WadiBorqeen	1701	14	2
11	Jabal Abo Dheer Low Area	842	11	1
12	Jabal Abo Dheer High Area	2366	15	3
13	Aljabriyat Area1	1564	14	2
14	Aljabriyat Area2	2160	15	3
15	Alghobas	1449	10	2
16	New Camp	938	5	1
17	Old Town	953	5	1
18	Abo Bakir& Faisal Street	1186	5	2
19	Alsebat	1237	5	2
20	Al Ibrahemia-WadiEzzdeen	2829	9	4
21	Alsharqiya	10271	6	15
22	Office Street &Hisba (Karaf	383	11	1
23	Alseka Street	3487	3	5
24	Nablus Street	675	11	1
25	Alnabatat	2592	9	4
26	Al Swatat& Marah Saad	6704	4	10
27	Almarah	1985	2	3
28	Al Marah High Area	1786	1	3
29	Alzahra	9294	6	14
30	Al BasateenAyyash Circle	425	11	1
31	Alhadaf	6162	4	9
32	University Street	730	11	1
33	Camp 1-JoratAldahab	2311	15	3
34	Camp-2- AbdullahAzzam	1971	2	3
35	Camp-3- Alsaha	2975	12	4
36	Camp-4- Aldamaj& Haw	2726	9	4
37	Camp-5- Al Alawni1	1166	5	2
38	Camp-6- Al Alawni2	1883	2	3
39	Camp-7-Hospital St	1922	2	3

Table 5-10: K-means clustering algorithm index result for November -2017

Original Data Set			Outputs Generated by K-Means	Proposed Number of Supply
No	Location	Con-Nov-M^3	Index / Cluster No	Days
1	Sabah Alkheir	4138	8	6
2	Kharooba	6806	5	11
3	NasiraStreet	2877	12	4
4	Industrial Area	970	7	1
5	Albasateen	3113	2	5
6	HalemaAlsa'diya	2187	4	3
7	Aldabous	1254	13	2
8	Alalmaniya	3296	2	5
9	KhaletAlsoha	1683	6	3
10	WadiBorqeen	1177	10	2
11	Jabal Abo Dheer Low Area	813	14	1
12	Jabal Abo Dheer High Area	1938	4	3
13	Aljabriyat Area1	991	7	1
14	Aljabriyat Area2	1178	10	2
15	Alghobas	1463	9	2
16	New Camp	838	14	1
17	Old Town	595	3	1
18	Abo Bakir& Faisal Street	1331	13	2
19	Alsebat	1076	11	2
20	Al Ibrahemia-WadiEzzdeen	845	14	1
21	Alsharqiya	8501	1	13
22	Office Street &Hisba (Karafat)	370	15	1
23	Alseka Street	3869	8	6
24	Nablus Street	839	14	1
25	Alnabatat	1645	6	3
26	Al Swatat& Marah Saad	6042	5	9
27	Almarah	1291	13	2
28	Al Marah High Area	2288	4	4
29	Alzahra	9603	1	15
30	Al BasateenAyyash Circle	493	3	1
31	Alhadaf	4904	8	8
32	University Street	543	3	1
33	Camp 1-JuratDahab	1884	6	3
34	Camp-2- AbdullahAzzam	2488	12	4
35	Camp-3- Alsaha	3006	2	5
36	Camp-4- Aldamaj&Hawasheen	2675	12	4
37	Camp-5- Al Alawni1	972	7	1
38	Camp-6- Al Alawni2	675	3	1
39	Camp-7-Hospital St	1641	6	2

Graphically we illustrate the clustering process more precisely and closer to understanding, so the following figures 5.14, 5.15 and 5.16 also show all clusters based on 39 neighborhoods for January, July, and November in 2017, respectively

