Using Markov Chains, Regression and Time Series Models to Predict Supply and Demand for Water and Their Behavior: Case Study Tulkarm City

By

Ghadeer Akram Al-Mallak

Supervisor

Prof. Dr. Saed Mallak

Cosupervisor

Prof. Dr. Saed Khayat

This Thesis was submitted in partial fulfillment of the requirements for the Master's Degree of Science in Mathematical Modeling

Faculty of Graduate Studies

Palestinian Technical University-Kadoorie

December, 2020

استخدام نماذج سلاسل ماركوف والانحدار والسلاسل الزمنية للتنبؤ بالعرض والطلب على المياه وسلوكها:دراسة حالة مدينة طولكرم

إعداد
غدير أكرم الملاك

المشرف
أ.د. سائد ملاك

المشرف المشارك
أ.د. سائد خياط

قدمت هذه الرسالة استكمالا لمتطلبات الحصول على درجة الماجستيرفي النمذجة الرياضية

كلية الدراسات العليا

جامعة فلسطين التقنية- خضوري

كانون أول،2020

جامعة فلسطين التقنية-خضوري

Palestine Technical University –Kadoorie

PTUK

Authorization form

I, Ghadeer Akram Abed Al-Fattah Al-Mallak, authorize the PTUK to supply copies of my thesis / dissertation to libraries or establishments or individuals on request, according to the university of PTUK. regulations

Signature:

Date:

جامعة فلسطين التقنية -خضوري

نموذج التفويض

أنا غدير أكرم عبد الفتاح الملاك أفوض جامعة فلسطين التقنية خضوري بتزويد نسخ من رسالتي / أطروحتي للمكتبات أوالمؤسسات أوالهيئات أوالأشخاص عند طلبهم حسب التعليمات النافذة في الجامعة.

التوقيع:

التاريخ:

الاقرار

انا الموقع ادناه مقدم الرسالة التي تحمل العنوان :

Using Markov Chains, Regression and Time Series Models to Predict Supply and Demand for Water and Their Behavior: Case Study Tulkarm City

اقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص ، باستثناء ما تم الإشارة اليه حيثما ورد ، وأن هذه الرسالة ككل، أو أي جزء منها لم يقدم من قبل لنيل درجة علمية أو بحث علمي لدى مؤسسة تعليمية أو بحثية اخرى.

Declaration

I hereby declare that this thesis is the product of my own efforts, except what has been referred to, and this thesis as a whole or any part of it has not been submitted as a requirement for attaining a scientific degree to any other educational or research institution.

Ghadeer Akram Al-Mallak	غدير اكرم عبد الفتاح الملاك
Signature: …………….	التوقيع: ...………….. ..
Date :…………………	التاريخ :........………

COMMITTEE DECISION

This thesis/dissertation Using Markov Chains, Regression and Time Series Models to Predict Supply and Demand for Water and Their Behavior : Case Study Tulkarm City

Examination committee	**Signature**
- Prof. Dr. Saed Mallak (Supervisor)	
- Prof. Dr. Saed Khayat (Cosupervisor)	
- Asst. Prof. Dr. Hassan Abu Hassan (External Examiner)	
- Asst. Prof. Dr. Mohammad Jarrar (Internal examiner)	

DEDICATION

I would like to dedicate my thesis to:

To my father, who always urged me for more work.

To my mother, whose prayers were with me all the way to success, who taught me to believe in myself.

To my brothers and sisters, whose support me in my life stages.

To my friends, who believe in me in all of these bad conditions.

To everyone, who always have inspired me.

AKNOWLEDGEMENT

First of all, praise and thanks to Allah, for giving me the courage and determination as well as guidance in my thesis.

I would to convey my deepest thanks for my thesis supervisor Prof. Dr. Saed Mallak and cosupervisor Prof. Dr. Saed Khayat for their trust that I am able to accomplish this work and for their support, encouragement and guidance for me throughout my work on this thesis. I am grateful to Mr. Nasouh Soos for his help and useful discussions and suggestions.

I am extremely grateful to my parents for their love, prayers and caring. Also, I express my thanks to my sisters, brothers, my friends for their support and valuable prayers.

Table of Contents

List of Tables

List of Figures

Using Markov Chains, Regression and Time Series Models to Predict Supply and Demand for Water and Their Behavior: Case Study Tulkarm City

By: Ghadeer Al-Mallak

Supervised by: Prof. Dr. Saed Mallak and Prof. Dr. Saed Khayat

Abstract

This study aimed to predict the amount of water supply and demand in Tulkarm municipality, based on monthly data within the period from January 2010 to December 2019.

The matrix of transitional possibilities was formed and then predicted values based on the probability transition matrix were obtained by using the MatLabstatistical program, the results showed that the supply chain stabilizes after (24 months) where we have reached a rise in the amount of water demand at the rate (55.8%). While the rate of decrease was (36.71%) and the rate of stability was (7.54%). Referring to water demand chain, the probability transition matrix stabilizes after (16) months, in which a rise in the amount of water demand has been reached also at the rate (55.46%) while the rate of decrease was (34.45%) and the rate of stability was (10.08%)

The study also tried to compare statistical models (Markov model, time series model, and simple linear regression model) to predict the future values of both series. This has been done through using the forecast accuracy scale to find the most suitable model for analyzing the data in the study.

The study found that the Markov model with the lowest criterion value (RMSE) is the most appropriate model to analyze the data of the water demand chain in Tulkarm municipality, but for the data of the water supply chain, it was found that the model of time series ARIMA (2, 1, 2) (0, 1, 1) is the most appropriate because it has the lowest forecast criterion value.

Based on previous models, the amount of water supply and demand in Tulkarmmunicipality were forecasted for 2030. The amount of Water supply in Tukaram is expected to increase to 1081098.91 MCM, while the total amount of demand in 2030 is expected to increase up to 680779.2 MCM.

The forecasted results emphasize the necessity for implementing an integrated water resources management plan in the district, thus utilize the excess supply and demand to ensure sustainable water for the local community on the future. The models shows good potential for expecting such critical water situation that needs an urgent action by the stakeholders through a comprehensive water management plan to ensure water security for the next generations.

الملخص باللغة العربية

تهدف هذه الدراسة إلى التنبؤ بكمية العرض والطلب على المياه في مدينة طولكرم، وذلك اعتمادا على البيانات الشهرية لكمية العرض والطلب على المياه خلال الفترة من يناير 2010 إلى ديسمبر 2019.

تم تكوين مصفوفة الانتقال الاحتمالية ، ثم تم الحصول على القيم المتوقعة على أساس مصفوفة الانتقال الاحتمالية باستخدام البرنامج الإحصائي MatLab ، وأظهرت النتائج أن سلسلة العرض تستقر بعد (24 شهرًا) حيث اشارت النتائج الى ان نسبة الارتفاع في كمية العرض على المياه كانت (55.8٪). في حين كانت نسبة الانخفاض (36.71%) ، و كانت نسبة الاستقرار (7.54٪). بينما اشارت نتائج سلسلة الطلب على المياه ان مصفوفة الانتقال الاحتمالية تستقر بعد (16) شهراً، حيث كانت نسبة الارتفاع في كمية الطلب على المياه (55.46%) في حين كانت نسبة الانخفاض (34.45%) وكانت نسبة الاستقرار (10.08%)

كما هدفت الدراسة الى مقارنة النماذج الإحصائية (نموذج ماركوف، ونموذج السلسلة الزمنية، ونموذج الانحدار الخطي البسيط) للتنبؤ بالقيم المستقبلية لكل من السلسلتين (الطلب والعرض على المياه). وذلك من خلال استخدام مقياس دقة التنبؤ للعثور على النموذج الأنسب لتحليل بيانات الدراسة.

وظهرت نتائج الدراسة أن نموذج الماركوف كانت قيمة المعيار (RMSE) له اقل من النماذج الاخرى وبالتالي يعتبر النموذج الأنسب لتحليل بيانات سلسلة الطلب على المياه في بلدية طولكرم والتنبؤ بالقيم المستقبلية ، أما بالنسبة لبيانات سلسلة العرض على المياه فقد تبين أن نموذج السلسلة الزمنية ARIMA (2, 1, 2) (0, 1, 1) هو الأنسب لأنه يحتوي على أدنى قيمة لمعيار التنبؤ.

واستناداً إلى النماذج السابقة، قمنا بالتنبؤ بكمية العرض والطلب على المياه في مدينة طولكرم لعام 2030.ومن المتوقع أن يرتفع حجم العرض على المياه في مدينة طولكرم إلى 1081098.91 مليون متر مكعب، في حين يتوقع أن يرتفع إجمالي الطلب في عام 2030 إلى 680779.2 مليون متر مكعب.

وتؤكد النتائج المتوقعة ضرورة تنفيذ خطة متكاملة لإدارة الموارد المائية في المنطقة، وبالتالي ضمان استدامة المياه للمجتمع المحلي في المستقبل واستقلال الفائض من العرض. وتظهر النماذج إمكانية جيدة لتوقع مثل هذه الحالة الحرجة للمياه التي تحتاج إلى إجراءات عاجلة من قبل أصحاب المصلحة من خلال خطة شاملة لإدارة المياه لضمان الأمن المائي للأجيال القادمة.

Chapter1

Introduction and Motivation

1.1Introduction

Nowadays, one of the serious financial issues facing many developing countries worldwide is water shortage and water production costs; especially potable water. As the many scientific reports indicate that by 2040 numerous countries including Palestine will face a severe deficiency in potable water.

The Israeli occupation and his sovereignty on West Bank water resources، the contamination of groundwater and the population growth as well as the economic and social development increase the pressure on available water resources in the West Bank and Gaza strip[12].

Compared to other countries in the Middle East and North Africa, water supply and sanitation in the Palestinian territories are characterized by severe water shortage, which is exacerbated by the effects of Israeli occupation. According to the WaSH Monitoring Program, Israelis use 87% of the water available from the mountain aquifer in the West Bank, and 82% of the water from the coastal aquifer under Gaza[10].

Generally, the availability and quality of services are considerably worse in the Gaza strip when compared to the West Bank. A 2011 survey shows that 70.9% of households in the West Bank consider the water quality to be good, while the share in the Gaza Strip is only 5.3% [8].

The lasting blockade of the Gaza Strip and the 2008-2009 Israel-Gaza conflict have caused severe damage to the infrastructure in the Gaza Strip[23].

About half of the delivered water in the Palestinian territories is unaccounted-for water (UFW) because of losses and billing deficiencies. With a share of 70%, agriculture is the sector which uses most of the scarce water resources. Concerning wastewater, the existing treatment plants do not have the capacity to treat all of the produced wastewater, causing severe water pollution [22]. The National Water Policy of 1995, the Water Resources Management Strategy of 1998 and the Water Law of 2002 aim at developing and improving the management of water resources. Furthermore, a new water pricing policy is being prepared [4].

The Palestinian Water Authority (PWA) was established following the Oslo interim agreement between Israel and the Palestinian authority (PA). With that, came the establishment of the Joint Water Committee (JWC) compromising of equal numbers of Palestinian and Israeli water experts whose role is to cooperatively manage the West Bank's water and wastewater systems and resources development. Nevertheless, any development water project in the West Bank requires prior approval from the JWC in addition to the approval of the higher authority of the Israeli Civil administration which is an Israeli military body which has been controlling the West Bank since 1967.

Therefore, Israel has a veto power and has hindered and constrained the Palestinian proposals for development of water infrastructure projects and even well construction and rehabilitation[6].

Moreover, the increase of population, the expansion of the agricultural and industrial activities, the different urban applications, the climate changes, and the dissimilarity in the climatic and timeline water distribution in addition to the water demands, all these factors have highlighted the importance of more water related studies to estimate and predict the total amount of water supply and demand, which will help the Palestinian to better manage their water resources and control their supply comping different stress and urgent obstacles in developing and reforming this vital sector [15].

The Palestinian Central Bureau of Statistics “PCBS”, houses connected to public water networks reached a percentage of 83.6% in 1997, then it increased in 2006 to achieve 90.8%. the PCBS also states that the daily water consumption per person reaches 135.8 litters according to 2007 statistics; 110.2 litter/day per person in West Bank، whereas in Gaza Strip 174.1 litter/day per person which is over 8% more than 2005 statistic and 15% more than 1997's [24].

In fact, 51% of this water is being purchased from the Israeli Water Corporation" Mekorot". The remaining 49% is coming from individual wells related to Water Local Authorities, besides to the springs and agricultural wells distributed all over west bank (Fig 1).

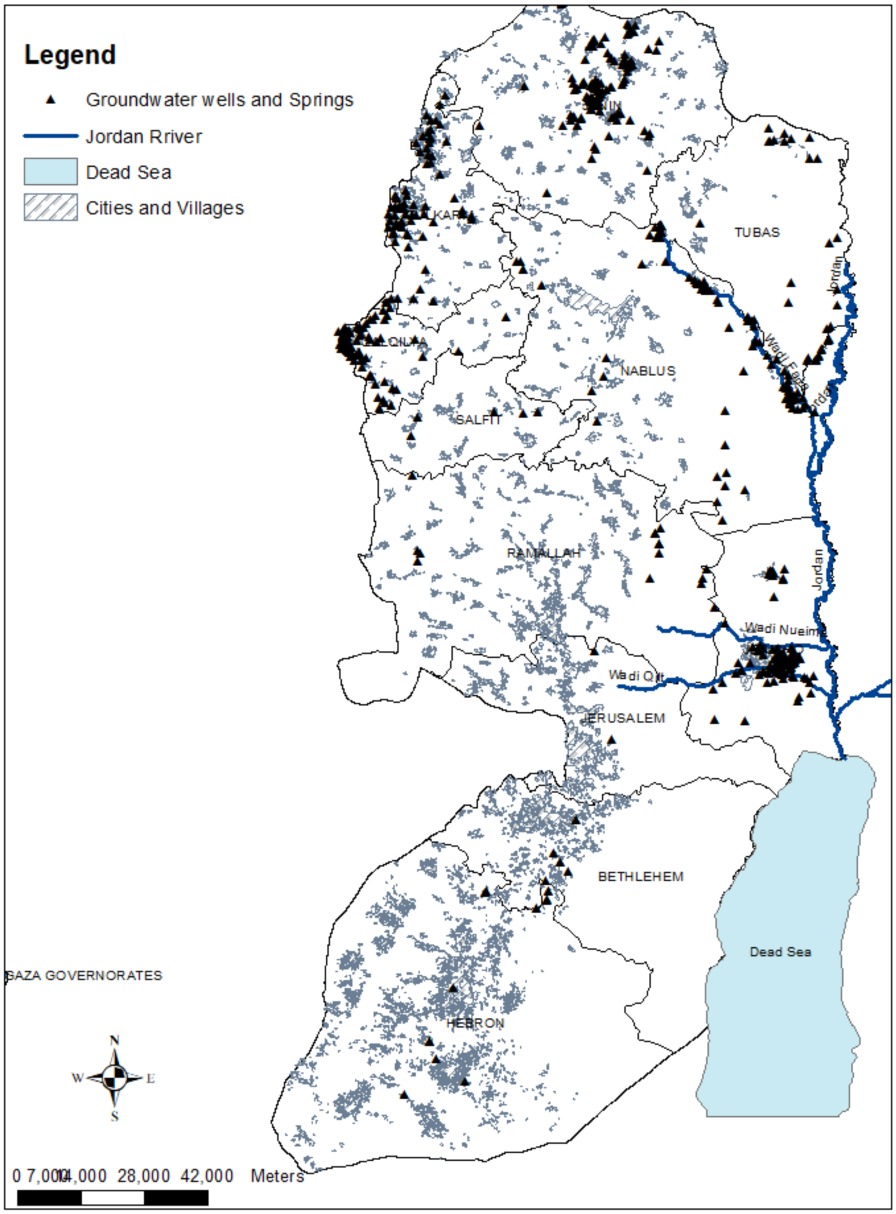


Figure 1: West Bank Map showing groundwater wells and springs all over West Bank that are used for different purposes (PTUK-GIS Lab, 2020)

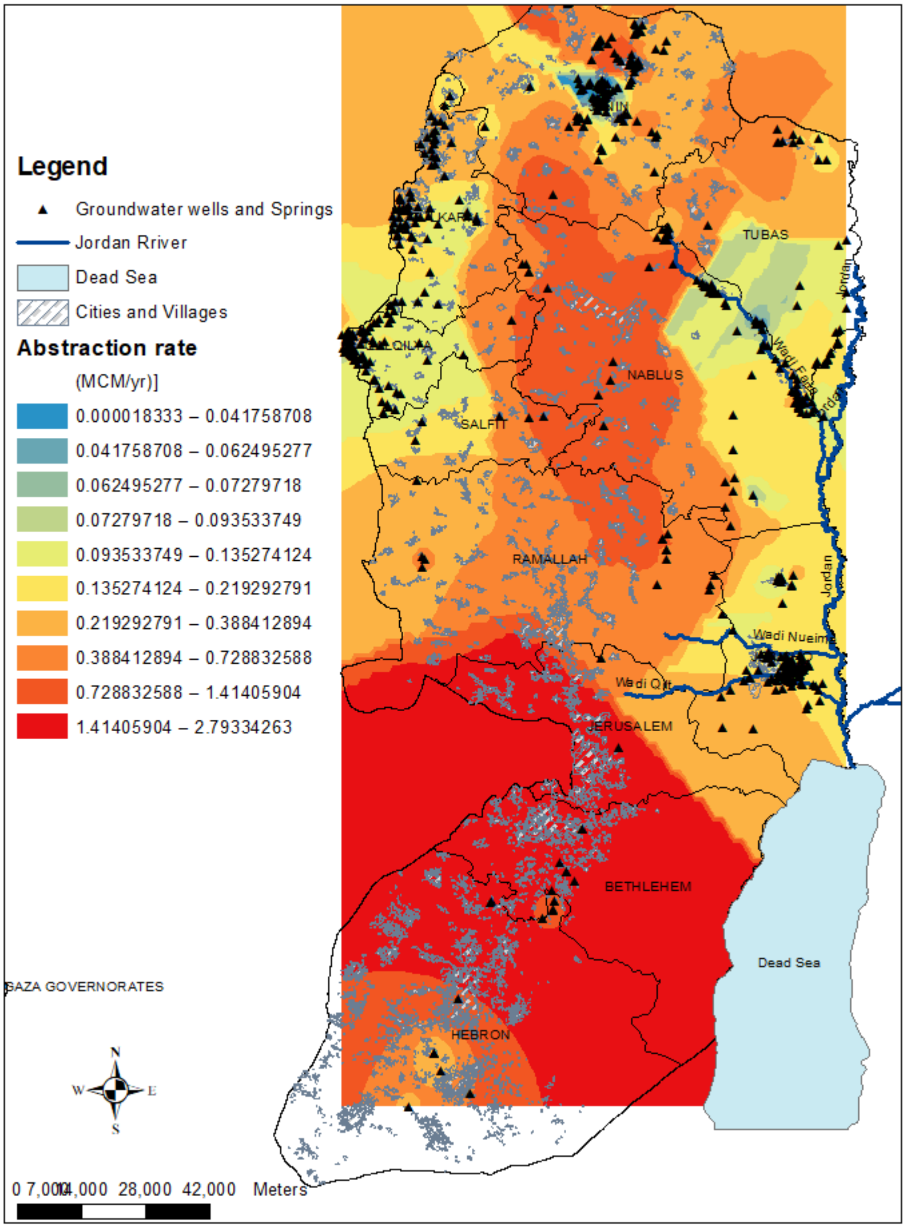


Fig. 2: Spatial Distribution of abstraction rate upon heavy demand in the West Bank (PTUK-GIS Lab, 2020)

Figure 2 shows the spatial distribution of abstraction rate in response to heavy demands distributed all over West Bank governorates. It shows that the heaviest demand is concentrated on the southern part of the West Bank mainly in Hebron and Bethlehem areas. These areas are moderately rich with agricultural activities, but with relatively 150% population density in relative to other areas in the northern model parts of the West Bank. This imply a high demand on domestic water supply that handled with overexploitation and heavy abstraction for the limited groundwater sources in the region.

The map shows also that the average annual abstraction rate in Tulkarm area range between 0.2-0.7 MCM/yr. This moderate value includes the abstraction for agricultural activities in the eastern and northern part, where map indicates relatively more abstraction activities.

In spite of moderate abstraction and pressure on the water resources in Tulkarm area, the increasing of population growth and creeping of building up areas mainly to the north and south of the district is expected to increase the pressure on available water resources, in a manner that the situation can reach similar situation in the southern part of the West Bank.

The best practice to save the water resources and to ensure its sustainability for the future generation, is to smartly manage the abstraction rate and monitoring the future demand that is mainly connected to population growth, urbanization and increasing of industrial and agricultural activities in the area.

Such smart management needs more information about the predicted demand in water resources thus to calculate exactly how the trend in water supply and abstraction rate from groundwater will change, and what are the best management scenarios needed to stop the worse situation due to expected increase in water demand to save and secure water supply that can meet such expected need.

1.2Research Problem

Currently there is no clear vision about the relation between Palestinian population growth and the prediction of water demand and supply. The annual crude birth rate for Palestinian community is about 27%. Based on PWA statistics the capita consumption of freshwater is 87.3 liter/Daily. This needs to be assess according the drastically increase in population rate and the quick change in the Palestinian style of life, which in role raise their daily water needs.

The PWA starts its sector reform strategy based on the available historical data since 20 years ago, however, predictions for different demand scenarios still not clear as it based on different visual statistical models. Using mathematical models is still under examination, and the use of these models approved its high importance in such predictions studies worldwide.

Therefore, the main issue of the research has been phrased into the following question: Is it possible to use Time Series, Regression Analysis & Markov Chain Models to predict the behavior of supply and demand for water in Palestine –Tulkarm city case study? Moreover, which one of the mentioned models has the best potential and capability to predict the real values?

Whereas, the statistical significancy of this study was obtained by finding out the most ultimate model of all statistical models which is used in prediction studies, this includes comparing Markov chain models with time series and regression analysis models in order to predict the amounts of supply and demand on water.

To date, there are no studies show comparative investigations between different statistical models. Such comparisons will give the researchers or sector modelers, more clear opinion about the best mathematical model that can be used to support the decision makers in preparing and drawing their sector strategies and policies.

1.3 The Most Important Characteristic of the Study

The outcome of this study is to provide recommendations for helping decision makers and dedicated authorities new guidelines tomanage and provide sufficient information about the future quantity of water demand for the public. In addition, the one of the main features for this study, to the best of our knowledge, is that it has been conducted in Palestine for the first time especially using the Markov chain forms, this could simplify the work of the researchers interested in similar future studies.

1.4 Objectives of the Study

Based on the subject problem, this study sought to achieve the following objectives:

1- Studying and predicting the behavior of Supply and Demand for Water by applying time series, regression analysis and Markov chains Models.
2- Comparing Markov chains models with other statistical models (time series and regression analysis) to find out which of these models will yield more accurate and better predictions results for the amount of water supply and demand and which is more practical.

1.6 Study Questions

1. What is the expected behavior of the water demand at long run?

2. What is the expected behavior of the water supply at long run?

3. What are the best statistical models among (time series, regression analysis and the Markov chain models to predict the quantity of the water supply and demand in Palestine -Tulkarm city case study?

1.7 Study Boundaries

Study aspects were divided into two; spatial and temporal aspects.

1- The spatial aspects: Tulkarm city
2- The temporal aspects: amount of supply and demand on water from January 2010 till May 2020.

1.8 Literature Review

It’s been a usual practice to use time series and regression analysis when performing the future predictions which aim at predicting the water future demand/supply for the upcoming years. In addition to the previous two methods, time series and regression analysis, Markov chains is a powerful tool to predict the behavior of the water future demand/supply for the upcoming years.

Cress well and Naser developed a time series model to forecast demands 24 hours in advance for the urban and agricultural zone in the South East Kelowna Irrigation District (British Columbia, Canada) [4]. Barhomi focused on developing a model using the MyWAS/WEAP tool which is an optimizing tool designed specifically for Palestine, Jordan and Israel. The model was a powerful and innovative tool that enables the cost- benefit analysis and can be used as a DSS(Decision Support System) to guide decision makers at all levels of water management[21].

Gagliardi, Alvisi, Kapelan and Franchini proposed a short-term water demand forecasting method based on the use of the Markov chain. The method provided estimates of future demands by calculating Probabilities that the future demand value would fall within pre-assigned intervals covering the expected total variability[8].

Haquea and Rahmana developed a principal component regression model by combining multiple linear regression and principal component analysis to forecast future water demand in the Blue Mountains Water Supply systems in New South Wales, Australia. In addition, performances of the developed principal component regression model were compared with multiple linear regression model by adopting several model evaluation statistics such as relative error, bias, Nash-Sutcliffe efficiency and accuracy factor [7].

Yalçıntaş, Bulu, Küçükvar and Samadi collected Istanbul's water supply and demand data for the period during 2006 and 2014. Then, using an autoregressive integrated moving average (ARIMA) model, the time series water supply and demand forecasting model was constructed for the period between 2015 and 2018 [27].

Abu-libda considered a model that was based on monthly data of U.S Dollar exchange rate to shekel NIS (New Israeli Shekels) in the Palestinian territory form January 2008 to December 2017. It was through the compared two models used for accuracy predication (RMSE, MAPE, MAD) to find out the best model for analyzing the concerned data, the study found that the ARIMA model (0, 2, 1) scored the lowest values of predication accuracy[28].

Chapter 2

Basic Concepts and Methodology

2.1. Markov Chains[28], [35-36-39-40]

2.1.1 Matrix of Transition Probabilities

Stochastic processes (Random phenomena) are the processes that randomly change with time and include a big set of forms, where one of the most used forms is the usual Markov forms which belong to the statistical forms. The Markov processes perform has big importance in analyzing the stochastic processes, this rank reinforces the variety of practical applications in our daily life, as well as its applications in the statistical and engineering forms, so it became a focal subject for many organizations and researchers dealing with this topic. Historically, Markov is one of the Soviet scientist whom discovered this theory which named by his name and improved it in the stochastic processes field.

Stochastic process of distinct variables is symbolized with:

$\{X_n : n = 0,1,2,3, \ldots\}$

While stochastic process of continuous variables is symbolized with:

$\{X_t : 0 \leq t < \infty\}$.

The transition in a single status is known as the Markov chain of discrete time $\{X_n : n = 0, 1, 2, 3 \ldots\}$, which takes a specific countable number of possible values represented in a set of non-negative integers.

The probability that (X_{n+1}) in the state (j) up to the condition that(X_n) in the case $(i)$is symbolized by the symbol $P_{ij}^{n,n+1}$:

$$P_{ij}^{(n,n+1)} = P\{X_{n+1} = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\}.$$

$$P_{ij}^{(n,n+1)} = P\{X_{n+1} = j | X_n = i\}. \qquad (1\text{-}2.1)$$

Moreover, the Markov chain of discrete time is stationary or homogeneous if the transitional probabilities doesn't depend on time, i.e. the stochastic characteristics does not change with time.

