Numerical Methods in Civil Engineering

Journal Homepage: https://nmce.kntu.ac.ir/

Development of a System Dynamics Model for Prediction of Karaj Reservoir Share in Tehran Water Supply

Zahra Sheikholeslami [*], Majid Ehteshami [**], Sara Nazif [***]

ARTICLE INFO

RESEARCH PAPER

Article history:
Received:
August 2023
Revised:
October 2023
Accepted:
January 2024

Keywords:
Karaj Dam,
System dynamics,
Time series analysis,
Tehran, water supply.

Abstract:

Tehran's water consumption (TWC) is rising as a result of rapid population growth, climate change, and precipitation decline. As water resources of Tehran are also affected by a variety of factors, the water supply scheme becomes so complicated and it is necessary to consider the complexity and dynamics interactions in water supply system before any decision making. In this study, Karaj reservoir as an important surface water resource of Tehran's water supply system was modeled through system dynamics (SD) approach for prediction of Karaj Dam share in Tehran water supply. The SD model was implemented in AnyLogic software using the historical data from April 2006 to March 2022, and the stock and flows and dynamics variables were predicted for April 2023 to March 2023. The novelty of this research is the development of SD model of Karaj Dam to simulate its relationships and interactions for prediction of Karaj Dam share in Tehran water supply in 2023. In this regard, the TWC and Karaj Dam inflow were predicted by using SARIMA$(1,0,0)(0,1,1)_{12}$ model for April 2023 to March 2023. Finally, to assess the precision of the results obtained from the SARIMA and SD models, the criteria of coefficient of determination (R^2), Error percentage (E%), and Nash–Sutcliffe model efficiency coefficient (NS%) was calculated. The results showed that the Karaj Dam inflow will be decreased during April 2023 to March 2023 due to the precipitation decline, consequently the Karaj Dam reservoir volume will be reduced and for this reason less water can be harvested from Karaj Dam reservoir for different applications. Therefore, it is clear that in the future we will have faced the challenge of water supply in Tehran.

1. Introduction

As a consequence of population growth, climate change, and urbanization, global water scarcity is one of the most important challenges to sustainable development in the world, especially in arid and semi-arid regions [1,2]. The largest and capital city of Iran, Tehran, is situated in semi-arid areas and is experiencing both water shortage and rapid population growth [3,4]. The population of Tehran in 2022 was 9'381'546 person, while in 2021, it was 9'259'000 person, which represented an annual change of 1.32%. By the latest revision of the UN World Urbanization Prospects, Tehran's population is predicted 9'729'742 person in 2025 [5], which caused challenges in Tehran's water supply due to its rapid population growth, climate change, and precipitation decline. All these challenges lead to an increase in the importance of considering the complexity and dynamics interactions in water resource systems and also regarding the importance of forecasting the water resources' behavior.

One of the popular methods to investigate the multiple interactions of water resource system components and deal with the system's complexity is the system dynamics (SD) method [6]. SD modeling as a powerful management tool, is used to consider the feedback loop processes of the water system by converting the whole water system into the interconnected flows and stocks for simulating the water

* *Corresponding Author: PhD Candidate of Environmental engineering, Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran. Email:* zsheikholeslami@email.kntu.ac.ir

