# Forecasting the Amount of Water Consumed in Erbil City using Time Series Model (SARIMA)

**ID No. 2594**

(PP 285 - 296)

https://doi.org/10.21271/zjhs.23.6.17

**Saman Hussein Mahmood**
Department of Statistics and Informatics / Administration and Economics College/ Salahaddin University-Erbil
saman.mahmood@su.edu.krd

**Imad Ali Omar**
Department of Civil Engineering / Engineering College / Salahaddin University-Erbil
imad.omar@ su.edu.krd

**Received: 16/01/2019**
**Accepted: 13/10/2019**
**Published:05/12/2019**

**Abstract**

Forecasting in the time series is one of the essential subjects in statistical science because its indicators helps the government to make a good plan to get an accurate decision in the future. This research was conducted to analyse the time series data for the amount of water consumed in Erbil City for the years proceeding the period between (1/1/2014) until of (30/6/2019). In the analysis, the study found non-stationary in the time series, so the necessary transformations were taken on the series to remove the non-stationary. Thus, the best model was found among the obtained models, SARIMA (0, 1, 2) $(1, 1, 1)_{12}$ produces the best results based on the minimum statistical criteria (RMSE, MAE, MAPE) used for comparison.

Finally, using the best model to forecast the water monthly average consumed in Erbil City for 12 months (6 months in 2019 and 6 months in 2020) and the results appeared to be consistent with those actual values in the time series.

**Keywords:** Forecasting, SARIMA, Water Consumed, Time series.

## 1- Introduction

Time series forecasting is a critical region of anticipating in which past perceptions of a similar variable assembled and broke down to produce a model are depicting the fundamental relationship. The model is then used to extrapolate the time arrangement into what's to come this displaying technique is especially helpful when little information is accessible on the basic information producing processor no agreeable illustrative model connects the forecast factors that relate the expectation variable to other logical factors (Zhang, 2003). There are numerous methodology and models to study the time-series information; subsequently, finding a fitting forecasting model for a particular water request isn't a simple undertaking (Ghiassi et al., 2008).

Time arrangement demonstrating is a dynamic research zone which has pulled in the convergence of analysts' group over the most recent couple of decades. The principle point of time series displaying is to precisely gather and rigorously consider the past estimations of a time series to build up a proper model which portrays the inalienable structure of the series (Adhikari and Agrawal, 2013). In spite of the fact that ARIMA models are extremely adaptable in that they can speak to a few different sorts of time series, i.e., straightforward autoregressive (AR), unadulterated moving average (MA) and combined AR and MA (ARMA) series, their significant restriction is the pre-accepted direct type of the model. That is, a linear correlation structure is expected among the time arrangement esteems and along these lines, no nonlinear examples can be caught by the ARIMA model. The estimation of direct models to a perplexing certifiable issue isn't generally tasteful (Zhang, 2003).

Water request forecasting is of incredible financial and natural significance. Numerous components can influence straightforwardly or indirectly water consumption. These incorporate precipitation, temperature, demography, arrive utilize, estimating, and control. Climate conditions have been broadly utilized as inputs of measurable multivariate models for hydrological time arrangement modelling and forecasting (Caiado, 2009). Efficiently working and directing a water supply structure consumed short-term water request estimates, and the estimation of future common water require is integral to the arranging of a territorial water-supply system (Zhou et al., 2002).

Long-term estimating is required generally for arranging and design; while short-term estimating is appropriate in operation and administration. As specified by( Bougadis, et al. 2005), short-term request figures offer assistance water directors make superior educated water administration choices when adjusting the necessities of water supply, residential/industrial requests, and stream flows for fish and other territories. Short-term request projections offer assistance utilities arrange and oversee water requests for near-term occasions (Jain and Ormsbee, 2002).

Forecasting water request is an extremely dynamic field of study. Early works tended to this inquiry by utilizing chiefly normal statistical models. Lately, more refined displaying strategies have been applied to this issue (Jain et al., 2001). (Zhou et al. 2002) created time series models for daily water utilization in Melbourne, Australia. These request levels can be communicated as an element of atmosphere factors, (for example, air temperature, volume, and the event of precipitation) and earlier water request. Introduced a regression methodology in light of fuzzy groups connected to assess hydrological informational indexes. (Msiza et al. 2007) clarified a comparable report with the application to water request of time series estimating. Two different strategies recommended by (Chen and Zhang 2006), whereby hourly water request is anticipated utilizing Bayesian and non-Bayesian slightest squares SVMs. A streamlining of pump- scheduling in light of estimating civil water interest for the city of Seoul was proposed by (Kim et al. 2007). (Razali et al. 2018) conducted a study to forecast the consumption of water at the University Tun Hussein Onn Malaysia (UTHM). The proposed Holt Winter's an Auto-Regressive Integrated Moving Average (ARIMA) models were applied to forecast the water consumption expenditure in Ringgit Malaysia. It is found the ARIMA model showed better results regarding the accuracy of the forecast with lower values of MAPE and MAD. Analysis showed that ARIMA (2,1,4) model provided a reasonable forecasting tool for university campus water usage. (Hasan and Hasan 2017)conducted a study of Forecasting the amount of water consumed in Erbil City using Time Series Model SARIMA, they collected the data from ministry of Agriculture and water resources, and their data are in $m^3/s$, they found that the best model was $ARIMA(1,1,1)x(3,1,0)_{12}$, also they recommended to study the forecasting with another methods in the future.

In this paper, our analysis is based on monthly drinking water consumed information depending on a database collected over a period of time from 2014 to end of 2018 from Erbil Water Directorate. The paper is organized to find the best model for forecasting water consumed in the future by ARIMA programme.

## 2. Methodology

This section introduces some basic definitions and concepts.

**2.1 Time Series:** It is an arrangement of information points, estimated ordinarily at progressive focuses in time dispersed at uniform time interims. Whenever series can contain a few or the greater part of the segments (Trend (T), Cyclical (C), Seasonal (S) Irregular (I)) (Kitagawa, 2010).

These components may be combined in different ways.

Multiplied $Zt = T \times C \times S \times I$ or added $Zt = T + C + S + I$

**2.2 Autocorrelation Function (ACF):** The correlation coefficient between Zt and Zt-k is given by

$$\rho_k = \frac{cov(Z_t - Z_t - k)}{\sqrt{var(Z_t)\, var(Z_t - k)}} = \frac{\gamma k}{\gamma o} \quad (1)$$

And by regarding it as a function of lag k, it is called the ACF (Kitagawa, 2010).