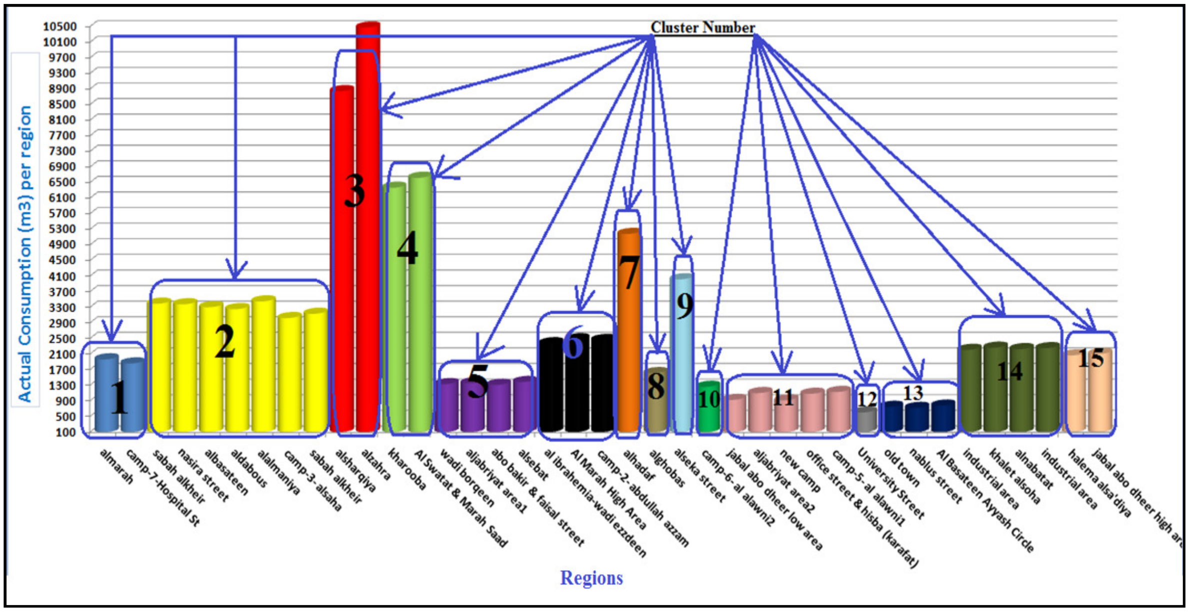

Figure 5-14: Distribution clusters for all neighborhoods January-2017

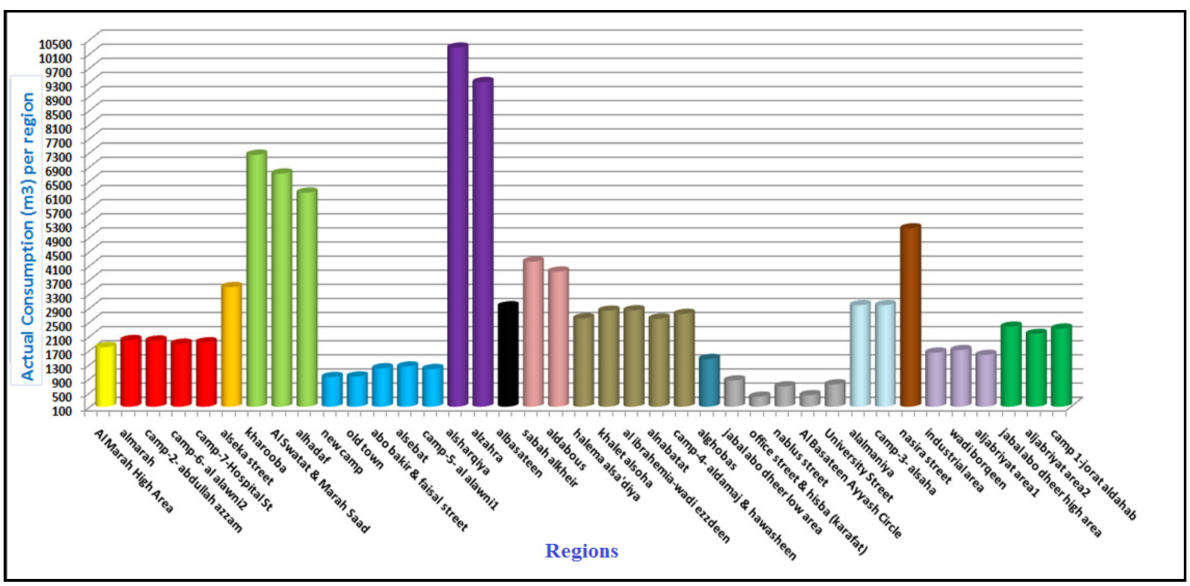

Figure 5-15: distribution clusters for all neighborhoods July-2017

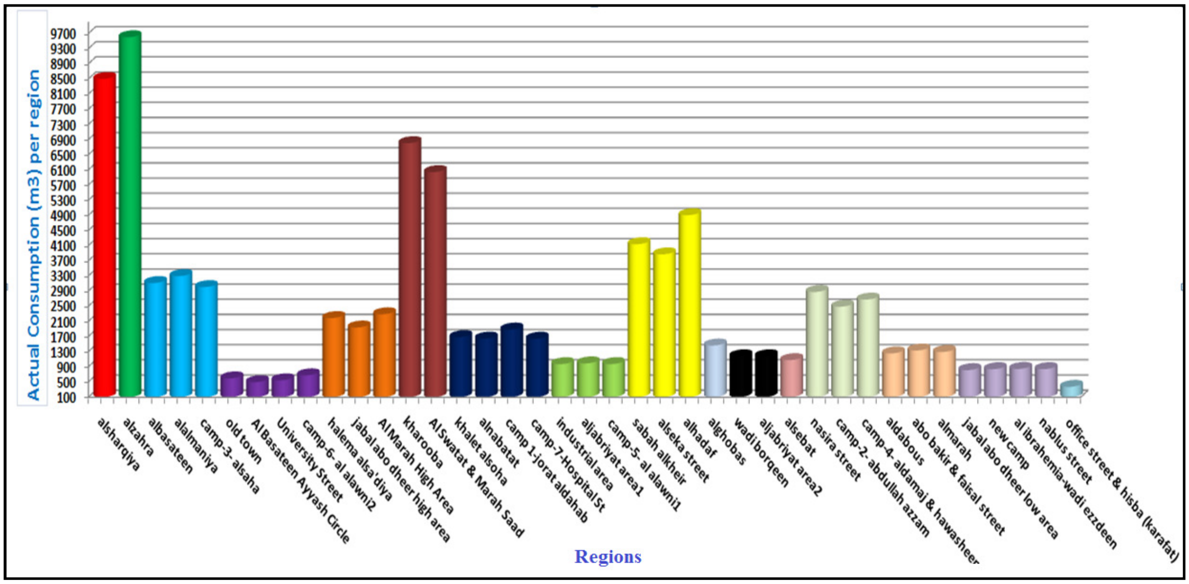

Figure 5-16: distribution clusters for all neighborhoods November -2017

According to table (5-8, 5-9 and 5-10) and figures (5.14, 5.15 and 5.16), it is clear that the K-means clustering algorithm performed an excellent result for aggregate the neighborhoods in highly accurate form according to an appropriate quantity of clusters (K=15). There is no any irregular data point or thumping inside the clusters. These excellent clusters distribution has helped us to understand the quantities of consumption and know of the characteristics of different neighborhoods and also represented the base stone of the proposal to distribute a fair and ideal to end the situation and the method of unfair distribution. The Proposed numbers of supply days represent how much days this neighborhood need of supply, each number obtained according to converted the Consumption into days by converting all quantities to fixed normalized values between

0 and 15, this procedure depends on each index value after arranging all clusters from smallest values to the largest values.