** *Associate Professor of Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran*

*** *Associate professor, School of Civil Engineering, College of Engineering, University of Tehran*

https://doi.org/10.61186/NMCE.2303.1029

resource systems [7,8]. To make strategic decisions, SD can therefore assist water managers in recognizing troubling trends, understanding their main causes, and assessing the most effective management practices [9,10]. As such, SD modeling attracted the attention of scientists and has become popular for water resource system analysis, hence several studies investigated the application of SD modeling in water resource management [6,9]. Abdolvandi et al. (2013) used the SD approach for modeling of Taleghan dam and groundwater resources to perform a comprehensive operational model that could be applied to similar conditions in the world [11]. Li et al. (2018) modeled the water resource of Zhengzhou, China including four subsystems as population, water supply, water demand, and economics through the SD approach, which provided future information of water distribution and water system cost to make optimal allocation [12]. Bakhshianlamouki et al. (2020) evaluated the effects of Urmia Lake Basin restoration proceeding through the SD model by considering climate change, return flow increment, and transferring the inter-basin water. The results show Urmia Lake Basin was susceptible to climate change [13]. Nozari et al. (2021) applied the SD model, to model the surface water allocation and control system operation of Dez Reservoir in Iran. The obtained results represented the high precision of the SD model in modeling the performance of Dez Reservoir behavior [7]. Also Babolhakami et al. (2023) implemented the same SD approach for analyzing water consumption of Gelevard Dam [14].

Tehran's water demand is supplied by surface and groundwater resources. Surface water resources are comprising, Lar, Karaj, Taleghan, Latian, and Mamlou Dams, and the Tehran-Karaj aquifer is its groundwater resource [3]. According to the water shortage in Tehran, studying the surface water resources of Tehran is necessary to calculate and estimate water supply resources capacity. In this respect, accurate prediction of dam inflow, provides information for reservoir operation strategies and generally for surface water resource planning systems. One of the methods applied for predicting is a time series modeling [15].

For a single time series, univariate Box–Jenkins methods such as autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA) were used to analyze and predict the observed data without considering the other variables [15,16]. The SARIMA model has the benefit of accounting for any seasonality patterns, making it more powerful than ARIMA, but it has some limitations as requiring at least 50 historical observations, and accurate prediction over a short period [17,18].

Several studies have been undertaken on predicting of time series through ARMA ARIMA, and SARIMA modeling. In this regard, Valipour et al. (2013) applied ARMA and SARIMA models to forecast monthly Dez dam reservoir inflow by using 42 years of historical data. The obtained results indicated that the SARIMA model had better efficiency than the ARMA model [19]. Wang et al. (2015) studied the effect of coupling ensemble empirical mode decomposition (EEMD) with ARIMA, to enhance the annual runoff prediction precision. Based on the results EEMD could be effectively enhanced the ARIMA model precision to predict the time series of annual runoff [20]. Adnan et al. (2017) applied ARIMA model to forecast the Astore River monthly flow. Based on common statistical efficiency assessment measures such as root mean square error, they concluded that ARIMA performed better results [21]. Tadesse et al. (2017) looked into how well the SARIMA model performed to forecast the monthly flow of Waterval River, South Africa. They found a good similarity between observed and SARIMA-modelled data [15]. Almanjahie et al. (2017) applied five SARIMA models to predict household water consumption in Saudi Arabia. The predicted values represented well harmony with its historical time series, which provided an approximative perspective on water consumption for decision-makers [22]. Aghelpour and Varshavian (2020) forecasted the Zilakirud River flow in Iran, through artificial intelligence and stochastic models. The results indicated that among the stochastic models, ARIMA, and among artificial intelligence, the Group Method of Data Handling had the best precision and performance in prediction [23]. Wang et al. (2021) compared the SARIMA and artificial neural network (ANN) models to forecast surface water quality time series as pH, dissolved oxygen (DO), chemical oxygen demand (COD), and ammonia nitrogen (NH3-N). The results showed that due to the nonlinear and complicated nature of the data in the water quality time series, the SARIMA model could not be analyzed in these time series as well [24]. As well as in the other study, Imran et al. (2023), achieved that the ANN model obtained more accurate results than the SARIMA model for analyzing and forecasting the flood events in Swat River Basin [16].

