

Using The ARIAM Model To Predict Future Values Of Outflow For The Kufa Barrage University Of Babylon- Civil Engineering College- Master Water Resources- Iraq –Babylon

Ali Salman Fahd Ibrahim¹, Zainab Ali Omran²

¹<https://orcid.org/0009-0008-8068-8023>

²<https://orcid.org/0000-0001-5012-2047>

Abstract

Water resource management for dam reservoirs requires comprehensive information on expected future water availability. Our current study focuses on the Kufa Dam, located on the Euphrates River in Najaf Governorate, specifically in the Kufa District. We analyzed the time series of outflow data for a period of 31 years (1994–2024) to study short-term hydrological trends. A linear trend model was applied to the historical data, resulting in a series of future outflows that were predicted using conventional time series modeling techniques. The Minitab 22 statistical software was used for all time series analyses, including model fit and prediction. This study presents the data, statistical results, and predictions in tables and graphs, providing a clear view for future water resource planning and reservoir management strategies. The study concluded that the ARIMA (2,1,2) model is the most suitable among the selected models for modeling and predicting the selected time series data.

Keywords: Minitab 22 Software, Kufa Barrage, ARIMA Model

1.Introduction

The ARIMA model developed by Box and Jenkins in the 1970, this is one of the most popular time series analysis models. This model can forecast future values by integrating three components: integration (I), moving average (MA), and auto regression (AR), which takes into account data non stationarity. The Box and Jenkins model is still used today, despite the scientific advances we have achieved so far due to its flexibility in dealing with different types of data the seasonal component (SARIMA) was added to the model in later periods, which contributed to the widespread use of this model in analyzing seasonal fluctuations. Ntebogang Dinah Moroke (2014) [3].

Time series analysis is widely used for predicting future values and monitoring operational systems, because it allows us to identify the underlying structure and influencing factors within the data. Time series analysis encompasses a variety of statistical methods for identifying, analyzing, and forecasting temporal patterns in data. Key objectives of this analysis include highlighting the main characteristics of the time series pattern, understanding how historical data impacts future outcomes, and studying how multiple interrelated time series interact to predict future values. Sara and Hanumanthappa (2017) [2].

2. Method and Material

2.1 Case Study

The Kufa barrage is located on the Euphrates River in the Najaf Governorate, specifically in the Kufa District. This barrage functions as the primary reservoir for the Al-Kafeel-Al-Shanafiyyah Irrigation Project, possessing an initial storage capacity of 75 million cubic meters. Data on outflows from the primary series, covering a 31-year period (1994–2024), were examined to evaluate short-term hydrological patterns.

2.2 Methodology

A well-known methodology was adopted in this research, where the ARIAM model was used for the purpose of predicting future values. Data on the discharge coming out of the Kufa barrage and recorded for the period from (1994 to 2024) were used. Some documents from General Authority of Dams and Reservoirs, Baghdad, Iraq, 2024 [4]

3. Time Series Analysis

A Time series plot of the average annual flow rates was created using Minitab 22 software, with the aim of Application of techniques used in time series analysis to study the water flow data of the Kufah barrage (Figure 1). This figure serves as a crucial initial step in highlighting trends, seasonal patterns, and anomalous behaviors. (Dr. Nigel Lee Forrester 2013) [5]

Then, the correlation between the outflow rate in a given year and its previous values was measured using the autocorrelation function (ACF). The resulting autocorrelation plot (Figure 2) clearly indicates how current outflows are statistically dependent on past lags a crucial step in identifying suitable autoregressive orders. (Spyros Makridakis et al 2018) [6].

The annual outflow rate data were analyzed using an autoregressive (AR) model to estimate future outflow values based on historical data. To overcome random fluctuations in outflow rates and achieve accurate predictions. The AR model was applied using an order determined by the partial autocorrelation function (PACF) and autocorrelation function (ACF) (Juan Frausto-Solis and Esmeralda Pita 2008) [7].

If the (Y) values are normally distributed, the partial autocorrelation between $(Y_t - u)$ and (Y_t) can be defined as

$$\phi(u) = \text{cor}(Y_t, Y_{t-u} | Y_{t-1}, \dots, Y_{t-u+1}).$$

A broader methodology using regression theory.