**2.3 Partial Autocorrelation Function (PACF):**

The conditional correlation Corr. ($X_t$, $X_{t-k}$ | $X_{t-1}$, $X_{t-2}$… $X_{t-k-1}$) is normally alluded to as the halfway autocorrelation in time arrangement investigation (Kitagawa, 2010).

$$\emptyset_{kk} = \frac{|R_k|}{|R^*_k|} \quad (2)$$

Where $R_k$ is equal to the matrix $R^*_k$, in which the $k_{th}$ column is replaced with $\rho_k$.

2.4 **Stationary Time Series Models**:

1- Autoregressive Model (AR (p))**:** At the point when an incentive from a period arrangement is relapsed on past qualities from that same time series (Palit and Popovic, 2005).

$$Z_{t=C+\emptyset_1} Z_{t-1+\emptyset_2} Z_{t-2+\cdots+\emptyset_p} Z_{t-p+e_t} \quad (3)$$

Where $e_t$ : White Noise, C: Constant, $\emptyset_1, \emptyset_2$ …,, $\emptyset_p$ : Parameters.

2- Moving Average Model (MA (q)): A moving average model is adroitly a direct relapse of the present estimation of the arrangement against present and past (in secret) background noise terms (Palit and Popovic, 2005).

$$Z_{t=C+\theta_1} e_{t-1+\theta_2} e_{t-2-\cdots-\theta_p} e_{t-q+e_t} \quad (4)$$

Where $e_t$ : White Noise, C: Constant, $\theta_1, \theta_2, \ldots,\, \theta_{2q}$: Parameters

3- Autoregressive-Moving Normal Models (ARMA (p, q))*:* The mix of the AR and MA models makes up the ARMA demonstrate.

$$Z_{t=C+\emptyset_1 Z_{t-1}+\emptyset_2 Z_{t-2}+\cdots+\,\emptyset_p Z_{t-p}\,+\,e_t\,-\theta_1\, e_{t-1}-\,\theta_2\, e_{t-2}\,-\cdots-\,\theta_q\, e_{t-q}} \quad (5)$$

The nearness of both autoregressive and moving-average terms in the ARMA demonstrate empowers the portrayal of complex time arrangement with fewer parameters that would be required utilizing a comparing AR model.

Demonstrating a pattern or occasional examples are non-stationary in nature. In such cases, differencing (d) and power changes are frequently used to evaluate the pattern and to make the series stationary. ARIMA (p, d, and q) is speculation of an ARMA model to incorporate the instance of non-stationary too (Adhikari and Agrawal, 2013).

**4- Seasonal Autoregressive Integrated Moving Average (SARIMA) Models**

The ARIMA demonstrate is for non-regular and non-stationary information. Box-Jenkins have summed up this model to manage regularity. Their proposed display is known as the Seasonal ARIMA (SARIMA) model. In this model regular differencing of a suitable request is utilized to expel non-stationary from the series. An order seasonal occasional distinction is a contrast between perception and the comparing perception from the earlier year and is computed as $\mathbf{z_t = y_t - y_{t-s}}$ for a month to month time series s= 12 and for quarterly time arrangement s = 4. This model is for the most part named as the SARIMA (p, d, q) × (P, D, Q) s demonstrate(Adhikari and Agrawal, 2013).

$$Z_{t=C+\emptyset_s Z_{t-s}+\cdots+\,\emptyset_{ps}\, Z_{t-ps}\,+\,e_t\,-\theta_s\, e_{t-s}-\cdots-\,\theta_{qs}\, e_{t-qs}} \quad (6)$$

**2.5 Box-Jenkins Procedure to modelling ARIMA:**

Box - Jenkins Analysis refers to a systematic method of identifying, fitting, checking, and using integrated autoregressive, moving average (ARIMA) time series models. The method is appropriate for time series of medium to long length. The Box-Jenkins model building process accepts that the time series to be demonstrated is stationary. Else, it ought to be differenced a few times or the proper change until the point when it ends up stationary. Further to this, the regularity of the time series must be evacuated. The demonstrating strategy includes an iterative four-organize process (Palit and Popovic, 2005).

i. Model identification phase: Recognizable proof of the potential models by taking a gander at the example autocorrelations and the fractional autocorrelations.

ii. Model forecasting phase: Estimation of the obscure parameters by some enhancement techniques.
iii. Model validation phase: Checking the sufficiency of a fitted model by performing ordinary likelihood plot, ACF, and PACF on demonstrating residuals.
iv. Model forecasting phase: Estimate future results in view of the known information.

## 3. Results and discussions

### 3.1 The Study Area

Erbil is a capital city in the Kurdistan region; The Population of Erbil City is estimated as based on population data of Central Statistical Office (CSO). Currently, the population is estimated to be 1,365,000 and is rapidly developing with annual growth rate of 3.5% (Figure 1, Erbil Central Statistical Office).

Erbil is an extremely old settlement, which has a dramatic and winding history. Erbil City is located between 36.191113_N to 36° 11' 28.0068"_N latitude and 44° 0' 33.0012"E to 44.009167_E longitude. The about 197 km2 with elevation 414 m above sea level (Figure 2). Erbil is situated in the Kurdistan district of Iraq, a region with a semi-parched mainland atmosphere. The late spring season (June-September) is hot and dry, while the winters are colder and wet. Erbil has two types of water resources, first from the wells and the second from surface water treatment from the Greater-Zab River.

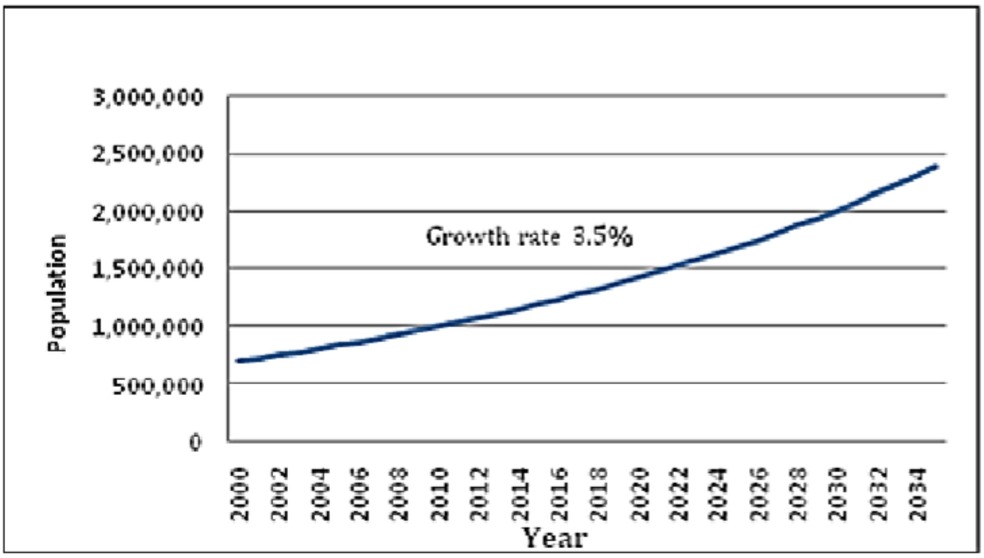


Figure 1: The Population of Erbil City and Growth Rate, (KOICA & K-Water, 2007)

The Greater-Zab River is the main wellspring of surface water in Erbil City for drinking and different purposes (AZIZ, 2007).