Graphically and to illustrate the best distribution clustering to be closer to understanding, the following figures (5.17, 5.18 and 5.19) show all neighborhoods and how much it needs from days for January, July, and November in 2017, respectively.

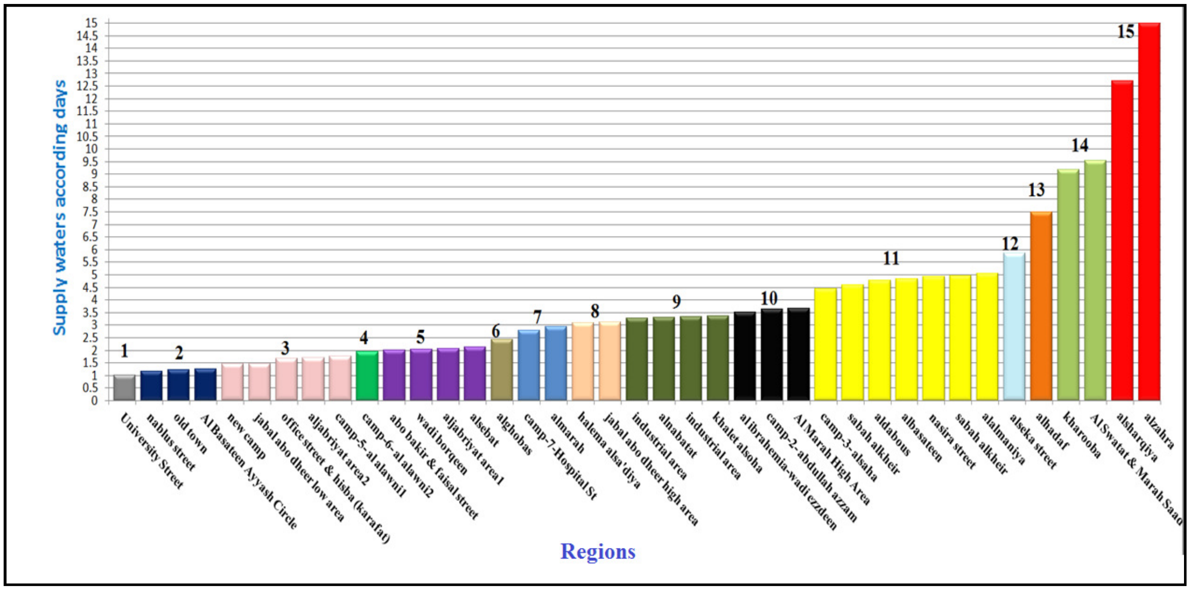

Figure 5-17: supply water According to the number of day's allocated- January-2017