The probability of X_{n+1} being in state j given that X_n is in state i is called the one -step transition probability and is denoted by $P_{ij}^{(n,n+1)}$. That is,

$$P_{ij} = P\{X_{n+1} = j | X_n = i\}. \qquad (2\text{-}2.1)$$

If (P_{ij}) represents the probability of transiting a phenomena from status (i) to status (j) in the certain period of time, and the Markov chain contains (N) of cases (where N is a positive integer), then the transitional probabilities can be placed in a matrix as follows:

$$P = \left(\begin{bmatrix} p_{11} & \cdots & p_{N1} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{bmatrix}\right) \qquad (3\text{-}2.1)$$

it is a square matrix with $(N * N)$ degree, its elements are not negative and the summation of each row in it equals 1.

The probability of transition from status (i) to status (j)in(m)steps is:

$$P_{ij}^{m} = P\{X_{n+m} = j | X_n = i\} \qquad (4\text{-}2.1)$$

If$n, m \epsilon N$ then:

$$P^{n+m} = P^n P^m \qquad (5\text{-}2.1)$$

represents the matrix of transitional probabilities of a Markov chain after $(n + m)$ of steps. The element in row (i) and column (j) in the matrix $P^{n+m}$is:

$$P_{ij}^{n+m} = \sum_{h=1}^{N} P_{ih}^{n} P_{hj}^{m}, \forall n, m > 0 \qquad (6\text{-}2.1)$$

The equation (6-2.1) is called Chapman Kolommogrov equation.

When:

$$\lim_{m \to \infty} P_{ij}^{m} = \pi_j \qquad (7\text{-}2.1)$$

The probability vector (π):

$$\pi = (\pi_1 \ , \pi_2 \ , \ldots , \pi_N)$$

$$\sum_{i=1}^{N} \pi_i = 1 \ , \qquad 0 < \pi_i < 1$$

$$\pi p = \pi$$

The Markov chain is stable and A positive recurrent, irreducible and a periodic Markov chain is called Ergodic.

2.1.2Using the Markov Model in Prediction [35]

The formula of the Markov model is based on the following procedures:

(1) Identifying Descriptive Statistical Measures and Markov Model Features:

There are four important features: the mean, the standard deviation, the twisting coefficient and the correlation coefficient, which can be found through the following formulas:

1-Mean:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (8\text{-}2.1)$$

2- Standard Deviation:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{X})^2} \qquad (9\text{-}2.1)$$

3- Twisting Coefficient:

$$g = \frac{n\sum_{i=1}^{n}(x_i-\bar{X})^3}{(n-1)(n-2)S} \qquad (10\text{-}2.1)$$

4-Correlation Coefficient:

$$r_k = \frac{n\sum_{i=1}^{n-k}(x_i-\bar{X})(x_{i+k}-\bar{X})}{(n-k)S^2} \quad ; k = 1 \qquad (11\text{-}2.1)$$

(2) Identifying Data Distribution:

In general, the natural distribution used for the data under investigation is the normal distribution, which is most commonly used in statistical applications, and before starting the model construction steps, we must verify that the time series used in the research is subject to normal distribution, and if this is not achieved, the conversion method will be used. And to know the nature of the data and how it follows the natural distribution should not be located on the straight line or close to it.

(3) Generating Random Numbers:

Random numbers are generated from the Statistical Program (Microsoft excel) or by the random number table and converted to track the normal distribution.

(4) Generating a Markov Model to Predict:

As for the monthly Markov model developed by (Thomas and Fiering), it takes the following formula:

$$q_{ij} = \bar{q}_j + b_j(q_{i-1,j-1} - \bar{q}_{j-1}) + S_j t_{ij}\sqrt{(1 - r_j^2)} \qquad (12\text{-}2.1)$$

i: The month symbol of the generated series and takes values from 1 to the length of the series

j: Month symbol in the series (From January to December)

$\bar{q}_j$: The mean of values of the month (jth) (From January to December)

b_j: Regression coefficient for values in two months jth, (j-1) th and it is calculated as follows: $b_j = \frac{r_j s_j}{s_{j-1}}$

$q_{i-1,j-1}$: Value at the month immediately before it.

$\bar{q}_{j-1}$: The mean of values the month before jth.

S_j: Standard deviation of values per month jth (From January to December).

t_{ij}: The random generated number follows the standard normal distribution.

r_j: Linear correlation coefficient for two months jth, $(j-1)th$.

(5) Forecasting Accuracy Standard:

It is intended to assess the suitability of the model for the pattern of the series data, and the accuracy of the model in predicting current and future series values.

Many measures that measure the suitability of the model depend on the degree of error, namely the difference between the actual value of the series at a given time and the string value that the model expects at that time.

In this study, we will rely on the following measurs to compare the statistical models used in the research and choose the best model, including:

1- Root Mean Squares of Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \qquad (13\text{-}2.1)$$

Where:

y_t:The real value.

$\hat{y}_t$:The expected value.

This standard is one of the most important measures of predictive accuracy of models, and in general, it is characterized by its easy statistical characteristics.

2- Mean Absolute Percentage Error (MAPE):

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} * 100 \qquad (7\text{-}2.2)$$

3- Mean Absolute Error(MAE):

$$\mathrm{MAP} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| = \frac{1}{n}\sum_{t=1}^{n}|e_t| \qquad (14\text{-}2.1)$$

Whenever the value of these criteria is small, this indicates that the expected values are close to the real values of the series data.

2.2 Time Series [20]

In the second half of the 20th century, the analysis of time series at the global level has certainly encountered a very important improvement, especially in the last three decades, and it is also certain that this improvement was due to the modern methodology presented by the scientists Bocce and Jenkins in the early 1970s, which since then has become the most accepted and common tool in the scientific, theoretical and practical means, especially in the developed world, as it has shown a high efficiency in modeling and predicting time data.

Box and Jenkins methodology is an unprecedented paradigm move in time data modeling and forecasting and the true entry point for modern analysis of time series, and has in a short period become the main reference for experts, researchers and academics inside and outside the corridors of universities, institutes, research centers and scientific consultancies as per which most recent studies are being evaluated.

The main foundations of this methodology were completed by statistical theories, numerical methods, graphical and mathematical instruments by the end of the 1970s. Forecasting has therefore been an important issue for a long time and has remained the focus of attention for researchers in other fields.

2.2.1. Time Series Analysis Objectives [20]

1- To determine the nature of the phenomenon through consecutive observations.
2- Create an interpretation model and explain the behavior of the series in terms of other variables by
3- Linking the view values to some of the rules of the series.
4- Predicting the future behavior of the chain based on past information.
5- Control the process from which the time series is generated.

2.2.2. Time Series Description [33]

The time series is usually described by graphs so as to obtaining a clear general picture of chain behavior in different time periods. In this study, the horizontal axis will represent the years 2010-2019, and the vertical axis represents the amount of water provided to Tulkarem governorate for the years mentioned and from the graphs, we realize the following:

1- The direction of the chain can be determined whether it is increasing, decreasing or constant.

2- Recognize the terminal values in the time series, repeated changes and describe these changes.

3- Knowing the relationship between the periods thru which this series has the highest (great value) and the bottom (small value).

4- To clarify the roughness of the chain and the extent to which values are related to each other.

2.2.3. Time Series Components [34]

Variables that affect the time series are classified into two categories:

First: regular changes: These changes that are repeated in the series in topics with specific characteristics, these changes are divided into:

1- General trend component, symbolized by :(Long Trend (t)).

2- Periodic component' Periodic Changes.

3- Seasonal component' Seasonal Changes: (Movements) S Seasonal.

Second: Irregular or Random Variation changes: Occasional or random changes.

2.2.4. General Trend Component [28]

The definition of a general trend component is the line that belongs to the time series data in a relatively long period, changes that occur with change in time.

The general trend determines the property of the time series increase, decrease or stability, so the general trend of the time series shows the growth or contraction of the series.

Several factors change over time and are the reason for determining the direction of the time series, these factors are:

1) Change in size, demographic characteristics and geographical distribution of population.

2) Technical developments in the country.

3) Economic developments.

4) Gradual changes in the habits and behaviors of individuals and communities.

This generally applies to any general trend line in addition to other factors that may appear with another particular series.

2.2.5. The Seasonal Component [28]

It's defined as variables that being repeated in subsequent periodical units, which occur from the external factors effect, or fluctuations being repeated in the same manner (yearly, monthly…), for instance; water and electricity consumption in the summer season.

2.2.6. The Periodical Component

This component is reflected in the long term time series, which expresses the financial situations transmittance, and it's defined as variations that are similar to the seasonal components, only it's being occurred in periods longer than the seasonal components, and in comparison with the seasonal components, the time period span is unknown, but it's ranges between 3 to 10 years, and thus it's difficult to identify the periodical fluctuations and their estimations, because it differs from a period to another.

2.2.7. Random Components

It expresses the irregular oscillations, in other words, it is the atypical variations occurs as a result of unpredictable urgent circumstances, for instance; earthquakes and labor strike.

2.2.8. Box-Jenkins Methodology in Time Series Analysis [20]

Time series analysis using the single variable ARIMA method is a method of extracting the expected variables of the data, where the time series is partitioned to many components called three linear filters; integrated filter, autoregressive filter and moving averages.

ARIMA methods are applied only on stationary series.

2.2.9. Time Series Stationaries [34]

To use the time series in analysis and prediction, the series must be in a stationary mode, the series is called stationary under the following conditions:

1- Constant mean

$E(Y_t) = \mu$ (1-2.2)

2- Constant variance

$var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$ (2-2.2)

3- Autocorrelation function dependent on the $lag\ (s, t)$ only $(t > s)$

$\rho_{(s,t)} = \frac{cov(Y_t, Y_s)}{var(Y_t)} = \frac{E(Y_t - \mu)(Y_s - \mu)}{\sigma^2}$ (3-2.2)

The autocorrelation depends only on the $lag\ (t - s)$

Proper transformation is done to the time series in case of non-constant variance, and there are many ways for transformation, (logarithmic transformation) is one of which the square root.

Differences method is also used to remove the general trend effect from the time series, until the time series is stationary.

Some tests that are used to make sure the time series is stable:

1. Fuller-dickey test DF

2. Augmented Dickey-Fuller Test ADF

3. Phillips and Perron Test

2.2.10. Auto-Correction Function[28],[33]

Autocorrelation function explains the correlations between the variables for different intervals, and it is a scale to measure the strength of internal correlation of the time series, as it is possible to distinguish the stationary time series from the non-stationary thru the autocorrelation functions factors values.

The autocorrelation function ACF is an important mean to determine the stationaries of the time series, as it tends to decay quickly towards zero with the increase of displacement intervals (k), or gets disconnected after numbers of displacement intervals $(k = q)$, i.e.:

$$\rho_k = 0 \quad ; \quad \forall k > q$$

Because the autocorrelation function is only estimations of autocorrelation, then it is possible for their numbers to be small but not zero's i.e.:

$$r_k \neq 0 \quad ; \quad \forall k > q$$

But when the time series is non-stationary because of an upward or downward trend in the average, then the autocorrelation function of the sample doesn't get disconnected or decayed slowly towards zero, this is because the variables tend to be on the same direction of the mean of the time series for many intervals, and as a result large correlations are obtained for long displacement intervals.

2.2.11. Partial Autocorrelation Function

Partial Autocorrelation Function Represents the relation between consequent values of some variable thru two different time p_{k1}, intervals, assuming constancy of other intervals, symbolized with p_{k1}, the factor of partial autocorrelation function between $Y_{t-k}$and Y_t indicates the correlation of both excluding values of Y_t that lie between the two intervals $(t-k)$, (t).

PACF used as a major mean in analysis of ARIMA models in addition to the ACF, where these two functions are used together to distinguish the ARIMA different models.

PACF of the stationary time series tends to quickly decay to zero with the increase of displacement intervals or disconnects after a certain number of the displacement intervals k.

2.2.12. Auto Regression and Integrated Moving Averages ARIMA[28],[34-37-38]

Box and Jenkins contributions showed in 1976 in time series field, where they derived the model with many advantages in analysis of time series with description of models in a whole manner, and they put a style for the correlated information with the stationaries and its processing of data, and they reached to the models called ARIMA (p, d, q).

Models (ARIMA) are generated from auto regression models and moving average models after taking the appropriate differences to make the time series stable, and these models are symbolized by ARIMA(p, d, q):

P: Represents the degree of auto-regression

d: Represents the number of differences needed when making the time series stable

q: Represents the degree of moving averages

ARIMA model is considered one of the most common models in predicting the economic variables values, e.g: prices of certain goods.

When the time series is not stationary, it must first be transformed to stationary, by taking the differences (d) or using one of the transformations and number of differences to transform the series into stationary called by its integration degree, as the model ARMA is transformed into ARIMA on the following form:

$\emptyset_p(B)y_t = \mu + \theta_q(B)\varepsilon_t$ (4-2.2)

$\emptyset_p(B)w_t = \theta_q(B)\varepsilon_t$ (5-2.2)

$y_t = (1-B)^d y_t$ (6-2.2)

Where:

$\emptyset_p(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$ (7-2.2)

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ (8-2.2)

Whereas:

$$w_t = \begin{cases} \nabla^d Y_t = (1-B)^d y_t, & d > 0 \\ Y_t & , d = 0 \end{cases}$$

y_t: Dependent variable represents the value at time t.

$\emptyset_1, \emptyset_2, \dots, \emptyset_p$: The autoregerssion model parameter.

$\theta_1, \theta_2, \dots, \theta_q$: The partial autoregerssion model parameter.

ε_t: The random error at time t, which the model does not explain.

2.2.13. Stages of Applying ARIMA Models in Predictions [28]

ARIMA Models are constructed thru the following stages:

(1) Data Configuration Stage**:**

It is the 1st stage to build ARIMA models, so if the data are stationary thru the graphs and the auto and partial correlations of them, then the data are ready and set for the next stage.

Whereas if the data were non-stationary in mean and variance, then these non-stationaries is processed in the mean by taking the 1st difference, and if it did not get stationary, the 2nd difference is then taken, but it often gets stationary after the 1st or 2nd differences. As for the non-stationaries in the variance, it's processed thru the proper transformation of data (the logarithmic or square root transformations) are the most common ways to do so.

(2) Identification Stage [19]:

After achieving the stationaries of the series, model identification stage starts, which means using the data or any other information about how the series is generated thru which, so the aim here is to get an idea of q, d, p values that is needed in the ARIMA models, and then getting basic estimations for the models' parameters.

The two means used to determine the model and its rank are ACF and PACF, and then their factors are matched with theoretical behavior, so if:

Table(1): Properties of the auto-regression function and the partial auto-regression function of the models (ARIMA)

Models	$\boldsymbol{\rho(k)}$	$\emptyset_{kk}$
AR(1)	Gradually approaching zero form exponentially or oscillating in the signal	Completely disconnected after the first lag time
AR(2)	Gradually approaching zero Form exponentially or oscillating in the signal or Sine waves.	Completely disconnected after the second lag time
AR(P)	Gradually approaching zero form exponentially or oscillating in the signal or sine waves.	Completely disconnected after p lag time
MA(1)	Completely disconnected after the first lag time	Gradually round to zero in exponential form
MA(2)	Completely disconnected after the second lag time	The sum of two functions gradually approaching zero in exponential form or waves simulating the sine function.
MA(q)	Completely disconnected after q lag time	A combination of functions gradually approaching zero in exponential form or waves simulating the sine function.
ARMA(p,q)	Gradually approaching zero after the first $(q-p)$ form exponentially or oscillating in the signal or sine waves	After the first $(q-p)$ of the lags is a combination of functions approaching zero in exponential form or waves from the sine function.

(3) Estimation Stage [19] [28]:

It's a difficult stage, aims to estimate the parameters of the model, means getting numerical values for the selected model in the identification stage, using one of the following estimation ways:

1- Yule- walker method which depends on the autocorrelation of the model, called moment method. (MM)
2- Maximum likelihood method, depends on signification of residues squares function $S\ (\varphi, M, \theta\ )$ to the minimum, and it's the most common way.
3- Non-linear least squares method, which works on minimizing the summation of residues squares estimation.

Diagnoses Stage [20]:

This stage is one of the most important stages, because the described and estimated model acceptance is determined is based on this stage. And according to this stage it's determined whether to continue in analysis and prediction processes or to go back to the 1st stage to select and estimate another model.

To test how suitable the model is, the residues analysis test is done, to make sure that the random variable(α_t)is independent and has a normal distribution with zero mean and variance:

$(\alpha_t \sim IIN(0, \sigma_a^2))$

Other tests:

1- Box-Jenkins (q) test

2- Residues randomness test (Runs test)

(4) Forecasting or Prediction Stage:

In this stage, the future values of the time series are obtained thru the suitable model from the last stages. The perfect prediction is the one by which the smallest residue and variance value is obtained.

(5) Criteria for Choosing a Model Rank:

To confirm the correctness of the model rank (ARIMA (p,d,q) a set of statistical criteria have been adopted that help in the trade-off between the nominated models, where the best model that has the lowest value of these criteria is chosen ,from these criteria:

1- Akaike criterion (AIC).

2- Bayesian criterion (BIC).

In our study, we will rely on these criteria to compare the proposed ARIMA models.

2.3. Regression Analysis [19][20-30-31]

Regression analysis is a statistical tool that establishes a statistical model to estimate the relationship between one quantum variable, which is the dependent variable, and another quantum variable, or several quantum variables, which are independent variables.

This relationship produces a statistical equation that reflects the relationship between variables.

This equation can be used to determine the type of relationship between variables and the estimation of the dependent variable using other variables.

When the relationship in the statistical model comes between a dependent variable and an independent variable, this model is the simplest regression model, which is the simple linear regression model.

But when several independent variables are more than one quantum variable, then the model is called the multiple linear regression model.

$$\mathrm{Y_i} = \beta_0 + \beta_1 \mathrm{X_{1i}} + \beta_2 \mathrm{X_{2i}} + \cdots \dots \dots \dots . + \beta_k \mathrm{X_{ki}} + \varepsilon_i \;, \quad for\ i = 1,2, \dots, n \qquad (1\text{-}2.3)$$

Whereas:

Y_i: Variable values of the dependent amiable.

$\beta_0, \beta_1\beta_2 \ldots \ldots \ldots \ldots \ldots \beta_k$: Parameters of the unknown model to be estimated.

$X_{1i}, X_{2i} \ldots \ldots \ldots X_{ki}$: values of the Independent variables

k: Number of independent variables.

n: Sample size.

Equation (1-2.3) can be written in the style of arrays as follows:

$$\begin{pmatrix} y1 \\ y2 \\ . \\ . \\ . \\ yn \end{pmatrix} = \begin{pmatrix} 1 & x11 & x21 \ldots & xN1 \\ 1 & x12 & x22 \ldots & xN2 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & x1k & x2k \ldots & xNk \end{pmatrix} \begin{pmatrix} \beta0 \\ \beta1 \\ . \\ . \\ . \\ \beta k \end{pmatrix} \begin{pmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \\ . \\ . \\ . \\ \boldsymbol{e_n} \end{pmatrix} \qquad (2\text{-}2.3)$$

Briefly as follows:

$$Y = X\beta + \varepsilon \qquad (3\text{-}2.3)$$

Whereas:

X: A class data matrix $n * (k + 1)$ containing views of independent variables $X_1, X_2, X_3, \ldots, X_k$ where the first column contains the values of the integer 1 to represent the fixed coefficient β_0.

β: Class vertical vector $(\mathrm{k}+1)*1$contains the parameters of the unknown Regression model $\beta_0, \beta_1, \ldots, \beta_k$ to be estimated.

ε : A class vertical vector that represents the error term.

2.3.1. Model Assumptions [19]

To use the usual Least squares method to estimate the parameters of the multiple linear regression analysis model, it depends on several basic hypotheses:

General assumptions:

1.The dependent variable y is a linear function in k of independent variables

2. nonexistence of linear overlap between independent variables, which means that the columns of matrix X are linearly independent.

3- Independent variables should be free of summation errors

4- The relationship to be assessed must be already identified and diagnosed

5. None existence of errors in measuring independent variables

6- Independent variables should be non-random, i.e. they contain constant values in repetitive previews and in order for the values of independent variables to be non-random, the researcher must control them empirically.

7. The model should include independent variables that contribute to the interpretation of the dependent variable

2.3.2. Random Error Term Obligations

1- The average random error term ε_i should be equal to zero, i.e.

$$E(\varepsilon_i) = 0\,, \quad \text{for all } i = 1,2,3,\dots\dots.., n \qquad (4\text{-}2.3)$$

2-The existence of the independence property between the ε error term and independent variables i.e.

$$E\left(\varepsilon_i, X_{ij}\right) = 0 \quad ,\text{for all} \quad i \neq j\ \&\ i\,, j = 1,2,3,\dots\dots\dots, n \qquad (5\text{-}2.3)$$

3- Fixed variation error term ε and the independence of error terms values from each other i.e.

$$Var(\varepsilon_i) = E\left(\varepsilon_i^2\right) = \boldsymbol{\sigma^2}\varepsilon \quad ,\text{for all} \quad i = 1,2,3,\dots\dots\dots\dots, n \qquad (6\text{-}2.3)$$

$$Cov\left(\varepsilon_i, \varepsilon_j\right) = 0 \quad , i \neq j$$

Random Variable ε_i is distributed according to the normal distribution with an average zero and constant variance:

$$\varepsilon_i \rightarrow N(0, \boldsymbol{\sigma^2})$$

2.3.3. Simple Linear Regression

A simple linear regression is a statistical tool used to determine the relationship between two quantum variables so that the value of the dependent variable y, depends on the value of a single independent variable x.

There are many regression models that may number up to hundreds of models but the most important and most common model in use is the simple linear regression model, in this model we have a y variable also known as the dependent variable and the independent variable x also known as the predictor variable, and the mathematical formula of the model as follows:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , \mathrm{i} = 1,2, \dots , \mathrm{n} \qquad (7\text{-}2.3)$$

Whereas:

y_i: the value of Dependent variable, for observation i.

x_i:the value of the Independent variable, for observation i.

β_0: The regression constant is the part that intersects the vertical axis y that reflects the value of the dependent variable y in the absence of a value for the independent variable x, in other words, when$x = 0$

β_1: The regression coefficient (slope) is the amount of change in y if x changes by one unit.

ε_i: Represents a random error that indicates the difference between the actual value of the y variable and the estimated value it symbolizes $\hat{y} = \beta_0 + \beta_1 x$, meaning that the random error equals $\varepsilon = y - (\beta_0 + \beta_1 x)$.

The model assumes that ε a random variable follows the normal distribution with an average of value 0 and a constant variationσ^2 for different values x ,so :

$$\varepsilon_i \sim N(0, \sigma^2) \qquad (8\text{-}2.3)$$

The parameters of the model can be estimated using the least-squares method.

2.3.4. Estimate of the Simple Linear Regression Model

There are several ways to estimate or calculate the simple linear regression model and all methods depend on the regression coefficients values calculation of (β_0 and β_1), and the least squares method is one of the best because it makes the sum of random error squares as little as possible, and to calculate the estimated simple regression equation value of the variable y as a function of the independent variable x, applied the following equation:

$$\hat{y}_\mathrm{i} = \hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i \qquad (9\text{-}2.3)$$

calculate the value of $\hat{\beta}_1$ as follows:

$$\hat{\beta}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \qquad (10\text{-}2.3)$$

calculate the value of $\hat{\beta}_0$ as follows:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (11\text{-}2.3)$$

2.3.5. Stages of the Formation of the General Linear Model

(1) The Stages of the Formation of the General Linear Model:

Identify the phenomenon of prediction: identifying its concept and its aftermath in time and space.

Determining the shape of the model: i.e. formation the propagation shape in order to confirm or deny the link between the points and determine the shape of the model i.e. determine the shape of the function is linear or non-linear.

Estimating the parameters of the model: The least-square method is the best way to estimate the parameters of regression models due to the accuracy of its results.

(2) Interpretation of Estimated Parameters:

It is interpreted ($\hat{\beta}_0$) as the location of the axis regression level intersection y and equal to the average dependent variable when all the values of the independent variables are equal to zero and ($\hat{\beta}_j$)as the partial regression coefficient of y on x_i when the rest of the independent variables are fixed and represent the amount of change that occurs in the expected value of the dependent variables as a result of the change of the variablex_i by one unit keeping all the residual independentvariables constant. (e.g. ($\hat{\beta}_1$)the change in the average value of the variable Y resulting from change in the variablex_1 one unit assuming the stability of the other variables (x_1 ,x_2 ,…,x_n). similarly,we explain the rest of the partial regression coefficients.

Positive coefficient indicates a direct relationship between the independent and the dependent variable, while the negative coefficient indicates that the relationship between the two variables is inverse.

The size of the coefficient also indicates the amount of change in the dependent variable resulting from an increase of one unit in the corresponding independent variable, assuming that the values of all other independent variables are consistent.

2.3.6. Overall Moral Test of the Model

A full test of the model's moral means testing the effect of independent variables (x_1 ,x_2 ,…,x_k) combined on the dependent variable, in other words, whether all the independent variables included in the regression model affect a moral level in predicting the values of the dependent variable and to answer this question the following hypothesis can be tested:

H_0: $\beta_1 \ldots = \beta_K = 0$

H_1: at least one β_i is not zero for i=1,2,…,k

The above hypothesis is tested using the F test, which takes the following formula:

$$F = \frac{SSR/K}{SSE/n-k-1} = \frac{R^2/K}{(1-R^2)/(n-k-1)} \qquad (12\text{-}2.3)$$

Where:

SSR: Represents the sum of slope squares

SSE: Represents the sum of the square residuals.

R^2: The coefficient of determination

K: Number of independent variables

n: number of observation.

After calculating the value of F, it compares with the F- table value $F_{\alpha,k,n-k-1}$, So if it's $F_{\alpha,k,n-k-1} \geq F_{cal}$we don't reject the null hypothesis H_0.

This means that there is no moral effect by the combined independent variables on the Dependent variable.

If $F_{\alpha,k,n-k-1} < F_{cal}$, we reject the null hypothesis and conclude the H1 alternative, which means that the combined independent variables affect the dependent variables .

2.3.7. Test of Partial Regression Coefficients ($\hat{\boldsymbol{\beta}}_J$)

If there is a Significant effect by the combined independent variables on the dependent variable, the effect of each independent variable can be tested in the presence of other independent variables, , this means testing the following hypothesis:

H_0: $\beta_j = 0$, $j = 1,2, \dots, k$

H_1: $\beta_j \neq 0$

(T-test) It is used to test this hypothesis, which is known as the following formula:

$$t_j = \frac{\widehat{\beta}_j}{S.E\ (\widehat{\beta}_j)} = \frac{\widehat{\beta}_j}{\sqrt{Var(\widehat{\beta}_j)}} \quad , \mathrm{j} = 1, 2, \dots, \mathrm{k} \qquad (13\text{-}2.3)$$

Whereas:

$S.E\ (\hat{\beta}_j)$: Represents the standard error of the estimator $(\hat{\beta}_j)$.