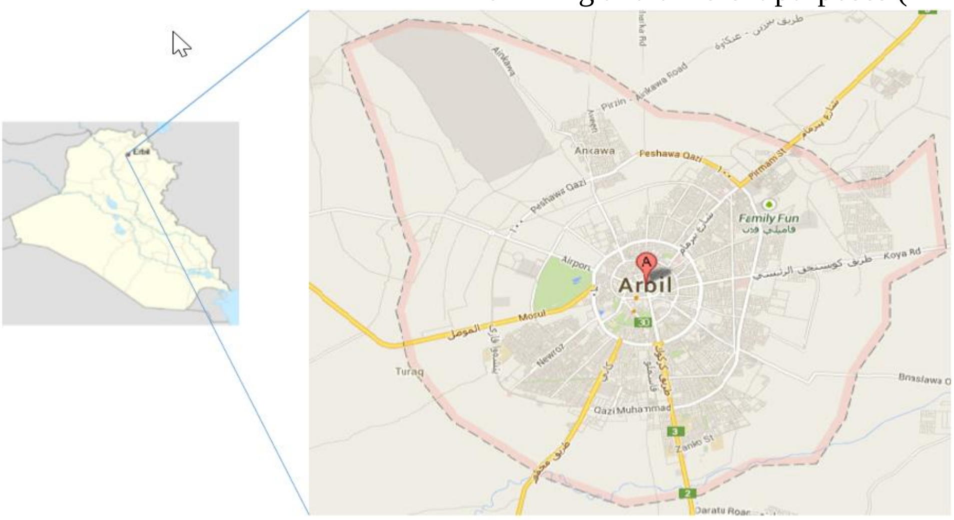


Figure 2: Location map of Erbil with administrative boundary.

Three water treatment plants (WTPs) were built on Greater-Zab River at Ifraz village (Figure 3). The Grater-Zab River together with Rawanduz Stream is the main wellspring of surface water accessible for giving water to drinking and different purposes. The More noteworthy Zab starts in Turkish Kurdistan and is incompletely directed by the Bekhma dam. Its length is 392 Km from the source to Almakhlut town in the south of Mosul (Shareef and Muhammad, 2008). Ifraz 1 (conventional WTP) built in 1968 with design capacity of 38400 m³/day.

This is an old WTP and most treatment units in this plant are in poor condition which needs repair and great upkeep (Shareef and Muhammad, 2008). Ifraz 2 constructed in 1985 with a design capacity of 72000m³/day. Which is situated on the Erbil-Ankawa road at right side, it supplies around 44000 m³/day (Shareef and Muhammad, 2008). Ifraz 3 built-in 2006 with a design capacity of 240000 m³/day.

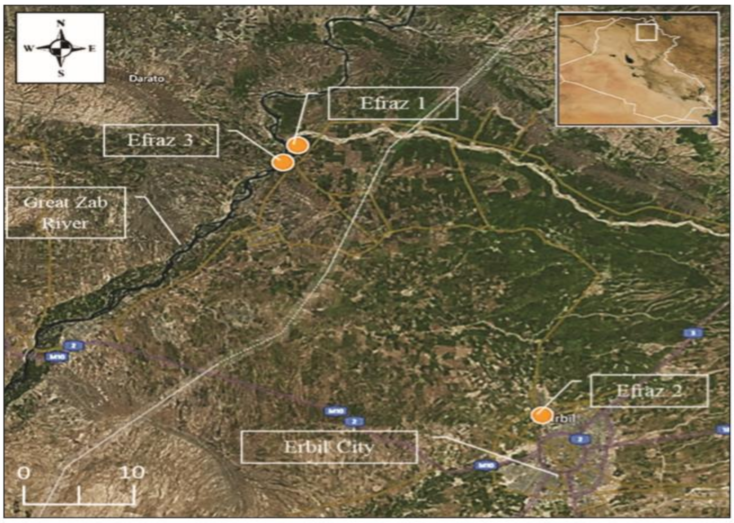


Figure 3: Satellite image of Ifraz 1, 2, and 3 at Greater-Zab River

## 3.2 Data Collection

As mentioned above, Erbil City has two main sources for potable drinking water, first from the wells (Groundwater) and the second from surface water treatment from Greater-Zab River.

From Erbil Water Directorate, Ministry of Municipality and Tourism we collect the annual productivity of drinking water in (m3) for all the projects which feed Erbil City, which contain of deep wells, Ifraz 1, Ifraz 2, and Ifraz 3. The data from 2014 to 2018 are (139,036,720) m3/year, (143,472,095) m3/year, (158,069,446) m3/year, (164,331,177) m3/year, and (172,687,525) m3/year respectively. Time series data used for a period five years (60 observations) represents the water monthly average water consumed measured in cubic meters in Erbil City, as shown in Table (1) below. The purpose of time series modelling, used the observations to fit the Box-Jenkins and SARIMA Models while (12) observations in the years 2019 were kept for the purpose of sample forecast accuracy checking and comparison. The data were analysed in the practical part depends on the programs (Statgraphics V.17).

