Scholars Journal of Physics, Mathematics and Statistics **ISSN 2393-8056 (Print)**
Sch. J. Phys. Math. Stat. 2014; 1(2):61-70 **ISSN 2393-8064 (Online)**
©Scholars Academic and Scientific Publishers (SAS Publishers)
(An International Publisher for Academic and Scientific Resources)

Forecasting Monthly Water Production in Gaza City Using a Seasonal ARIMA Model

Mahmoud K. Okasha*, Deaa M.M Abu Shanab
Department of Applied Statistics, Al-Azhar University – Gaza, Gaza, Palestine

***Corresponding Author:**
Mahmoud K. Okasha
Email: m.okasha@palnet.com

Abstract: This paper aimed to analyze seasonal time series data on monthly water production in the Gaza Strip during the period between 2006 and 2012 using the Box-Jenkins methodology. A model was proposed that could forecast future monthly water production in the Gaza Strip, one of the mostly densely populated areas with one of the highest natural growth rates in the world and very limited water resources. The increasing demand for domestic water consumption requires Palestinian officials and international donors to take suitable measures to conserve or allocate water supplies. A model was found that best represented the time series data according to different criteria; the Box-Ljung test identified a seasonal model of lag 12 (SARIMA $(1,1,1) \times (1,1,1)_{12}$). Monthly water production in Gaza Strip was forecast for the period from January 2013 to December 2013 using this model. A comparison of these forecasts with observed values over this time period indicated that the model was highly accurate.

Keywords: Box-Jenkins methodology, Seasonal models, Forecasting, Box-Ljung test, Augmented Dickey-Fuller test, Kwiatkowski-Phillips-Schmidt-Shin test.

INTRODUCTION

Gaza City is one the most densely populated cities in the world. One of the biggest concerns for planners in terms of social and economic policies in Gaza city is to provide water to the population for the coming year. Water scarcity in the Gaza Strip means that planners need to decide how best to fill the gap between supply and demand. This requires building a good seasonal model that is able to forecast future monthly water production. The most prominent of these models is the Seasonal Autoregressive Integrated Moving Average (SARIMA), as proposed by Box and Jenkins [1]. The effect of seasonal fluctuations and changing trends in domestic use mean that establishing a seasonal model based on monthly time series data can be rather complex. In this study, we aimed to use the Box-Jenkins approach to establish a reliable seasonal time series model that can be used for forecasting future monthly water production for domestic use in Gaza city.

A time series is as an ordered sequence of observations taken at equally spaced intervals. It is said to be stationary if the joint distribution of $X_{t1,} X_{t2}, \ldots , X_{tn}$ is the same as the joint distribution of $X_{t1+T,} X_{t2+T}, \ldots, X_{tn+T}$ for all values of τ . In other words, shifting the time origin by an amount τ has no effect on the joint distribution, which must therefore depend only on the intervals between $t_{1,} t_2, \ldots, t_n$ [2]. Stationarity can be assessed using a run sequence plot. The run sequence plot should have a constant location and scale. This can often be detected from an autocorrelation plot as it indicates very slow decay. The first step in developing a Box-Jenkins model is to determine whether the time series is stationary and whether there is any significant seasonality that needs to be modeled.

Important stages in the Box-Jenkins methodology include a diagnostic check of the residuals and tests of the model's adequacy. This is commonly done in ARIMA modeling by tentatively fitting more than one model to the data, estimating the parameters for each model, and performing a diagnostic check to test the validity of each model. The model which best fits the data, according to various statistical tests, is then selected for forecasting. In particular, a study of the residual series obtained after fitting the model to the data is needed to see if any pattern remains unaccounted for. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) plots of the residual series help in detecting any unaccounted patterns[3,4]. The residuals should ideally be just random noise (white noise) with zero mean and constant variance. Some statistical tests for lack of fit may be used to test for the randomness of the residuals.

In this study, past data on monthly water production for domestic use in Gaza city was obtained from the Coastal Municipality Water Utility (CMWU) for the period from January 2006 to December 2012. An optimal model that best fitted monthly water production values for this period was identified. We compared the accuracy of different models using different criteria, such as AIC, BIC, MSE, RMSE, MAPE and MAE, as well as the Box-Ljung test. The final goal was to accurately forecast future monthly water production in Gaza city using a seasonal time series model to help planners to meet the future needs of the population.

SEASONAL TIME SERIES MODELS

The term "seasonal time series" is used to refer to the similar patterns that appear in a time series in corresponding months over successive years. These trends are usually due to recurring events which may take place annually or quarterly. By plotting the series against time checks for seasonal changes it is possible to reveal non-stationarity in the data. Seasonal models have pronounced regular ACF and PACF patterns with a periodicity equal to the order of seasonality. It is also important to note the number of times per year that seasonal variations occur. If the seasonality is annual, seasonal variation in ACF spikes denotes heightened patterns of seasonal lags over and above the regular non-seasonal variation once per year. If the seasonality is quarterly, there will be prominent ACF spikes four times per year.