The calculated (t- value) is compared with the table value $t_{\frac{\alpha}{2}, n-k-1}$ and if the calculated is greater than the table value, it means rejecting the null- hypothesis and don't rejects the alternative hypothesis, i.e. there is a significant effect by the independent variable x_j on the dependent variable in the presence of the other independent variables.

If the calculated (t- value) is lower than the table value, this means that the independent variable x_j has no significant effect on the dependent variable in the presence of the other independent variables.

2.3.8. Forecasting

At this stage, the previous values of the phenomenon are plugged in the estimated model to obtain the predicted values.

2.4. Method of Work

To achieve the objective outlined for this study, the following steps of work were followed:

2.4.1. Data Source and Processing

Data was acquired, (monthly amount of supply and demand on water in Tulkarm city) from the municipality of Tulkarm, as we attained the monthly data for the years 2018, 2019, and only five months of 2020. However, due to covid-19 obstacles, we were not able to collect the data appropriately.

As a result of insufficient data collection from Tulkarm municipality-water department (29 values), and in order to be able to implement the three models "time series, regression analysis, and the Markov chains" we generated extra monthly data values from year 2010 to year 2017 (96 additional values), by which total values of 125 records has been reached; 120 of which were used to generate the three previously mentioned models and the other 5 from year 2020 were used to validate the estimated forms.

2.4.2. Statistical Methods

The main idea in this study concentrated on comparison between statistical analysis types for predicting the amount of supply and demand of water in Tulkarm city in the year 2030, the statistical techniques were selected in the following forms: Markov chains form, Time series form (ARIMA) in addition to regression analysis form as these forms are applicable on such studies and approved as ideal technique in analysis of phenomenon being studied.

One of the most common used statistical type for prediction is ARIMA, as it was successfully used in prediction of linear time series analysis. This method was developed by Box and Jenkins.

Regression analysis form is also considered one of the common statistical techniques, as it's considered a very important statistical technique, which clearly determines the relationship between dependent and independent variables in the form of equation. This analysis can be inferred to differentiate the importance, strength and direction of this relation, it also reveals the estimation of response and prediction of this relation for what is too important in planning and decision making in its regard.

With regard to the Markov series forms, it's considered one of the mathematical model that has applications in so many different fields, and it's one of the practical stochastic applications which helps describing, monitoring and predicting, which depends in its structure on statistical technique, with increases accuracy in description the studied phenomenon.

Statistical software "SPSS -E-Views – Matlab   Excel" were all used in analyzing the study data.

Chapter3

Data Analysis and Results Discussion

3.1 Results Overview

In this chapter the results of the practical aspect of the study will be presented, through the detailed demonstration for the steps of statistical analysis of the studied data, which is the amount of supply and demand for water in Tulkarm municipality during the time period 2010-2020.

The chapter begins with a simplified statistical description of the studied data through statistical measures and graphs in order to give an overview of the nature of the data to be represented using statistical models (Markov chains, time series (ARIMA) and regression analysis). This is based on statistical software (SPSS), (E Views) and (Microsoft Excel).

Finally, we present some statistical criteria (RAMSE-MAPE-MAE) in order to compare the resulting statistical models and to find out the forecast accuracy of each model.

3.2 Data Analysis

3.2.1 Data Source

Data source was acquired (amount of supply and demand on water in Tulkarm city monthly) from the municipality of Tulkarm. These data were the monthly data for the years 2018, 2019, and only five months of 2020 as shown in table (2):

Table(2): supply and demand for water in the municipality of Tulkarem in the period (2018to 2020)

Month	Demand			Supply		
	2020	**2019**	**2018**	**2020**	**2019**	**2018**
Jan	342116	325261	355628	552465	595379	585333
Feb	329008	335053	325052	548691	568120	588242
Mar	367220	367566	366762	598768	619416	612514
Apr	372342	377204	374223	637289	642554	641932
May	416814	379944	412416	702951	727326	717864
Jun		392458	401228		751413	744319
Jul		405507	443773		767656	762135
Aug		458690	427765		758925	762183
Sep		496451	424208		752563	759489
Oct		471374	411882		737737	742549
Nov		450015	404332		720426	718011
Dec		358512	384651		658395	611725

Source: Department of Water-municipality of Tulkarem

3.2.2 Data Processing

Because there was data deficit for demand and supply acquired from Tulkarm municipality-water sector (29 values). In order to apply the three mentioned models, time series, regression analysis and the Markov chains, we generated extra monthly data values from year 2010 to year 2017 (96 values), using Microsoft Excel and according to the following equation:

$$y = 6 * (1.27x + x)$$

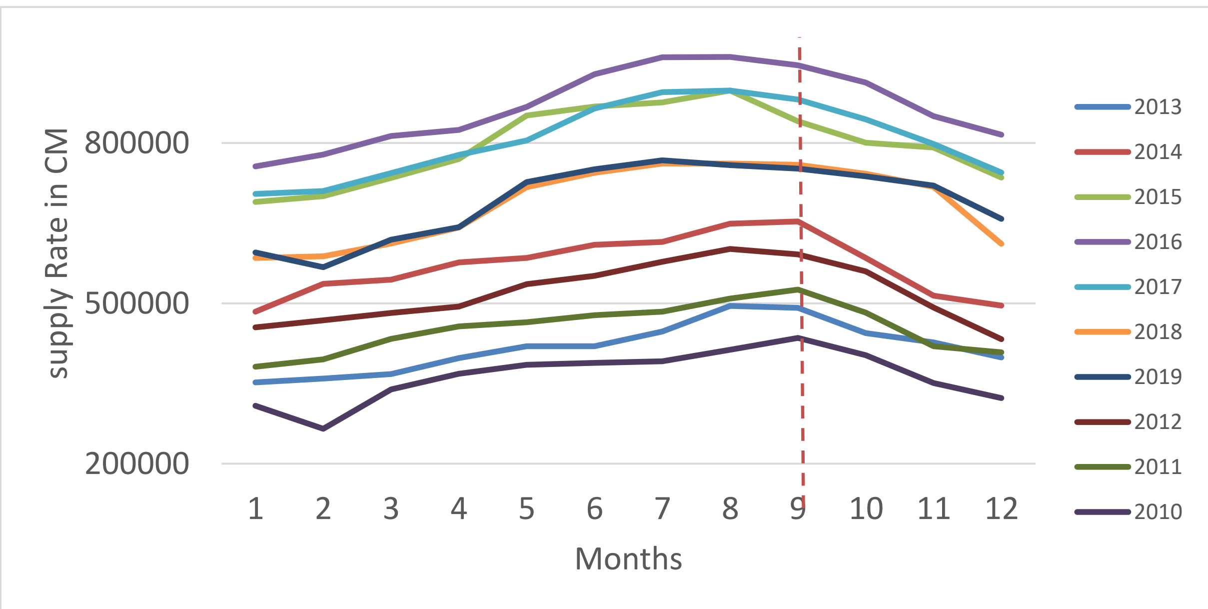
y = Annual value of the amount of water (Supply /demand).

x = Amount of water supply/demand in winter.

to reach a total number of values 125 as shows in tables (3 & 4) and charts (3 & 4); from which 120 were used to generate the three previously mentioned models. The rest 5 values from year 2020 were used to validate the estimated models.

Table (3): Represents monthly amount of water supply in the period (2010-2020)

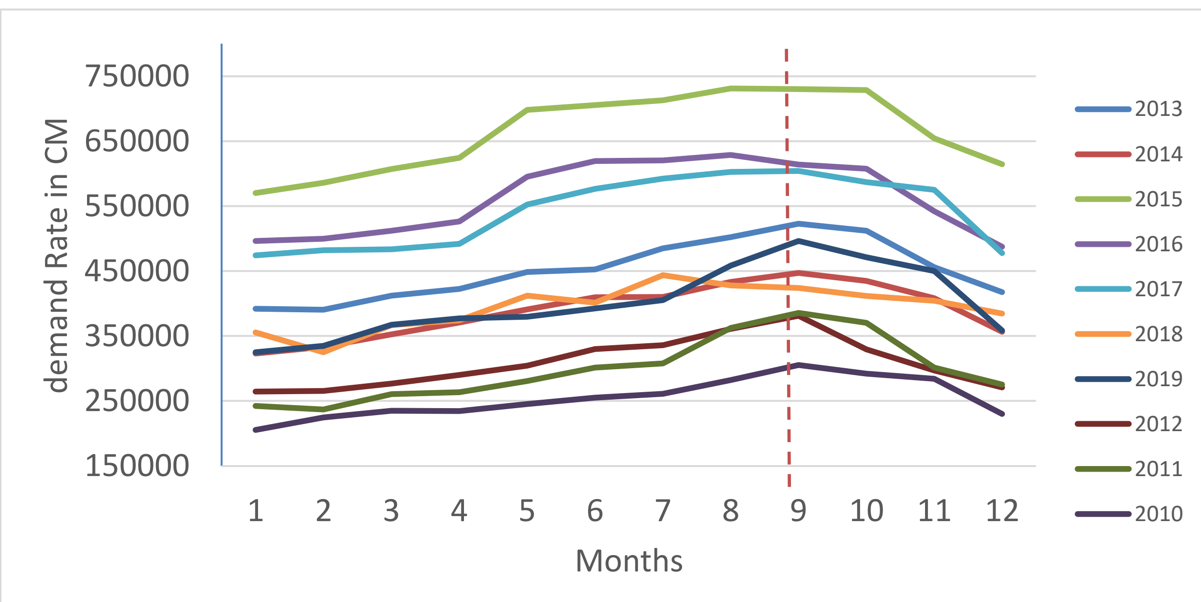
Month	Supply										
	2020	**2019**	**2018**	**2017**	**2016**	**2015**	**2014**	**2013**	**2012**	**2011**	**2010**
Jan	552465	595379	585333	704846	756241	690162	484582	352423	455213	381791	308370
Feb	548691	568120	588242	710589	778959	701025	536532	359360	468287	395356	265396
Mar	598768	619416	612514	744060	813692	734705	544416	368014	482480	433635	339034
Apr	637289	642554	641932	778198	824657	770247	577179	398014	494512	456955	368588
May	702951	727326	717864	805049	867905	851445	585349	420004	536278	464844	385417
Jun	.	751413	744319	864928	929148	868626	609802	420142	551539	478006	388766
Jul	.	767656	762135	895155	960426	876505	615419	447577	578121	484875	391629
Aug	.	758925	762183	898219	961316	898322	649195	495486	602118	509311	413148
Sep	.	752563	759489	881662	945954	840802	653307	491972	592118	525985	435835
Oct	.	737737	742549	844589	913692	800774	585219	444391	560416	483071	403230
Nov	.	720426	718011	798219	850504	792046	514679	427118	492568	420041	351312
Dec	.	658395	611725	745136	815978	735505	496307	398821	432908	408529	323230



Graph (3): represents supply rate [cm] by of time [months]

Table (4): Represents monthly amount of water demand in the period (2010-2020)

Month	Demand										
	2020	**2019**	**2018**	**2017**	**2016**	**2015**	**2014**	**2013**	**2012**	**2011**	**2010**
Jan	342116	325261	355628	474074	496296	570370	323054	391791	264317	242291	205580
Feb	329008	335053	325052	482136	499642	586104	333214	390471	265715	237007	224407
Mar	367220	367566	366762	483498	512257	607105	352675	412375	276771	260385	235017
Apr	372342	377204	374223	491967	526192	624305	370798	422464	289870	263366	234212
May	416814	379944	412416	552274	595481	698371	391146	448811	304135	280533	245456
Jun	.	392458	401228	576705	619596	705798	409768	452683	329866	301345	255055
Jul	.	405507	443773	592593	620370	712963	410279	484876	335682	307709	261086
Aug	.	458690	427765	602669	628764	731346	433555	502410	360790	362629	281991
Sep	.	496451	424208	604372	614308	730403	447273	522947	381180	385586	305262
Oct	.	471374	411882	586740	607506	728614	434996	512282	329588	370533	291991
Nov	.	450015	404332	575180	541964	654671	408136	456255	296781	301287	283950
Dec	.	358512	384651	477491	487549	614718	356378	417443	271122	275482	230215



Graph (4): represents demand rate [cm] by of time [months]

3.2.3 Descriptive Statistic

Table (5) shows statistical standards of time series (demand /supply); where it's clear that the amount of supplied water through the time interval (265396 - 961316) with a mean value of (617780.93) and standard deviation of(179138.816), whereas for amount of water demand it's clear that it ranges between (205580-731346) with a mean of (426655.158) and a standard deviation of (131827.987).

Table (5): Descriptive statistics of water supply/demand data

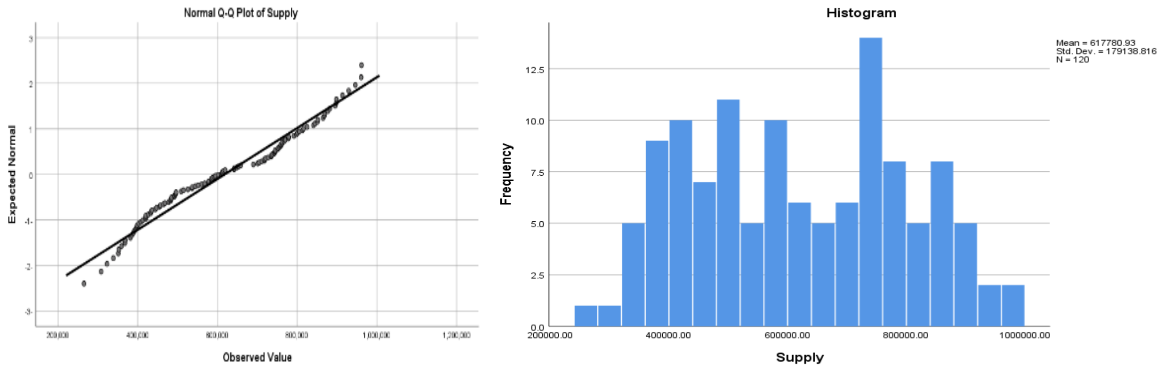
	N	Mean	Std. Deviation	Minimum	Maximum
Supply	120	617780.933	179138.816	265396	961316
Demand	120	426655.158	131827.987	205580	731346

3.2.4 Data Distribution Recognition

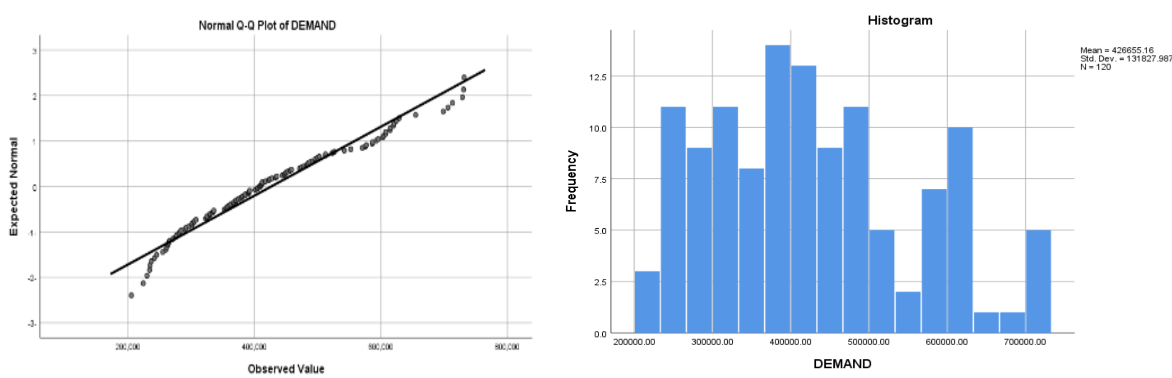
To validate if the data, follow normal distribution or not, we used (Kolmogrov-Smirnov test), where:

H_0 : sample data following the normal distribution

H_1 : sample data do not follow the normal distribution



Graph (5): represents normal distribution test of supply rate [cm]



Graph (6): represents normal distribution test of demand rate [cm]

Table (6): showing the results of kolmogrov-smirov test.

	k-s	N	Sig
Supply	0.101	120	**0.04**
Demand	0.085	120	**0.034**

Table (6) shows that both of P-value (supply) = 0.04 and P- value (Demand) = 0.034 are less than level of significance value 0.05, this means the rejection of null hypothesis, i.e. data does not follow the normal distribution.

3.3.1 Behavior of the Water Supply in the Long Run

In order to follow the behavior of the water supply at long run, the following steps were followed:

Data related to the amount of water supplied for the domestic sector in Tulkarm municipality and a sample of (120) months was taken, so that it dealt with production of the transition matrix.

Table (7): Represents the amount of water supply from 2010-2019

month	supply	month	supply	Month	Supply	month	supply	month	Supply
Jan-10	308370	**Jan-12**	455213	**Jan-14**	484582	**Jan-16**	756241	**Jan-18**	**585333**
Feb-10	265396	**Feb-12**	468287	**Feb-14**	536532	**Feb-16**	778959	**Feb-18**	**588242**
Mar-10	339034	**Mar-12**	482480	**Mar-14**	544416	**Mar-16**	813692	**Mar-18**	**612514**
Apr-10	368588	**Apr-12**	494512	**Apr-14**	577179	**Apr-16**	824657	**Apr-18**	**641932**
May-10	385417	**May-12**	536278	**May-14**	585349	**May-16**	867905	**May-18**	**717864**
Jun-10	388766	**Jun-12**	551539	**Jun-14**	609802	**Jun-16**	929148	**Jun-18**	**744319**
Jul-10	391629	**Jul-12**	578121	**Jul-14**	615419	**Jul-16**	960426	**Jul-18**	**762135**
Aug-10	413148	**Aug-12**	602118	**Aug-14**	649195	**Aug-16**	961316	**Aug-18**	**762183**
Sep-10	435835	**Sep-12**	592118	**Sep-14**	653307	**Sep-16**	945954	**Sep-18**	**759489**
Oct-10	403230	**Oct-12**	560416	**Oct-14**	585219	**Oct-16**	913692	**Oct-18**	**742549**
Nov-10	351312	**Nov-12**	492568	**Nov-14**	514679	**Nov-16**	850504	**Nov-18**	**718011**
Dec-10	323230	**Dec-12**	432908	**Dec-14**	496307	**Dec-16**	815978	**Dec-18**	**611725**
Jan-11	381791	**Jan-13**	352423	**Jan-15**	690162	**Jan-17**	704846	**Jan-19**	**595379**
Feb-11	395356	**Feb-13**	359360	**Feb-15**	701025	**Feb-17**	710589	**Feb-19**	**568120**
Mar-11	433635	**Mar-13**	368014	**Mar-15**	734705	**Mar-17**	744060	**Mar-19**	**619416**
Apr-11	456955	**Apr-13**	398014	**Apr-15**	770247	**Apr-17**	778198	**Apr-19**	**642554**
May-11	464844	**May-13**	420004	**May-15**	851445	**May-17**	805049	**May-19**	**727326**
Jun-11	478006	**Jun-13**	420142	**Jun-15**	868626	**Jun-17**	864928	**Jun-19**	**751413**
Jul-11	484875	**Jul-13**	447577	**Jul-15**	876505	**Jul-17**	895155	**Jul-19**	**767656**
Aug-11	509311	**Aug-13**	495486	**Aug-15**	898322	**Aug-17**	898219	**Aug-19**	**758925**
Sep-11	525985	**Sep-13**	491972	**Sep-15**	840802	**Sep-17**	881662	**Sep-19**	**752563**
Oct-11	483071	**Oct-13**	444391	**Oct-15**	800774	**Oct-17**	844589	**Oct-19**	**737737**
Nov-11	420041	**Nov-13**	427118	**Nov-15**	792046	**Nov-17**	798219	**Nov-19**	**720426**
Dec-11	408529	**Dec-13**	398821	**Dec-15**	735505	**Dec-17**	745136	**Dec-19**	**658395**

- **Modelation the Transition Matrix**

The transition matrix is 3*3, because there are three cases in which the rate of water supply is affected with (increase-stability-decrease).

From this, the transition probability matrix will be as follows:

$$P_{ij} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}$$

This matrix reflects changes in the amount of water supply (increase-stability- decrease) as:

P_1: Represents the state of increased amount of water supply.

P_2 :Represents the stability of the amount of water supply.

P_3 :Represents a decrease in the amount of water supply.

P_{11} :Represents the probability of increasing in the amount of water supply after the previous increase.

P_{12} :Represents the probability of increasing in the amount of water supply after it was stable.

P_{13}:Represents the probability of increasing in the amount of water supply after it decreased.

P_{21}: Represents the probability of stability in the amount of water supply after it increased.

P_{22}: Represents the probability of stability the amount of water supply after it was stable.

P_{23}:Represents the probability of stability in the amount of water supply after it decreased.

P_{31}: Represents the probability of decreasing in the amount of water supply after it increased.

P_{32}: Represents the probability of decreasing in the amount of water supply after it was stable.

P_{33}: Represents the probability of decreasing in the amount of water supply after it decreased.

- **Determine the Transition Matrix (P)**

To determine the transition matrix (P) Related to the amount of water supply during the period (2010-2019) the following steps has been taken:

Table (8): Represents the labeled and counted of water quantity data supply during the period (2010-2019)

Month	Supply	States	Transition counts
Jan-10	308370	P1	
Feb-10	265396	p3	**p13**
Mar-10	339034	p1	**p31**
Apr-10	368588	p1	**p11**
May-10	385417	p1	**p11**
Jun-10	388766	p2	**p12**
Jul-10	391629	p1	**p21**
Aug-10	413148	p1	**p11**
Sep-10	435835	p1	**p11**
Oct-10	403230	p3	**p13**
Nov-10	351312	p3	**p33**
Dec-10	323230	p3	**p33**
Jan-11	381791	p1	**p31**
Feb-11	395356	p1	**p11**
Mar-11	433635	p1	**p11**
Apr-11	456955	p1	**p11**
May-11	464844	p1	**p11**
Jun-11	478006	p1	**p11**
Jul-11	484875	p1	**p11**
Aug-11	509311	p1	**p11**
Sep-11	525985	p1	**p11**
Oct-11	483071	p3	**p13**
Nov-11	420041	p3	**p33**
Dec-11	408529	p3	**p33**

Jan-12	455213	p1	**p31**
Feb-12	468287	p1	**p11**
Mar-12	482480	p1	**p11**
Apr-12	494512	p1	**p11**
May-12	536278	p1	**p11**
Jun-12	551539	p1	**p11**
Jul-12	578121	p1	**p11**
Aug-12	602118	p1	**p11**
Sep-12	592118	p3	**p13**
Oct-12	560416	p3	**p33**
Nov-12	492568	p3	**p33**
Dec-12	432908	p3	**p33**
Jan-13	352423	p3	**p33**
Feb-13	359360	p2	**p32**
Mar-13	368014	p1	**p21**
Apr-13	398014	p1	**p11**
May-13	420004	p1	**p11**
Jun-13	420142	p2	**p12**
Jul-13	447577	p1	**p21**
Aug-13	495486	p1	**p11**
Sep-13	491972	p2	**p12**
Oct-13	444391	p3	**p23**
Nov-13	427118	p3	**p33**
Dec-13	398821	p3	**p33**
Jan-14	484582	p1	**p31**
Feb-14	536532	p1	**p11**
Mar-14	544416	p1	**p11**
Apr-14	577179	p1	**p11**
May-14	585349	p1	**p11**
Jun-14	609802	p1	**p11**
Jul-14	615419	p1	**p11**
Aug-14	649195	p1	**p11**
Sep-14	653307	p1	**p11**
Oct-14	585219	p3	**p13**

Nov-14	514679	p3	**p33**
Dec-14	496307	p3	**p33**
Jan-15	690162	p1	**p31**
Feb-15	701025	p1	**p11**
Mar-15	734705	p1	**p11**
Apr-15	770247	p1	**p11**
May-15	851445	p1	**p11**
Jun-15	868626	p1	**p11**
Jul-15	876505	p1	**p11**
Aug-15	898322	p1	**p11**
Sep-15	840802	p3	**p13**
Oct-15	800774	p3	**p33**
Nov-15	792046	p3	**p33**
Dec-15	735505	p3	**p33**
Jan-16	756241	p1	**p31**
Feb-16	778959	p1	**p11**
Mar-16	813692	p1	**p11**
Apr-16	824657	p1	**p11**
May-16	867905	p1	**p11**
Jun-16	929148	p1	**p11**
Jul-16	960426	p1	**p11**
Aug-16	961316	p2	**p12**
Sep-16	945954	p3	**p23**
Oct-16	913692	p3	**p33**
Nov-16	850504	p3	**p33**
Dec-16	815978	p3	**p33**
Jan-17	704846	p3	**p33**
Feb-17	710589	p1	**p31**
Mar-17	744060	p1	**p11**
Apr-17	778198	p1	**p11**
May-17	805049	p1	**p11**
Jun-17	864928	p1	**p11**
Jul-17	895155	p1	**p11**
Aug-17	898219	p2	**p12**

Sep-17	881662	p3	**p23**
Oct-17	844589	p3	**p33**
Nov-17	798219	p3	**p33**
Dec-17	745136	p3	**p33**
Jan-18	585333	p3	**p33**
Feb-18	588242	p2	**p32**
Mar-18	612514	p1	**p21**
Apr-18	641932	p1	**p11**
May-18	717864	p1	**p11**
Jun-18	744319	p1	**p11**
Jul-18	762135	p1	**p11**
Aug-18	762183	p2	**p12**
Sep-18	759489	p3	**p23**
Oct-18	742549	p3	**p33**
Nov-18	718011	p3	**p33**
Dec-18	611725	p3	**p33**
Jan-19	595379	p3	**p33**
Feb-19	568120	p3	**p33**
Mar-19	619416	p1	**p31**
Apr-19	642554	p1	**p11**
May-19	727326	p1	**p11**
Jun-19	751413	p1	**p11**
Jul-19	767656	p1	**p11**
Aug-19	758925	p3	**p13**
Sep-19	752563	p2	**p32**
Oct-19	737737	p3	**p23**
Nov-19	720426	p3	**p33**
Dec-19	658395	p3	**p33**

Number of points in p1=69	
Number of points inp2=9	
Number of points in p3=42	
Counting of p11=56	**Probability of P11=0.81159**
Counting of p12=6	**Probability of P12=0.08696**
Counting of p13=7	**Probability of P13=0.1015**
Counting of p21=4	**Probability of P21=0.44445**
Counting of p22=0	**Probability of P22=0**
Counting of p23=5	**Probability of P23=0.55556**
Counting of p31=8	**Probability of P31=0.195122**
Counting of p32=3	**Probability of P32=0.0732**
Counting of p33=30	**Probability of P33=0.73171**

From Table (8) above the following transition matrix is produced:

$$N = \begin{bmatrix} 56 & 6 & 7 \\ 4 & 0 & 5 \\ 8 & 3 & 30 \end{bmatrix}$$

This is followed by dividing the elements of each row by the total sum of that row, producing the transition probability matrix as follows:

$$P_{ij} = \begin{bmatrix} 56/69 & 6/69 & 7/69 \\ 4/9 & 0 & 5/9 \\ 8/41 & 3/41 & 30/41 \end{bmatrix} = \begin{bmatrix} 0.81159 & 0.08696 & 0.1015 \\ 0.4445 & 0 & 0.5556 \\ 0.19512 & 0.0732 & 0.73171 \end{bmatrix}$$

The transition probability matrix indicates:

(0.81159): The probability of transition from an increase to an increase

(0.08696): The probability of transition from an increase to a stability

(0.1015): The probability of transition from an increase to a decrease

(0.4445): The probability of transition from a stable to an increase

(0.5556): The probability of transition from a stable to a decrease.