Generally, urban water supply systems include different parts with complicated relationships and interactions which can even change through the time and under different conditions. Therefore, by simulating this system through SD approach, the importance of spatial and temporal changes and variabilities are taken into account. In this study, Karaj Reservoir as one of the main surface water resources of Tehran, was modeled through SD approach by using AnyLogic software. The novelty of this research is the development of a system dynamics model of Karaj Dam to simulate its relationships and interactions for prediction of

Karaj Dam share in Tehran water supply in 2023, to assist water managers in recognizing future trends, probable challenges, and assessing the management practices to choose the most effective one.

2. Case Study

Karaj Dam is one of the main surface water resources of Tehran, which supplies 25% of Tehran's water demand. It is located on the Karaj River, northwest of Tehran City, at geographical coordinates of 51°5'E, 23°57'N; 23 km from the city of Karaj [25]. The Total characteristics of Karaj Dam are represented in Table 1.

Table 1: Total characteristics of Karaj Dam

River	Karaj
Main Water Supply Goal	Agricultural Water Supply, Tehran's Water Supply
Total Capacity	205 MCM
Water Supply for the Treatment Plant	Jalalie, Kan, and number 6 Treatment plant
Minimum volume	60 MCM

3. Methodology

In this study, Karaj reservoir as an important surface water resource of Tehran's water supply system was modeled through system dynamics (SD) approach for prediction of Karaj Dam share in Tehran water supply. In this regard the required data was prepared from the National Water and Wastewater Engineering Company of Iran, then the TWC and Karaj Dam inflow was predicted by SARIMA models. The SD model was implemented by using the historical data from April 2006 to March 2022, and the results of SARIMA models. The precision of the model and the formulas obtained from the data were checked with the given criteria in Table 2. It is worth noting that the closer values of coefficient of determination (R^2) and Nash Sutcliffe model efficiency coefficient (NS) to one, and closer values of Error (E%) to 0, indicate the higher precision of the model.

Table 2: The model precision criteria

Coefficient of determination	$R^2 = \frac{(\sum_1^n X_{pi}X_{mi})^2}{\sum_1^n X_{pi}^2 \sum_1^n X_{mi}^2}$
Error	$\%E = \frac{\sum_1^n X_{mi} - X_{pi}}{\sum_1^n X_{mi}}$
Nash Sutcliffe model efficiency coefficient	$NS = 1 - \frac{\sum_1^n (X_{mi} - X_{pi})^2}{\sum_1^n (X_{pi} - \overline{X}_p)^2}$
X_p =estimated values of the model X_m =observed values,	$\overline{X}_p$= the average of the observed values N = the number of data.

3.1 Data Preparation

The data of Karaj dam consisting of inflow, outflow, evaporation, household water supply, agricultural, and TWC, prepared from the National Water and Wastewater Engineering Company of Iran. The data was from April 2006- March 2022 .

Since evaporation is low in Karaj dam in the last few years, in this study it is considered the monthly average of April 2006- March 2022 for the SD model (Table 3).

Table 3: Average monthly evaporation of Karaj Dam in April 2006- March 2022 (MCM)

Month	Evap	Month	Evap
April	0.16	October	0.27
May	0.31	November	0.11
June	0.49	December	0.05
July	0.58	January	0.02
August	0.58	February	0.01
September	0.45	March	0.06

3.2 System Dynamics (SD)

Urban water supply systems include different parts with various relationships and interactions which should be simulated to use in the design, operation, and management of the urban water supply system. Rarely are these relationships and interactions in water supply systems taken into account using a dynamics approach, and usually, the models were used in, were explicit in the relationships among significant parameters, and the importance of spatial and temporal dynamics was not regarded [26–29]. SD is a problem-solving approach to evaluate the physical and behavioral of the systems, which is based on feedback loops and interaction between stocks and flows. The state of the system is described by stock variables, and the stock variables changes are described by flow variables [14]. In this study, the SD model of Karaj Dam was applied via AnyLogic V8.7.7 software as shown in Figure 1.