Seasonal autoregressive models are built using seasonal autoregressive (SAR) parameters that represent autoregressive relationships between time series data separated by multiples of the number of periods per season. For example, a model with one SAR parameter is written as

$$X_t = \phi_s X_{t-s} + \varepsilon_t \tag{1}$$

This model may be expressed as ARIMA(1, 0, 0)^s where s is the number of periods per season. The parameter is called the SAR parameter with order s. A general seasonal autoregressive model with p SAR parameters is written as follows:

$$X_t = \sum_{i=1}^{p} \phi_{is} X_{t-is} + \varepsilon_t \tag{2}$$

where X_{t-s} is of order s, X_{t-2s} is of order 2s, and X_{t-ps} is of order p_s [5].

The time sequence plot of an ACF or PACF can be used as a primary instrument for identifying seasonal autoregressive models. A multiplicative seasonal autoregressive model contains both non-seasonal autoregressive factors. A simple example of a seasonal autoregressive model would be one with a regular first order autoregressive term and a seasonal term of order 12. Seasonal autoregressive models can also be defined. Consider

$$x_t = \phi x_{t-12} + \varepsilon_t \tag{3}$$

where $|\phi| < 1$ and ε_t is independent of $x_{t-1}, x_{t-2}, \ldots$. It is obvious that $|\phi| < 1$ ensures stationarity. Thus, it is easy to argue that $E(X_t) = 0$. Multiplying equation (3) by X_{t-k}, taking expectations and dividing by γ_0, where γ_k is the autocorrelation function at lag *k*, yields:

$$\rho_k = \phi \rho_{k-12} \quad for\ k \geq 1 \tag{4}$$

It is clear that

$$\rho_{12} = \phi \rho_0 = \phi$$

so more generally,

$$\rho_{12} = \phi^k \quad for\ k = 1, 2, \ldots$$

and

$$\rho_{24} = \phi \rho_{12} = \phi^2$$

Furthermore, setting k = 1 and then k = 11 in Equation (4) and using $\rho_k = \rho_{k-1}$ gives:

$$\rho_1 = \phi \rho_{11}$$

and

$$\rho_{11} = \phi \rho_1$$

This implies that $\rho_1 = \rho_{11} = 0$. Similarly, it can be shown that $\rho_k = 0$, except at the seasonal lags 12, 24, 36, etc. At these lags, the autocorrelation function decays exponentially like an AR(1) model.
Generally, a seasonal AR(p) model with *s* seasonal periods is given as:

$$x_t = \phi_1 x_{t-s} + \phi_2 x_{t-2s} + \ldots + \phi_p x_{t-ps} + \varepsilon_t \tag{5}$$

With a seasonal characteristic polynomial

$$\phi(x) = 1 - \phi_1 x^s + \phi_2 x^{2s} - \ldots - \phi_p x^{ps} \tag{6}$$

ε_t is required to be independent of $X_{t-1}, X_{t-2}, \ldots$ and, for stationarity, the roots of $\Phi(x) = 0$ should be greater than 1 in absolute value [6].

The Seasonal Moving Average (SMA) model with Q parameters is given by

$$X_t = \sum_{i=1}^{Q} \phi_{is} e_{t-is} + e_t \tag{7}$$

For $ARIMA(0,\ 0,\ 1)^4$, the model is a quarterly seasonal moving average of order one, which means that it has one seasonal moving average parameter. An MA model with one seasonal moving average parameter is written as [7]:

$$X_t = \phi_s e_{t-s} + e_t \tag{8}$$

MODEL ACCURACY

Two methods commonly used to assess the accuracy of predictive models are the mean absolute deviation (MAD) and the mean squared error (MSE)[8]. Other methods include the mean absolute percentage error (MAPE), mean absolute error (MAE), and the root mean square error (RMSE).

Mean Absolute Deviation (MAD)

The median absolute deviation (MAD) is a robust measure of the variability of a univariate sample of quantitative data. This term can also refer to the population parameter that is estimated by the MAD calculated from a sample. For a univariate data set $X_1, X_2, ..., X_n$, the MAD is defined as the median of the absolute deviations from the median [9-11].

$$MAD = \frac{1}{n}\sum_{t=1}^{n}\left|e_t\right| = \frac{\sum_{i=1}^{n}|Z_t - \hat{Z}_t|}{n} \tag{9}$$

Mean Square Error (MSE)

Mathematically, the mean square error (MSE) is defined by:

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \tag{10}$$

where X_t is the actual observation for time t, F_t is the forecast value for the same period, $e_t = X_t - F_t$ is the error term, and n is the number of forecasting values [5,12].

Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of accuracy of a method for constructing fitted time series values. It usually expresses accuracy as a percentage and is defined by the formula:

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{X_t - F_t}{X_t}\right| \tag{11}$$

The difference between A_t and F_t is divided by the actual value A_t. The absolute value in this calculation is summed for every fitted or forecasted point in time and divided by the number of fitted points n. Multiplying by 100 makes it a percentage error[13].