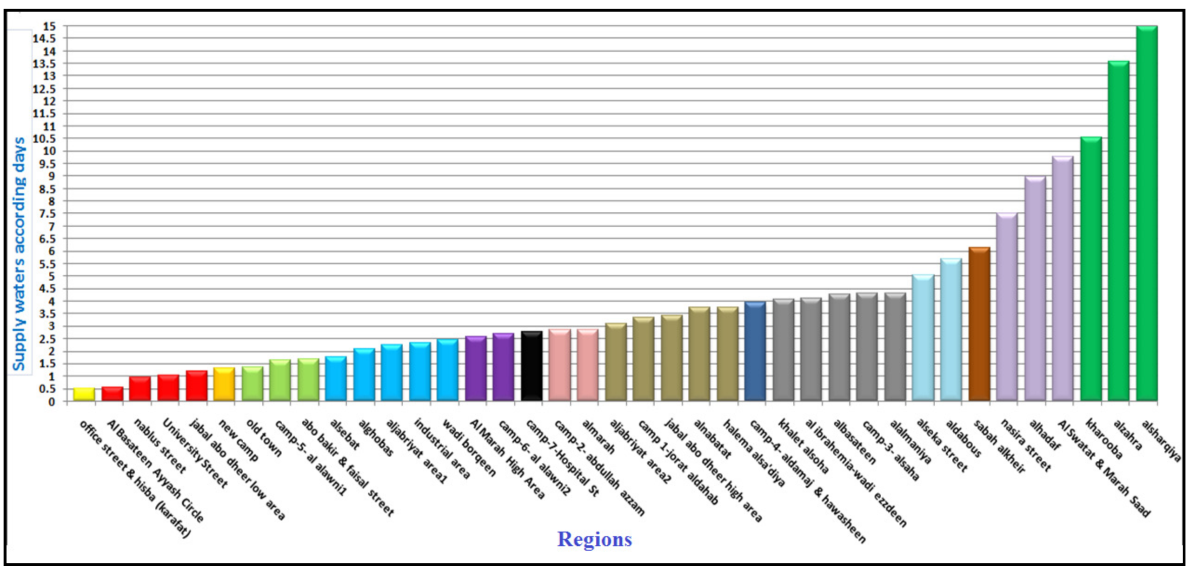

Figure 5-18: supply water According to the number of day's allocated- July-2017

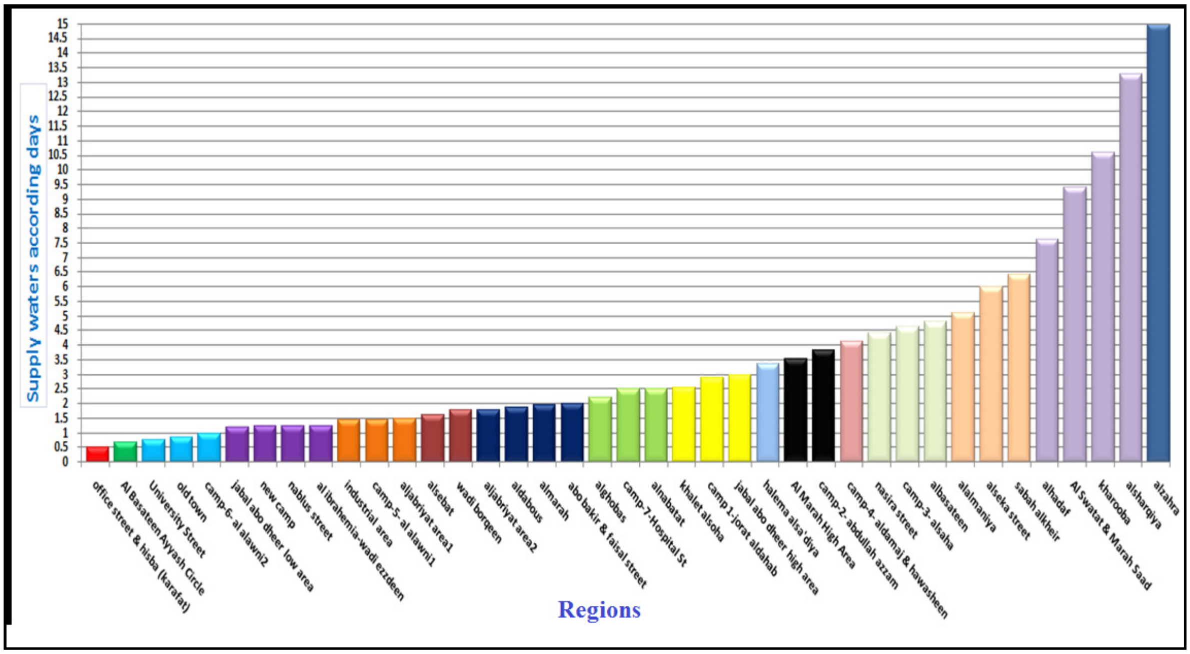

Figure 5-19: supply water According to the number of day's allocated- November-2017

According to previous figures, you can note how much the clusters is accurate and it is a regularity, add to that the neighborhoods and data points values are near together in the same cluster and there are no large differences.

Depending on the experimental result of clustering, the features computed on the water consumption data for each region in Jenin city as shown in the map actually represented consumers with similar properties to get "naturally" grouped together.

The presented result illustrates in graphs in 3 months January, July, and November. As shown in the figures above the k-means clustering algorithm classifies each similar gropes of regions in one cluster depends on the consumption. This process aims to help in the control of the water distribution for each region.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6 Conclusion and Future Work

This chapter summarizes the thesis results and offers the main important points gained during applying our experiments. In our thesis, different prediction algorithms (ARIMA, Multi-Layer Perceptron Feed Forward Back Propagation, Radial Basis Function with genetic algorithms, K means clustering), are used to model and predict the future water demand and consumption, also to model a new way for water distribution and supply widely in a Palestinian city Jenin.

The modeling was based on collected data from all neighborhoods of Jenin city including the consumption as well as these available in the department of water database that follow the municipality. It is clearly observed that in our country there are no any applied models that predicted the water demand and consumption or discussed until now. The algorithms were applied to water demand to predict the future of water needs according to the breadth and increase of demand that happen with the time passage. The major interest of this thesis was to assess these algorithms over our datasets, then select and determine the most suitable and accurate one to predict the needs of the coming years.

6.1 Conclusions

The applied Neural Networks models have shown a very good prediction of the monthly demand of urban water in the Jenin city. Considering as input variables which present the time data series of the water demands and the objective data is the real water consumption during the time series data. The applied models used to recognize the patterns in the historical data of the water demand, and use this pattern to predict the future data of water demand. The results of the NNs, in general, are superior to those obtained with multiple regressions and with the ARIMA models of time series. This characterization of short-term demand the term has as its main objective its use as an entry into methods and/or management programs in a real-time of water distribution systems, but also, it supposes a better adjustment of the distribution dais depends on smart clustering algorithm which use to group the regions depends on its water needed. Which finally pump the real demand of the network of distribution and therefore a more rational use of water resource.

According to all discussions of this thesis, the following conclusions can be expressed:

- The data for water demand and consumption is non-stationary and changed over time Series.
- The models treat seven years (84 months and 39 neighborhoods) of about 7500 subscribers based on historical water consumptions.
- After a number of modeling trials, ANNs particularly is the best model for predicting the water demand and consumption especially the MLPFFBP algorithm and achieve a better result than ARIMA and also RBF with genetic algorithms.
- The new model produced very good results depending on the high correlation between the actual and predicted values of water demand based on the MLP model.
- The K means clustering algorithm achieved superior results in adjusting, rearranging and clarifying the characteristics of water consumption by regrouping similar objects according to quantities and pattern of consumption within clear and organized clusters.
- Based on the results of all the applied models and the variance of their predictions, it is clear to us that there is a severe shortage of supplies needed to meet the needs

of the consumers due to the fluctuation in consumption and the continuous interruption in service delivery and the large disparity between the neighborhoods with little consumption and neighborhoods with large consumption and the quantities of water supplied and unfair.