**Table 1: monthly average water consumption in (m3) in Erbil City for the period 1/1/2014 - 30/6/2019**

| Months | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|---|
| January | 11,720,657 | 11,780,382 | 12,437,872 | 13,350,310 | 12,687,874 | 12,564,654 |
| Feb. | 10,135,833 | 10,076,975 | 12,103,376 | 12,021,906 | 12,142,126 | 12,347,657 |
| March | 11,454,758 | 11,447,125 | 13,284,950 | 13,541,255 | 13,676,668 | 13,456,543 |
| Apr. | 11,250,530 | 11,487,768 | 13,119,297 | 12,858,007 | 14,178,024 | 13,765,768 |
| May | 11,829,197 | 13,171,573 | 14,133,266 | 14,126,739 | 15,256,879 | 15,654,678 |
| Jun. | 11,945,210 | 11,889,242 | 13,290,897 | 13,451,770 | 15,646,045 | 15,876,889 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Jul. | 12,728,210 | 14,086,748 | 14,915,663 | 16,442,989 | 17,758,429 | - |
| Aug. | 12,946,109 | 13,657,322 | 14,228,038 | 14,633,460 | 17,019,721 | - |
| Sept. | 12,429,278 | 13,101,712 | 14,180,530 | 14,425,968 | 16,116,830 | - |
| Oct. | 11,169,685 | 12,206,497 | 12,028,590 | 11,720,685 | 13,729,380 | - |
| Nov. | 10,744,891 | 10,628,615 | 11,555,484 | 10,887,597 | 12,259,213 | - |
| Dec. | 10,682,361 | 10,575,202 | 11,325,315 | 11,063,882 | 11,781,725 | - |
| Total | 139,036,720 | 143,472,095 | 158,069,446 | 164,331,177 | 172,252,91 | |

The first step is the description and analysis a time series of water consumed graphically to asses as examples and conduct an information after later time, as in Figure (4), where we note that there is an increasing trend over time as well as high fluctuations in the consumed series for water when it reaches the peak (every summer) with the minimum consumed in (every winter). These fluctuations are repeated regularly at the same time every year with a different time at which the increase from year to year. These changes indicate a seasonal component which repeats itself every 12 months and the gradual increasing trend can be seen also through Figure (4). This indicates a non-stationary and to make sure that the time series was not stationary we used a (Box-Pierce) test as in Table (2) where we reject the null hypothesis (P = 0.00 <0.05).

**Table (2) Hypothesis and P-value for time series**

| Hypothesis | P-value |
|---|---|
| Null hypothesis ($H_0$): data it is stationary. | **0.00** |
| Alternative hypothesis ($H_1$): data it is non-stationary | |

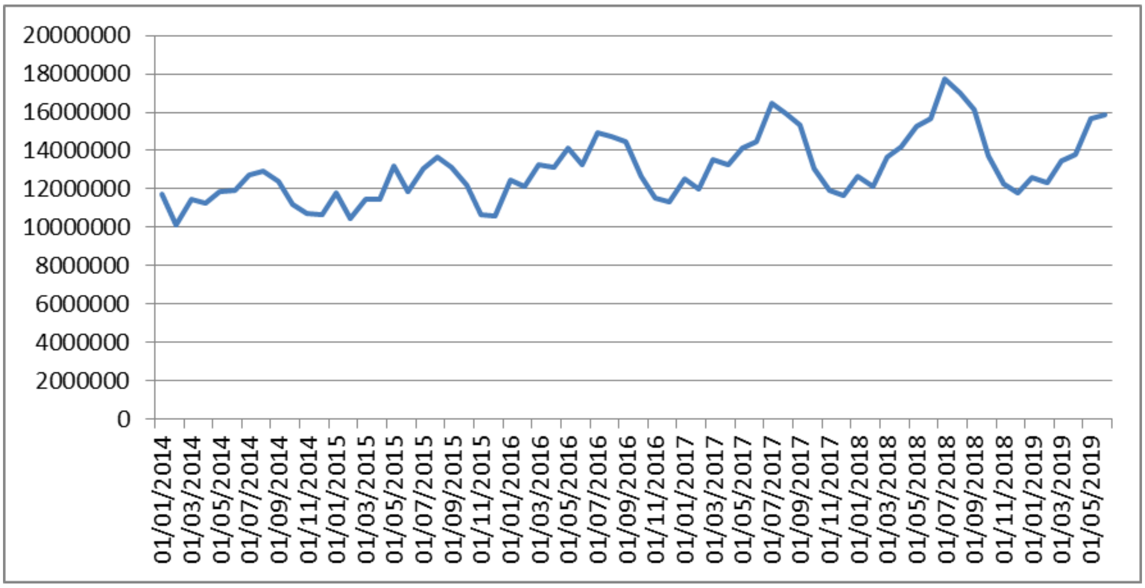


Figure 4: Water consumed from 1/1/2014 to 30/6/2019

### 3.3 Verifying Stationary

Stationary data is the first stage to build models by the Box-Jenkins approach by transforming data until it looks stationary, in order for stationary series must take the first difference seasonal for removing a non-stationary about mean (trend) and variance, after using transformation like taking the first difference you get the stationary about the mean and it means no trend, while the p-value =0.00 of the test (Box-Pierce) but there is no stationary about the variance, it confirms the seasonal affected. It can be plainly observed that the information is non-stationary and contains seasonal by the behaviour of autocorrelation functions and partial autocorrelation functions after transformations.

To remove the seasonal effects were taken the first difference seasonal (S=12) it was found that the series has become stationary by (P-value = 0.664) for the test (Box-Pierce), which is bigger than 0.05. So, we can't dismiss the null hypothesis that all autocorrelation function coefficients equal to zero, and then redraw plot time series modified , ACF and PACF, as in Figures (5,6) and (7), with tables(3,4) for ACF and PACF.