Mean Absolute Error (MAE)

The mean absolute error (MAE) is defined mathematically by:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{12}$$

where $e_t = X_t - F_t$ is the error term [5].

Root Mean Square Error (RMSE)

The formula for computing RMSE is:

$$MAE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{X}_i - \hat{X}_i)^2} \tag{13}$$

where $\hat{X}_i$ is the predicted value.

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)

The final model can be selected using a penalty function statistic, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC). The AIC and BIC are measures of the goodness of fit of an estimated statistical model. Given a data set, several competing models may be ranked according to their AIC or BIC, the best being the one having the lowest information criterion value. These information criteria judge a model by how close its fitted values tend to be to the true values, in terms of a certain expected value. The criterion value of a model is only meant to rank competing models and to tell which is the best model among the given alternatives. The criteria attempt to find the model that best explains the data with minimum free parameters but also includes a penalty that is an increasing function of the number of estimated parameters. In general, the AIC and BIC are calculated as:

$$\begin{aligned} AIC &= 2k - 2\log(L) \quad OR \quad 2k + n\log\left(\frac{RSS}{n}\right) \\ BIC &= -2\log(L) + k\log(n) \quad OR \quad \log(\sigma_e^2) + \frac{k}{n}\log(n) \end{aligned} \tag{14}$$

where
k is the number of parameters in the statistical model,
L is the maximized value of the likelihood function for the estimated model,
RSS is the residual sum of squares of the estimated model,
n is the number of observations or, equivalently, the sample size,
σ_e^2 is the error variance.

Overall, the model that achieves the lowest values of AIC, BIC, MSE, RMSE, MAPE, and MAE criteria would be the most efficient model in terms of these accuracy measures.

MODEL FORECASTS

The main goal of building a time series model is to make predictions for future observations of a given phenomenon with minimum errors. Seven features of a good ARIMA model should be considered[14]. First, a good model is parsimonious. That is, it should have the smallest possible number of coefficients. Second, a good autoregressive (AR) model must be stationary. Third, the moving average (MA) of the model should be invertible. Fourth, a good model should have statistically significant estimates of its coefficients (AR and MA). Fifth, the residuals should be independent. Sixth, the residuals should be normally distributed. Lastly, the model should give acceptable forecasts [15].

CASE STUDY

Preliminary Investigation of the Data

The data on monthly water production used in this study was obtained from the Coastal Municipality Water Utility (CMWU) in Gaza city. Water production in wells was regularly measured by cup (cubic meters) over the period from 2006 up to 2012. Gaza city contains 71 water wells for domestic consumption. The amount of water that can be extracted from wells is constrained by the power of the pumps available and the CMWU's regulations. The monthly data on water production meets all the requirements of a seasonal time series, a plot of which is shown in Fig 1. The figure illustrates that the time series is not stationary; it exhibits a general trend and a consistent pattern of short-term fluctuations, which suggests that there are seasonal variations in water production.

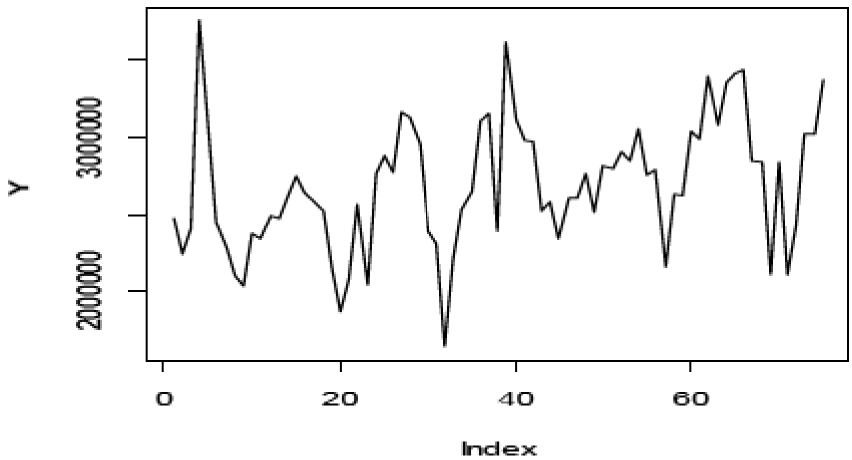

Fig-1:Time series plot of monthly water production in Gaza city.

Table 1 shows that the p-value of the augmented Dickey-Fuller (ADF) test was 0.11, which indicates that the time series of monthly production of water is not stationary. The table also shows that the p-value to the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was 0.01, confirming that the time series is not stationary [16].

Table-1:Results of augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests for stationarity of monthly water production in Gaza city.

Test	t-Statistic	P-value
ADF	1.3943	0.11
KPSS	0.0263	0.01

The sharp decrease in water production in winter every year shows that there is also a seasonal component that makes the time series non-stationary. The seasonal component was also investigated by examining the autocorrelation and partial autocorrelation functions as shown in Fig. 2 and Fig. 3.