6.2 Recommendations

- Different water sources should be constructed, such as building additional tanks and taking advantage of rainwater and collecting them into large wells to fill the deficit, especially in the summer.
- Rehabilitate and renew distribution networks, improve the capacity of the pumps, increase their numbers and arrange the valves to go on with changes in the geographical nature of the city.
- The development of new methods to control and regulate the use and improve methods of collecting the prices and eliminate the unauthorized use by installing prepaid meters and thus benefit from the revenue in the development of the whole water sector in the city.

6.3 Future Work

We plan to search for new sources of information and data is available such as temperature factors, population growth, and other factors in order to link them to prediction processes using the modern artificial networks methodology. Models will be expanded and disseminated in different sectors in cooperation with the Palestinian Water Authority and other applications will be developed in the field of future predictions to improve conditions in our lives and to contribute and to enrich the scientific research fields in our country.

Another important thing that could be based on this study is to predict water losses in the city of Jenin.

7 References

[1] Mekonnen, Mesfin, and Arjen Y. Hoekstra. "Four billion people facing severe water scarcity." *Science advances* 2.2 (2016): e1500323.

[2] Ghiassi, Manoochehr, David K. Zimbra, and HassineSaidane. "Urban water demand prediction with a dynamic artificial neural network model." Journal of Water Resources Planning and Management 134.2 (2008): 138-146.

[3] Kofinas, et al. "Urban water demand prediction for the island of Skiathos." Procedia Engineering 89 (2014): 1023-1030.

[4] Sampathirao, Ajay Kumar, et al. "Water demand prediction for the optimal operation of large-scale drinking water networks: The Barcelona case study." IFAC, 2014.

[5] Herrera, Manuel, et al. "Predictive models for prediction hourly urban water demand." Journal of Hydrology 387.1-2 (2010): 141-150.

[6] Fekri, Hassan. "Water history for our times IHP essays on water history." (2010).

[7] http://geocurrentscommunity.blogspot.com/2010/12/water-politics-in-Israeli-palestinian.html. Accessed [14 09 2018].

[8] https://www.ampalestine.org/palestine-101/key-topics/water-resources/west-bank-water-resources-map. Accessed [14 09 2018].

[9] http://www.pcbs.gov.ps/post.aspx?lang=en&ItemID=3149. Accessed [14 09 2018].

[10] http://www.pcbs.gov.ps/Downloads/book1632.pdf. Accessed [14 09 2018].

[11] unesdoc.unesco.org/images/0021/002154/215492e.pdf. Accessed [14 09 2018].

[12] https://www.ncbi.nlm.nih.gov/books/NBK83724/. Accessed [14 09 2018].

[13] Dziegielewski, Ben, and John J. Boland. "PREDICTION URBAN WATER USE: THE IWR□MAIN MODEL 1." JAWRA Journal of the American Water Resources Association 25.1 (1989): 101-109.

[14] Yang, Bohan, Weiwei Zheng, and XinliKe. "Prediction of Industrial Water Demand Using Case-Based Reasoning—A Case Study in Zhangye City, China." Water 9.8 (2017): 626.

[15] Liu, Yu, et al. "Improving Agricultural Water Use Efficiency: A Quantitative Study of Zhangye City Using the Static CGE Model with a CES Water− Land Resources Account." Sustainability 9.2 (2017): 308.

[16] Casillas, Myrna V., et al. "Leak signature space: An original representation for robust leak location in water distribution networks." Water 7.3 (2015): 1129-1148.

[17] https://blogs.worldbank.org/opendata/chart-globally-70-freshwater-used-agriculture. Accessed [14 09 2018].

[18] Zhang, G. Peter. "Time series prediction using a hybrid ARIMA and neural network model." Neurocomputing 50 (2003): 159-175.

[19]http://www.arij.org/files/admin/19964_Environmantal_Profiles_for_the_West_Bank_Volume_7_Jenin_District.pdf. Accessed [14 09 2018].

[20] Wagner, Neal, et al. "Intelligent techniques for prediction multiple time series in real-world systems." International Journal of Intelligent Computing and Cybernetics 4.3 (2011): 284-310.

[21] Reale, Marco. "A graphical modeling approach to time series." Ph.D.diss., University of Lancaster, 1998.

[22] Brockwell, Peter J., Richard A. Davis, and Matthew V. Calder. Introduction to time series and forecasting. Vol. 2. New York: Springer, 2002.

[23] Kihoro, J., R. Otieno, and C. Wafula. "Seasonal time series forecasting: A comparative study of ARIMA and ANN models." AJST 5.2 (2004).

[24] Adhikari, Ratnadip, and R. K. Agrawal. "An introductory study on time series modeling and forecasting." arXiv preprint arXiv:1302.6613 (2013).

[25] Zhang, Guoqiang, B. Eddy Patuwo, and Michael Y. Hu. "Prediction with artificial neural networks:: The state of the art." International journal of prediction 14.1 (1998): 35-62.