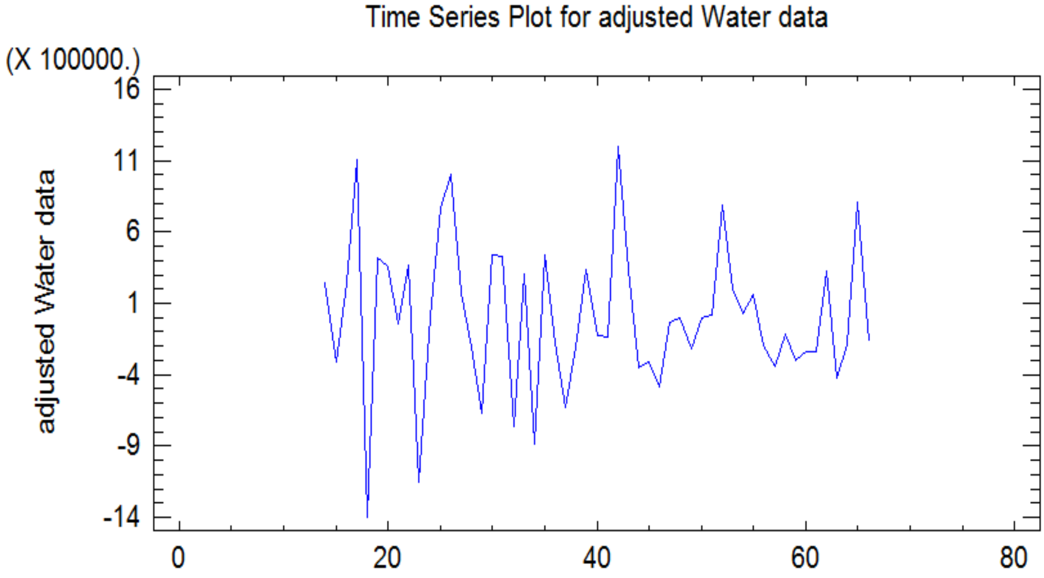


Figure 5: Time series for Water consumed after transfers

**Table (3) Estimated Autocorrelations for adjusted Water data**

| *Lag* | *Autocorrelation* | *Stnd. Error* | confidence interval *Lower bound 95.0%* | confidence interval *Upper bound 95.0%* |
|---|---|---|---|---|
| 1 | -0.15201 | 0.137361 | -0.269222 | 0.269222 |
| 2 | -0.164372 | 0.140499 | -0.275373 | 0.275373 |
| 3 | 0.00749083 | 0.144081 | -0.282395 | 0.282395 |
| 4 | -0.1443 | 0.144089 | -0.282409 | 0.282409 |
| 5 | 0.184072 | 0.14679 | -0.287704 | 0.287704 |
| 6 | -0.10324 | 0.151082 | -0.296117 | 0.296117 |
| 7 | -0.196491 | 0.152408 | -0.298714 | 0.298714 |
| 8 | -0.128421 | 0.157115 | -0.30794 | 0.30794 |
| 9 | 0.132369 | 0.159083 | -0.311797 | 0.311797 |
| 10 | 0.0229633 | 0.161148 | -0.315844 | 0.315844 |
| 11 | 0.110388 | 0.161209 | -0.315965 | 0.315965 |
| 12 | -0.0713816 | 0.162629 | -0.318748 | 0.318748 |
| 13 | 0.0538338 | 0.163219 | -0.319905 | 0.319905 |
| 14 | 0.115155 | 0.163554 | -0.320561 | 0.320561 |
| 15 | -0.16375 | 0.165077 | -0.323545 | 0.323545 |

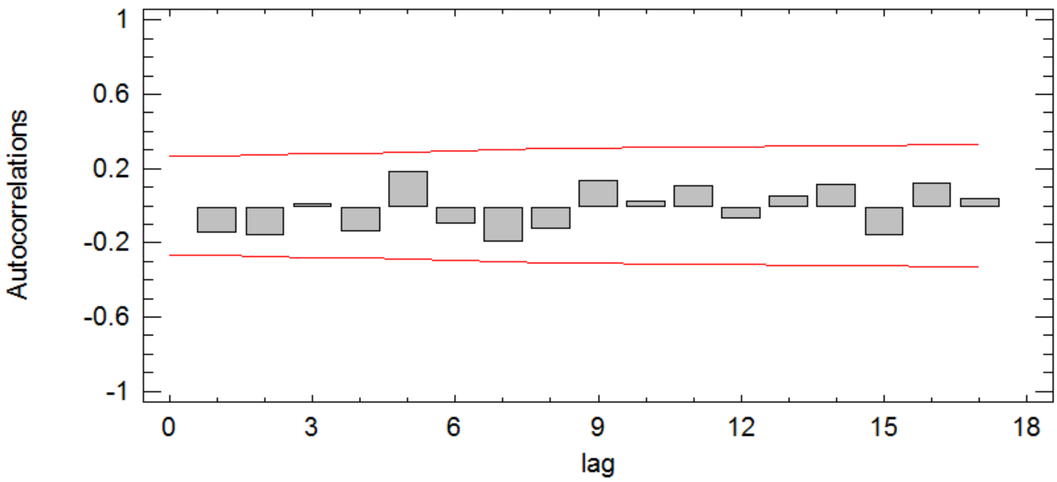


Figure 6: ACF Water consumed after transfers

**Table (4)Estimated Partial Autocorrelations for adjusted Water consumed**

| *Lag* | *Partial Autocorrelations* | *Stnd. Error* | confidence interval *Lower bound 95.0%* | confidence interval *Upper bound 95.0%* |
|---|---|---|---|---|
| 1 | -0.186942 | 0.137361 | -0.269222 | 0.269222 |
| 2 | -0.238004 | 0.137361 | -0.269222 | 0.269222 |
| 3 | -0.109165 | 0.137361 | -0.269222 | 0.269222 |
| 4 | -0.223127 | 0.137361 | -0.269222 | 0.269222 |
| 5 | 0.145536 | 0.137361 | -0.269222 | 0.269222 |
| 6 | -0.0846095 | 0.137361 | -0.269222 | 0.269222 |
| 7 | -0.184666 | 0.137361 | -0.269222 | 0.269222 |
| 8 | -0.263349 | 0.137361 | -0.269222 | 0.269222 |
| 9 | 0.0764794 | 0.137361 | -0.269222 | 0.269222 |
| 10 | -0.14541 | 0.137361 | -0.269222 | 0.269222 |
| 11 | 0.0848157 | 0.137361 | -0.269222 | 0.269222 |
| 12 | -0.172413 | 0.137361 | -0.269222 | 0.269222 |
| 13 | 0.125324 | 0.137361 | -0.269222 | 0.269222 |
| 14 | -0.0620747 | 0.137361 | -0.269222 | 0.269222 |
| 15 | -0.107412 | 0.137361 | -0.269222 | 0.269222 |

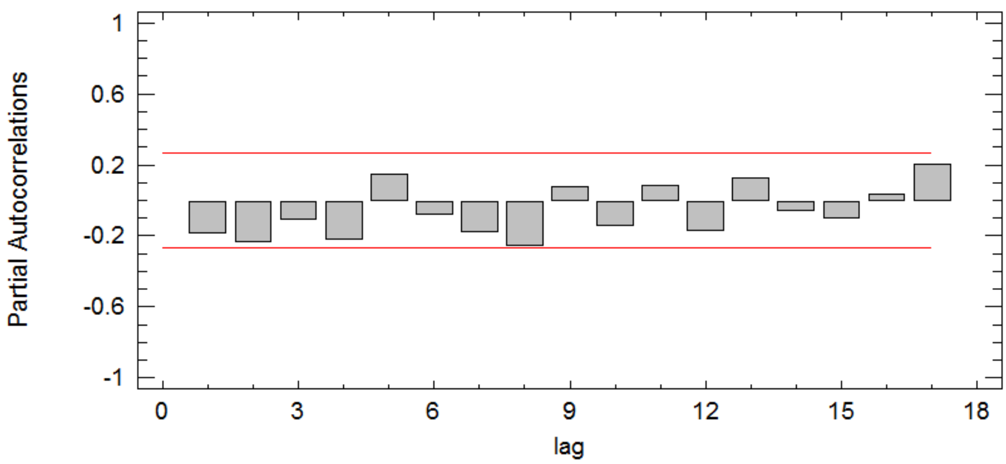


Figure 7: PACF Water consumed after transfers

### 3.4 Selecting Fitting Pattern

Box and Jenkins exhibit an intuitive approach for fitting ARIMA models to time series. It comes after the stationary of time-series around the mean and the variance by transformation. Choose the appropriate model for time series data by building five models in Table (5) and comparing them based on the minimum statistical criteria (RMSE, MAE, MAPE), and study their behaviour to reach an appropriate model to represent the data in table (1), where the proposed model is seasonal multiplicative model SARIMA (0, 1, 2) $(1, 1, 1)_{12}$, the results. Parameters estimated values of the model SARIMA in table (6)