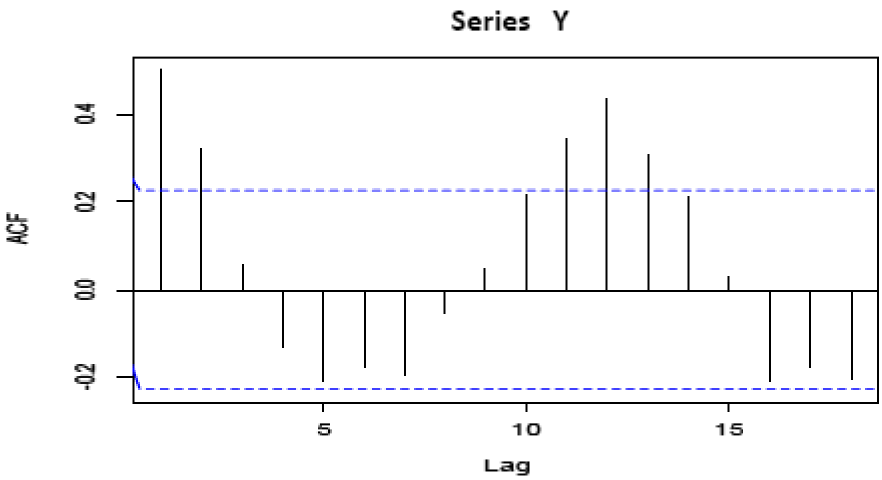

Fig- 2: The autocorrelation function for monthly water production in Gaza city.

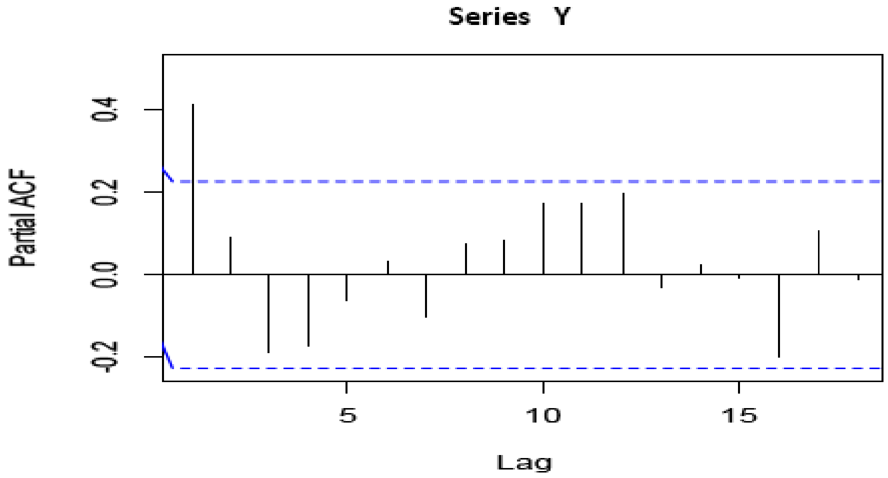

Fig- 3:The partial autocorrelation function for monthly water production in Gaza city.

The two plots above confirm that the time series is not stationary and the data would need to be transformed to make it stationary before carrying further certain analyses. The most suitable transformation for this series is to find the first differences, in order to detrend the original series and achieve stability. The first difference series in Fig. 2 exhibits seasonal patterns. By estimating the autocorrelation function (ACF) for the first difference series in Fig, 2, we note that the autocorrelation coefficients are high at lag 12; therefore, we may conclude that the series has a seasonal component of length 12.

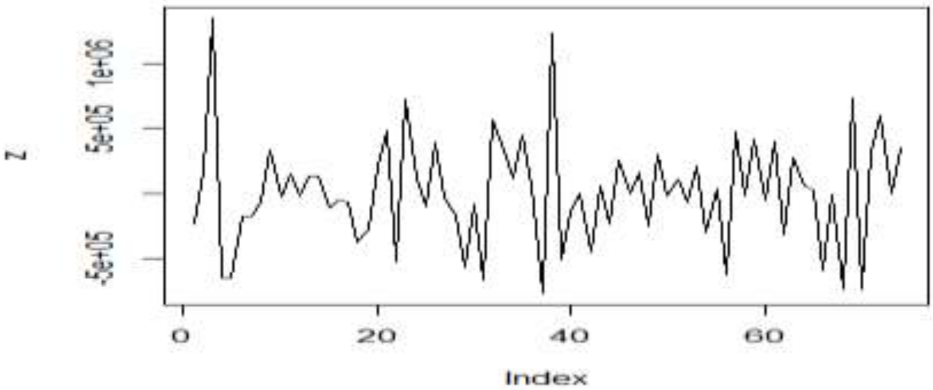

Fig-4: Time series plot of first differences for monthly water production in Gaza city.