(0.19512): The probability of transition from a decrease to an increase

(0.0732): The probability of transition from a decrease to a stable

(0.73171): The probability of transition from a decrease to a decrease

For the probability vector (π_0), we get it from the sum of each row is divided by the total as follows:

$$\pi_0 = P_0 = (\frac{69}{119}\frac{9}{119}\frac{41}{119})$$

$$\pi_0 = (0.5798 \quad 0.07563 \quad 0.34454)$$

The cases under study can be predicted by multiplying the initial vector π_0 in the matrix of transitional probability P_{ij}, as follows:

$$\pi_i = \pi_i P_{ij}$$

$$\pi_1 = (0.571404 \quad 0.07564 \quad 0.352973)$$

$$\pi_2 = (0.56624 \quad 0.075527 \quad 0.358297)$$

$$\pi_3 = (0.563037 \quad 0.075468 \quad 0.361606)$$

$$\pi_4 = (0.561057 \quad 0.075431 \quad 0.363668)$$

$$\pi_5 = (0.559088 \quad 0.075398 \quad 0.365764)$$

- **Steady States (Stability)**

To reach the stability of the matrix, we applied the rule (1-3.3) and it turns out that the matrix becomes stable at (n=24), we used Mat lab to reach the results described below:

$$\lim_{n\to\infty} p^n = \pi \quad (1\text{-}3.3)$$

$$P_{ij}^{10} = \begin{bmatrix} 0.5613 & 0.0754 & 0.3634 \\ 0.5559 & 0.0753 & 0.3690 \\ 0.5529 & 0.0753 & 0.3721 \end{bmatrix}$$

$$P_{ij}^{23} = \begin{bmatrix} 0.5579 & 0.0754 & 0.3670 \\ 0.5580 & 0.0754 & 0.3671 \\ 0.5580 & 0.0754 & 0.3671 \end{bmatrix}$$

$$P_{ij}^{24} = \begin{bmatrix} 0.5580 & 0.0754 & 0.3671 \\ 0.5580 & 0.0754 & 0.3671 \\ 0.5580 & 0.0754 & 0.3671 \end{bmatrix}$$

$$P_{ij}^{28} = \begin{bmatrix} 0.5580 & 0.0754 & 0.3671 \\ 0.5580 & 0.0754 & 0.3671 \\ 0.5580 & 0.0754 & 0.3671 \end{bmatrix}$$

We note from the above matrices, that a steady state is reached after (24) months.

To be sure, the ensuing linear system of equations was solved to find the steady states.

$$(\pi_1 \quad \pi_2 \quad \pi_3)\begin{pmatrix} 0.81159 & 0.08696 & 0.1015 \\ 0.4445 & 0 & 0.5556 \\ 0.19512 & 0.0732 & 0.73171 \end{pmatrix} = (\pi_1 \quad \pi_2 \quad \pi_3)$$

$0.81159\pi_1 + 0.4445\pi_2 + 0.19512\pi_3 = \pi_1$---------------------(1)

$0.08696\pi_1 + 0\pi_2 + 0.0732\pi_3 = \pi_2$------------------------(2)

$0.1015\pi_1 + 0.5556\pi_2 + 0.73171\pi_3 = \pi_3$------------------------(3)

$\pi_1 + \pi_2 + \pi_3 = 1$---------------------------(4)

(1),(2) & (4) :

$\pi_1 = 0.558 = 55.8\%$

$\pi_2 = 0.0754 = 7.54\%$

$\pi_3 = 0.3671 = 36.71\%$

(0.558):The probability of transition to increase after (24) months.

(0.0754): The probability of transition to stable after (24) months.

(0.3671): The probability of transition to decrease after (24) months.

According to the above stated results we conclude that the rate of increase in the amount of water supply (55.8%) is the highest, followed by the rate of decrease 36.71%. Therefore, the extreme stability situation (7.54%).

The amount of supply to water at long run will be stabilized after (n=24 months) as long as there are no interfered circumstances that affect it.

3.3.2. Behavior of the Water Demand in the Long Run

In order to trace the behavior of the water demand in the long run, the following steps were followed:

Data on the amount of water demand from the domestic sector in Tulkarem municipality and a sample of (120) months were taken, which were used to produce the transition matrix.

Table (9): Represents the amount of water demand from 2010-2019

month	demand	Month	Demand	month	demand	Month	demand	month	demand
Jan-10	205580	**Jan-12**	264317	**Jan-14**	323054	**Jan-16**	496296	**Jan-18**	**355628**
Feb-10	224407	**Feb-12**	265715	**Feb-14**	333214	**Feb-16**	499642	**Feb-18**	**325052**
Mar-10	235017	**Mar-12**	276771	**Mar-14**	352675	**Mar-16**	512257	**Mar-18**	**366762**
Apr-10	234212	**Apr-12**	289870	**Apr-14**	370798	**Apr-16**	526192	**Apr-18**	**374223**
May-10	245456	**May-12**	304135	**May-14**	391146	**May-16**	595481	**May-18**	**412416**
Jun-10	255055	**Jun-12**	329866	**Jun-14**	409768	**Jun-16**	619596	**Jun-18**	**401228**
Jul-10	261086	**Jul-12**	335682	**Jul-14**	410279	**Jul-16**	620370	**Jul-18**	**443773**
Aug-10	281991	**Aug-12**	360790	**Aug-14**	433555	**Aug-16**	628764	**Aug-18**	**427765**
Sep-10	305262	**Sep-12**	381180	**Sep-14**	447273	**Sep-16**	614308	**Sep-18**	**424208**
Oct-10	291991	**Oct-12**	329588	**Oct-14**	434996	**Oct-16**	607506	**Oct-18**	**411882**
Nov-10	283950	**Nov-12**	296781	**Nov-14**	408136	**Nov-16**	541964	**Nov-18**	**404332**
Dec-10	230215	**Dec-12**	271122	**Dec-14**	356378	**Dec-16**	487549	**Dec-18**	**384651**
Jan-11	242291	**Jan-13**	391791	**Jan-15**	570370	**Jan-17**	474074	**Jan-19**	**325261**
Feb-11	237007	**Feb-13**	390471	**Feb-15**	586104	**Feb-17**	482136	**Feb-19**	**335053**
Mar-11	260385	**Mar-13**	412375	**Mar-15**	607105	**Mar-17**	483498	**Mar-19**	**367566**
Apr-11	263366	**Apr-13**	422464	**Apr-15**	624305	**Apr-17**	491967	**Apr-19**	**377204**
May-11	280533	**May-13**	448811	**May-15**	698371	**May-17**	552274	**May-19**	**379944**
Jun-11	301345	**Jun-13**	452683	**Jun-15**	705798	**Jun-17**	576705	**Jun-19**	**392458**
Jul-11	307709	**Jul-13**	484876	**Jul-15**	712963	**Jul-17**	592593	**Jul-19**	**405507**
Aug-11	362629	**Aug-13**	502410	**Aug-15**	731346	**Aug-17**	602669	**Aug-19**	**458690**
Sep-11	385586	**Sep-13**	522947	**Sep-15**	730403	**Sep-17**	604372	**Sep-19**	**496451**
Oct-11	370533	**Oct-13**	512282	**Oct-15**	728614	**Oct-17**	586740	**Oct-19**	**471374**
Nov-11	301287	**Nov-13**	456255	**Nov-15**	654671	**Nov-17**	575180	**Nov-19**	**450015**
Dec-11	275482	**Dec-13**	417443	**Dec-15**	614718	**Dec-17**	477491	**Dec-19**	**358512**

- **Modelation The Transition Matrix**

The transition matrix is 3*3, because there are three cases in which the rate of water demand is affected by (increase-stability-decrease).

From this, the transition probability matrix will be as follows:

$$P_{ij} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}$$

This matrix reflects changes in the amount of water demand (increase-stability- decrease) .

- **Determine the Transition Matrix(p)**

To determine the transition matrix (p) Related to the amount of water demand during the period (2010-2019) the following steps has been taken:

Table (10): Represents the labelled and counted of the amount of demand to water during the period (2010-2019)

month	Demand	States	Transition counts
Jan-10	205580	p3	
Feb-10	224407	p1	**p31**
Mar-10	235017	p1	**p11**
Apr-10	234212	p2	**p12**
May-10	245456	p1	**p21**
Jun-10	255055	p1	**p11**
Jul-10	261086	p1	**p11**
Aug-10	281991	p1	**p11**
Sep-10	305262	p1	**p11**
Oct-10	291991	p3	**p13**
Nov-10	283950	p3	**p33**
Dec-10	230215	p3	**p33**
Jan-11	242291	p1	**p31**
Feb-11	237007	p3	**p13**
Mar-11	260385	p1	**p31**
Apr-11	263366	p2	**p12**
May-11	280533	p1	**p21**
Jun-11	301345	p1	**p11**
Jul-11	307709	p2	**p12**
Aug-11	362629	p1	**p21**
Sep-11	385586	p1	**p11**
Oct-11	370533	p3	**p13**
Nov-11	301287	p3	**p33**
Dec-11	275482	p3	**p33**
Jan-12	264317	p3	**p33**
Feb-12	265715	p2	**p32**
Mar-12	276771	p1	**p21**
Apr-12	289870	p1	**p11**
May-12	304135	p1	**p11**

Jun-12	329866	p1	**p11**
Jul-12	335682	p1	**p11**
Aug-12	360790	p1	**p11**
Sep-12	381180	p1	**p11**
Oct-12	329588	p3	**p13**
Nov-12	296781	p3	**p33**
Dec-12	271122	p3	**p33**
Jan-13	391791	p1	**p31**
Feb-13	390471	p2	**p12**
Mar-13	412375	p1	**p21**
Apr-13	422464	p1	**p11**
May-13	448811	p1	**p11**
Jun-13	452683	p1	**p11**
Jul-13	484876	p1	**p11**
Aug-13	502410	p1	**p11**
Sep-13	522947	p1	**p11**
Oct-13	512282	p3	**p13**
Nov-13	456255	p3	**p33**
Dec-13	417443	p3	**p33**
Jan-14	323054	p3	**p33**
Feb-14	333214	p1	**p31**
Mar-14	352675	p1	**p11**
Apr-14	370798	p1	**p11**
May-14	391146	p1	**p11**
Jun-14	409768	p1	**p11**
Jul-14	410279	p1	**p11**
Aug-14	433555	p1	**p11**
Sep-14	447273	p1	**p11**
Oct-14	434996	p3	**p13**
Nov-14	408136	p3	**p33**
Dec-14	356378	p3	**p33**
Jan-15	570370	p1	**p31**
Feb-15	586104	p1	**p11**
Mar-15	607105	p1	**p11**

Apr-15	624305	p1	**p11**
May-15	698371	p1	**p11**
Jun-15	705798	p1	**p11**
Jul-15	712963	p1	**p11**
Aug-15	731346	p1	**p11**
Sep-15	730403	p2	**p12**
Oct-15	728614	p3	**p23**
Nov-15	654671	p3	**p33**
Dec-15	614718	p3	**p33**
Jan-16	496296	p3	**p33**
Feb-16	499642	p2	**p32**
Mar-16	512257	p1	**p21**
Apr-16	526192	p1	**p11**
May-16	595481	p1	**p11**
Jun-16	619596	p1	**p11**
Jul-16	620370	p1	**p11**
Aug-16	628764	p2	**p12**
Sep-16	614308	p3	**p23**
Oct-16	607506	p3	**p33**
Nov-16	541964	p3	**p33**
Dec-16	487549	p3	**p33**
Jan-17	474074	p3	**p33**
Feb-17	482136	p1	**p31**
Mar-17	483498	p2	**p12**
Apr-17	491967	p1	**p21**
May-17	552274	p1	**p11**
Jun-17	576705	p1	**p11**
Jul-17	592593	p1	**p11**
Aug-17	602669	p1	**p11**
Sep-17	604372	p2	**p12**
Oct-17	586740	p3	**p23**
Nov-17	575180	p3	**p33**
Dec-17	477491	p3	**p33**
Jan-18	355628	p3	**p33**

Feb-18	325052	p3	**p33**
Mar-18	366762	p1	**p31**
Apr-18	374223	p1	**p11**
May-18	412416	p1	**p11**
Jun-18	401228	p3	**p13**
Jul-18	443773	p1	**p31**
Aug-18	427765	p3	**p13**
Sep-18	424208	p2	**p32**
Oct-18	411882	p3	**p23**
Nov-18	404332	p3	**p33**
Dec-18	384651	p3	**p33**
Jan-19	325261	p3	**p33**
Feb-19	335053	p1	**p31**
Mar-19	367566	p1	**p11**
Apr-19	377204	p1	**p11**
May-19	379944	p2	**p12**
Jun-19	392458	p1	**p21**
Jul-19	405507	p1	**p11**
Aug-19	458690	p1	**p11**
Sep-19	496451	p1	**p11**
Oct-19	471374	p3	**p13**
Nov-19	450015	p3	**p33**
Dec-19	358512	p3	**p33**

Number of points in p1=66	
Number of points in p2=12	
Number of points in p3=42	
Counting of p11=48	**Probability of P11=0.72727**
Counting of p12=9	**Probability of P12=0.13636**
Counting of p13=9	**Probability of P13=0.13636**
Counting of p21=8	**Probability of P21=0.6667**
Counting of p22=0	**Probability of P22=0**
Counting of p23=4	**Probability of P23=0.3333**
Counting of p31=10	**Probability of P31=0.2439**
Counting of p32=3	**Probability of P32=0.073171**
Counting of p33=28	**Probability of P33=0.68293**

From Table (10) above the following transition matrix is produced:

$$N = \begin{bmatrix} 48 & 9 & 9 \\ 8 & 0 & 4 \\ 10 & 3 & 28 \end{bmatrix}$$

This is followed by dividing the elements of each row by the total sum of that row, producing the transition probability matrix as follows:

$$P_{ij} = \begin{bmatrix} \frac{48}{66} & \frac{9}{66} & \frac{9}{66} \\ \frac{8}{12} & 0 & \frac{4}{12} \\ \frac{10}{41} & \frac{3}{41} & \frac{28}{41} \end{bmatrix} = \begin{bmatrix} 0.72727 & 0.13636 & 0.13636 \\ 0.6667 & 0 & 0.3333 \\ 0.2439 & 0.073171 & 0.68293 \end{bmatrix}$$

The transition probability matrix indicates:

(0.72727): The probability of transition from an increase to an increase.

(0.13636): The probability of transition from an increase to a stable.

(0.13636: The probability of transition from an increase to a decrease.

(0.6667): The probability of transition from a stable to an increase.

(0.3333: The probability of transition from a stable to a decrease.

(0.2439): The probability of transition from a decrease to an increase

(0.073171): The probability of transition from a decrease to a stable.

(0.68293): The probability of transition from a decrease to a decrease.

For the probability vector (π_0), we get it from the sum of each row is divided by the total as follows:

$$\pi_0 = P = (\frac{66}{119}\frac{12}{119}\frac{41}{119})$$

$$\pi_0 = (0.5546 \quad 0.1008 \quad 0.33613)$$

The cases under study can be predicted by multiplying the initial vector (π_0) in the matrix of transitional probability (P_{ij}), as follows:

$$\pi_1 = \pi_0 P_{ij} = \left(\frac{67}{119}\frac{12}{119}\frac{40}{119}\right) * P_{ij}$$

$$\pi_1 = (0.55256 \quad 0.10022 \quad 0.33875)$$

$\pi_2 = (0.551327 \quad 0.100133 \quad 0.340059)$

$\pi_3 = (0.55069 \quad 0.100061 \quad 0.34076)$

$\pi_4 = (0.550355 \quad 0.100026 \quad 0.341128)$

$\pi_5 = (0.550175 \quad 0.100007 \quad 0.341322)$

- **Steady States (Stability)**

To reach the stability of the matrix, we applied the rule (2-3.3) and it turns out that the matrix becomes stable at (n=16), we used Mat lab to reach the results described below:

$$\lim_{n\to\infty} p^n = \pi \qquad (2\text{-}3.3)$$

$$P_{ij}^{10} = \begin{bmatrix} 0.5552 & 0.1009 & 0.3439 \\ 0.5547 & 0.1008 & 0.3444 \\ 0.5536 & 0.1007 & 0.3456 \end{bmatrix}$$

$$P_{ij}^{15} = \begin{bmatrix} 0.5546 & 0.1008 & 0.3445 \\ 0.5546 & 0.1008 & 0.3445 \\ 0.5546 & 0.1008 & 0.3446 \end{bmatrix}$$

$$P_{ij}^{16} = \begin{bmatrix} 0.5546 & 0.1008 & 0.3445 \\ 0.5546 & 0.1008 & 0.3445 \\ 0.5546 & 0.1008 & 0.3445 \end{bmatrix}$$

$$P_{ij}^{25} = \begin{bmatrix} 0.5546 & 0.1008 & 0.3445 \\ 0.5546 & 0.1008 & 0.3445 \\ 0.5546 & 0.1008 & 0.3445 \end{bmatrix}$$

We note from the above matrices, that a steady state is reached after (16) months.

To ensure the above results, the ensuing linear system of equations was solved to find the steady states.

$$(\pi_1 \quad \pi_2 \quad \pi_3)\begin{pmatrix} 0.72727 & 0.13636 & 0.13636 \\ 0.6667 & 0 & 0.3333 \\ 0.2439 & 0.073171 & 0.68293 \end{pmatrix} = (\pi_1 \quad \pi_2 \quad \pi_3)$$

$0.72727\pi_1 + 0.6667\pi_2 + 0.2439\pi_3 = \pi_1$ (1)

$0.13636\pi_1 + 0\pi_2 + 0.073171\pi_3 = \pi_2$ (2)

$0.13636\pi_1 + 0.3333\pi_2 + 0.68293\pi_3 = \pi_3$ (3)

$\pi_1 + \pi_2 + \pi_3 = 1$ (4)

(1),(2) & (4) :

$\pi_1 = 0.5546 = 55.46\%$

$\pi_2 = 0.1008 = 10.08\%$

$\pi_3 = 0.3445 = 34.45\%$

(0.5546):The probability of transition to increase after (16) months.

(0.1008): The probability of transition to stable after (16) months.

(0.3445): The probability of transition to decrease after (16) months.

We conclude from these results that the rate of increase in the amount of water demand (55.46%) is the highest, followed by the rate of decrease of 34.45%. As a result, the extreme stability situation was (10.08%).

The amount of water demand at long run will stabilize after (n=16 months) as long as there are no interfered circumstances that affect it.

3.3.3 The Best Statistical Models to Predict the Quantity of the Water Supply and Demand

3.3.3.1 Markov Chains Model Construction Stages

- **Determination of Descriptive Statistical Standards and Markov Model Marks**

Referring to table (5) in which descriptive statistical standards were calculated and tabulated.

According to Markov chains model (Thomas-Fiering) of the following formula:

$$\mathbf{q_{ij} = \bar{q}_j + b_j\left(q_{i-1,J-1} - \bar{q}_{J-1}\right) + t_{ij}s_j(1 - r_j^2)^{1/2}}$$

Where:

i: The month symbol of the generated series and take values from 1 to the length of the series

j: Month symbol in the series (From January to December)

$\bar{q}_J$: The mean of values the month (jth) (From January to December)

b_j: Regression coefficient for values in two months jth, (j-1)th And it's calculated as follows: $b_j = \frac{r_j s_j}{s_{j-1}}$

$q_{i-1,j-1}$: Value at the month immediately before it.

$\bar{q}_{j-1}$: The mean of values the month before jth.

S_j: Standard deviation of values per month jth (From January to December).

t_{ij}: The random generated number follows the standard natural distribution.

r_j: Linear correlation coefficient for two months jth, (j-1)th.

To find the monthly data of Markov chain marks from 2010 to 2019 which are 120 value, tables (11 & 12) are followed:

Table (11): monthly data of Markov chain marks for water supply amount from 2010 to 2019

	$\bar{q}_j$	s_j	s_j^2	r_j	b_j	$\overline{q_{j-1}}$
JAN	531434	158113.5	2.5E+10	0.991843	0.991843	562653.4
FEB	537186.6	166428.1	2.77E+10	0.98857	1.040555	531434
MAR	569196.6	164014.5	2.69E+10	0.992655	0.978259	537186.6
APR	595283.6	163644.7	2.68E+10	0.998595	0.996344	569196.6
MAY	636148.1	181066.7	3.28E+10	0.989977	1.095373	595283.6
JUN	660668.9	197745	3.91E+10	0.997811	1.089721	636148.1
JUL	677949.8	202137	4.09E+10	0.998719	1.020901	660668.9
AUG	694822.3	190500.4	3.63E+10	0.997736	0.940299	677949.8
SEP	687968.7	174907.2	3.06E+10	0.996448	0.914885	694822.3
OCT	651566.8	179390.1	3.22E+10	0.996632	1.022176	687968.7
NOV	608492.4	185464.3	3.44E+10	0.992017	1.025607	651566.8
DEC	562653.4	172588.2	2.98E+10	0.990514	0.921746	608492.4

Table (12): monthly data of Markov chain marks for water demand amount from 2010 to 2019

	$\bar{q}_j$	s_j	s_j^2	r_j	b_j	$\overline{q_{j-1}}$
JAN	364866.2	118352.6	1.4E+10	0.990163	0.990163	387356.1
FEB	367880.1	120559.5	1.45E+10	0.993345	1.011868	364866.2
MAR	387441.1	118947	1.41E+10	0.995394	0.98208	367880.1
APR	397460.1	122680.8	1.51E+10	0.999268	1.030635	387441.1
MAY	430856.7	145780.5	2.13E+10	0.996375	1.183984	397460.1
JUN	444450.2	145979.9	2.13E+10	0.996825	0.998188	430856.7
JUL	457483.8	145969.7	2.13E+10	0.995481	0.995412	444450.2
AUG	479060.9	138796.8	1.93E+10	0.990653	0.941972	457483.8
SEP	491199	129299.9	1.67E+10	0.995479	0.927365	479060.9
OCT	474550.6	136177.7	1.85E+10	0.995985	1.048964	491199
NOV	437257.1	125359.6	1.57E+10	0.98293	0.904845	474550.6
DEC	387356.1	117037.5	1.37E+10	0.978291	0.913347	437257.1

- **Random Figures Generation Stage**

In this stage random numbers from the uniform distribution U(0,1) has been generated using Microsoft excel, through the command RAND (), to get the random variable (T) with mean =0 and standard deviation =1 from the normal distribution N(0,1) inverse error function (erf-1) is used according to the following formula:

$$erf^{-1}(z) = \frac{1}{2}\sqrt{\pi}(z + \frac{\pi}{12}z^3 + \frac{7\pi^2}{480}z^5 + \frac{127\pi^3}{40320}z^7 + \cdots)$$

While the value Z can be found through cumulative distribution function for logarithm distribution as follows:

$$CDF = \frac{1}{2} + \frac{1}{2}erf\left[\frac{lnx - \mu}{\sqrt{2\sigma^2}}\right] = RAND()$$

$$erf(z) = erf\left[\frac{lnx - \mu}{\sqrt{2\sigma^2}}\right] = [RAND(\quad) - 0.5] * 2$$

$$z = [RAND() - 0.5] * 2$$

The mean value of the normal logarithm of the random figures equals its standard deviation, i.e. $\mu = \sigma = 1$, then:

$$\text{erf}\left[\frac{lnx - 1}{\sqrt{2}}\right] = z$$

$$erf^{-1}(z) = \frac{lnx - 1}{\sqrt{2}}$$

$$lnx - 1 = \sqrt{2}\, erf^{-1}(z)$$

$$t = lnx = \sqrt{2}\, erf^{-1}(z) + 1$$

Note: the random numbers in the table (13) were used from the (Abu Libda,2018)[28], these values give better results when applied on the data used in this study.

Table (13): represents the steps followed to generate random numbers

	Rand	Z	$\mathbf{erf^{-1}}$	$\mathbf{t_{ij}}$
1	0.699645	0.399289	0.370085	1.523379
2	0.45481	-0.090379	-0.08027	0.886483
3	0.63732	-0.872536	-1.0558	-0.49313
4	0.224711	-0.50577	-0.53482	0.243657
5	0.236038	-0.527923	-0.50840	0.280915
6	0.471912	-0.56176	-0.04983	0.929536
7	0.999341	0.998683	1.443813	3.041859
8	0.533139	0.066278	0.58805	1.083163
9	0.095672	-0.808656	-0.91763	-0.29772
10	0،044676	-0.910651	-1.15355	-0.63136
11	0.997494	0.994989	1.429319	3.021363
12	0.407816	-0.184368	-0.1687	0.766834

- **Markov Model Construction Stage for Prediction**

Markov model consists of two parts; the first one deterministic part, which considers the effect of previous value in the model. The other part is random part which represents the random model. By combining the two parts, the monthly model of Markov's for prediction is constructed according to the following formula:

$$q_{ij} = \bar{q}_J + b_j\left(q_{i-1,J-1} - \bar{q}_{J-1}\right) + t_{ij}s_j(1 - r_j^2)^{1/2}$$

Where:

i: The month symbol of the generated series and take values from 1 to the length of the series

j: Month symbol in the series (From January to December)

$\bar{q}_J$: The mean of values the month (jth) (From January to December)

b_j: Regression coefficient for values in two months jth, (j-1)th And it's calculated as follows:$b_j = \frac{r_j s_j}{s_{j-1}}$

$q_{i-1,j-1}$: Value at the month immediately before it.

$\bar{q}_{j-1}$: The mean of values the month before jth.

S_j: Standard deviation of values per month jth (From January to December).

t_{ij}: The random generated number follows the standard natural distribution.

r_j: Linear correlation coefficient for two months jth, (j-1)th.

The following table (14 &15) shows the prediction of water supply/demand for 2019, depending on which month of the year can be predicted based on condition where the previous values for this month is known.