The volume of water in Karaj Dam is an integrated stock, which can be calculated as Equation 2.

$$\frac{\mathrm{d(Karaj\ Dam)}}{\mathrm{dt}} = \mathrm{Karaj\ Inflow} - \mathrm{overflow\ Karaj} - \mathrm{EvapKaraj} - \mathrm{KarajOut} + V_0(98\ \mathrm{MCM}) \quad (2)$$

$$\mathrm{Overflow\ Karaj} = \mathrm{Karaj\ Dam\ volume} - \mathrm{Total\ capacity\ of\ Karaj\ Dam}$$

To estimate the volume of Karaj Dam share in Tehran water supply, a linear regression (Eq 3-4) was formed between the reservoir volume, Tehran's water consumption, and the volume of Karaj Dam share in Tehran water supply. It is worth mentioning that the R^2 coefficients value for Equation 3 and Equation 4 are equal to 0.7 and 0.5, respectively.

$$\mathrm{KJK1} = \mathrm{KJK1\_A1} - (\mathrm{KJK1\_A2} \times \mathrm{Tehran\ Water\ Consumption}) + (\mathrm{KJK1\_A3} \times \mathrm{KarajDam}) \quad Time < 24 \quad (3)$$

$$\mathrm{Overflow\ Karaj} = \mathrm{Karaj\ Dam\ volume} - \mathrm{Total\ capacity\ of\ Karaj\ Dam} \quad Time \geq 24 \quad (4)$$

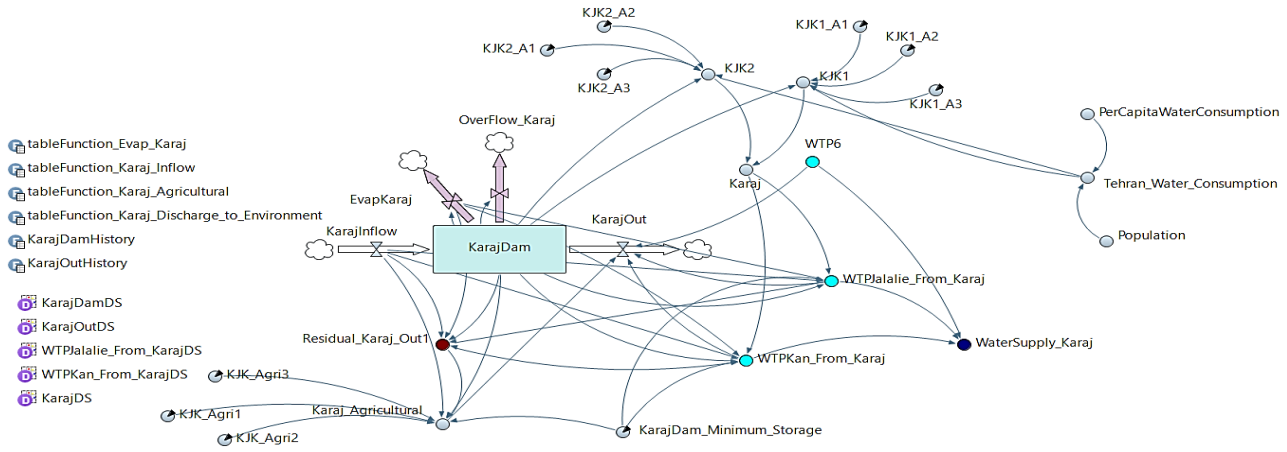


Fig. 1: SD Model of Karaj Dam

where

Tehran_Water_Consumption =*TWC* per capita × Population

KJK1_A1 = 9.3

KJK1_A2 =0.094

KJK1_A3 =0.023

KJK2_A1 =46.48

KJK2_A2 =0.18

KJK2_A3 =0.002

To estimate the agricultural water supply from Karaj Dam as shown in Equation 5, a linear regression was applied between agricultural water supply as a dependent variable and volume reservoir and Karaj Dam inflow as independent variables. The value of the R^2 coefficient for Equation 5 is equal to 0.76.

$$\text{Karaj\_Agricultural} = \text{KJK\_Agri1} + (\text{KJK\_Agri2} \times \text{KarajInflow}) + (\text{KJK\_Agri3} \times \text{KarajDam}) \quad (5)$$

where.