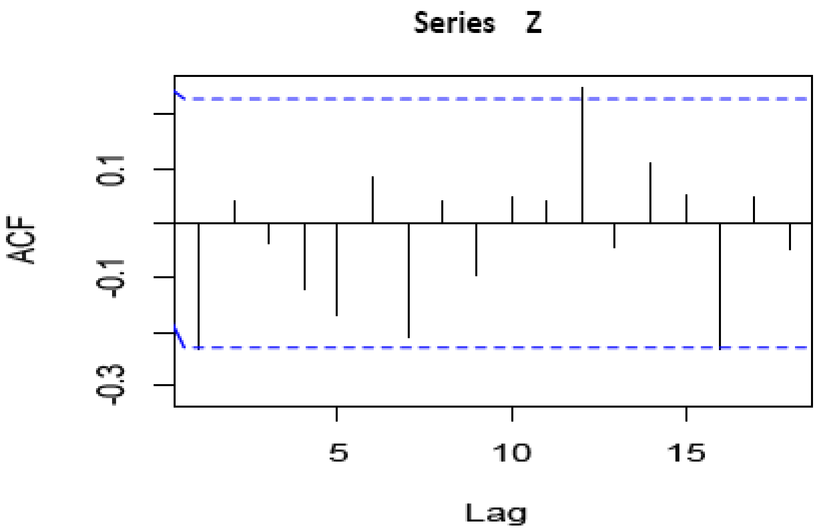

Fig-5: The autocorrelation function of first differences for monthly water production in Gaza city.

To remove the effect of the seasonal component and achieve a stationary series, we calculated the differences for the first difference series at lag 12. Figure 6 displays the time series plot of the data after the transformation at lag 12, illustrating that the time series is now stationary. To further verify the stationarity of the time series, we conducted a unit roots test (augmented Dickey-Fuller, ADF) and a Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test using the first difference series, the results of which are shown in Table 2.

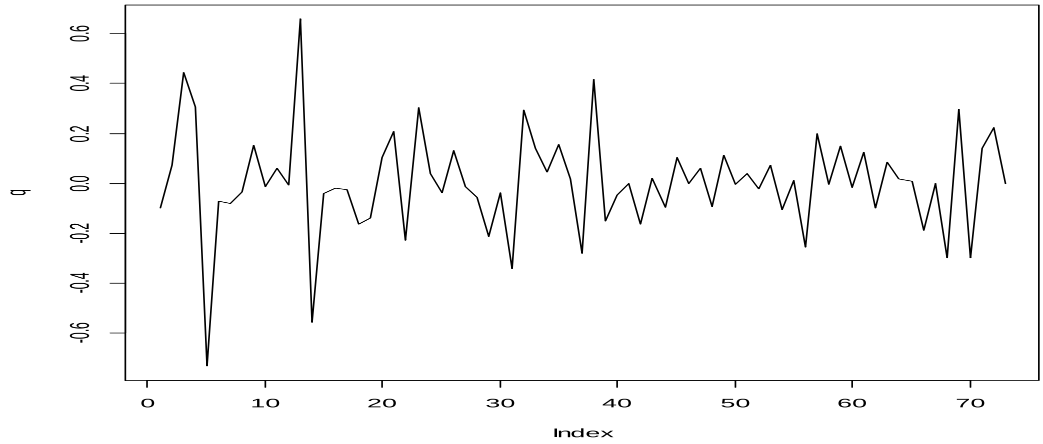

Fig-6:Time series plot for the first difference series at lag 12.

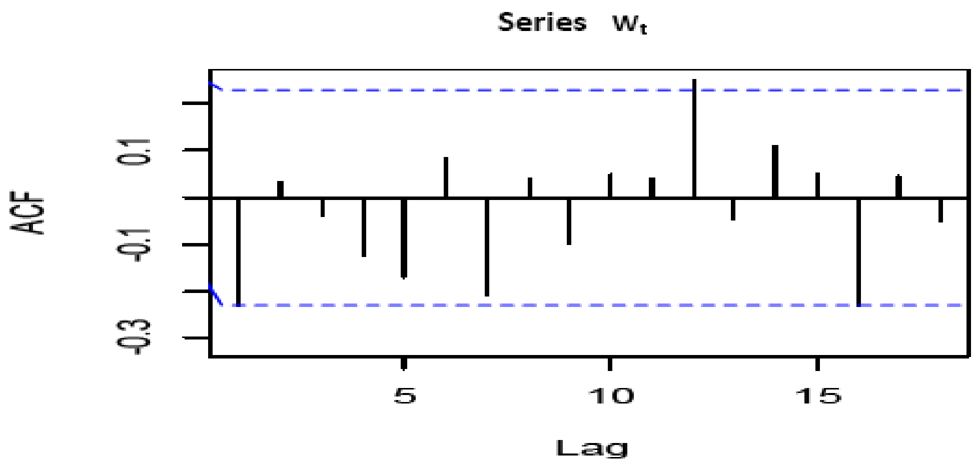

Fig-7: Autocorrelation function for the first difference series at lag 12.