Y_t based on $Y_{t-1}, \dots, Y_{t-u+1}$. The prediction is

$$\hat{Y}_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} \dots \beta_{u-1} Y_{t-u+1}$$

In the case of thinking back and predicting Y_{t-u} , we can use the same set of predictors. While originally proposed in Bras & Rodríguez-Iturbe (1985), this concept was later formalized in a time-symmetric moving-average representation by Koutsoyiannis (2000), which provides a general stochastic framework suitable for both backward and forward forecasting in hydrologic time series. (Demetris Koutsoyiannis 2000) [8]. the best predictor will be

$$\hat{Y}_{t-u} = \beta_1 Y_{t-u+1} + \beta_2 Y_{t-u+2} \dots \beta_{u-1} Y_{t-1}$$

A resulting formula was applied to all data using Minitab22 built-in functions. The results are explained in Table (2) and are also presented in the solid and dashed graphs in Figure (4).

Figure (1), a graph of the total outflows data for cofa barrage

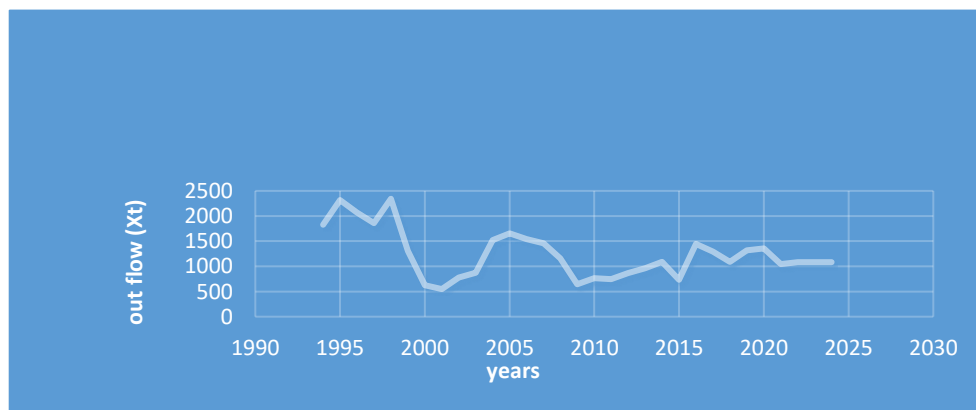
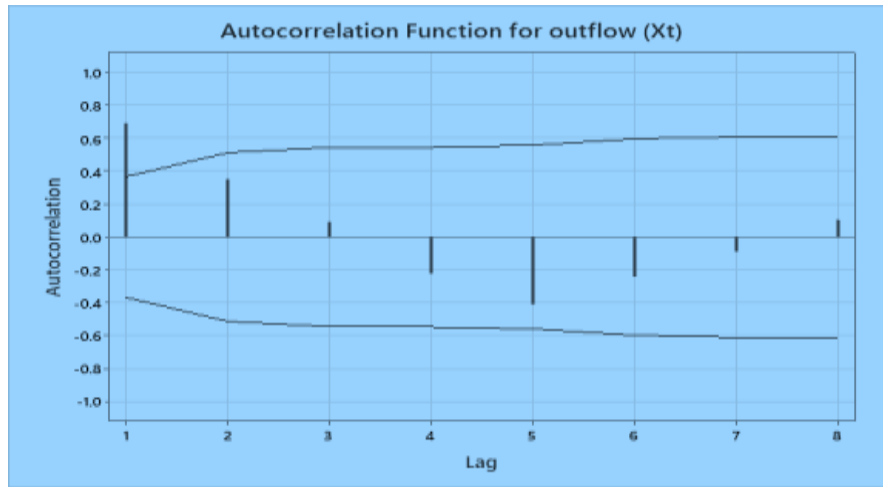


Table (1): Data on the outflow from the Kufa barrage for the period (1994-2024)