[26] Liu, Junguo, Hubert HG Savenije, and Jianxin Xu. "Forecast of water demand in Weinan City in China using WDF-ANN model." Physics and Chemistry of the Earth, Parts A/B/C 28.4-5 (2003): 219-224.

[27] Hawley, Delvin D., John D. Johnson, and Dijjotam Raina. "Artificial neural systems: A new tool for financial decision-making." Financial Analysts Journal 46.6 (1990): 63-72.

[28] Dreiseitl, Stephan, and LucilaOhno-Machado. "Logistic regression and artificial neural network classification models: a methodology review." Journal of biomedical informatics 35.5-6 (2002): 352-359.

[29] Wan, Eric A. "Neural network classification: A Bayesian interpretation." IEEE Transactions on Neural Networks 1.4 (1990): 303-305.

[30] Karlik, Bekir, and A. VehbiOlgac. "Performance analysis of various activation functions in generalized MLP architectures of neural networks." International Journal of Artificial Intelligence and Expert Systems 1.4 (2011): 111-122.

[31] Grothmann, Ralph. Multi-Agent Market Modeling Based on Neural Networks. Diss. University of Bremen, Germany, 2003.

[32] Willmott, Cort J., and Kenji Matsuura. "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance." Climate Research 30.1 (2005): 79-82.

[33] Hu, Michael Jen-Chao. Application of the Adaline system to weather forecasting. Diss. Department of Electrical Engineering, Stanford University, 1964.

[34] Baudat, Gaston, and FatihaAnouar. "Kernel-based methods and function approximation." Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on. Vol. 2. IEEE, 2001.

[35] Awad, Mohammed. "Optimization RBFNNs parameters using genetic algorithms: applied on function approximation." International Journal of Computer Science and Security (IJCSS) 4.3 (2010): 295.

[36] Orr, Mark JL. "Introduction to radial basis function networks." (1996).

[37] Yu, Bing, and Xingshi He. "Training radial basis function networks with differential evolution." In Proceedings of IEEE International Conference on Granular Computing. 2006.

[38] CÖMERT, Zafer, and Adnan Fatih KOCAMAZ. "A study of artificial neural network training algorithms for classification of cardiotocography signals." BitlisEren University Journal of Science and Technology 7.2 (2017): 93-103.

[39] Sundar, C., M. Chitradevi, and G. Geetharamani. "Classification of cardiotocogram data using neural network based machine learning technique." International Journal of Computer Applications 47.14 (2012).

[40] Fritzke, Bernd. "Growing cell structures—a self-organizing network for unsupervised and supervised learning." Neural networks 7.9 (1994): 1441-1460.

[41] Lourakis, Manolis IA. "A brief description of the Levenberg-Marquardt algorithm implemented by levmar." Foundation of Research and Technology 4.1 (2005): 1-6.

[42] Ekonomou, L. "Greek long-term energy consumption prediction using artificial neural networks." Energy 35.2 (2010): 512-517.

[43] Ediger, Volkan Ş., and SertacAkar. "ARIMA prediction of primary energy demand by fuel in Turkey." Energy Policy 35.3 (2007): 1701-1708.

[44] Chen, Ching-Fu, Yu-Hern Chang, and Yu-Wei Chang. "Seasonal ARIMA prediction of inbound air travel arrivals to Taiwan." Transportmetrica 5.2 (2009): 125-140.

[45] Valipour, Mohammad, Mohammad EbrahimBanihabib, and Seyyed Mahmood Reza Behbahani. "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in prediction the monthly inflow of Dez dam reservoir." Journal of Hydrology 476 (2013): 433-441.

[46] Memarian, Hadi, and Siva Kumar Balasundram. "Comparison between multi-layer perceptron and radial basis function networks for sediment load estimation in a tropical watershed." Journal of Water Resource and Protection 4.10 (2012): 870.

[47] Gagliardi, Francesca, et al. "A comparison between pattern-based and neural network short-term water demand prediction models." Water Science and Technology: Water Supply 17.5 (2017): 1426-1435.

[48] Jain, Ashu, Ashish Kumar Varshney, and Umesh Chandra Joshi. "Short-term water demand forecast modeling at IIT Kanpur using artificial neural networks." Water resources management 15.5 (2001): 299-321.

[49Msiza, Ishmael S., Fulufhelo Vincent Nelwamondo, and TshilidziMarwala. "Water demand prediction using artificial neural networks and support vector regression." (2008).

[50] Xing, Ying, et al. "City water demand prediction based on improved BP neural network." Journal of Residuals Science & Technology (2017).

[51] Perea, Rafael González, et al. "Optimisation of water demand prediction by artificial intelligence with short data sets." Biosystems Engineering (2018).

[52] Deng, Xiao, et al. "Hourly Campus Water Demand Prediction Using a Hybrid EEMD-Elman Neural Network Model." Sustainable Development of Water Resources and Hydraulic Engineering in China. Springer, Cham, 2019. 71-80.

[53] Jayalakshmi, T., and A. Santhakumaran. "Statistical normalization and backpropagation for classification." International Journal of Computer Theory and Engineering 3.1 (2011): 1793-8201.

[54]Jayalakshmi, T., and A. Santhakumaran. "Statistical normalization and backpropagation for classification." International Journal of Computer Theory and Engineering 3.1 (2011): 1793-8201.

[55] Wong, Tzu-Tsung. "Performance evaluation of classification algorithms by k-fold and leave-one-out cross-validation." Pattern Recognition 48.9 (2015): 2839-2846.