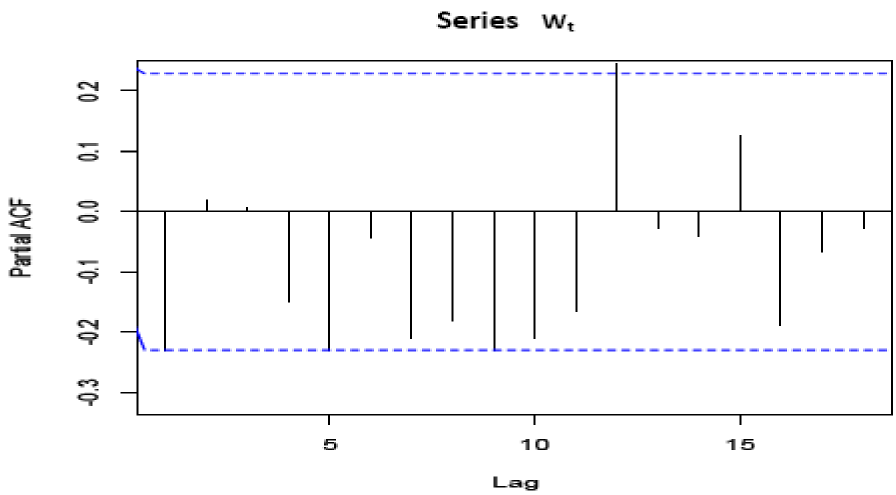

Fig-8:Partial autocorrelation function for the first difference series at lag 12.

Table 2: Results of augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin tests after taking the differences to get rid of the effect of the general trend and seasonal fluctuations.

	Z	
Test	**T-Statistic**	**P-value**
ADF	-5.6166	0.01
KPSS	0.6846	0.1

Table 2 shows that the p-value of the ADF test was 0.01, which is less than the required value of $\alpha = 0.05$, indicating that the hypothesis that the transformed time series is stationary is rejected. This demonstrates that the transformation successfully changed the stationarity of the data on monthly water production. In addition, the p-value of the KPSS test was 0.1, which is greater than the significance level $\alpha = 0.05$. This indicates that the null hypothesis of stationarity for the transformed time series is not rejected, confirming that the transformation has caused the time series to become stationary.

Model Identification

All relevant criteria that were discussed in section 3 were computed to select the best seasonal ARIMA model for the data in the water production time series. These criteria were the ACF and PACF, in addition to the AIC, BIC, MSE, RMSE, MAPE, and MAE, as well as the Box-Ljung test, In Fig. 7 and Table 2, the ACF starts from the ρ_1 value; this means that the series is just the AR, as we see that the ACF cuts off after lag 1. We note that the correlation coefficient of the seasonal gaps is interrupted after the first seasonal gap, but it is not possible to make a decision about the optimal model before looking at all the plots of the autocorrelation function. By examining the partial autocorrelation function (Fig. 8 and Table 2), it is clear that the PACF of the stationary series cuts off after time lag 1. Again, we note that the partial autocorrelation coefficient of seasonal gaps is interrupted after the first seasonal gap, but it is still too early to make a decision on the optimal model until we have looked at all the plots of the partial autocorrelation function. From the autocorrelation and partial autocorrelation coefficients of the series we see that it is necessary to consider the seasonal changes when identifying and estimating the model. As such, the best seasonal model was chosen based on the lowest values of the AIC, BIC, MSE, RMSE, MAPE, and MAE criteria, as shown in Table 3.

It is shown in Table 3 that the SARIMA$(1,1,1)(1,1,1)_{12}$ model produced the smallest values for the AIC, BIC, MSE, RMSE, MAPE, and MAE criterion. This means that the SARIMA$(1,1,1)(1,1,1)_{12}$ model is the best of all the models analyzed and is suitable for predicting monthly domestic water production in Gaza city.

Table-3: SARIMA model criteria for the monthly water production in Gaza city.