T	year	outflow (Xt)	Xt*T	T^2
1	1994	1827	1827	1
2	1995	2318	4636	4
3	1996	2070	6210	9
4	1997	1859	7436	16
5	1998	2343	11715	25
6	1999	1307	7842	36
7	2000	625	4375	49
8	2001	548	4384	64
9	2002	779	7011	81
10	2003	872	8720	100
11	2004	1525	16775	121
12	2005	1657	19884	144
13	2006	1543	20059	169
14	2007	1458	20412	196
15	2008	1159.76	17396.4	225
16	2009	645.17	10322.72	256
17	2010	761.99	12953.83	289
18	2011	743.78	13388.04	324
19	2012	867.25	16477.75	361
20	2013	960.71	19214.2	400
21	2014	1089.06	22870.26	441
22	2015	729.95	16058.9	484
23	2016	1442.82	33184.86	529
24	2017	1293.35	31040.4	576
25	2018	1086.35	27158.75	625
26	2019	1315.8	34210.8	676
27	2020	1355.6	36601.2	729
28	2021	1045.47	29273.16	784
29	2022	1083.66	31426.14	841
30	2023	1085	32550	900
31	2024	1080	33480	961

Figures (2). The autocorrelation coefficient for the outflow data



Figures (3). Partial autocorrelation coefficient for the outflow data

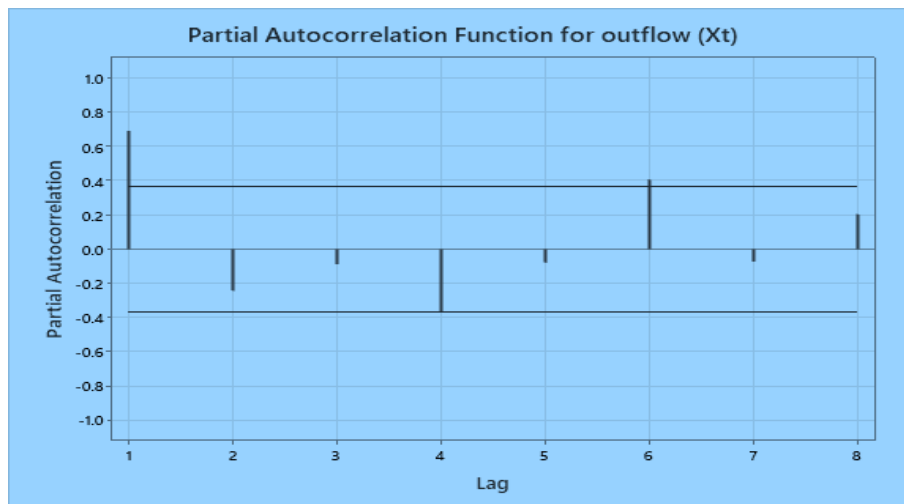


Figure (4), Time series plot for outflow kufa barrage (actual and forecasted) (1994-2024)

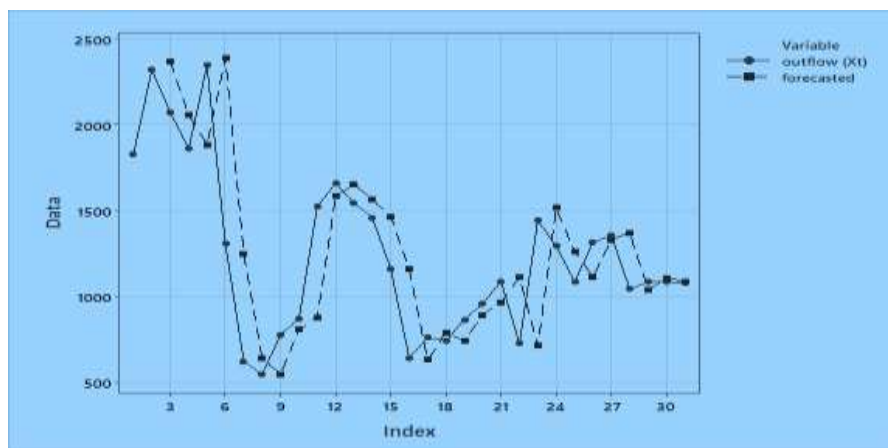


Figure (5), Trend Analysis Plot for outflow (Xt)