[56] Goyal, P., Andy T. Chan, and Neetu Jaiswal. "Statistical models for the prediction of respirable suspended particulate matter in urban cities." Atmospheric Environment 40.11 (2006): 2068-2077.

[57] Campisi-Pinto, Salvatore, Jan Adamowski, and Gideon Oron. "Prediction urban water demand via wavelet-denoising and neural network models. Case study: city of Syracuse, Italy." Water resources management 26.12 (2012): 3539-3558.

[58] Hipel, Keith W., and A. Ian McLeod. Time series modeling of water resources and environmental systems. Vol. 45. Elsevier, 1994.

[59] DeLurgio, Stephen A. Prediction principles and applications. Irwin Professional Publishing, 1998.

[60] Jiang, Heng, Michael Livingston, and Robin Room. "Alcohol consumption and fatal injuries in Australia before and after major traffic safety initiatives: a time series analysis." Alcoholism: clinical and experimental research 39.1 (2015): 175-183.

[61] Flaherty, John, and Ric Lombardo. "Modelling Private New Housing Starts In Australia." A paper presented in the Pacific-Rim Real Estate Society conference. 2000.

[62] Lee, Junsoo. "Univariate time series modeling and prediction (Box-Jenkins Method)." Econ 413, lecture 4.

[63] Faraway, Julian, and Chris Chatfield. "Time series prediction with neural networks: a comparative study using the airline data." Journal of the Royal Statistical Society: Series C (Applied Statistics) 47.2 (1998): 231-250.

[64] Kihoro, J., R. Otieno, and C. Wafula. "Seasonal time series forecasting: A comparative study of ARIMA and ANN models." AJST 5.2 (2004).

[65] Nagahamulla, Harshani RK, Uditha R. Ratnayake, and AsangaRatnaweera. "Artificial neural network ensembles in time series forecasting: an application of rainfall prediction in Sri Lanka." ICTer 6.2 (2014).

[66] Shekhar, Shashi, Minesh B. Amin, and Prashant Khandelwal. "Generalization performance of feed-forward neural networks." Neural Networks. 1992. 13-38.

[67] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research 15.1 (2014): 1929-1958.

[68] Deb, Chirag, et al. "Prediction diurnal cooling energy load for institutional buildings using Artificial Neural Networks." Energy and Buildings 121 (2016): 284-297.

[69] Vaisla, Kunwar Singh, and Ashutosh Kumar Bhatt. "An analysis of the performance of artificial neural network technique for stock market forecasting." International Journal on Computer Science and Engineering 2.6 (2010): 2104-2109.

[70] Gardner, Matt W., and S. R. Dorling. "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences." Atmospheric Environment 32.14-15 (1998): 2627-2636.

[71] Riedmiller, Martin, and Heinrich Braun. "A direct adaptive method for faster backpropagation learning: The RPROP algorithm." Neural Networks, 1993., IEEE International Conference on. IEEE, 1993.

[72] Kişi, Özgür, and ErdalUncuoğlu. "Comparison of three back-propagation training algorithms for two case studies." (2005).

[73] HAMDAN, IHAB, MOHAMMED AWAD, and WALID SABBAH. "SHORT-TERM PREDICTION OF WEATHER CONDITIONS IN PALESTINE USING ARTIFICIAL NEURAL NETWORKS." Journal of Theoretical & Applied Information Technology 96.9 (2018).

[74] Wilamowski, Bogdan M. "Neural network architecturesand learning algorithms." IEEE Industrial Electronics Magazine 3.4 (2009).

[75] Moody, John, and Christian J. Darken. "Fast learning in networks of locally-tuned processing units." Neural Computation 1.2 (1989): 281-294.

[76] Harpham, C., Christian W. Dawson, and Martin R. Brown. "A review of genetic algorithms applied to training radial basis function networks." Neural Computing & Applications 13.3 (2004): 193-201.

[77] Aishwarya, K. M., et al. "Lossy image compression using svd coding algorithm." Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on. IEEE, 2016.

[78] Vance, Ashlee. "Data analysts captivated by R's power." New York Times 6.5.4 (2009).

[80 www.pcbs.gov.ps/portals/_pcbs/PressRelease/Press_En_Preliminary_Results_Report-en-with-tables.pdf.page-65. Accessed [14 09 2018].

[81] Pandey, Abhishek, and AnkitaMalviya. "Optimizing Map Reduce Scheduling Using Parallel Processing Model On Data Nodes With K-Mean Clustering Algorithm In Hadoop Environment." (2018).

[82] Ding, Chris, and Xiaofeng He. "K-means clustering via principal component analysis." Proceedings of the twenty-first international conference on Machine learning. ACM, 2004.

[83] Kanungo, Tapas, et al. "An efficient k-means clustering algorithm: Analysis and implementation." IEEE Transactions on Pattern Analysis & Machine Intelligence 7 (2002): 881-892.

[84] Maggiori, Emmanuel, et al. "Convolutional neural networks for large-scale remote-sensing image classification." IEEE Transactions on Geoscience and Remote Sensing 55.2 (2017): 645-657.

[85] Grassia, Filippo, et al. "Spike pattern recognition using artificial neuron and Spike-Timing-Dependent Plasticity implemented on a multi-core embedded platform." Artificial Life and Robotics 23.2 (2018): 200-204.

[86] Xiao, Xuefeng, et al. "Building fast and compact convolution neural networks for offline handwritten Chinese character recognition." Pattern Recognition 72 (2017): 72-81.

[87] Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N.,& Hullender, G. (2005, August). Learning to rank using gradient descent. In Proceedings of the 22nd international conference on Machine learning (pp. 89-96). ACM.