SARIMA Models	AIC	BIC	MSE	RMSE	MAPE	MAE
SARIMA$(0,1,1)(1,1,1)_{12}$	2090.75	25.879	82434038292	366846.2	10.347	277404.0
SARIMA$(1,1,0)(1,1,1)_{12}$	2090.29	26.016	95818917496	392729.1	11.562	307831.7
SARIMA$(1,1,1)(0,1,1)_{12}$	2083.56	25.866	90744681401	364402.1	10.265	276135.0
SARIMA$(1,1,1)(1,1,0)_{12}$	2085.56	25.944	118366131270	378813.7	10.541	283327.9
SARIMA$(0,1,0)(1,1,1)_{12}$	2095.86	26.203	136178174767	445290.2	12.981	344618.3
SARIMA$(1,1,1)(1,1,1)_{12}$	2082.56	25.836	81078653662	364339.0	10.160	272226.5
SARIMA$(2,1,1)(1,1,1)_{12}$	2094.28	26.000	97017582377	365771.0	10.250	274446.2
SARIMA$(1,1,2)(1,1,1)_{12}$	2094.27	26.006	93653117357	366791.6	10.161	272598.5
SARIMA$(1,1,1)(2,1,1)_{12}$	2094.28	26.012	81176084961	367971.1	10.182	273619.8
SARIMA$(1,1,1)(1,1,2)_{12}$	2094.27	26.014	81595051126	368248.1	10.179	273800.7
SARIMA$(2,1,1)(2,1,1)_{12}$	2088.25	26.075	173258435018	367942.4	10.234	273887.0
SARIMA$(2,1,1)(1,1,2)_{12}$	2089.22	26.077	108617601582	368340.7	10.231	273851.0
SARIMA$(1,1,2)(1,1,2)_{12}$	2089.29	26.115	112742109970	375219. 2	10.415	278826.5
SARIMA$(2,1,2)(1,1,1)_{12}$	2090.28	26.088	128396840523	370210.3	10.219	275146.4
SARIMA$(2,1,2)(2,1,1)_{12}$	2093.21	26.165	106588142997	372885.8	10.207	274413.4
SARIMA$(2,1,2)(2,1,2)_{12}$	2094.28	26.239	99119309493	374810.1	10.178	273621. 3
SARIMA$(1,1,1)(1,1,2)_{12}$	2092.20	26.013	95702398019	368248. 1	10.179	273800.7
SARIMA$(1,1,1)(2,1,2)_{12}$	2094.22	26.084	96702609971	369641.7	10.153	272463.6
SARIMA$(2,1,1)(2,1,2)_{12}$	2093.25	26.149	176472086892	369805.3	10.172	273003.1
SARIMA$(1,1,2)(2,1,2)_{12}$	2094.26	26.186	99468092035	376716.4	10.407	277743.9

Parameter Estimation

Maximum likelihood was used to estimate the parameters of the SARIMA(1,1,1)(1,1,1)$_{12}$ model; the results are shown in Table 4. The p-values for the coefficients of AR1, SAR12, MA1, and SMA12 are significantly different to zero. The Box-Ljung test for the SARIMA(1,1,1)(1,1,1)$_{12}$ model produced a p-value greater than $\alpha = 0.05$, which supports the finding that this model is appropriate.

Table 4: Parameter estimates for the SARIMA(1,1,1)(1,1,1)$_{12}$ model.

SARIMA(1,1,1)(1,1,1)$_{12}$ Model					
AIC	**BIC**	**MSE**	**RMSE**	**MAPE**	**MAE**
2082.56	25.836	81078653662	364338.998	10.160	272226.510
Coefficient	**Estimate**		**T-value**		**P-value**
AR 1	0.0890		-1.68		0. 040
SAR 12	-0.3536		-2.38		0.020
MA 1	0.9993		56.60		0.000
SMA 12	0.8227		6.60		0.000
Box-Ljung test : P-value = 0.706					

Based on these results, the final model SARIMA(1,1,1)(1,1,1)$_{12}$ can be expressed as:

$$(1-0.089B)(1+0.354B_{12})(1-B_{12})(1-B)y_t = (1-0.9993\,B)(1-0.823B_{12})\varepsilon_t \qquad (15)$$

The quality of the above model has been assessed, the model diagnostics have been checked, and the results show that the SARIMA(1,1,1)(1,1,1)$_{12}$ model shown above is an adequate model for the data.

Forecasting

Using the final model, SARIMA (1,1,1) (1,1,1)$_{12}$, as expressed in Eq. (15), we forecast future quantities of monthly water production for domestic use in Gaza city for 12 months in 2013; the last 4 actual values were not included in the original series so that they could be compared with the forecasted values. The forecast time series plot for monthly water production in Gaza city is shown in Fig. 9. The series of the forecasted values appears to follow the same behavior of the original series. All forecasted values for the year 2013 lie between the upper and lower boundaries of the 95% confidence intervals, indicating that the forecasting was accurate.

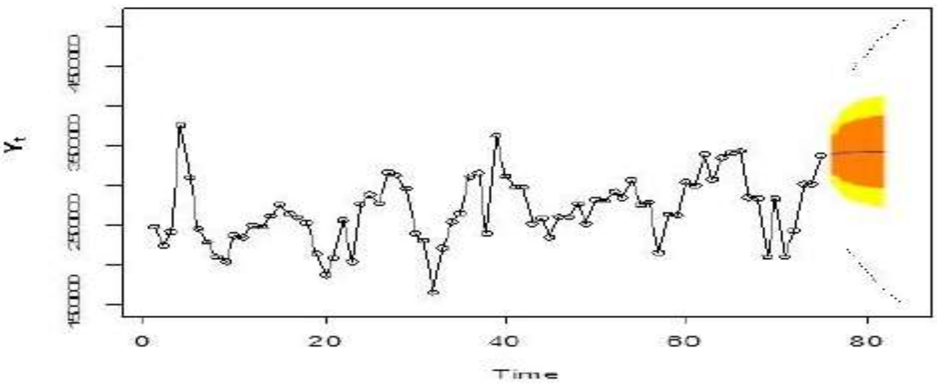

Fig- 9: A plot of time series data for monthly water production in Gaza city with forecast production and 95% confidence interval for predicted values.