Table (14): Markov model construction for prediction of water supply amount for 2019

$q_{i-1,j-1}$	$\bar{q}_j + b_j(q_{i-1,j-1} - \overline{q_{j-1}})$	t_{ij}	$t_{ij}s_j(1-r_j^2)^{1/2}$	q_{ij}
611725	580105.3	1.523379	30702.26	610807.6
585333	593271.5	0.886483	22243.2	615514.7
588242	619142	-0.49313	-9784.95	609357
612514	638442.6	0.243657	2112.548	640555.2
641932	687245.5	0.280915	7183.508	694429
717864	749716.4	0.929536	12154.6	761871
744319	763348.3	3.041859	31111.69	794460
762135	773981.5	1.083163	13876.02	787857.5
762183	749596	-0.29772	-4385.11	745210.9
759489	724673.2	-0.63136	-9287.44	715385.7
742549	701804.4	3.021363	70662.08	772466.5
718011	663601.7	0.766834	18186.36	681788.1

Table (15): Markov model construction for prediction of water demand amount for 2019

$q_{i-1,j-1}$	$\bar{q}_j + b_j(q_{i-1,j-1} - \overline{q_{j-1}})$	t_{ij}	$t_{ij}s_j(1-r_j^2)^{1/2}$	q_{ij}
355628	362187.7	1.523379	30702.26	392890
325052	358532.3	0.886483	22243.2	380775.5
366762	345380.5	-0.49313	-9784.95	335595.5
374223	376147.5	0.243657	2112.548	378260
412416	403344.4	0.280915	7183.508	410527.9
401228	426042.9	0.929536	12154.6	438197.5
443773	414459.9	3.041859	31111.69	445571.6
427765	466145.7	1.083163	13876.02	480021.7
424208	443629	-0.29772	-4385.11	439243.9
411882	404279.5	-0.63136	-9287.44	394992
404332	380551.7	3.021363	70662.08	451213.8
384651	357284.1	0.766834	18186.36	375470.4

In the same manner, amount of water (supply/demand) can be predicted for 2020 year, as explained in the following table (16) & (17):

Table (16): Markov model construction for prediction of water supply amount for 2020

$q_{i-1,j-1}$	$\bar{q}_j + b_j(q_{i-1,j-1} - \overline{q_{j-1}})$	t_{ij}	$t_{ij}s_j(1-r_j^2)^{1/2}$	q_{ij}
658395	626394.6	1.523379	30702.26	657096.9
595379	603724.9	0.886483	22243.2	625968.1
568120	599457.5	-0.49313	-9784.95	589672.5
619416	645319.4	0.243657	2112.548	647432
642554	687926.8	0.280915	7183.508	695110.3
727326	760027.3	0.929536	12154.6	772181.9
751413	770590.6	3.041859	31111.69	801702.3
767656	779172.9	1.083163	13876.02	793048.9
758925	746615.3	-0.29772	-4385.11	742230.2
752563	717593.6	-0.63136	-9287.44	708306.1
737737	696869.2	3.021363	70662.08	767531.2
720426	665827.7	0.766834	18186.36	684014.1

Table (17): Markov model construction for prediction of water demand amount for 2020

$q_{i-1,j-1}$	$\bar{q}_j + b_j(q_{i-1,j-1} - \overline{q_{j-1}})$	t_{ij}	$t_{ij}s_j(1-r_j^2)^{1/2}$	q_{ij}
358512	336305.8	1.523379	30702.26	367008.1
325261	327804.9	0.886483	22243.2	350048.1
335053	355202.3	-0.49313	-9784.95	345417.3
367566	376976.1	0.243657	2112.548	379088.7
377204	406873.8	0.280915	7183.508	414057.3
379944	393629.8	0.929536	12154.6	405784.4
392458	405730.2	3.041859	31111.69	436841.9
405507	430100.2	1.083163	13876.02	443976.2
458690	472307.7	-0.29772	-4385.11	467922.6
496451	480059.8	-0.63136	-9287.44	470772.3
471374	434382.8	3.021363	70662.08	505044.8
450015	399008.5	0.766834	18186.36	417194.8

Table (18): the predictive value (supply) of Markov model.

Month	y_t	$\hat{y}_t$	Relative error
JAN(1/2020)	552465	657096.9	0.18939=18.9%
FEB(2/2020)	548691	625968.1	0.14084=14%
MAR(3/2020)	598768	589672.5	0.01519=1.5%
APR(4/2020)	637289	647432	0.01592=1.6%
MAY(5/2020)	702951	695110.3	0.011154=1.1%

Table (19): the predictive value (demand) of Markov model.

Month	y_t	$\hat{y}_t$	Relative error
JAN(1/2020)	325261	367008.1062	0.12835=12.8%
FEB(2/2020)	335053	350048.0712	0.04475=4.5%
MAR(3/2020)	367566	345417.3019	0.060258=6%
APR(4/2020)	377204	379088.6677	0.005=0.5%
MAY(5/2020)	379944	414057.3141	0.08979=8.9%

Prediction accuracy criteria MAE, MAPE, RMSE are considered an indication of how efficient it is in prediction as explained in the following table (20):

Table (20): Represents the results of the test of statistical criteria(MAE) (MAPE) (RMSE)

Criteria	**RMSE**	**MAPE**	**MAE**
Supply	11960.58	0.310408	1741.568
Demand	5496.595941	0.273453	957.4071438

3.3.3.2. Time Series (ARIMA) Model Application Stages for Prediction

- **Time Series Graph**

Time series in table (3) & (4) was graphed as shows in graph (7) & (8) using the statistical software (SPSS) so as to realize its basic characteristics. The general trend for the chain increasing with time, in addition to oscillations represented in concaves and protrusions (graph 7), these oscillations are repetitive regularly and in the same pace they increase by each year along the 12 months, this gives an indicator about the presence of general trend and seasonal component.

For further clarification, Dickey –Fuller and Phillips perron tests were used to determine the degree of stationary time series for each series alone. Tests' results are shows in table (21).

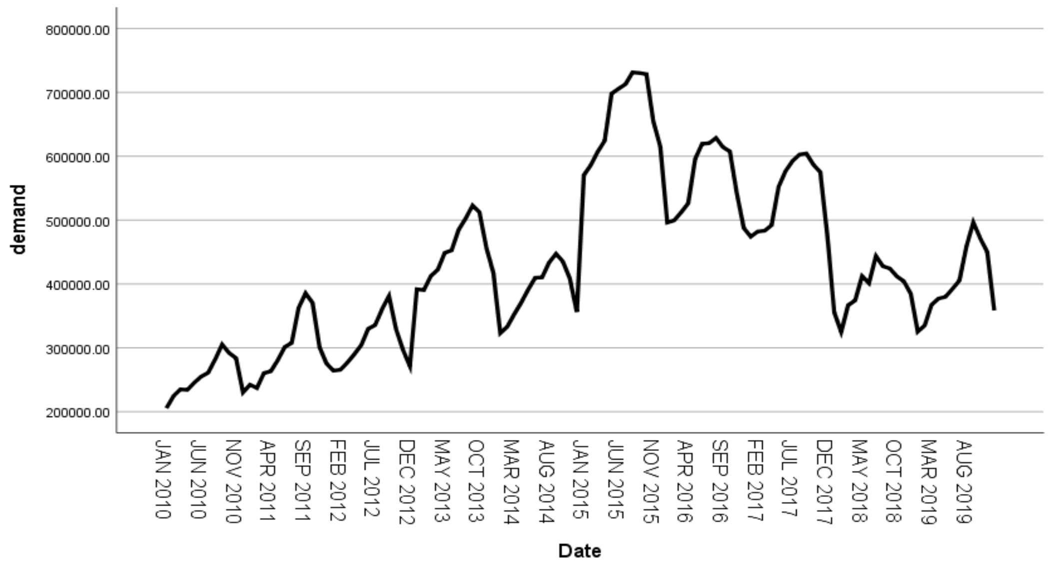
$$H_0: \rho = 1$$

$$H_1: \rho < 1$$

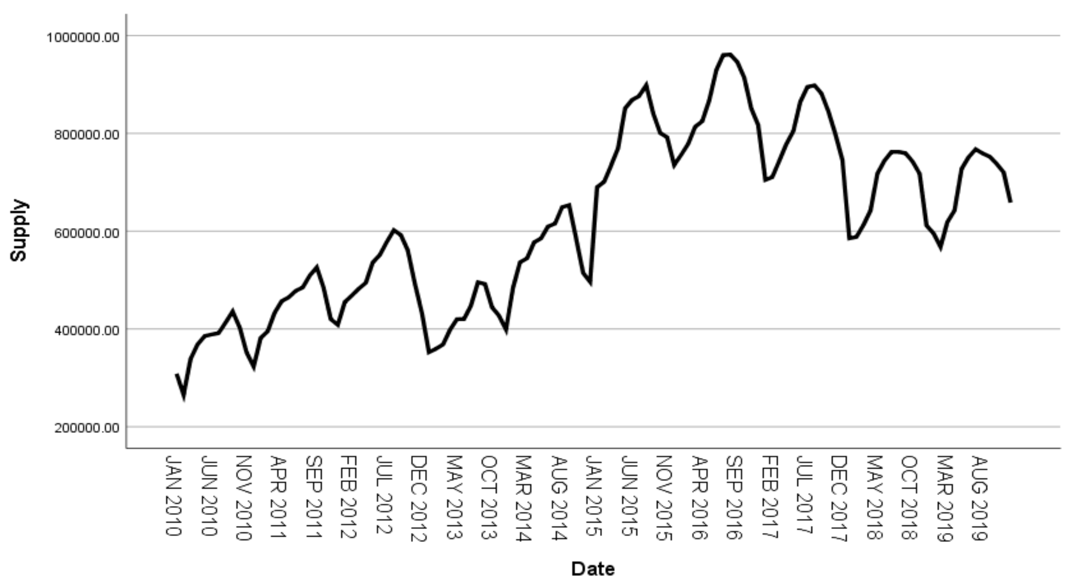
Table (21): results of Dickey -Fuller and Philips perron tests to validate time series stationariness.

Time Series	Dickey -Fuller Test				Phillips-Perron Test			
	Level		1st Difference		Level		1st Difference	
	Statistical	P-Value	Statistical	P-Value	Statistical t-test	P-Value	Statistical	P-Value
Supply	-1.599562	0.4793	-7.275279	0.000	-2.063031	0.2600	-7.280118	0.000
Demand	-2.471361	0.1251	-8.141707	0.000	-2.101992	0.2444	-8.101797	0.000

The results in table 21 show that the time series (supply/demand) are non-stationary at the level of significance 0.05, as the p-value (supply) and p-value (demand) is greater than 0.05.



Graph (7): represents the curve of time series of demand amount in terms of time in months



Graph (8): represents the curve of time series of supply amount in terms of time in months

- **Time Series Stationaries Test**

For getting stationaries in the variance, data were processed by taking the natural logarithm, this is shows in graphs (9) & (10), which show that stationariness in the variance is achieved by taking the natural logarithm for the data, therefore, it was considered when applying the models.

From the same graphs (9) & (10), it's clear that there is a general trend in the data. In order to confirm that trend, and to recognize the nature of the series, auto and partial correlation were extracted as shows in graphs (11) & (12) which also show that factors of auto correlation function differs significantly from zero, and factors of auto correlation function within the confidence interval $-0.23 < R_k < 0.23$.

The table (22) below shows results of (ljung& Box -test) of the total significance of auto correlation function.

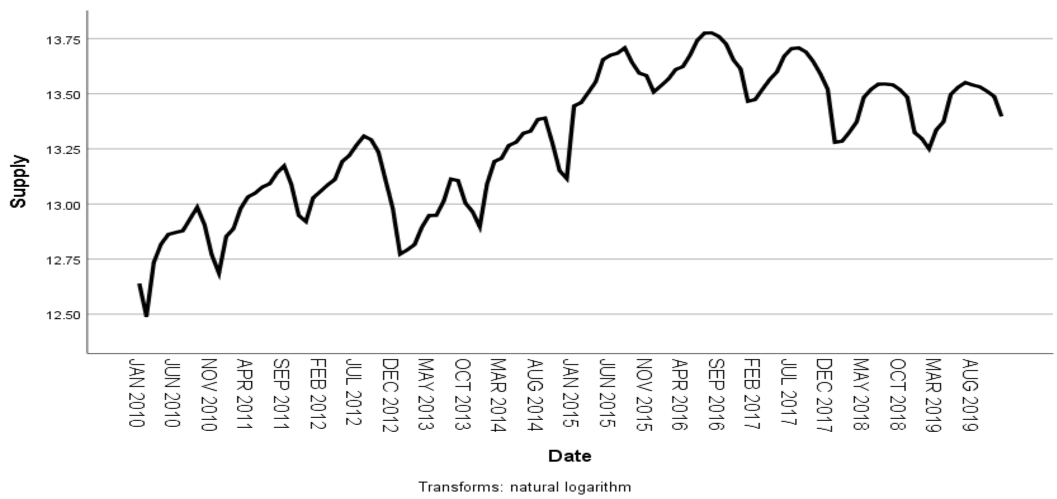
Table (22): results of Ljung&Box tests to validate time series stationariness.

	N	LBQ(Q^*. stat)	$\chi^2_{(24,0.05)}$
Supply	120	1068.428	36.42
Demand	120	888.409	36.42

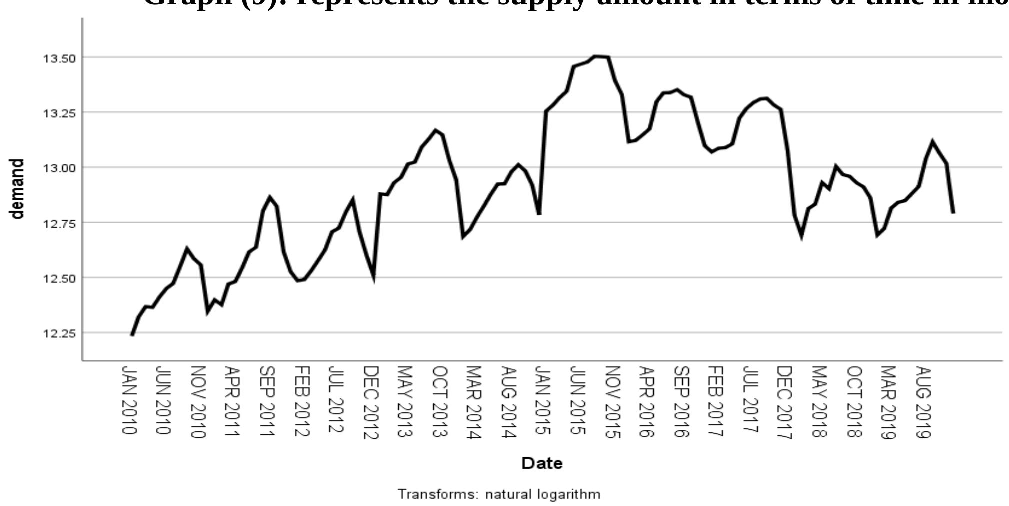
Therefore, null hypothesis which states that all factors of autocorrelation function equal zero is rejected.

$$H_0: \rho_1 = \rho_2 = \rho_3 = \cdots = \rho_k = 0$$

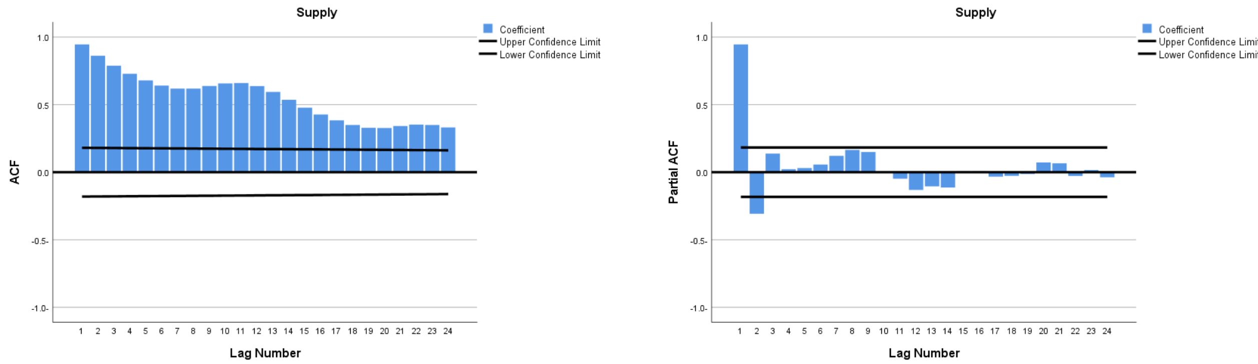
So, the alternative hypothesis is accepted which means that the time series is non-stationary.



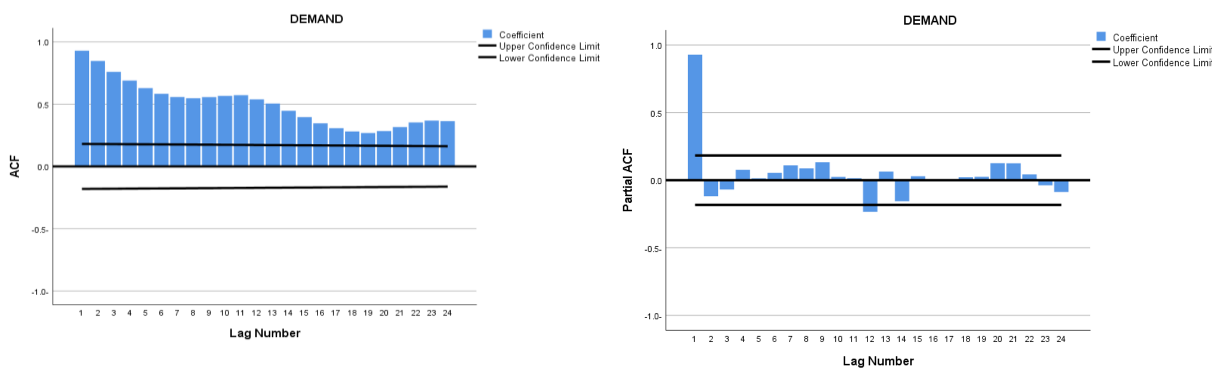
Graph (9): represents the supply amount in terms of time in months.



Graph (10): represents the demand amount in terms of time in months.



Graph (11): represents the coefficients of auto correlation and partial correlation after taking the natural log of the time series(supply series)



Graph (12): represents the coefficients of auto correlation and partial correlation after taking the natural log of the time series (demand series)

- **Removal of Non-Stationaries of Time Series**

1- Removal of General Trend

In order to remove the general trend, the differences of 1st degree were taken then the modified series was obtained, where ($\nabla y_t = y_t - y_{t-1}$) (the logarithmic series), graphs (13&14) shows the modified time series curve after taking the first difference of which.

By looking at the same graph, it's also noted that the general trend is not there anymore in the series, but the seasonal component is still there, i.e. the series is non-stationary.

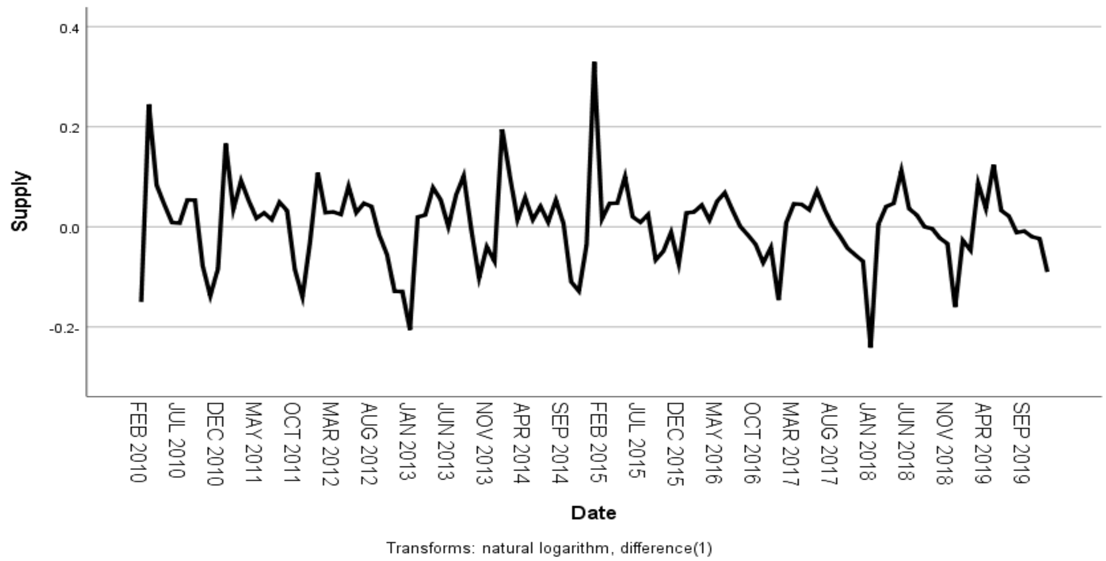
Table (23) below shows results of (ljung& Box -test) of the total significance of auto correlation function.

Table (23): results of Ljung&Box tests to validate time series stationariness.

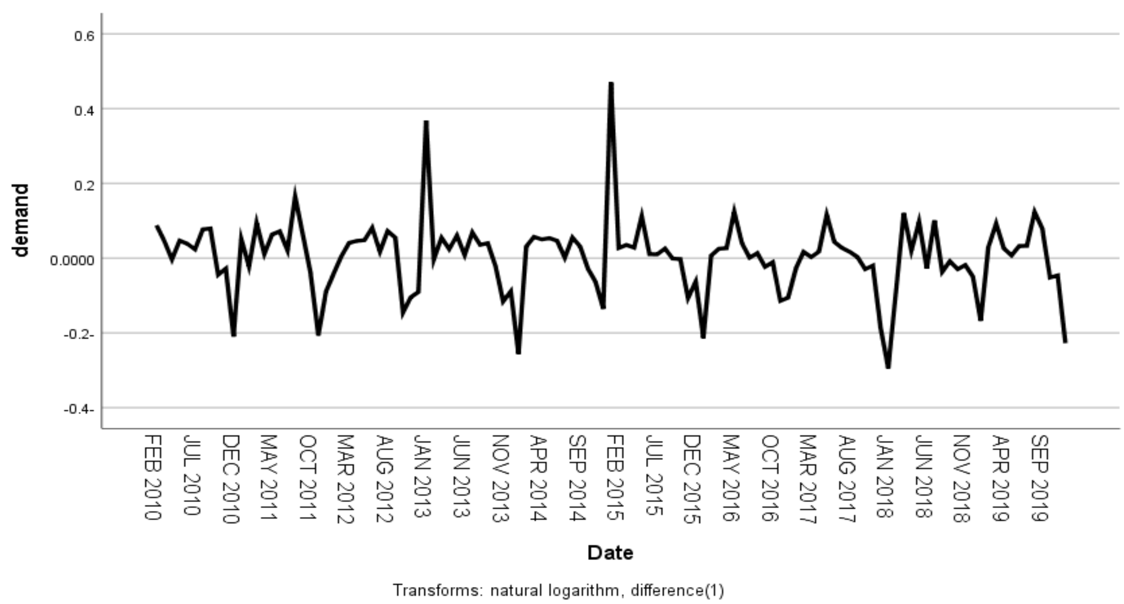
	N	LBQ(Q^*. stat)	$\chi^2_{(24,0.05)}$
Supply	120	114.782	35.17
Demand	120	53.973	35.17

For this reason, null hypothesis which assumes that all factors of autocorrelation function are insignificant is rejected.

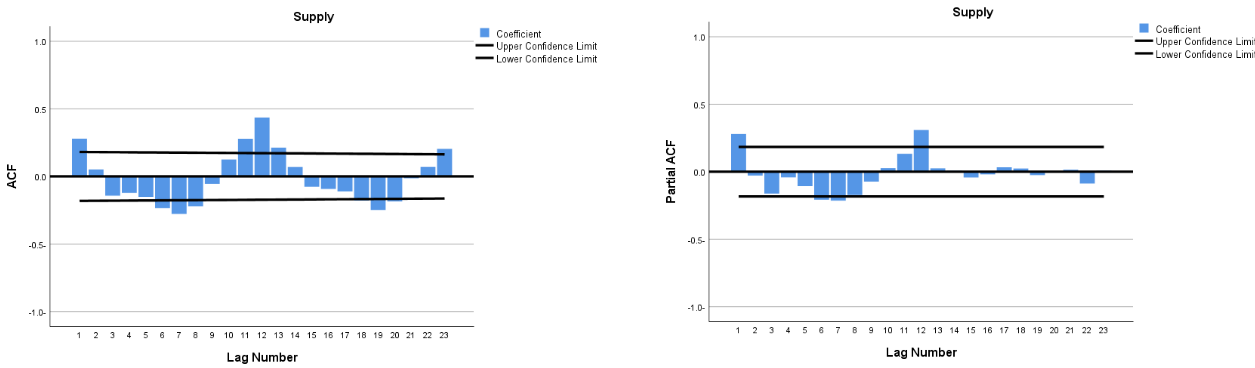
After taking the first-degree differences of the time series, the following graphs were generated:



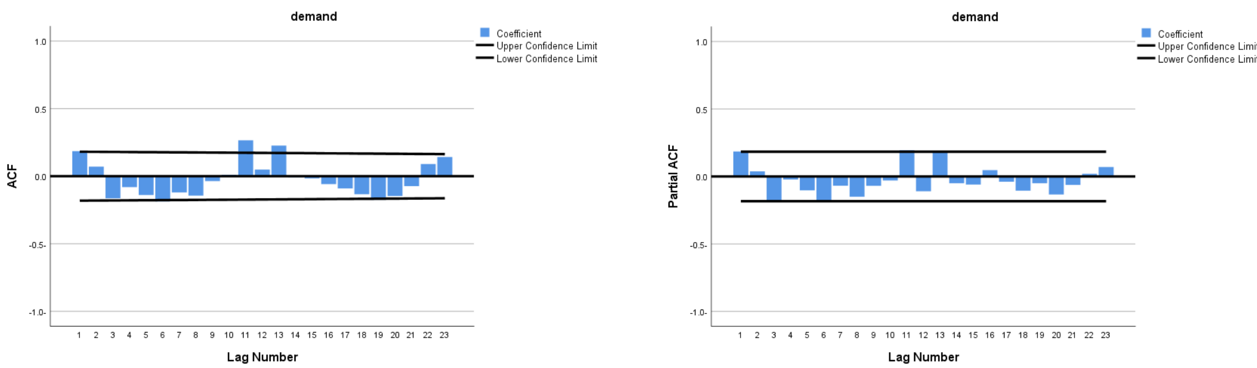
Graph (13): represents the first difference in terms of time in months



Graph (14): represents the first difference in terms of time in months.