[88] Battiti, Roberto. "First-and second-order methods for learning: between steepest descent and Newton's method." Neural Computation 4.2 (1992): 141-166.

[89] Cui, Miao, et al. "A new approach for determining damping factors in the Levenberg-Marquardt algorithm for solving an inverse heat conduction problem." International Journal of Heat and Mass Transfer 107 (2017): 747-754.

من المعروف بان سلطة المياه في جنين تقوم بتنظيم الإمداد بالمياه يومًا واحدًا في الأسبوع لكل منطقة في المدينة، وتم تطبيق فكرة إبداعية لتنظيم توزيع المياه لكل منطقة في مدينة جنين اعتمادًا على الاستهلاك السابق. تعتمد هذه الطريقة على استخدام خوارزمية التجميع K-means لتصنيف المناطق، بحيث يمكن للسلطات توفير المياه لكل منطقة مع عدد من الأيام للحصول على العدل في التوزيع.

يتم تطبيق نوعين من الشبكات العصبية الاصطناعية بيرسبترون متعدد الطبقات MLPNN وخوارزمية تدريب ليفينبيرق ماركاردت LMA والشبكات العصبية الاصطناعية الشعاعية (RBFNNs) مع الخوارزميات الجينية للتدريب وللتعلم، بالإضافة لنوع واحد من الطرق الإحصائية الخطية يسمى المتوسط المتحرك المتكامل للانحدار (ARIMA) على بيانات الطلب على المياه التي تم جمعها من مدينة جنين للتنبؤ بالطلب على المياه في المستقبل. تظهر النتائج التجريبية أن (MLPNN) يتفوق على نماذج (RBFNN) و (ARIMA) في التنبؤ بقيم الطلب على المياه. وعلاوة على ذلك، تم استخدام خوارزمية تجميع لبناء نظام توزيع عادل للمياه في مدينة جنين بالاعتماد على الاستهلاك السابق حيث كانت النتيجة التجريبية لتجميع المناطق وتصنيفها في المدينة فعالة لتوفير المياه بطريقة عادله.

الملخص

الماء مهم جدا للإنسان، في كل مجالاتنا الحياتية، الزراعة، الصناعة، والصحة، حيث ازداد استهلاك المياه بطريقة كبيرة. يعتبر استخدام المياه المنزلية من أهم العوامل الأساسية في سلسلة الطلب على المياه واستهلاكها في المناطق الحضرية. تم اقتراح العديد من النماذج والتقنيات الحسابية في العقود الأخيرة للتنبؤ بالطلب على المياه في سلاسل زمنية قصيرة أو طويلة الأمد، مثل الطرق الإحصائية والطرق الرياضية للتنبؤ بالكميات المستقبلية من الطلب اليومي والشهري والسنوي. تم تطبيق أساليب أخرى أكثر تقدمًا مثل مقاربات تحليل سلاسل الوقت. تعتمد غالبية هذه الطرق على استقراء الاتجاهات التاريخية وربط الطلب بالمتغيرات الاجتماعية والاقتصادية. من الضروري تقييم قدرة الموارد الحالية لتلبية الاحتياجات المستقبلية وتوفير الأساس لتخطيط النظام المستقبلي وتحسينه للحد من أوجه عدم اليقين في توقعات الطلب مثل النمو السكاني والتغير الاقتصادي والتغيرات في عادات الاستهلاك وتغير المناخ خاصة في منطقتنا، وبالتالي فإن التنبؤ بالطلب على المياه يساعد شركات توزيع المياه والحكومة على معرفة والتحقيق في الطلب المتوقع والآثار على تخطيط التنمية المستدامة. وفقا لمجموعة من التحديات التي تواجه سلطات المياه والبلديات وقطاع المياه، بصفة عامة، فإن أهم الأشياء هي نقص الأدوات التي يمكن أن تتوقع الطلب على المياه في المستقبل .

في هذه الرسالة، نستخدم طرقًا خطية وغير خطية للتنبؤ بالطلب على المياه في مدينة جنين، لهذا الغرض، نستخدم أنواعًا مختلفة من الشبكات العصبية الاصطناعية (ANNs) مع طرق تعليمية مختلفة للتنبؤ بالطلب على المياه، مقارنة مع نوع معروف من الأساليب الإحصائية. يتم استخدام تقنية حسابية تعتمد على الشبكات العصبية الاصطناعية (ANNs) وطريقة هجينة من NNs مع خوارزمية التحسين للتنبؤ بالطلب على المياه في المناطق الحضرية. تعتمد مجموعة البيانات على مجموعات البيانات المجمعة (المستخرجة من قواعد بيانات البلديات) خلال فترة زمنية محددة، ومن ثم نقترح نموذجًا غير خطي للتنبؤ بالطلب على المياه شهريًا وسنوياً، وفي النهاية توفير نموذج التنبؤ الأكثر دقة مقارنة بالطرق الخطية وغير الخطية الأخرى. نحن نهدف إلى إنشاء نموذج قادر على إجراء تنبؤ دقيق للطلب على المياه في المستقبل لمدينة جنين. يتم إجراء هذا التنبؤ بأشهر زمنية أو سنوات زمنية، اعتمادًا على البيانات المستخرجة، سيتم استخدام هذه البيانات لتغذية نموذج الشبكة العصبية لتنفيذ آليات ونظام يمكن استخدامه للتنبؤ بأمد قصير لطلب المياه.