**Table (5) Proposed models for monthly average water consumed**

| *Model* | *RMSE* | *MAE* | *MAPE* |
|---|---|---|---|
| **SARIMA(0,1,2)x$(1,1,1)_{12}$** | **436720** | **325830** | **2.49968** |
| SARIMA(1,1,2)x$(1,1,0)_{12}$ | 483585 | 372701 | 2.83737 |
| SARIMA(2,1,2)x$(0,1,1)_{12}$ | 456312 | 357155 | 2.75837 |
| SARIMA(1,1,1)x$(1,1,0)_{12}$ | 480001 | 374885 | 2.85256 |
| SARIMA(1,1,2)x$(1,1,0)_{12}$ | 482457 | 369520 | 2.81268 |

**Table (6) Parameters estimated values of the model SARIMA (0, 1, 2) $(1, 1, 1)_{12}$**

| *Parameter* | *Estimate* | *stander Error* | *T* | *P-value* |
|---|---|---|---|---|
| MA(1) | 0.308776 | 0.136305 | 2.26533 | 0.027948 |
| MA(2) | 0.342359 | 0.137833 | 2.48388 | 0.016464 |
| SAR(1) | 1.18588 | 0.154595 | 7.67085 | 0.000000 |
| SMA(1) | 1.27162 | 0.0990348 | 12.8401 | 0.000000 |

Since the p-value of all models less than 0.05, this means all the estimated parameters have a statistically significant effect.

### 3.5 Model Testing

When we have distinguished and assessed the hopeful SARIMA (0, 1, 2) (1, 1, 1)12 models, we need to survey the ampleness of the choose models to the information. After estimating the model parameters, the diagnostic checking is applied to see if the model is adequate or not.

Diagnostic checking of the residuals for SARIMA (0, 1, 2) $(1, 1, 1)_{12}$ model by ACF and PACF plots for residual, as in Figure (8 and 9), we conclude that all the ACF and PACF of the residuals values is no statistically huge at the 95.0% certainty level (within confidence limits). This means that residuals are random white noise, and it is appropriate model for this data.

Use Box-Pierce test to confirm the appropriate user model, and testing no significant autocorrelation of the residuals. The (P-value = 0.43) for this test is greater than (0.05), this means there is no critical autocorrelation in the residuals, which demonstrates that the residuals white noise.

We conclude that the model SARIMA (0, 1, 2) $(1, 1, 1)_{12}$ the fitting demonstrates to the accessible data because it succeeded in diagnostic tests to building model.

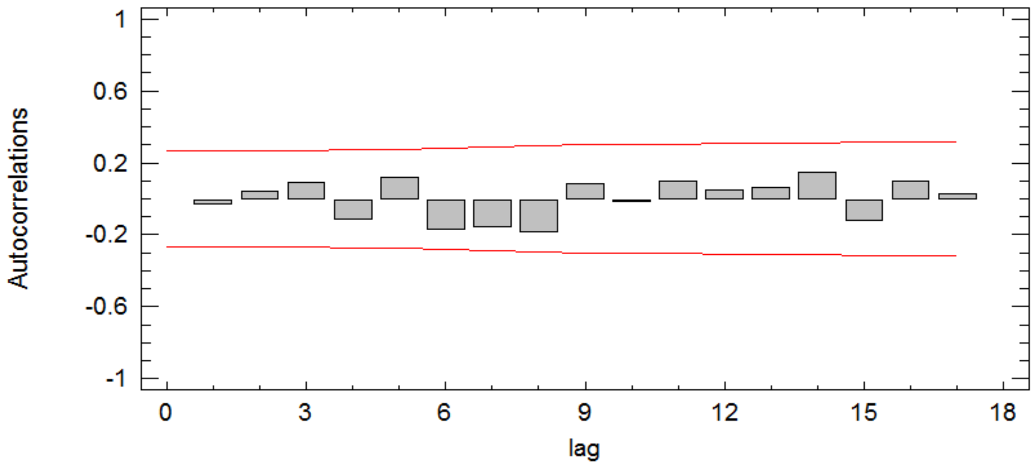


Figure 8: Residual ACF SARIMA (0, 1, 2) $(1, 1, 1)_{12}$

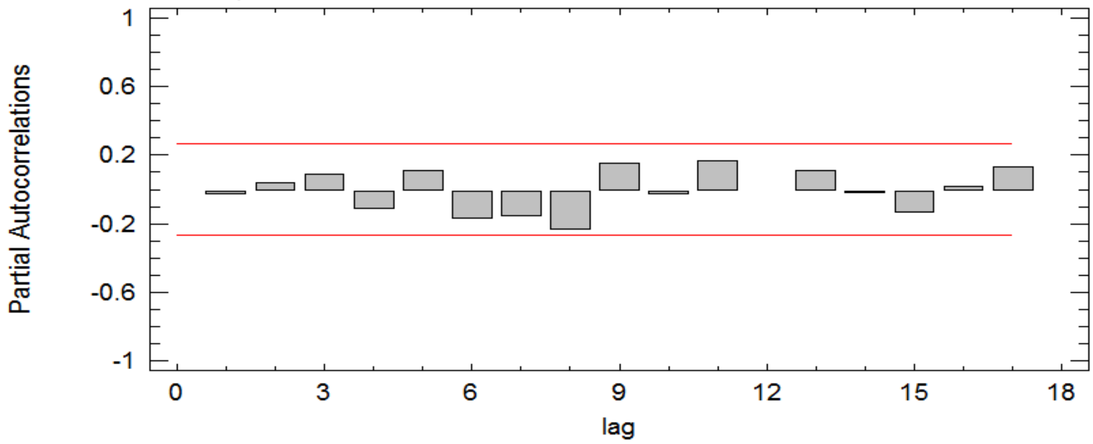


Figure 9: Residual ACF for SARIMA (0, 1, 2) $(1, 1, 1)_{12}$

**3.6 Forecasting:**

Using the SARIMA (0, 1, 2) $(1, 1, 1)_{12}$ model for forecasting values of the monthly average water consumed from 1/7/2019 in Erbil city to 30/6/2020 in Erbil City. To building SARIMA model for the purpose of forecasting used all available data from January 1, 2014 to Jun. 30, 2019. Table (7) and Figure (10) show the forecasting monthly average water consumed in Erbil City. Figure (11) shows the original values of the data and forecasting values estimated by the model, and we note a convergence between two values