Graph (15): represents the coefficients of auto correlation and partial correlation After taking the first-degree differences of the time series (supply series)



Graph (16): represents the coefficients of auto correlation and partial correlation After taking the first-degree differences of the time series (demand series)

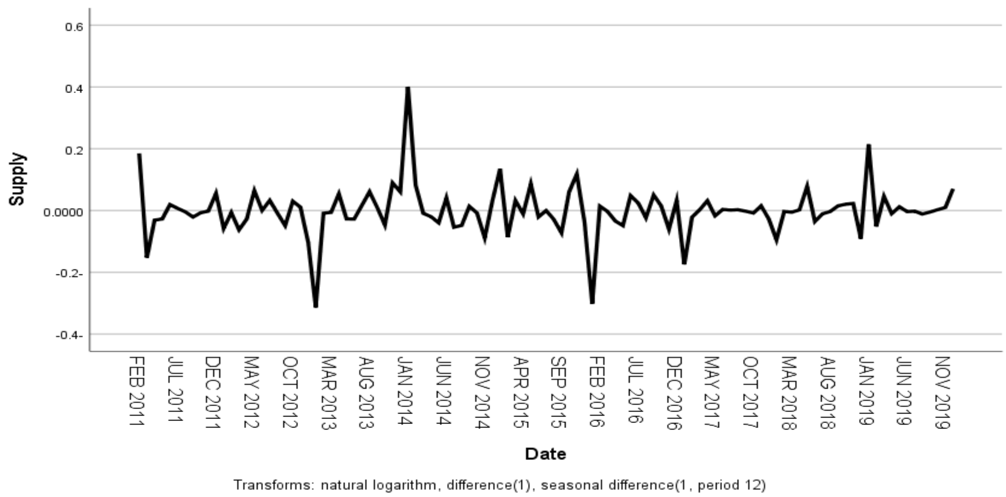
2- Removal of the Seasonal Component

By noting the values of autocorrelation of the modified time series after taking the first degree differences of it which is shows in graphs (15) & (16), it's clear that these values are significant in lag (12). So to get rid of the seasonal component, the seasonal differences of the second degree were taken to get the following modified series:

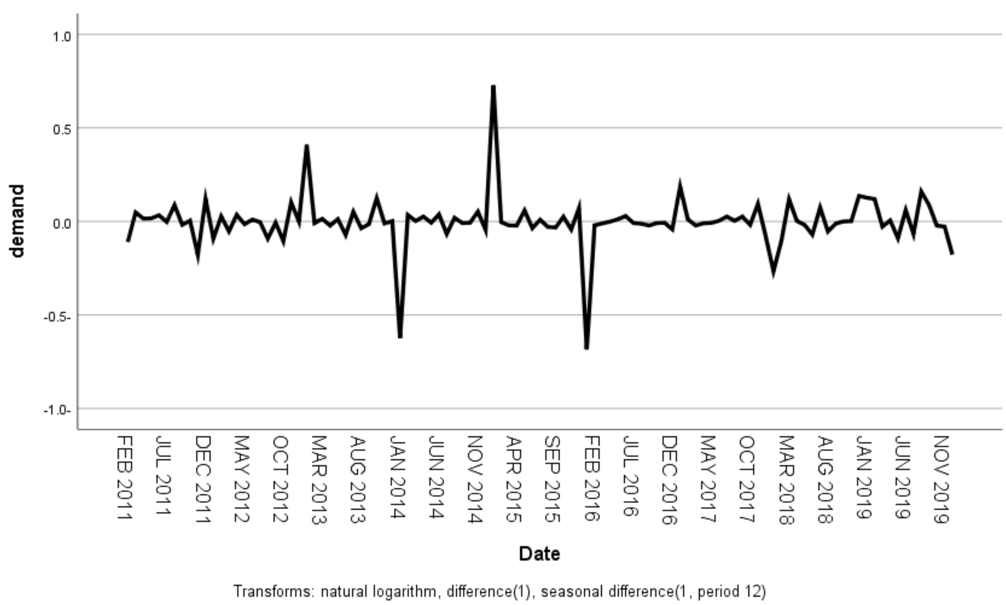
$$\nabla\nabla_{12}y_t = y_{t-1} - y_{t-12}$$

Graph (17&18) shows the modified time series curve after taking the seasonal differences, and the factors of the partial and autocorrelation function were taken and graphed as shows in graphs (19) & (20) respectively.

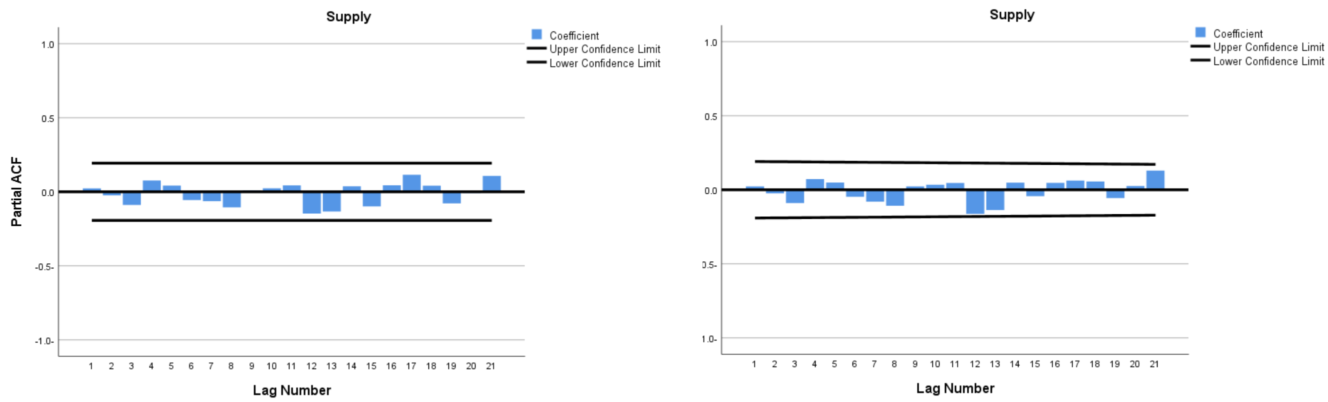
It's also noted from the same graphs that factors of the autocorrelation function are within the confidence level after the seasonal displacement (12) which indicates the stationaries of the time series.



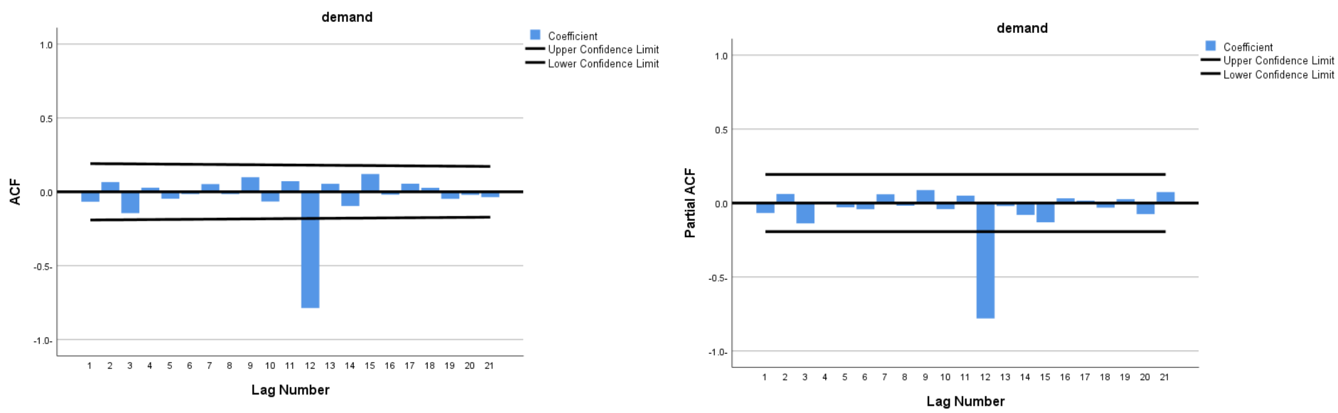
Graph (17): represents the amount of supply in terms of time in months



Graph (18): represents the amount of demand in terms of time in months



Graph (19): represents the coefficients of auto correlation and partial correlation after the seasonal displacement of the time series(supply series).



Graph (20): represents the coefficients of auto correlation and partial correlation after the seasonal displacement of the time series(demand series)

- **Estimation and Definition Stage**

1- Model Selection:

Means the proper model identification by determining the model rank (MA)(AR) through the shape of partial and autocorrelation functions, and when the partial & autocorrelation factors values of the time series are matched after taking the 1st and seasonal differences as in graphs (19 &20) with its theoretical behavior, it's clear that the partial and autocorrelation functions of the sample is gradually decreasing with increase of displacement intervals (K) (sinusoidal oscillations behaviour), through this indicator, it's concluded that this model is the seasonal multiple model (SARIMA) written as follows:

The appropriate model for the amount of water supply:

ARIMA (2,1,2) (0,1,1)

The appropriate model for the amount of water demand:

ARIMA (2,1,2) (1,1,1)

2- **Estimation of Model Parameters:**

- **Estimation of model parameters for water supply amount:**

After reviewing all the possible models using SPSS/E-Views, the following model was concluded: SARIMA, through the criteria AIC, RMSE, BIC, MAPE, R2, MAE and applying it on the time series data being studied, where the following was also concluded as in the table (24) below:

Table (24): comparison of candidate model and the directly upper/lower models.

ARIMA Model	AIC	BIC	R^2	RMSE	MAPE	MAE
(2,1,2)(0,1,1)	23.799252	23.96273	0.930	44732.252	4.509	26718.260
(2,1,3)(0,1,1)	23.815267	24.00209	0.929	45116.583	4.605	27301.057
(3,1,2)(0,1,1)	23.815409	24.00224	0.927	45698.304	4.638	27679.550
(4,1,2)(0,1,1)	23.823477	24.03366	0.932	44372.712	4.510	26518.790
(2,1,4)(0,1,1)	23.824780	24.03497	0.932	44525.819	4.590	27197.791

It's noted from the above table (24) that the best model is (ARIMA (2,1,2) (0,1,1)), because it has least value of criteria (AIC), (BIC) and (MAPE) so the final usable model for prediction is (ARIMA (2,1,2) (0,1,1)).

After determining the suggested model (ARIMA (2,1,2) (0,1,1)), the model parameters will be estimated, and by applying the least square method on the series data, the model parameters estimations were obtained as in above table.

Table (25): represents the model (ARIMA (2,1,2) (0,1,1)) parameters estimates.

					Estimate	SE	T	Sig.
Supply-Model_1	Supply	Natural Logarithm	Constant		-.002-	.003	-.507-	.613
			AR	Lag 1	.245	.051	4.759	.000
				Lag 2	-.996-	.046	-21.697-	.000
			Difference		1			
			MA	Lag 1	.226	.150	1.501	.137
				Lag 2	-.998-	1.294	-.771-	.442
			Seasonal Difference		1			
			MA, Seasonal	Lag 1	.606	.102	5.950	.000

- **Estimation of model parameters for water demand amount:**

After reviewing all the possible models using SPSS/E-Views, the following model was concluded: SARIMA, through the criteria AIC RMSE BIC MAP, and applying it on the time series data being studied, and the following was also concluded as in the table (26) below:

Table (26): comparison of candidate model and the directly upper/lower models.

ARIMA Model	AIC	BIC	R^2	RMSE	MAPE	MAE
(2,1,2)(1,1,1)	-1.98638	-1.799546	0.923	35265.387	5.052	21278.331
(0,1,0)(1,1,1)	-1.873855	-1.78044	0.922	34973.143	5.182	21838.477
(1,1,0)(1,1,1)	-1.85889	-1.742115	0.921	35173.088	5.194	21861.451
(0,1,1)(1,1,1)	-1.85888	-1.742105	0.921	35181.345	5.198	21877.065
(2,1,1)(1,1,1)	-1.84800	-1.684522	0.923	35213.412	5.095	21467.106

It’s noted from the above table that the best model is (ARIMA (2,1,2)(1,1,1)), because it has least value of criteria(AIC), (BIC),(R^2),(MAPE) and (MAE) so the final usable model for prediction is (ARIMA(2,1,2)(1,1,1)).

After determining the suggested model (SARIMA (2,1,2) (1,1,1)), the model parameters will be estimated, and by applying the least square method on the series data, the model parameters estimations were obtained as in above table.

Table (27): represents the model (ARIMA(2,1,2)(1,1,1)) parameters estimates

					Estimate	SE	T	Sig.
DEMAND-Mod_1	DEMAND	Natural Logarithm	Constant		-.003-	.003	-1.298-	.197
			AR	Lag 1	-.048-	.210	-.226-	.821
				Lag 2	.788	.214	3.686	.000
			Difference		1			
			MA	Lag 1	-.081-	.906	-.089-	.929
				Lag 2	.917	.870	1.054	.295
			AR, Seasonal	Lag 1	-.721-	.086	-8.369-	.000
			Seasonal Difference		1			
			MA, Seasonal	Lag 1	.189	.147	1.285	.202

- **Diagnoses stage:**

  Testing how suitable the model is for the data:

This term means diagnosing and testing the suggested model which has been determined and its parameters were estimated, as the Assumption of the residues are being tested, and they are:

$$a_t \sim IIN(0, \sigma_a^2)$$

1- Residues Randomness Test

It's tested via (runs test), which tests the null hypothesis:

H_0 = residues randomness distribution

H_1 = residues non-randomness distribution

The below table (28) shows the significance of the test; thus, the null hypothesis is accepted and the residues are random.

Table (28): represents the residues randomness test results:

Runs test		
Time series	**Demand**	**supply**
The observed number of runs	53	53
The expected number of runs	54	54
p-value(sig)	0.381	0.922

2- Residues Independence Test:

Assumes research and tests null hypothesis

H_0= Values of random residues estimations

H_1 = values of non-random (independent) residues estimations

i.e.:

$$H_0 = \rho_S(a_t) = 0$$

$$H_1 = \rho_S(a_t) \neq 0$$

The hypotheses are tested using (box-pierce) test.

When comparing test results with the value of $(\chi^2_{\alpha,n-k-1})$, results came as follows:

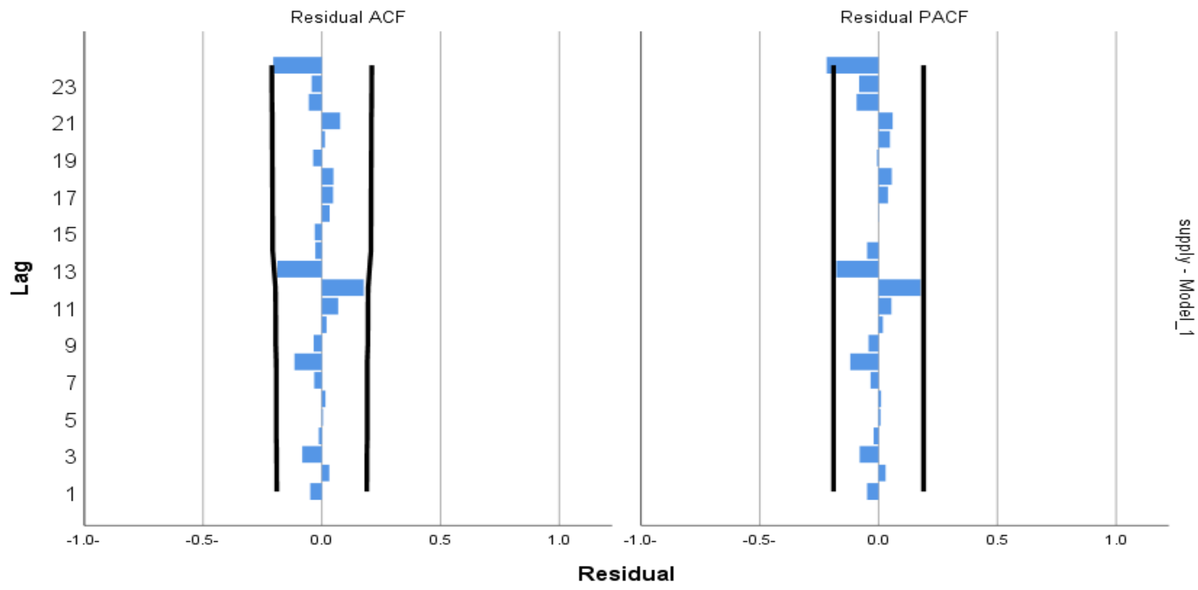
(supply series):

$$\text{LBQ}(Q^*.\text{stat}) = 12.731 < \boldsymbol{\chi}^{\mathbf{2}}_{(\ 13,\mathbf{0.05})} = 22.36$$

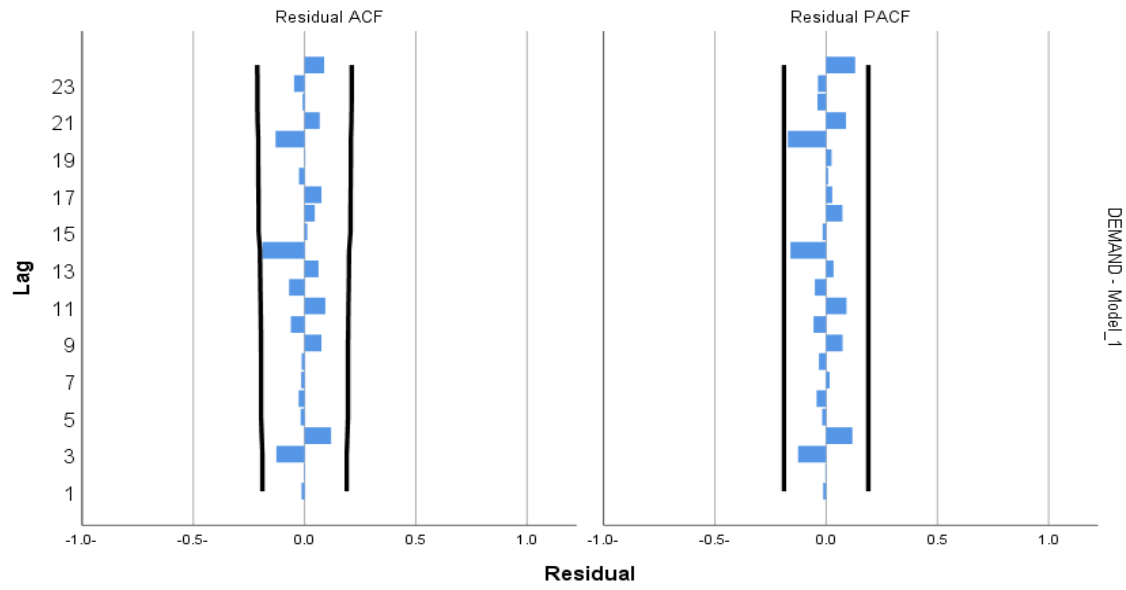
(demand series):

$$\text{LBQ}(Q^*.\text{stat}) = 12.297 < \boldsymbol{\chi}^{\mathbf{2}}_{(\ 12,\mathbf{0.05})} = 21.03$$

Therefore, we accept the null hypothesis, which means the randomness of the residues.



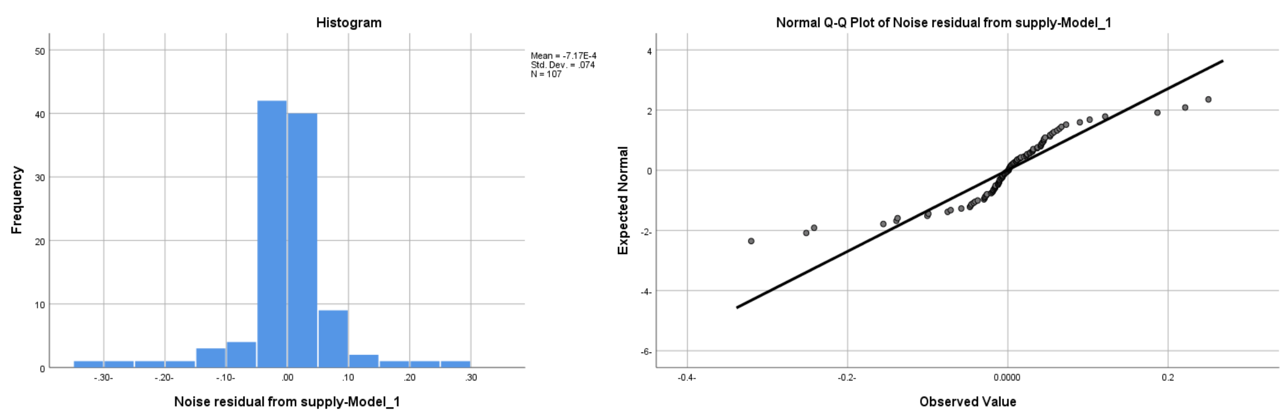
Graph (21): represents the coefficients of auto correlation and partial correlation after the seasonal displacement of the time series(supply series)



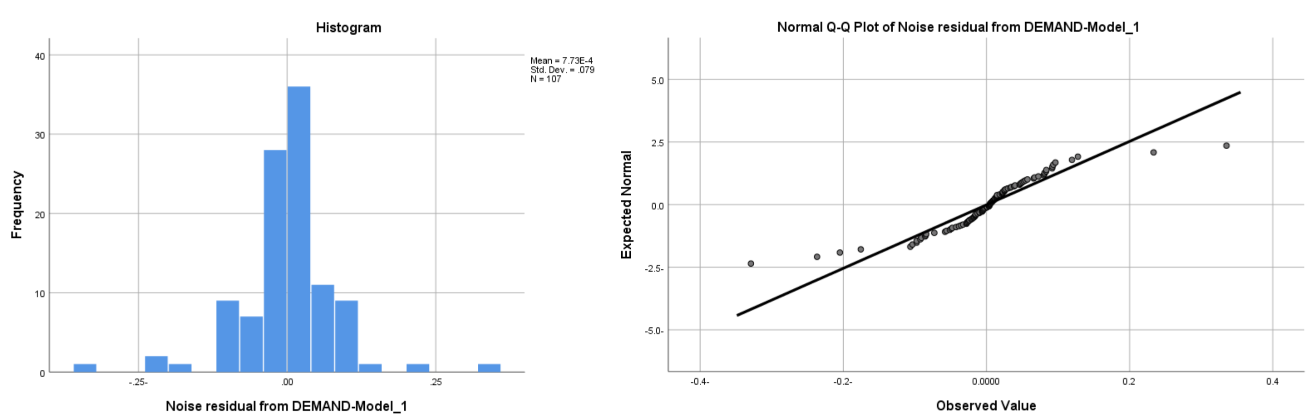
Graph (22): represents the coefficients of auto correlation and partial correlation after the seasonal displacement of the time series(demand series)

3- Residues Distribution Test:

The purpose of this test is to confirm that residues distribution follows the normal distribution, this can be known through graphing the natural probability (normal probability plot) and the (histogram) shows in the graphs (23) & (24), and from these graphs it's clear that the residues are normal.



Graph (23): represents the normal Q-Q plot of noise residual from supply –model-



Graph (24): represents the normal Q-Q plot of noise residual from demand –model.

- ## Prediction Stage:

After passing through proper models of supply/demand amount data identification steps, and the estimation of their parameters and testing the models to make sure it is suitable, the prediction models are used to predict the future values (supply and demand).

Table (29): the predictive value(supply) of ARIMA(2,1,2)(0,1,1)) model.

Month	y_t	$\hat{y}_t$	Error
JAN(1/2020)	552465	620054.28	0.12234=12.2%
FEB(2/2020)	548691	624498.51	0.13816=13.8%
MAR(3/2020)	598768	661672.85	0.10505=10.5%
APR(4/2020)	637289	675468.57	0.05991=5.9%
MAY(5/2020)	702951	737730.02	0.0495=4.9%

Table (30): the predictive value(demand) of ARIMA(2,1,2)(1,1,1) model.

Month	y_t	$\hat{y}_t$	Error
JAN(1/2020)	325261	279692.45	0.140098413=14%
FEB(2/2020)	335053	261745.43	0.218793952=21.8%
MAR(3/2020)	367566	286662.38	0.220106375=22%
APR(4/2020)	377204	291026.77	0.228463192=22.8%
MAY(5/2020)	379944	316817.82	0.166146011=16.6%

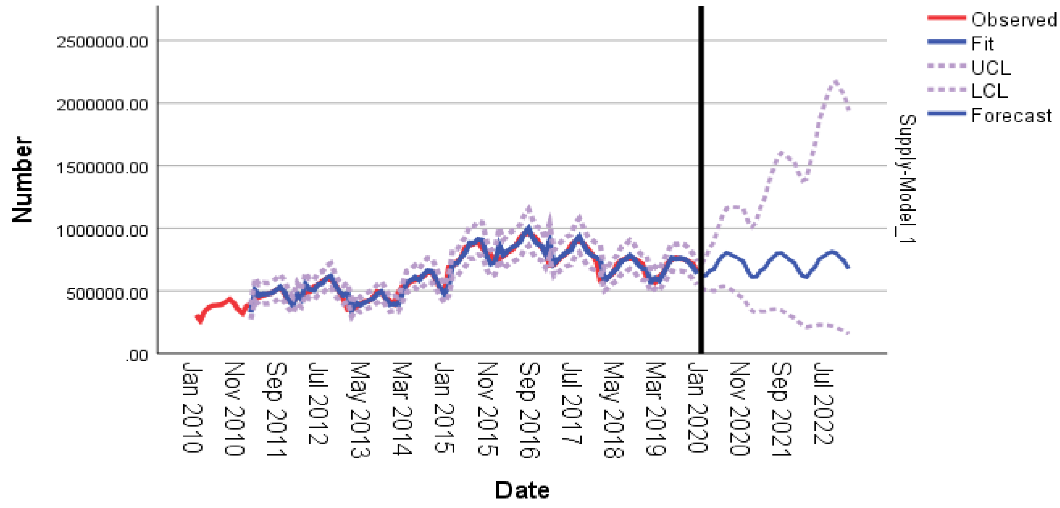
From the tables (29) & (30), comparison will be made between the original value and the predictable values from month 1 to 5 of 2020, these values shows harmony up to a certain extent with the original values.

Prediction accuracy criteria MAE, MAPE, RMSE are considered an indication of how efficient it is in prediction as explained in the following table (31):

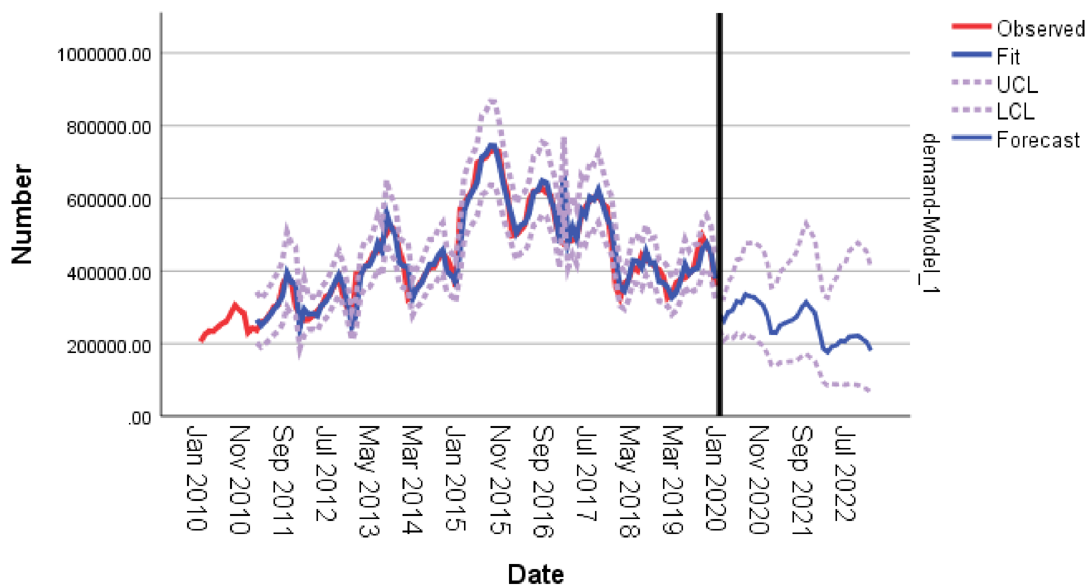
Table (31): Represents the results of the test of statistical criteria (MAE)(MAPE)(RMSE)

Time Series	Criteria		
	RMSE	MAPE	MAE
Supply	11881.1368	0.39579	2327.1686
Demand	14550.8642	0.811339953	2909.02625

The graphs (25) & (26) show the behaviour of time series curves (supply /demand) the future of the years (2020-2021-2022).