CONCLUSION

From the previous discussion, the following conclusions may be drawn:

1- Statistical tests show that the time series of monthly water production in Gaza city involves a general trend and seasonal patterns at a lag of 12 months. The difference transformation at a lag of 1 followed by a lag of 12 was used to achieve stability in the series.

2- The most suitable model for the time series of the monthly water production in Gaza city was found to be the seasonal model SARIMA (1 , 1 , 1)×(1 , 1 , 1)$_{12}$. This model was selected as it had the smallest values for the

AIC, BIC, MSE, RMSE, MAPE, and MAE criteria, as well as the Box-Ljung test. The final model can be expressed as:

$$(1-0.089B)(1+0.354B_{12})(1-B_{12})(1-B)y_t = (1-0.9993B)(1-0.823B_{12})\varepsilon_t$$

3- Using the final model, monthly water production in Gaza city was forecast for 12 months in the year 2013. The forecast values for 2013 were in keeping with the original series values. Moreover, forecast values for the year 2013 were all within the upper and lower boundaries of the 95% confidence intervals. The forecast values showed an increasing trend for the monthly water production in the city. Overuse of water is likely to be a real problem for the city as there will be increasing demand for domestic water in the face of water scarcity. Therefore, decision-makers should take suitable measures to tackle the problem of increasing monthly water production in Gaza city.

4- The increasing demand for water production in the coming years in Gaza city is due to the increasing population, in addition to the shortage of ground water and pollution of ground water reservoirs. Other sources and alternatives for water production, such as building desalination plants for sea water, should be investigated in the immediate future.

RECOMMENDATIONS

1. The model identified here should be used by officials and decision makers in forecasting future water needs in the coming months in Gaza city. These predictions should be taken on board so that the problems of future water production in the city are not aggravated.
2. We recommend the use of this forecasting approach. The model can be updated regularly using more up to date time series data to improve the model's performance.
3. By generalizing the results of this study, similar models can be used to predict water production in other cities and comparisons may be drawn.
4. Other forecasting methods should be examined and so that comparisons may be drawn between the predictions made.

REFERENCES

1. Box GEP, Jenkins GM; Time Series Analysis, Forecasting and Control (3rd ed.). Englewood Clifs, New Jersey: Prentice Hall. 1976.
2. Hamilton JD; Time series analysis. Princeton, NJ: Princeton University Press. 1994.
3. Paretkar PS; Short-Term Forecasting of Power Flows over Major Pacific Northwestern Interties: Using Box and Jenkins ARIMA Methodology. Electrical Engineering Department, Virginia Polytechnic Institute and State University, Virginia. 2008.
4. Asomaning SS; Analysis of Maternal Mortality with Time; A Case Study of the Okomfo Anokye Teaching Hospital – Kumasi (2000-2010) (Unpublished BSc. Mathematics and Statistics thesis). Kwame Nkrumah University of Science and Technology, West Africa. 2012.
5. Spyros M, Wheelwright SC, Hyndman RJ; Forecasting: Methods and Applications (3rd ed.). USA: John Wiley & Sons, Inc. 1998.
6. Akinmutmi C; A seasonal Arima Modelling of Residential Coal Consumption Series in Nigeria. Department of Mathematics, Ahmadu Bello University, Nigeria. 2011.
7. Gyasi-agyei K; Analysis and modeling of prevalence of measles in the Ashanti region of Ghana. Kwame Nkrumah University of Science and Technology, West Africa. 2012.
8. Render B, Stair R, Hanna M; Quantitative Analysis for Management (9th ed.). Boston, MA: Allyn & Bacon. 2006.
9. Hoaglin DC, Mosteller F, John W; Understanding Robust and Exploratory Data Analysis. Hoboken, NJ: John Wiley & Sons, 1983; 404–414,
10. Russell RS, Taylor BW; Operations Management, John Wiley & Sons, 2006; 497-498.
11. Venables WN, Ripley BD; Modern Applied Statistics with S-PLUS. Springer. 1999; 128.
12. Stram DO, Wei WWS; Temporal Aggregation in the ARIMA Process. Journal of Time Series Analysis, 1986; 7:279-292.
13. Mynsbrugge JV; Bidding Strategies Using Price Based Unit Commitment in a Deregulated Power Market. K.U.Leuven, 2010; 62: 24-28.
14. Pankratz A; Forecasting with Univariate Box-Jenkins Models: Concepts and Cases. New York: John Wiley & Sons. 1983.
15. Kibar M; Building Cost Index Forecasting with Time Series Analysis (unpublished MSc thesis). Civil Engineering Department, Middle East Technical University, Turkey. 2007.
16. Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y; Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. Journal of Econometrics, 1992; 54:159-178.