Table (7) Forecasting monthly average water consumed in 2019 and 2020

| ***Period*** | ***Forecast*** | ***Lower limit 95 %*** | ***Upper Limit 95%*** |
|---|---|---|---|
| **7/2019** | 18000700 | 17022800 | 18978500 |
| **8/2019** | 17386700 | 16172000 | 18601400 |
| **9/2019** | 16585600 | 15311100 | 17860200 |
| **10/2019** | 14251900 | 12958700 | 15545200 |
| **11/2019** | 12838700 | 11538200 | 14139100 |
| **12/2019** | 12405200 | 11101300 | 13709100 |
| **1/2020** | 13236100 | 11930100 | 14542100 |
| **2/2020** | 12968500 | 11660900 | 14276100 |
| **3/2020** | 14143700 | 12834800 | 15452600 |
| **4/2020** | 14464400 | 13154200 | 15774500 |
| **5/2020** | 16220700 | 14909400 | 17532000 |
| **6/2020** | 16463800 | 15151400 | 17776300 |

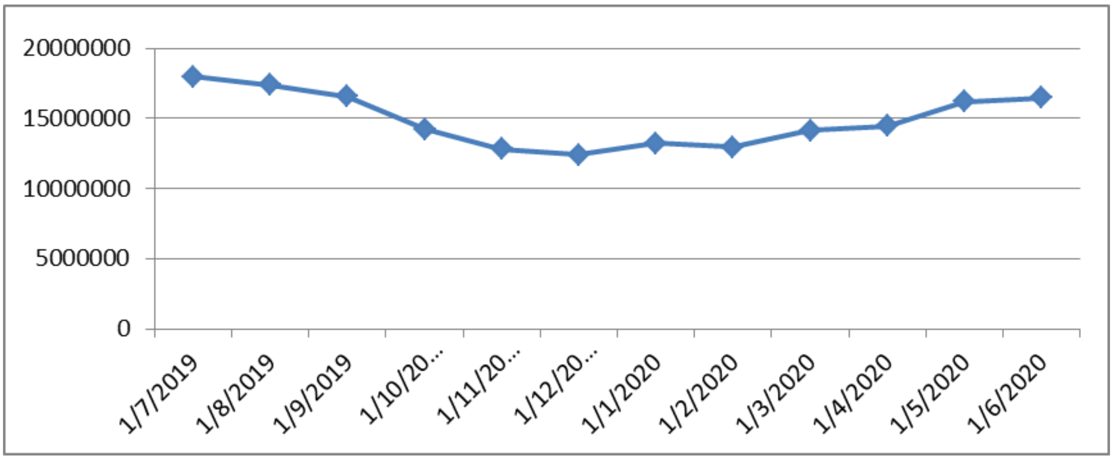


Figure 10: Forecasting plot for SARIMA from 7/2019 to 6/2020

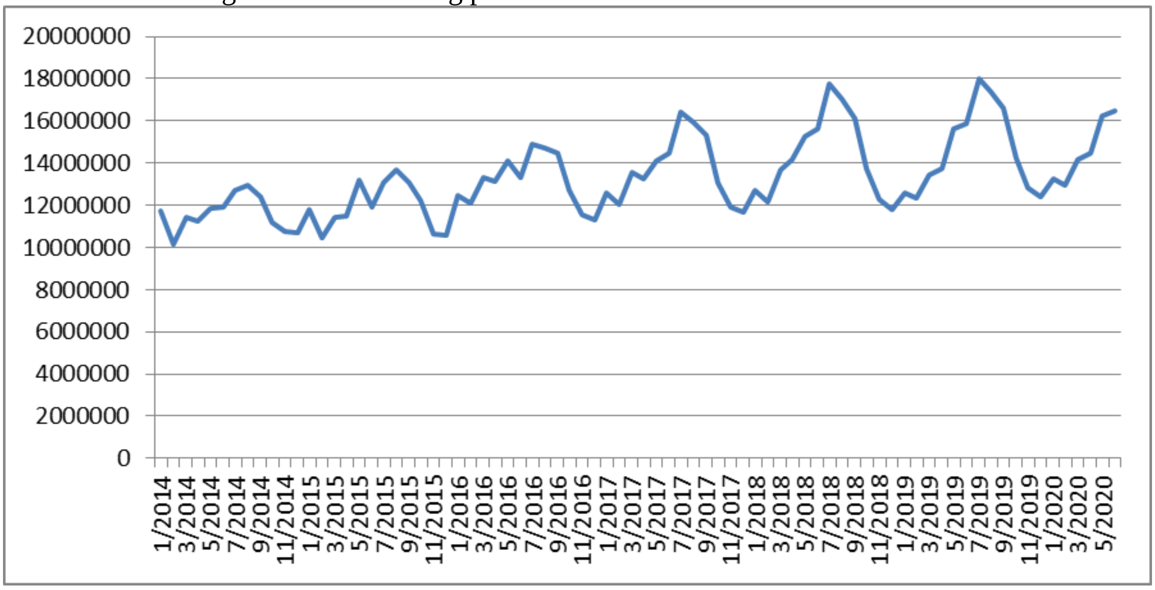


Figure 11: Plot for (Original data and Forecasting) by SARIMA from 1/2014 to 6/2020

## 4. Conclusion

Seasonal multiplicative model ARIMA, It was found that SARIMA (0, 1, 2) (1, 1, 1)12 gives the best result based on the minimum statistical criteria (RMSE, MAE, MAPE) used to compare estimated models, so it is a better model to represent and forecast the water monthly average consumed in Erbil City, due to ARIMA models are particularly suited to short term forecasting. Using the SARIMA model to forecast the monthly average for water consumed in Erbil city for a period (12 months)(6 months in 2019 and 6 months in 2020) is the appropriate model to the available data because it succeeded in diagnostic tests to the building model. We conclude that the consumed for water in Erbil City is constantly increasing from one year to the other, and more demand for water in Erbil city for next period, so it is good indicator to the local authority and water directorate to put their plan so that they can cover the amount of consumed water particularly in peak season (summer), this due to the forecasting of water consumed for Erbil city for the year of 2019. From the data and the forecasting data, it is obvious that there is an increasing trend over time as well as high fluctuations in the consumed series for water when it reaches the peak (every summer) with the minimum consumed in (every winter).