Graph (25): Supply curve for a year (2010-2021-2022)

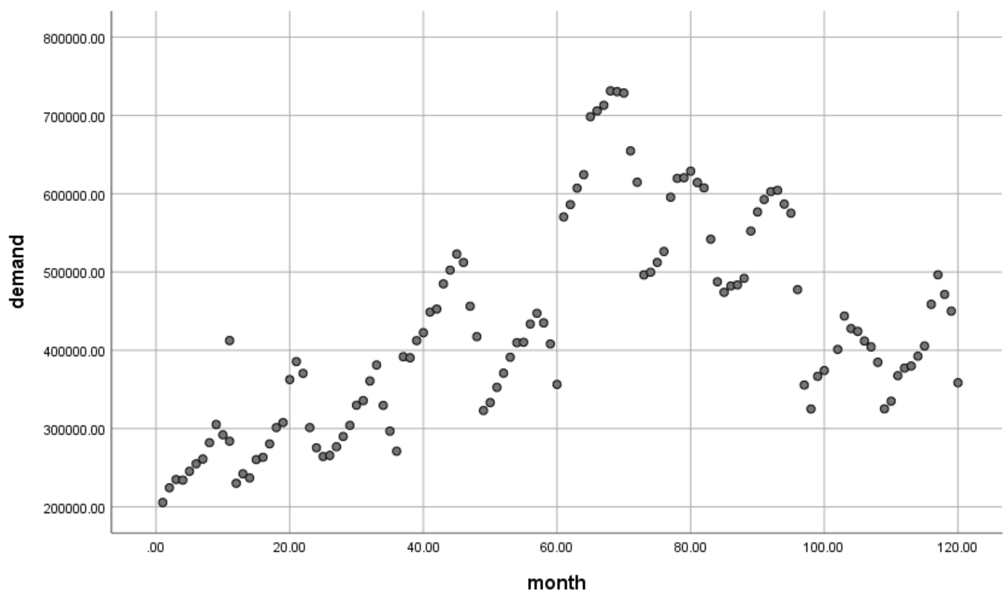


Graph (26): demand curve for a year (2010-2021-2022)

3.3.3.3. Regression Analysis Model Application Stages for Prediction

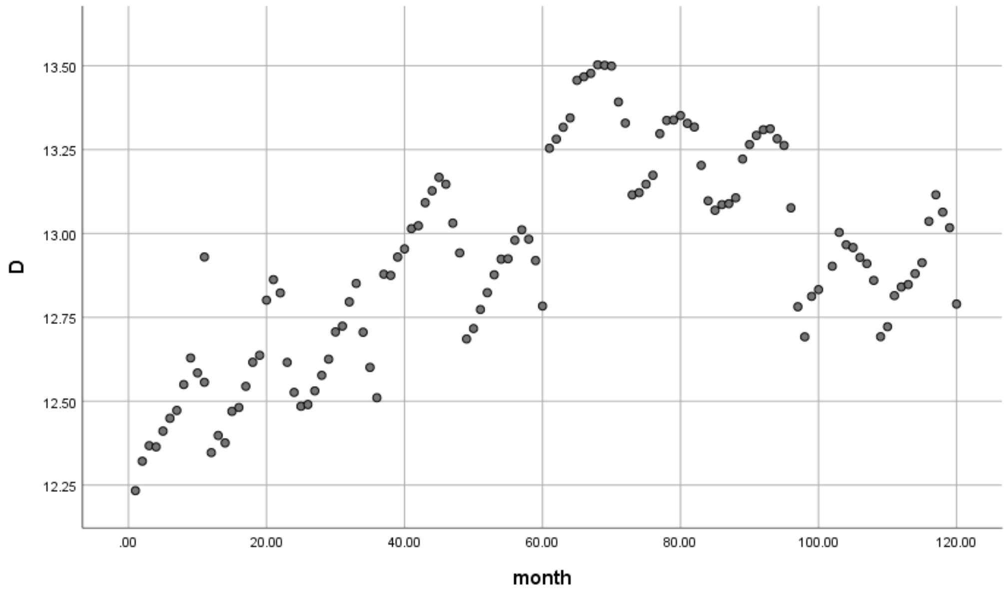
- **Data Series Scatter Plot**

Scatter plot between the two variables (the amount of water demand - the months) is drawn to determine the correlation and direction of the relationship as the following graph (27):

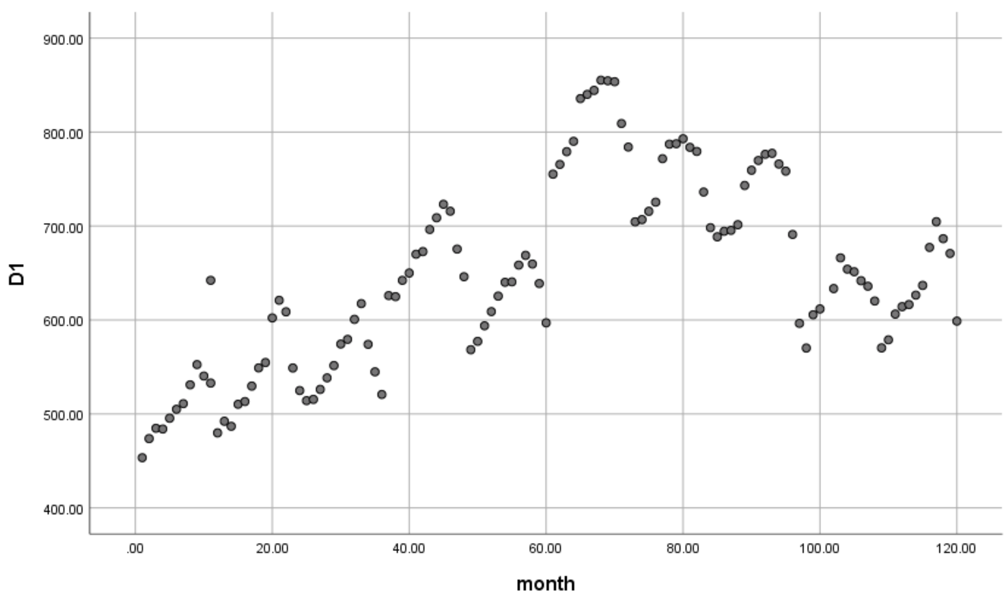


Graph (27): represents demand rate [cm] in terms of time [months]

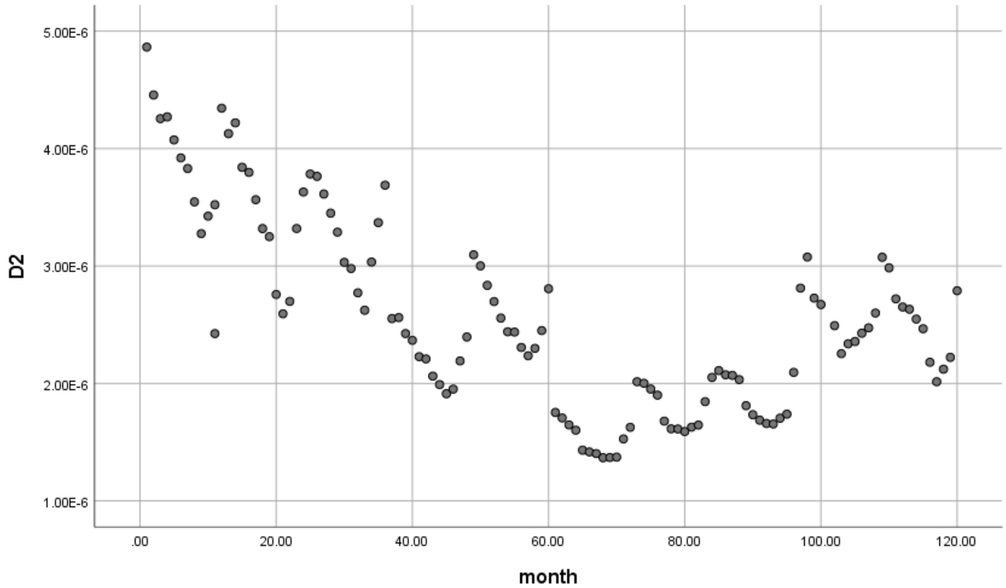
From the above scattered plot we noticed that the data appears as a curve not as linear model, this indicates a problem in the data variance, and the reason is due to the fact that the data is not subject to normal distribution, so in order to solve this problem we have made some transformation such as (logarithmic - square root - 1/Y) and graphs (28-29-30) showing the results we obtained:



Graph (28): represents scatter plot after transmodeling data to LN (Logarithm)



Graph (29): represents scatter plot after transmodeling data to square root (√)



Graph (30): represents scatter plot after transmodeling data to 1/y.

Graph 30 shows that the problem remains the same and none of the previous transactions can be applied to the studied data, so we resort to the use of non-linear regression.

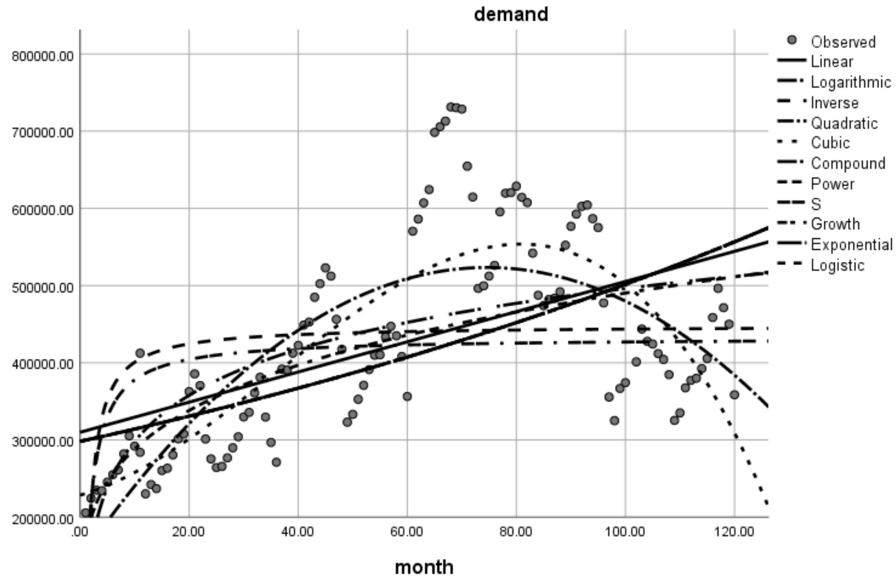
To find out the appropriate regression equation of studied data, we examined the value of the Pearson correlation coefficient$R$, and the selection coefficientR^2, and it was shows that the cubic nonlinear regression model was the best; for having the highest value for the correlation coefficient R and the selection coefficientR^2, and the table (32) showing the selection coefficient for the proposed models:

Table (32): Models Summary and Parameter Estimates

Dependent Variable: demand

	Model Summary					Parameter Estimates			
Equation	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.267	42.990	1	118	.000	309941.918	1953.360		
Logarithmic	.378	71.564	1	118	.000	100201.153	85985.287		
Inverse	.154	21.467	1	118	.000	448331.879	-477296.418-		
Quadratic	.577	79.712	2	117	.000	145910.499	10114.771	-67.756-	
Cubic	.629	65.451	3	116	.000	228497.328	2041.557	98.637	-.917-
Compound	.339	60.493	1	118	.000	298183.723	1.005		
Power	.486	111.746	1	118	.000	169478.782	.231		
S	.223	33.927	1	118	.000	12.978	-1.359-		
Growth	.339	60.493	1	118	.000	12.605	.005		
Exponential	.339	60.493	1	118	.000	298183.723	.005		
Logistic	.339	60.493	1	118	.000	3.354E-6	.995		

The independent variable is month.



- **Interpretation of the Relationship Between Variables (Correlation Coefficient) (R):**

  The model explains 62.9% of the variation in the amount of water demand.

Table (33): Cubic nonlinear regression Model Summary

.

R	R Square	Adjusted R Square	Std. Error of the Estimate
.793	.629	.619	81369.035

The independent variable is month.

- **Regression Model Estimation:**

Table (34): Model Coefficients Estimation

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Month	2041.557	2180.797	.540	.936	.351
month ** 2	98.637	41.902	3.246	2.354	.020
month ** 3	-.917-	.228	-3.456-	-4.026-	.000
(Constant)	228497.328	30289.907		7.544	.000

From the table (34), we can model a predictive regression equation:

$$\hat{Y}_i = 228497.328 + 2041.557X + 98.637X^2 + -0.917\ X^3$$

- **Models' Component Interpretation:**

β_0: The dependent variable average (the amount of water demand) is equal (228497.328)

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

We note from the table (34) that the value of the test is equal to (0.936) and that (p-value = 0.351) is greater than 0.05, which means don't reject the null hypothesis and that (β_1) it is not significant

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

We note from the table (34) that the value of the test is equal to (2.354) and that (p-value = 0.020) is less than 0.05, which means reject the null hypothesis and that (β_2) it is significant.

$$H_0: \beta_3 = 0$$

$$H_1: \beta_3 \neq 0$$

We note from the table (34) that the value of the test is equal to (-4.026) and that (p-value = 0.000) is less than 0.05, which means reject the null hypothesis and that (β_3) it is significant.

We note from the results of the hypothesis test the estimation of the models' parameters (β_0)(β_2)(β_3) are significance but the parameter (β_1) is not significance.

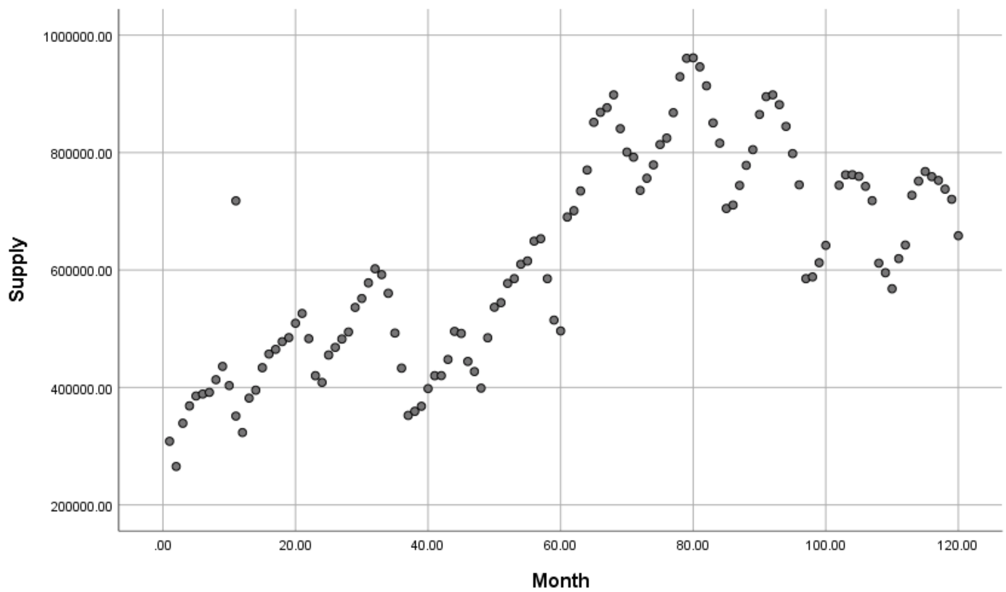
Table (35): the predictive value (demand) of Cubic nonlinear regression Model.

Month	y_t	$\hat{y}_t$	Error
JAN(1/2020)	325261	295148.1486	0.092581=9.3%
FEB(2/2020)	335053	280547.774	0.16268=16.3%
MAR(3/2020)	367566	265472.973	0.27775=27.7%
APR(4/2020)	377204	249918.7	0.33744=33.7%
MAY(5/2020)	379944	233879.453	0.3844=38.44%

When extracting the predictive values for 5 months, the results were as shows in the table (35), where it shows a little relatively error.

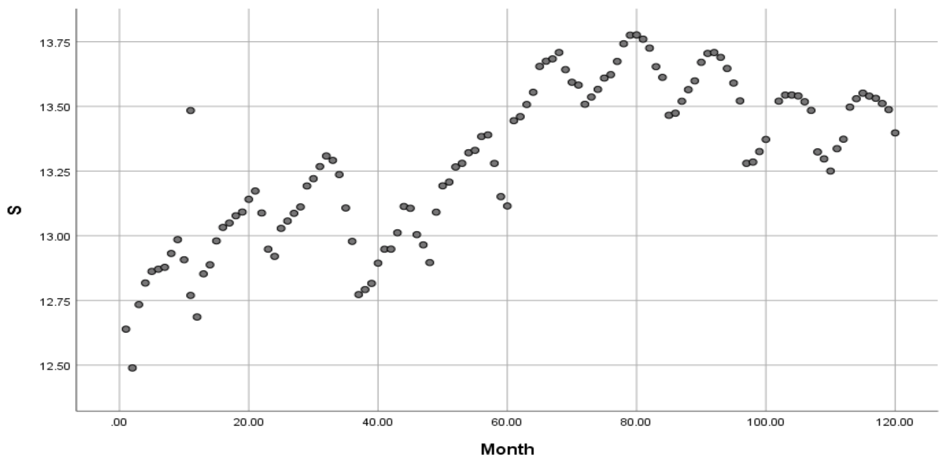
- **Data series (supply) scatter plot:**

Scatter plot between the two variables (the amount of water supply - the months) is drawn to determine the correlation and direction of the relationship as the following graph (31):

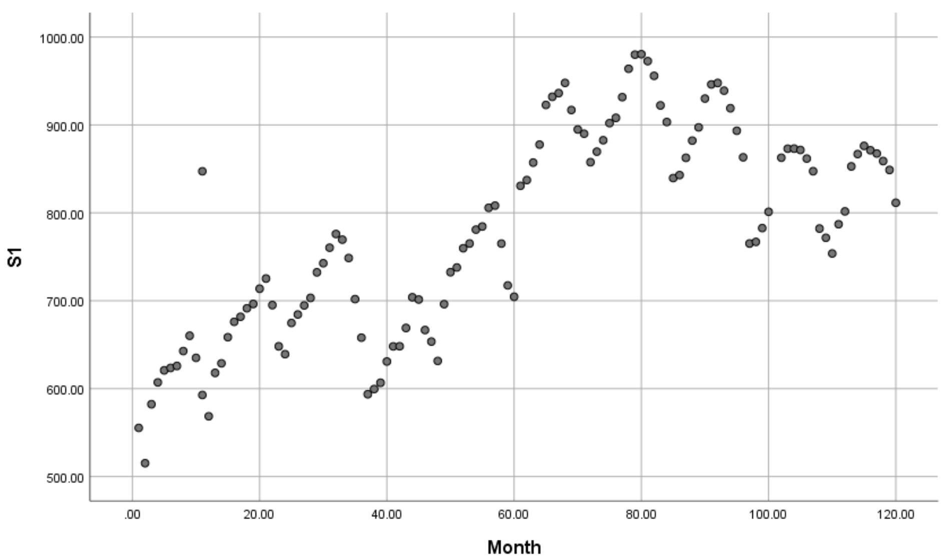


Graph (31): represents supply rate [cm] in terms of time [months]

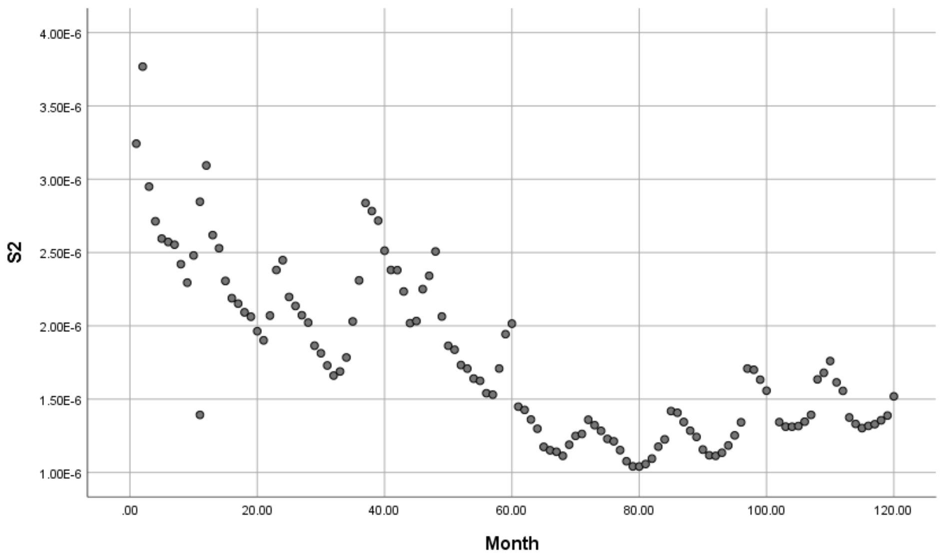
From the above scattered plot, we noticed that the data appeared as a curve not as linear model, this indicates a problem in the data variance, and the reason is due to the fact that the data is not subject to normal distribution. In order to solve this problem, we have made some transformation such as (logarithmic - square root - 1/Y) and graphs (32-33-34) showing the results we obtained:



Graph (32): represents scatter plot after transmodeling data to LN (Logarithm)



Graph (33): represents scatter plot after transmodeling data to square root (√)



Graph (34): represents scatter plot after transmodeling data to 1/y.

From the graphs we note that the problem remains the same and none of the previous transactions can be applied to the studied data, so we resort to the use of non-linear regression.

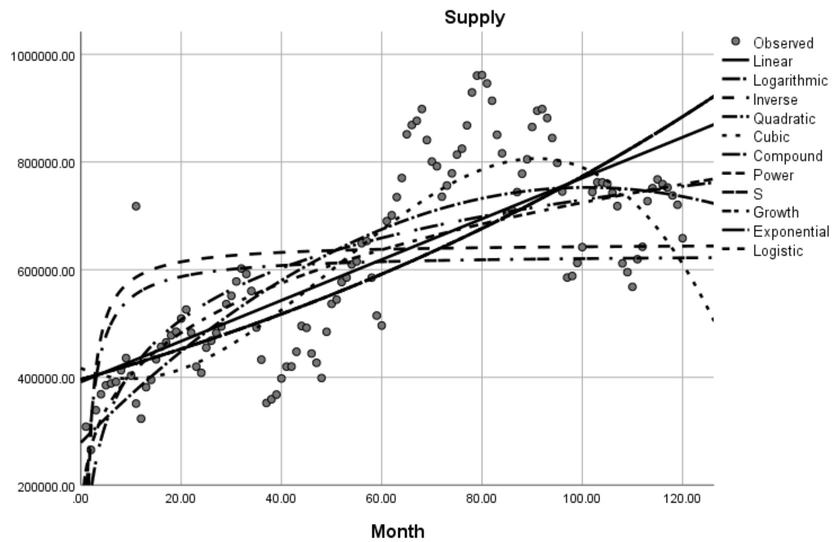
To find out the appropriate regression equation of studied data, we examined the value of the Pearson correlation coefficient R, and the selection coefficient R^2, and it was shows that the cubic nonlinear regression model was best; for having the highest value for the correlation coefficient R and the selection coefficient R^2, and the table(36) showing the selection coefficient for the proposed models:

Table (36): Models Summary and Parameter Estimates

Dependent Variable: Supply

	Model Summary					Parameter Estimates			
Equation	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.543	140.323	1	118	.000	391576.730	3785.844		
Logarithmic	.529	132.560	1	118	.000	92629.713	138320.491		
Inverse	.178	25.638	1	118	.000	649501.010	-698439.542-		
Quadratic	.622	96.387	2	117	.000	278929.048	9390.649	-46.531-	
Cubic	.701	90.620	3	116	.000	417070.664	-4113.282-	231.792	-1.534-
Compound	.581	163.765	1	118	.000	396145.654	1.007		
Power	.603	179.397	1	118	.000	226672.456	.252		
S	.240	37.273	1	118	.000	13.352	-1.384-		
Growth	.581	163.765	1	118	.000	12.890	.007		
Exponential	.581	163.765	1	118	.000	396145.654	.007		
Logistic	.581	163.765	1	118	.000	2.524E-6	.993		

The independent variable is Month.



- **Interpretation of the Relationship Between Variables (Correlation Coefficient) (R):**

The model explains 70.1% of the variation in the amount of water supply .

Table (37): Cubic nonlinear regression Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.837	.701	.693	99226.071

The independent variable is MONTH.

- **Regression Model Estimation:**

Table (38): Model Coefficients Estimation

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
MONTH	-4113.282-	2659.389	-.801-	-1.547-	.125
MONTH ** 2	231.792	51.098	5.613	4.536	.000
MONTH ** 3	-1.534-	.278	-4.254-	-5.522-	.000
(Constant)	417070.664	36937.251		11.291	.000

From the table (38) above we can model a predictive regression equation:

$$\hat{Y}_i = 417070.664 - 4113.282X + 231.792X^2 - 1.534X^3$$

- **Models' Component Interpretation:**

β_0: The dependent variable average (the amount of water supply) is equal (417070.664).

$$H_0 = \beta_1 = 0$$

$$H_1 = \beta_1 \neq 0$$

We note from the table (38) that the value of the test is equal to (-1.547) and that (p-value = 0.125) is greater than 0.05, which means don't reject the null hypothesis and that (β_1) it is not significant.

$$H_0 = \beta_2 = 0$$

$$H_1 = \beta_2 \neq 0$$

We note from the table (38) that the value of the test is equal to (4.536) and that (p-value = 0.000) is less than 0.05, which means reject the null hypothesis and that (β_2) it is significant.

$$H_0 = \beta_3 = 0$$

$$H_1 = \beta_3 \neq 0$$

We note from the table (38) that the value of the test is equal to (-5.522) and that (p-value = 0.000) is less than 0.05, which means reject the null hypothesis and that (β_3) it is significant.

We note from the results of the hypothesis test the estimation of the models' parameters (β_0)(β_2)(β_3) are significance but the parameter (β_1) is not significance.

Table (39): the predictive value (supply) of the Cubic nonlinear regression model.

$Month$	y_t	$\hat{y}_t$	error
JAN(1/2020)	552465	595455.64	0.077816=7.8%
FEB(2/2020)	548691	579731.556	0.056572=5.7%
MAR(3/2020)	598768	563348.168	0.059155 =5.9%
APR(4/2020)	637289	546296.272	0.142781=14.3%
MAY(5/2020)	702951	528566.664	0.24807=24.8%

When extracting the predictive values for 5 months, the results were as shown in the table (39). where it shows a little relatively error.

3.4.Comparison of the Markov Model, Time Series Model (ARIMA) and Regression Analysis Model

- **Comparison of the Markov Model and Time Series Model (ARIMA):**

For comparison between the three models, accuracy measures (RMSE – MAPE -MAE) were used, which are useful in developing a real vision for future estimates to predict the behaviour of the phenomenon to take the appropriate decision.