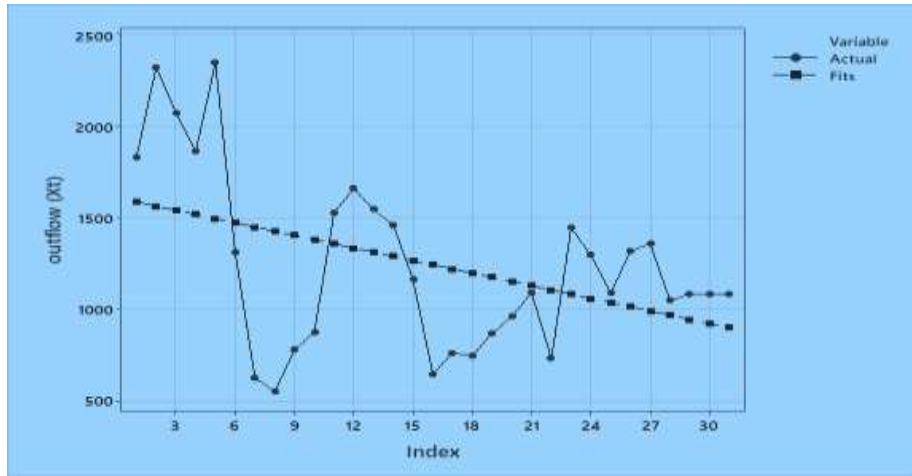


Table (2), the outflow rates for kufa barrage (actual and forecasted) (1994-2024)

year	outflow (Xt)	Forecasted	e
1994	1827		
1995	2318		
1996	2070	2364.484253	-294.4842529
1997	1859	2057.584083	-198.5840825
1998	2343	1883.408079	459.591921
1999	1307	2385.946936	-1078.946936
2000	625	1247.515586	-622.5155859
2001	548	642.9698639	-94.9698639
2002	779	550.0056983	228.9943017
2003	872	809.3971993	62.60280073
2004	1525	877.8794194	647.1205806
2005	1657	1581.585183	75.41481684
2006	1543	1654.654729	-111.6547293
2007	1458	1562.898075	-104.898075
2008	1159.76	1464.855298	-305.0952975
2009	645.17	1160.67421	-515.5042102
2010	761.99	633.8737109	128.1162891
2011	743.78	793.4348088	-49.65480876
2012	867.25	741.0797655	126.1702345
2013	960.71	894.3581834	66.35181662
2014	1089.06	968.2946236	120.7653764
2015	729.95	1111.589057	-381.6390572
2016	1442.82	713.7736384	729.0463616

2017	1293.35	1514.47958	-221.1295802
2018	1086.35	1262.243031	-175.893031
2019	1315.8	1114.274276	201.5257241
2020	1355.6	1332.448035	23.15196464
2021	1045.47	1369.154794	-323.6847939
2022	1083.66	1039.385671	44.27432909
2023	1085	1108.921421	-23.9214211
2024	1080	1088.684093	-8.684092728

4. Results and Discussions

4.1 In this paper, two approaches were applied, visual inspection through graphical representation of the data and the Augmented Dickey–Fuller (ADF) statistical test. The results revealed a discrepancy between the two methods. While the visual inspection suggested that the series was non-stationary due to the presence of An Downward trend, the ADF test indicated that the series is stationary. Specifically, the test produced a p-value of (0.010), which is less from the 5% significance level, Which led to the rejection of the null hypothesis and acceptance of the alternative hypothesis of stationarity. Accordingly, reliance it placed on the statistical test, as it provides a more objective and rigorous criterion compared to visual inspection, which serves only as a preliminary indicator. Thus, the time series under study can be considered stationary based on statistical evidence. This confirms that the time series is stationary. Three models were identified using (PACF) and (ACF). Of the three selected models, *ARIMA* (2, 1, 2) was selected based on information criteria and standard error estimates. The model passed all diagnostic tests and was subsequently used to produce time-series forecasts of outflow from kufa barrage.

4.2 the model coefficients were estimated using Minitab 22 and its suitability was evaluated using the Akaike Information Criterion (AIC) as shown in Table 3.