## 5. References

Adhikari, R and Agrawal, R. 2013. An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613.

Aziz, S. Q. 2007. Liminological observations on the water treatment plants in Efraz. 2nd Environmental Conference-Water, Duhok, Kurdistan, Iraq. April.

Bougadis, J., Adamowski, K. & Diduch, R. 2005. Short term municipal water demand forecasting. Hydrological Processes, 19, 137-148.

Caiado, J. 2009. Performance of combined double seasonal univariate time series models for forecasting water demand. Journal of Hydrologic Engineering, 15, 215-222.

Chen, L. & Zhang, T. 2006. Hourly water demand forecast model based on bayesian least squares support vector machine. Tianjin University Science and Technology, 39, 1037-1042.

Ghiassi, M., Zia, D. K. & Saidane, H. 2008. Urban water demand forecasting with a dynamic artificial neural network model. Journal of Water Resources Planning and Management, 134, 138-146.

Hasan, M. T. & Hasan, S. K. 2017. Forecasting the Amount of Water Consumed in Erbil City using Time Series Model (SARIMA). Zanko Journal for Human science, 21, 415-425.

Jain, A. & Ormsbee , L. E. 2002. Short-term water demand forecast modeling techniques—Conventional methods versus AI. Journal (American Water Works Association), 94, 64-72.

Jain, A., Varshney, A. K. & Joshi, U. C. 2001. Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. Water resources management, 15, 299-321.

Kim, S., Koo, J., Kim, H. & Choi, Y. 2007. Optimization of pumping schedule based on forecasting the hourly water demand in Seoul. Water Science and Technology: Water Supply, 7, 85-93.

Kitagawa, G. 2010. Introduction to time series modeling, CRC press.

Msiza, I. S., Nelwamondo, F. V. & Marwala, T. Artificial neural networks and support vector machines for water demand time series forecasting. Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on, 2007. IEEE, 638-643.

Palit, A. K. & Popovic, D. 2005. Computational Intelligence in Time Series Forecasting, Springer, Germany.

Shareef, K. M. & Muhammad, S. G. 2008. Natural and drinking water quality in Erbil, Kurdistan. Current World Environment. 3(2), 227-238.

Zhang, G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.

Zhou, S., Mcmahon, T., Walton, A. & Lewis, J. 2002. Forecasting operational demand for an urban water supply zone. Journal of hydrology, 259, 189-202.

## پێشبینی کردنی بری ئاوی بەکارهاتوو لە شاری هەولێر بە بەکارهێنانی مۆدێلی زنجیرەی کاتیەکان (SARIMA)

**سامان حسین محمد**
بەشی ئامار – کۆلێژی بەڕێوەبردن و ئابوری / زانکۆی سەلاحەددین-هەولێر

**عماد علی عمر**
بەشی شارستانی – کۆلێژی ئەندازیاری / زانکۆی سەلاحەددین-هەولێر

**پوختە**

پێشبینی کردن لە زنجیرەکاتیەکان یەکێکە لەو بابەتە گرنگانەی زانستی ئامار کە یارمەتی دەزگاکانی حکومەت دەدات بۆ پلاندانان و بڕیاردانی گونجاو لە دوا روزژە. وئامانج لەم توێژنەوەیە شیکاري کردنە بۆ داتای زنجیرەکاتیەکانی ئاوی بەکارهاتوو لە شاری هەولیر لە ماوەی (2014/1/1) تا کوتای (2020/6/30) ولە شیکردنەوەکەدە کەیشتین بەوەی کە ناجێگیری لە زنجیرەکاتیەکاندە هەیە هەربویە هەستایەن بە گۆ ريني زنجيرەکە بۆ بەدەست هینانی جێگیری لە زنجیرەکاتیەکان. وبۆ دیاریکردني باشترین مؤدیل لە نیوان مودێلەکان، بۆمان دەرکەوت کە مودێلی وەرزی $SARIMA(0,1,2)(1,1,1)_{12}$ گونجاو بوو بۆ پێشبینی کردن و پشت بەستن بەکەمترین پێورەکانی ئاماری ( RMSE، MAE، MAPE) کە بۆ بەراورد کردنی مۆدێلەکان بە کاردەهێنریت.

لە کۆتای دا گەیشتین بە باشترین مۆدێلی پیشبێنی کراو بۆ تێکرای مانگانە ئاوی بەکارهاتوو بۆ ماوەی 12 مانگ (6 مانگ لە ساڵی2019 و6 مانگ لە ساڵی2020) و نرخەکان کونجاون لەگەل نرخە راستەقینەکان.

## التنبؤ بكمية المياه المستهلكة في مدينة أربيل باستخدام نموذج السلاسل الزمنية (SARIMA)

**سامان حسين محمد**
قسم الاحساء – كلية ادارة والاقتصاد / جامعة صلاح الدين-اربيل

**عماد علي عمر**
قسم المدنى – كلية الهندسة / جامعة صلاح الدين-اربيل

**ملخص**

يعد التنبؤ في السلسلة الزمنية أحد الموضوعات الأساسية في العلوم الإحصائية لان مؤشراتها تساعد الحكومة على وضع خطة جيدة للحصول على قرار دقيق في المستقبل. وقد اجرى هذا البحث بهدف تحليل بيانات السلسلة الزمنية الخاصة بكمية المياه المستهلكة لمدينة أربيل للسنوات السابقة للفترة بين (2014/1/1) حتي (2020/6/30). في التحليل، وجدت الدراسة عدم استقرارية في السلسلة الزمنية، لذلك تم إجراء التحويلات اللازمة على السلسلة لإزالة عدم استقرارية. ومن ثم اختيار النموذج الافضل من بين للنماذج المتحصل عليها. وجد أن $SARIMA(0,1,2)(1,1,1)_{12}$ يعطي أفضل النتائج بناءً على اقل المعايير الإحصائية ( RMSE، MAE، MAPE) المستخدمة للمقارنة.

أخيراً، باستخدام أفضل نموذج تم التنبؤ بالمعدل الشهري للماء المستهلك في مدينة أربيل لفترة (12) شهرًا (6 اشهرفي سنة 2019 و6 اشهرفي سنة 2020) وبدا النتائج بانة تتسق مع تلك القيم الفعلية في السلاسل الزمنية.