KJK_Agri1 =-27.4

KJK_Agri2 =0.47

KJK_Agri3 =0.14

3.1 SARIMA Modelling

In this study, to predict the TWC per capita and Karaj Dam inflow, the SARIMA model was used, also in recent researches, it was applied to predict short-term water demand [17]. SARIMA model is a linear regression model that incorporates both seasonal and non-seasonal time series components. The main condition of using time series analysis such as ARMA, ARIMA, and SARIMA models are data stationarity and homogenous, having no trend, and following the normal distribution [22,24,30]. To check the data stationarity the Augmented Dickie-Fuller statistical test (ADF) was used. The null hypothesis of ADF test is the non-stationarity of the time series, when the p-value is lower than 0.05, the null hypothesis is rejected [15]. To check the homogeneity and trend of the time series data, Pettit and Mann-Kendall tests was used, which their null hypothesis is that the time series data are homogeneous and have no trend. The p-value larger than 0.05 is a confirmation of the null hypothesis [15]. The normality of the data was checked using the Anderson-Darling test. The null hypothesis of the normality test is normality of the data distribution, which is confirmed due to the p-value being larger than 0.05.[31]

The SARIMA model is denoted by SARIMA (p, d, q) (P, D, Q)s where p, d, and q are respectively represented non-seasonal autoregressive (AR) term, non-seasonal differences term, and non-seasonal moving average term (MA). Also, P, D, and Q are respectively represented the seasonal AR term, seasonal differences term, and seasonal MA term [15,32]. The SARIMA model is defined in Equation 6 [15].

$$\Phi(BS)\,\varphi(B)(1-BS)^D(1-B)^d\,y_t = \Theta(BS)\,\theta(B)\varepsilon_t \quad (6)$$

where

φ, Φ = AR coefficients for the non-seasonal and seasonal component

θ, Θ= MA coefficients for the non-seasonal and seasonal component

B= backward operator

(1–BS)D = Dth, a seasonal difference of season s

(1 – B)d = dth, non-seasonal difference

εt = a distributed independent random variable

The autocorrelation function (ACF) and partial autocorrelation function (PACF) curves were employed as useful tools to verify stationarity, identify the SARIMA model structures, and specify the order of p, P, q, and Q [22,24]. Repetitive procedures by applying different orders were examined and the best model was chosen via the least value of the Akaike Information Criterion (AIC) indicator [33]. The goodness of the fitted model must be looked into after selecting the best model. The diagnostic check was done by the p-values of the modified Ljung-Box test, as well as examining the residuals main characteristics (independence, randomness, normal distribution, and constant variance) by using the ACF and PACF plots, Anderson-Darling test, and Run Test [21,32,34,35]. Finally after selecting the best SARIMA models, the precision of the models have been investigated by the R squared, error percentage, and Nash Sutcliffe (NS) efficiency coefficient.

4. Results

4.1 TWC Per Capita Prediction

First, Pettit and Mann-Kendall tests were used to examine the homogeneity and trend of the time series data of water consumption per capita in Tehran. The p-values obtained from the statistical tests are given in Table 4. According to the null hypothesis in Pettitt's test, the p-value of 0.4 confirmed the homogeneity of TWC per capita. Similarly, according to the result of the Mann-Kendall test, the p-value of 0.31 that is larger than 0.05, the null hypothesis of the Mann-Kendall test is approved and the data have no trend.