Table (3) Comparison table between ARIMA models

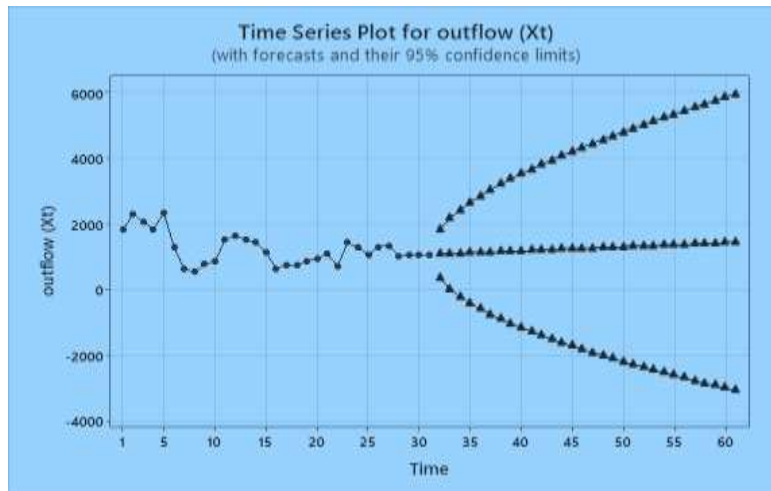
MODLE	AICc	P-value (Lag 12)	P-value (Lag 24)	Overall Rating
ARIMA(1,1,2)	448.02	0.013	0.012	Weak – because it has a higher value for the Corrected AICc Information Criterion and a high autocorrelation in the residuals
ARIMA(2,1,2)	437.679	0.106	0.132	Best in terms of AICc Corrected Information Standard and Residual Independence
ARIMA(2,2,2)	440.341	0.068	0.209	Acceptable equilibrium; the residuals are not correlated, but the AICc value is higher than (2,1,2).

In Table (4), the forecast values demonstrate a distinct upward trend in the anticipated discharge over time, with values progressively increasing from one period to the next. The ARIMA model has effectively captured the presence of a trend component in the time series, as evidenced by this behavior. This suggests that the trend is continued. future periods of discharge will increase provided that the influencing conditions remain essentially unchanged. The accuracy of the model decreases in long-term forecasting due to the wide range of uncertainties, while it is considered suitable for short-term forecasting, where the confidence intervals are comparatively narrow.

Therefore, it is advisable to use near-term forecast results to support operational decisions, while exercising caution when relying on long-term values shown in Figure (6). We observe that SE Forecast values shown in Table (4) increase as we move further away from the last actual observation, a normal behavior in time series models. This reflects the increasing uncertainty in the forecast over time, as estimation errors accumulate as the forecast horizon lengthens, thus reducing the reliability of long-term forecasts compared to short-term forecasts.

Table (4) Forecasts values after year 2024

years	Forecast	SE Forecast	Lower	Upper
2025	1096.62	378.28	355.04	1838.19
2026	1106.40	551.86	24.55	2188.26
2027	1120.05	676.68	-206.51	2446.60
2028	1131.51	786.28	-409.91	2672.94
2029	1144.21	881.51	-583.89	2872.31
2030	1156.21	969.12	-743.65	3056.07
2031	1168.60	1049.72	-889.27	3226.48
2032	1180.77	1125.52	-1025.70	3387.24
2033	1193.07	1197.06	-1153.65	3539.79
2034	1205.30	1265.27	-1275.13	3685.72
2035	1217.56	1330.55	-1390.84	3825.96
2036	1229.81	1393.37	-1501.76	3961.37
2037	1242.06	1454.03	-1608.42	4092.54
2038	1254.31	1512.81	-1711.40	4220.02
2039	1266.56	1569.91	-1811.08	4344.20
2040	1278.82	1625.51	-1907.83	4465.46
2041	1291.07	1679.76	-2001.93	4584.07
2042	1303.32	1732.80	-2093.66	4700.29
2043	1315.57	1784.73	-2183.21	4814.35
2044	1327.82	1835.65	-2270.78	4926.42
2045	1340.07	1885.64	-2356.53	5036.68
2046	1352.32	1934.78	-2440.62	5145.27
2047	1364.58	1983.14	-2523.16	5252.31
2048	1376.83	2030.76	-2604.26	5357.92
2049	1389.08	2077.70	-2684.04	5462.20
2050	1401.33	2124.02	-2762.59	5565.25
2051	1413.58	2169.75	-2839.98	5667.14
2052	1425.83	2214.92	-2916.29	5767.95
2053	1438.09	2259.58	-2991.59	5867.76
2054	1450.34	2303.76	-3065.94	5966.61

Figure (6), Time Series plot for outflow (xt)

Conclusion.