Table (40): Represents the results of the test of statistical criteria (MAE) (MAPE) (RMSE)

Time Series	Markov chains			Time Series (ARIMA)		
	RMSE	MAPE	MAE	RMSE	MAPE	MAE
Supply	11960.58	0.3104	1741.568	11881.137	0.396	2327.169
Demand	5496.596	0.2735	957.407	14550.864	0.8114	2909.026

Depending on the forecast measures as shows in the table (40), it appears that the Markov model for demand is out performs both the time series model (ARIMA(2,1,2)(0,1,1) & ARIMA (2,1,2)(1,1,1) because it has the lowest value of prediction accuracy criteria (MAPE & MAE) .on the other hand ARIMA model for supply has accuracy close to that of Markov chain model because it has the lowest value of prediction accuracy criteria (RMSE).

- **Comparison of Time Series Model (ARIMA) and Regression Analysis Model:**

Table (41): Represents the results of the test of statistical criteria (R^2)(RMSE)

Time Series	Time Series (ARIMA)		Regression Analysis	
	RMSE	R^2	RMSE	R^2
Supply	11881.137	0.930	99226.071	0.543
Demand	14550.864	0.923	81369.035	0.267

Based on the forecast metrics as shows in Table (41), it appears that the time series outperform the simple linear regression model because it has the highest value of R^2 and the lowest value of criterion (RMSE).

For comparison between the three models, accuracy measure (RMSE) was used, which is useful in developing a real vision for future estimates to predict the behaviour of the phenomenon which can be used in appropriate decision support for water resources management.

Table (42): Represents the results of the test of statistical criterion (RMSE)

Time Series	Markov chains	Time Series (ARIMA)	Regression Analysis
	RMSE	RMSE	RMSE
Supply	11960.58	11881.14	99226.071
Demand	5496.596	14550.86	81369.035

Depending on the forecast measure as shows in the table (42), it appears that the Markov model is outperforms both the time series model ARIMA (2,1,2)(1,1,1) and the simple regression model in the demand series because it has the lowest value of prediction accuracy criterion.

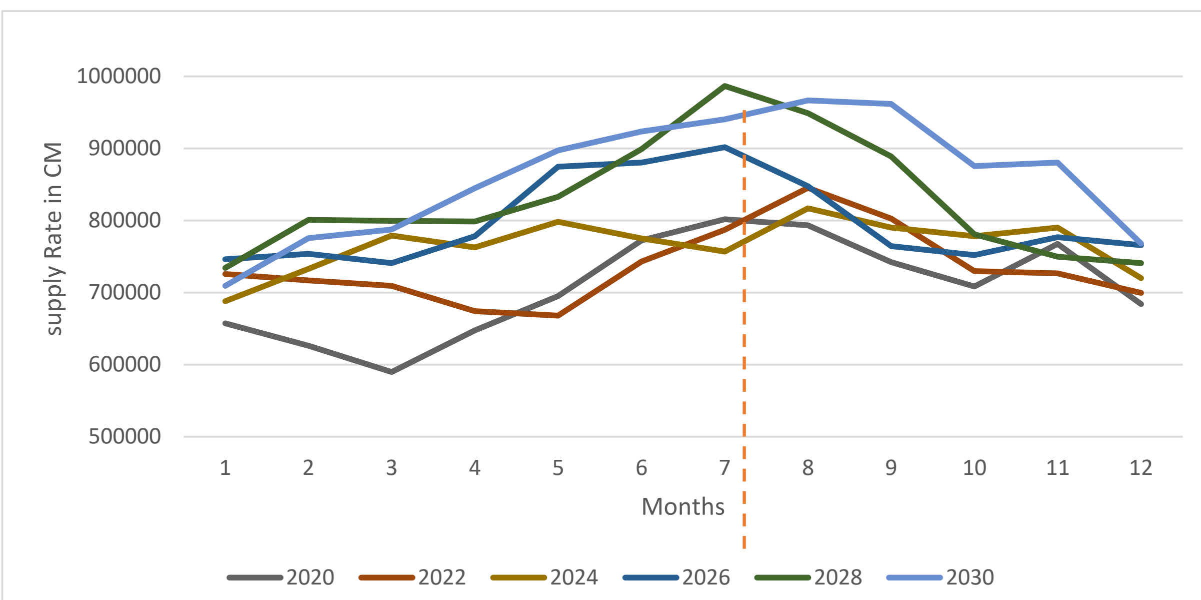
And it appears that the time series ARIMA (2,1,2)(0,1,1) model is outperforms both the Markov chains model and the simple regression model in the supply series because it has the lowest value of prediction accuracy criterion.

Therefore, the Markov model is the best and most accurate model in predicting the future of the series under study.

Based on these results, we can predict the amount of supply and demand for water for the period of time (2022-2024-2026-2028-2030), using the Markov prediction model.

Table (43): Represents the Predicted values of the amount of water supply during the years (2020-2022-2024-2026-2028-2030) by Markov prediction model.

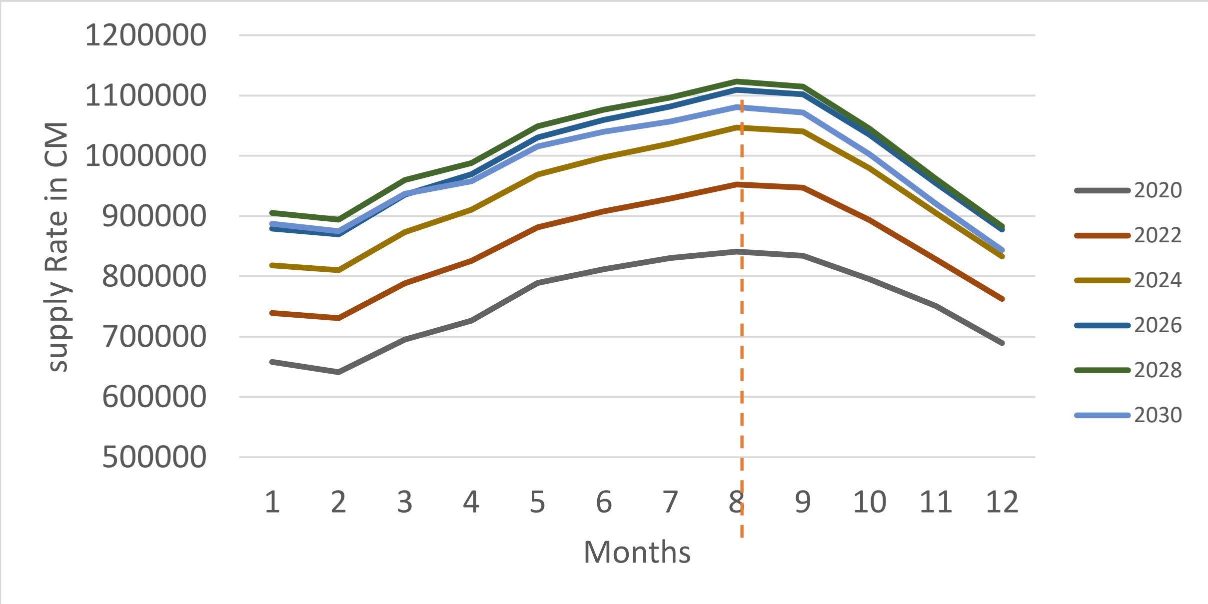
Month	Supply					
	2020	2022	2024	2026	2028	2030
JAN	657096.9	725571.8	688073.1	746066.1	733998	709123.8
FEB	625968.1	716629.6	732726.3	753556.5	800957.2	775308.7
MAR	589672.5	709087.6	778790.4	740619.3	799652.2	787367.6
APR	647432	674180.7	762546.9	778236.1	798539	844739.6
MAY	695110.3	667992.4	798318.2	874389.6	832730.9	897157.5
JUN	772181.9	742898.4	774827.1	880305.3	899032.8	923267.3
JUL	801702.3	787065.4	756896.8	901883.9	986513.2	940167.9
AUG	793048.9	844999.8	816889.1	847539	948793.1	966770.5
SEP	742230.2	802738.5	790146.9	764193.9	888921.2	961724.8
OCT	708306.1	729655.9	778238.9	751950.6	780613.7	875303.8
NOV	767531.2	726514.3	789948.2	776747.8	749540	880298
DEC	684014.1	699610.7	719793.7	765721.6	740870	767966.5



Graph (35): represents supply rate [cm] in terms of time [months]

Table (44): Represents the Predicted values of the amount of water supply during the years (2020-2022-2024-2026-2028-2030) by ARIMA (2,1,2) (0,1,1) prediction model.

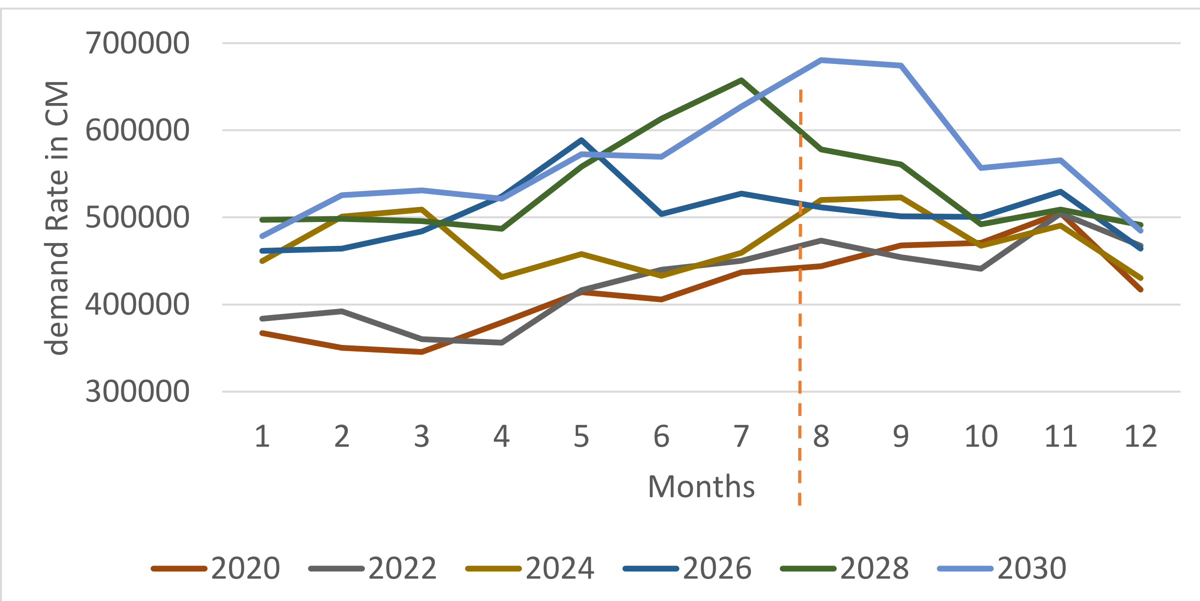
Month	Supply					
	2020	**2022**	**2024**	**2026**	**2028**	**2030**
JAN	658106.85	739427.62	818233.26	879011.26	905081.41	887030.57
FEB	641095.89	730836.24	810481.25	869890.87	894352.46	874976.47
MAR	695330.98	788726.15	873433.41	935711.48	959913.28	936888.29
APR	726226.35	825436.68	910368.25	969567.00	987953.75	957531.63
MAY	789110.39	881366.57	968803.22	1030624.83	1049133.55	1015668.49
JUN	811778.53	908067.47	997740.83	1059723.00	1076705.54	1040170.52
JUL	830478.45	929221.67	1020305.65	1081779.52	1096760.82	1057075.15
AUG	841149.30	952606.17	1047071.36	1109364.68	1123334.73	1081098.91
SEP	834001.80	947133.33	1040771.20	1101958.15	1114788.03	1071630.30
OCT	795382.26	893351.73	979308.05	1035082.30	1045260.59	1002832.75
NOV	750668.98	828498.89	905114.22	954551.29	961806.35	920607.17
DEC	689513.77	762742.12	833008.92	877576.90	882996.53	843807.02



Graph (36): represents supply rate [cm] in terms of time [months].

Table (45): Represents the Predicted values of the amount of water demand during the years (2020-2022-2024-2026-2028-2030)

MONTH	Demand					
	2020	2022	2024	2026	2028	2030
JAN	367008.1	383799.7	449900.2	461683.6	497046.1	478572.3
FEB	350048.1	392290.6	500945.1	464314.8	498154.5	525554.2
MAR	345417.3	360143.7	508825.9	484002	495711.7	530852.6
APR	379088.7	356261.4	431393.1	524171.9	487096	521347.3
MAY	414057.3	416288.8	457854.5	588661.9	558370.4	572659.2
JUN	405784.4	439835.8	433022.5	503878.6	613528.1	569710.4
JUL	436841.9	450107.1	459365.1	527519.6	657490.9	627393
AUG	443976.2	473492.8	520005.2	511528.1	577966.3	680779.2
SEP	467922.6	454277.6	522847.3	501325.5	560861.9	674398.7
OCT	470772.3	440847.1	467402.1	500492.8	492246.5	556875.8
NOV	505044.8	504500.4	490452.3	529554.5	509127	565636
DEC	417194.8	467456.1	430398.6	464173.6	491521	484706



Graph (37): represents demand rate [cm] in terms of time [months]

Chapter 4

Conclusion and Recommendations

4.1. Conclusions

This study aimed at predicting the amount of water (supply and demand) at long run in Tulkarm city based on the actual data. For this purpose, Markov chains model has been chosen and tested the best and valid model to undergoes such prediction.

Upon the results of the time series data that has been generated for the amount of water (supply, demand) based on the not actual municipality data within the period from (JAN, 2010) to (DEC, 2019). The following remarks can be concluded:

1- The results indicated that the amount of water supply will stabilize at (n=24), where the value of increase state (0.558), decrease state (0.3671) , and stability state (0.0754). This means that there will not be a significant change in water supply during each 24 months, in this aspect Palestinian Water Authority and Tulkarem Municipality must seek new or alternative water resources that must be compensate the decremental rate in yield each 24 months which is about 36.71%/ year.

2- The water demand series show stability at (n=16) where the value of increase state (0.5546), decrease state (0.3445), and stability state(0.1008). This indicate a significant change in water demand each 16 months at a rate of 34.45%/year. This incremental change in water demand followed many factors including population growth and the yearly variation in rainfall amount.

3- The transitional probability matrix for the amount of water supply (stable distribution), shows that: the ratio of increase in water supply is 55.8% at long run, while the decrease ratio 36.71%, and the stability ratio 7.54%.

4- The transitional probability matrix for the amount of water demand also shows that the ratio of increase in water demand is 55.46%, while decrease ratio 34.45%, and the stability ratio (10.08%), as long as there are no interfered circumstances that affect these ratios. Such results statistically mean that we predict an increase in the amount of water supply and demand in the future at a ratio of 55.8%%/24 months for water supply and 55.46%/16 months for water demand.

- The study also aimed to choose the best model for the water (supply, demand) in Tulkarem city, three statistical models has been tested (Markov chains - time series model (ARIMA) - simple regression model), where these models were tested on water supply and demand data. The comparison study for the three models reflect the following:

  1- Time series for water supply and demand represents an unstable time series.

  i. The amount of water supply series shows unstable trend, however, taking the first difference and the first seasonal component, make the series more stable. This mean that the model is susceptible for several factors that may affect its stability such as seasonal variation and fluctuation in supply.

  ii. The amount of water demand series shows unstable trend, however, taking the first difference and the first seasonal component, make the series also stable. The same as for supply, this also mean that the model is susceptible for several factors that may affect its stability such as seasonal variation and fluctuation in demand.

5- After examining several models and through the evaluation criteria for comparing the proposed ARIMA models, it was found that the best amount of water supply series was by using ARIMA (2,1,2) (0,1,1), and the best amount of water supply series model was by using ARIMA (2,1.2) (1,1,1).

6- ARIMA (2,1,2) (0,1,1) & (ARIMA) (2,1,2) (1,1,1)) was found to be the best tool for the proposed models because they have less value of the evaluation criteria (AIC)(BIC), which exceed all other models by passing all tests and residual check properties of the model.

7- It was also found that the cubic nonlinear regression analysis model ($\hat{Y}_i = 228497.328 + 2041.557\text{X} + 98.637X^2 - 0.917X^3$) was the most outperforms model for the water demand series than the water supply series, where the most appropriate model for this purpose was the ($\hat{Y}_i = 417070.664 - 4113.282\text{X} + 231.792X^2 - 1.534X^3$) model.

8- The results of the models were also compared according to the predictive accuracy criterion (RAMS) between the three used models (Markov model, the model (ARIMA) and the regression model). The comparison results show that the Markov model outperformed the other models in demand series. In addition, results shows that the (ARIMA) model outperformed the other models in supply series.

9- In general, it was found that the Markov model is the best applied tool to predict the time series for the amount of water demand in 2021-2025-2030 in Tulkarem city. While, the time series ARIMA (2,1,2) (0,1,1) found to be the best applied model to predict the time series for the amount of water supply in 2021-2025-2030 in Tulkarem city.

4.2. Recommendations

Based on the models results and the comparison between models and based on information collected from various and several sources. The following recommendations has been made to the stakeholders in water and agricultural sectors, thus to draw their attention to take the best decision that leads to ensure the sustainability of water resources at long run in Tulkarem city. Therefore, the following recommendations have been highlighted:

1- PWA- Palestinian Water authority has to take into account the amount of increase in water demand that faced by the decrease in water supply, specially that the results show more decrease in water supply in a shorter period that reached 16 months while the increase on demand take relatively longer period of 24 months.

2- Researchers are encouraged to increase their interests and concern on studying the time series related to the prediction of the amount of water supply and demand at long run, using Markov model and ARIMA model respectively, because of its great importance and great impact on society.

3- Markov chains is recommended to be used in creating a transition probability matrix for other fields that can have a transition probability matrix.

4- Researchers are also encouraged to conduct studies that aim to compare Markov models with other advanced statistical methods in many fields.

5- Palestinian Water Authority PWA is invited to continue the application of these mathematical models, that can give significant results and information about the changes in supply and demand, in order to predict its changes with incremental population growth and climate change conditions, thus to take the proper action that can help the PWA in better management for the available water resources to ensure its sustainability for the next generations.

References

1- Alrubaiee, L., & Al-Nazer, N. (2010). Investigate the impact of relationship marketing orientation on customer loyalty: The customer's perspective. International Journal of Marketing Studies, 2(1), 155.

2- Alsiddiky, A., Awwad, W., Bakarman, K., Fouad, H., & Mahmoud, N. M. (2020). Magnetic resonance imaging evaluation of vertebral tumor prediction using hierarchical hidden Markov random field model on Internet of Medical Things (IOMT) platform. Measurement, 159, 107772.

3- Bata, M. T. H., Carriveau, R., & Ting, D. S. K. (2020). Short-term water demand forecasting using nonlinear autoregressive artificial neural networks. Journal of Water Resources Planning and Management, 146(3), 04020008.

4- Cresswell, M., & Naser, G. (2013). A water resources management strategy for small water districts—a case study of the South East Kelowna irrigation district. Canadian Journal of Civil Engineering, 40(6), 499-507.

5- Du, H., Zhao, Z., &Xue, H. (2020). ARIMA-M: A New Model for Daily Water Consumption Prediction Based on the Autoregressive Integrated Moving Average Model and the Markov Chain Error Correction. Water, 12(3), 760.

6- Hatemi, G., Christensen, R., Bang, D., Bodaghi, B., Celik, A. F., Fortune, F., ... &Mahr, A. (2018). 2018 update of the EULAR recommendations for the management of Behçet’s syndrome. Annals of the Rheumatic Diseases, 77(6), 808-818.

7- Haquea , Rahmana, jotham , Lydia: Principal Component Regression Analysis in Water Demand Forecasting. Australia . Uganda . Blue Mountains 2013 .

8- Gagliardi, F., Alvisi, S., Kapelan, Z., &Franchini, M. (2017). A probabilistic short-term water demand forecasting model based on the Markov Chain. Water, 9(7), 507.

9- Ghalehkhondabi, I., Ardjmand, E., Young, W. A., &Weckman, G. R. (2017). Water demand forecasting: review of soft computing methods. Environmental monitoring and assessment, 189(7), 1-13.

10- Melloul, D., Marshak, S., &Cerasi, E. (2002). Regulation of insulin gene transcription. Diabetologia, 45(3), 309-326.

11- Marshak, S., Benshushan, E., Shoshkes, M., Havin, L., Cerasi, E., &Melloul, D. (2000). Functional Conservation of Regulatory Elements in thepdx-1 Gene: PDX-1 and Hepatocyte Nuclear Factor 3β Transcription Factors Mediate β-Cell-Specific Expression. Molecular and cellular biology, 20(20), 7583-7590.

12- Nazer, D. W., Siebel, M. A., Van der Zaag, P., Mimi, Z., &Gijzen, H. J. (2010). A financial, environmental and social evaluation of domestic water management options in the West Bank, Palestine. Water resources management, 24(15), 4445-4467.

13- Papadopoulos, A., Fatta, D., Parperis, K., Mentzis, A., Haralambous, K. J., &Loizidou, M. (2004). Nickel uptake from a wastewater stream produced in a metal finishing industry by combination of ion-exchange and precipitation methods. Separation and Purification Technology, 39(3), 181-188.

14- Pacchin, E., Gagliardi, F., Alvisi, S., &Franchini, M. (2019). A comparison of short-term water demand forecasting models. Water resources management, 33(4), 1481-1497.

15- Rosegrant, M. W. (1997). Water resources in the twenty-first century: Challenges and implications for action (Vol. 20). Intl Food Policy Res Inst

16- Storelli, M. M. (2008). Potential human health risks from metals (Hg, Cd, and Pb) and polychlorinated biphenyls (PCBs) via seafood consumption: estimation of target hazard quotients (THQs) and toxic equivalents (TEQs). Food and Chemical Toxicology, 46(8), 2782-2788.

17- Sempewo, J. I., & Kyokaali, L. (2016). Prediction of the future condition of a water distribution network using a Markov based approach: a case study of Kampala water. Procedia Engineering, 154, 374-383

18- Shahraki, A. F., Yadav, O. P., & Liao, H. (2017). A review on degradation modelling and its engineering applications. International Journal of Performability Engineering, 13(3), 299-314.

19- Tuama, S. A. (2012). Using Analysis of Time Series to Forecast numbers of The Patients Malignant Tumors in Anbar Provinc with. AL-Anbar University journal of Economic and Administration Sciences, 4(8).

20- Tayeb, S. E., Abdalla, S., Mørkve, O., Heuch, I., & Van den Bergh, G. (2014). Injuries in Khartoum state, the Sudan: a household survey of incidence and risk factors. International journal of injury control and safety promotion, 21(2), 144-153.

21- Tiehatten, B., Assaf, K., Barhumi, H., Bastiaanssen, W., Ghanem, M., Jayyousi, A., ... & Zayed, O. Paper 3 Evaluation of Water Harvesting and Managed Aquifer Recharge Potential in Upper Fara'basin in Palestine: comparing MYWAS and Water Productivity approaches.

22- Wang, B., Liu, L., Huang, G. H., Li, W., & Xie, Y. L. (2016). Forecast-based analysis for regional water supply and demand relationship by hybrid Markov chain models: a case study of Urumqi, China. Journal of Hydroinformatics, 18(5), 905-918.

23- World Water Assessment Programme (United Nations), & UN-Water. (2009). Water in a changing world.

24- West, C. P., Dyrbye, L. N., &Shanafelt, T. D. (2018). Physician burnout: contributors, consequences and solutions. Journal of internal medicine, 283(6), 516-529.

25- Xenochristou, M., Hutton, C., Hofman, J., &Kapelan, Z. (2020). Water demand forecasting accuracy and influencing factors at different spatial scales using a Gradient Boosting Machine. Water Resources Research, 56(8), e2019WR026304.

26- Yang, Z., Ce, L., & Lian, L. (2017). Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. Applied Energy, 190, 291-305.

27- Yalçıntaş, M., Bulu, M., Küçükvar, M., & Samadi, H. (2015). A framework for sustainable urban water management through demand and supply forecasting: The case of Istanbul. Sustainability, 7(8), 11050-11067.

28- اياد يونس عبد الحميد ابو لبدة. (2018). استخدام نموذجي ماركوف و ARIMA في التنبؤ بسعر صرف الدولار مقابل الشيكل (Doctoral dissertation).

29- - أبو فخيدة، إسماعيل. (2013 .(إستخدام الأساليب الإحصائية في دراسة دور الأطفال في البيئة المنزلية في محافظة رام الله و البيرة، فلسطين (Master dissertation, Birzeit University).

30- احمد & ,حسن الجيلي البشير. (2016). دراسة العوامل المؤثرة على ضغط الدم باستخدام اختبار مربع كاي (دراسة حالة مستشفى الداير بني مالك العام بمنطقة جازان) بالمملكة العربية السعودية.

31- أمينه سليمان حسين، & احمد محمد عبدالله حمدي. (2017). دراسة مقارنة للتنبؤ بتقدير وتحليل العوامل المؤثرة في حوادث المرور باستخدام السلسلة الزمنية التدخلية في ولاية الخرطوم (Doctoral dissertation، جامعة السودان للعلوم والتكنولوجيا).

32- بن حملة، شادية، صرارمة & ,عبد الوحيد. (2016). إستخدام السلاسل الزمنية في التنبؤ بمبيعات مؤسسة الإقتصادية.

33- بوساق عبد المجيد. (2017). التنبؤ بالمبيعات باستخدام السلاسل الزمنية دراسة حالة مؤسسة (Doctoral dissertation، جامعة محمد بوضياف).

34- شلبي، زلفي عبد الفتاح. (2002). دراسة مقارنة لبعض أساليب التنبؤ بإستخدام السلاسل الزمنية.

35- عمار فردريك جون البازي & ,صبا زكي محمد العباسي. (2006). تطبيقات سلاسل ماركوف في كلية التمريض–جامعة بغداد.(9) ,AL-MANSOUR JOURNAL .

36- عبد القادر بوالسبت. (2015). -في التنبؤ بإنتاجية القمح في الجزائر-Chaînes Markov- استخدام سلاسل ماركوف .مجلة العلوم الإنسانية.183-171 ,

37- مراس & ,محمد. (2017). أهمية نمذجة التنبؤ باستخدام نماذج السلاسل الزمنية على معطيات اتخاد القرار في المؤسسة الاقتصادية الجزائرية _ دراسة حالة القطاع الخدماتي._

38- HADRI, K. (2015). استخدام السلاسل الزمنية من خلال منهجية بوكس جينكينز في اتخاذ القرارالانتـاجي دراسـة حالـة مطـاحن الريـاض سـطيف وحـدة تقـرت–فـي الفتـرة (2008-2013) (Master dissertation, Université Mohamed Khider-Biskra).

39- م. م. اوات سردار وادي م. م. سـامية خالـد حسـن. (2013). التنبـؤ بأسـعار الـذهب وذلـك بأستخدام سلاسل ماركوف وطريقة الأمكان الأعظم (MLE). AL-dananeer, 1(4).

40- https://www.researchgate.net/publication/329308860_astkhdam_slasl_markwf_lltnbw_balarqam_alqyasyt_lasar_almsthlk_fy_alraq