The ARIMA (2,1,2) model outperforms the others in terms of residual independence and model fit, according to Table 3. results of the comparison of the models. The model yields the lowest corrected Akaike Information Criterion (AICc) value of 437.679. Moreover, no statistically significant autocorrelation is detected in the residuals based on the Ljung–Box test, with p-values of 0.106 at lag 12 and 0.132 at lag 24, which indicates that the model provides an adequate fit. On the other hand. The analysis indicates that the ARIMA(1,1,2) model does not adequately represent time series behavior, as evidenced by a high residual autocorrelation ($p < 0.05$) and a high AICc (Accurately Corrected Information Criterion) of 448.020. In contrast, the ARIMA(2,2,2) model outperforms ARIMA(2,1,2) with an AICc of 440.341, despite achieving an acceptable degree of residual independence ($p = 0.068$ and 0.209). Based on the final comparison of the models, ARIMA(2,1,2) demonstrates the best fit and accuracy in representing the time data and in its predictive power compared to the other tested models. The results of the ARIMA (2,1,2) model show that the predicted discharge follows an upward trend, with a clear increase in uncertainty as the forecast period lengthens. Despite the model's statistical efficiency, the interpretation of the results must be within a physical and hydrological framework, taking into account the limitations inherent in the nature of the data.

References;

- [1] N. Naresh and K. Soorya, "A study on forecasting models," International Journal of Scientific and Research Publications, vol. 6, no. 5, pp. 1–5, 2016.
- [2] S. Sara and M. Hanumanthappa, "A comparative study of time series forecasting models," International Journal of Computer Applications, vol. 161, no. 9, pp. 1–6, 2017.
- [3] N. D. Moroke, "Comparative analysis of univariate time series forecasting models for short-term rainfall prediction," International Journal of Statistics and Applications, vol. 4, no. 6, pp. 285–294, 2014.
- [4] Some documents from General Authority of Dams and Reservoirs, Baghdad, Iraq, 2024
- [5] N. L. Forrester, "Forecasting models and techniques: A review," Journal of Business and Economics, vol. 5, no. 12, pp. 2351–2360, 2013.
- [6] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," PLoS ONE, vol. 13, no. 3, e0194889, 2018.

- [7] J. Frausto-Solis and E. Pita, "Time series prediction using evolutionary neural networks," in Proc. IEEE Congress on Evolutionary Computation (CEC), 2008, pp. 209–214.
- [8] D. Koutsoyiannis, "Uncertainty, entropy, scaling and hydrological stochasticity. 1. Marginal distributional properties of hydrological processes and state scaling," Hydrological Sciences Journal, vol. 45, no. 4, pp. 505–524, 2000.

تحليل سلسلة زمنية للتدفقات الخارجة من سد الكوفة

علي سلمان فهد إبراهيم

جامعة بابل - كلية الهندسة المدنية - ماجستير موارد مائية - العراق - بابل

ali.ibrahim990@student.uobabylon.edu.iq

د. زينب علي عمران

Eng.zainabali@uobabylon.edu.iq

ملخص

تتطلب إدارة موارد المياه لخزانات السدود معلومات شاملة حول توافر المياه المتوقع في المستقبل. تركز دراستنا الحالية على سد الكوفة، الواقع على نهر الفرات في محافظة النجف، وتحديدًا في قضاء الكوفة. قمنا بتحليل السلاسل الزمنية لبيانات التدفق الخارج على مدى 31 عامًا (1994-2024) لدراسة الاتجاهات الهيدرولوجية قصيرة الأجل. طُبِّق نموذج اتجاه خطي على البيانات التاريخية، مما أسفر عن سلسلة من التدفقات الخارجة المستقبلية التي تم التنبؤ بها باستخدام تقنيات نمذجة السلاسل الزمنية التقليدية. استخدم برنامج Minitab 22 الإحصائي لجميع تحليلات السلاسل الزمنية، بما في ذلك ملائمة النموذج والتنبؤ. تعرض هذه الدراسة البيانات والنتائج الإحصائية والتنبؤات في جداول ورسوم بيانية، مما يوفر رؤية واضحة لتخطيط موارد المياه المستقبلية واستراتيجيات إدارة الخزانات. لخصت الدراسة إلى أن نموذج $ARIMA(2,1,2)$ يعد الأكثر ملائمة من بين النماذج المختبرة لتمثيل بيانات السلاسل الزمنية المختارة والتنبؤ بقيمتها المستقبلية بدقة.

الكلمات المفتاحية: برنامج Minitab 22 ، سدة الكوفة، نموذج ARIMA.