Table 4: P-values of statistical tests of TWC per capita time series

Pettitt's Test	Mann-Kendall Tests	Augmented Dickey-Fuller Test	Anderson-Darling normality test
0.4	0.31	0.03	0.0005

According to Table 4, the p-value of the ADF test is equal to 0.03, which indicated the stationarity of the time series. After checking the homogeneity, trend, and stationarity in the time series data, the normality of the data was checked using the Anderson-Darling test. Given that the p-value of Anderson-Darling test for TWC per capita time series was lower than 0.05, data do not follow the normal distribution, hence according to the principle of normality of the data, in analyzing and predicting time series, normalizing the time series is necessary. To normalize the time series some transformation functions were checked as shown in Table 5. Due to the higher p-value of Anderson-Darling test (p-value>0.05), Johnson Transformation function was selected as the best transformation functions to normalize data by following Equation 7

Table 5: Transformation identification for TWC per capita

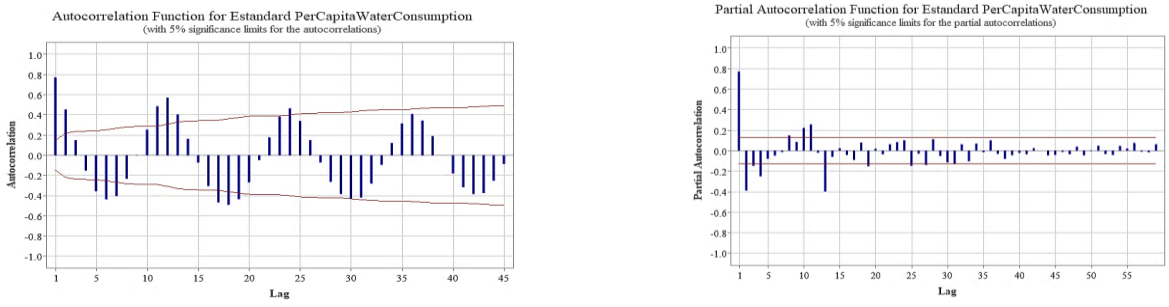
Transformation	P-Value
Normal	<0.005
Box-Cox Transformation	0.045
Lognormal	0.022
Exponential	<0.003
Gamma	0.047
Johnson Transformation	0.372

$$\text{Johnson transformation function} = -7.02 + 3.56 \times \text{Ln}(\text{X} - 4.2) \quad (7)$$

In Equation 7, X is TWC per capita.

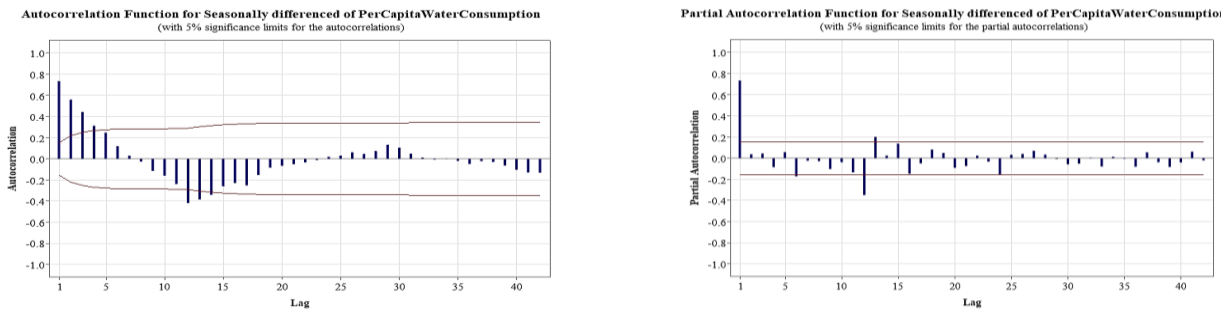
After Johnson Transformation, the results show that, the p-value (0.372) is larger than 0.05 and the null hypothesis of the Anderson-Darling test is rejected and the transformed data follows the normal distribution. Finally, all the essential requirements of the time series for analyzing and forecasting time series have been met. To prepare the data for SARIMA modeling, the transformed data was standardized, and then the proper model was fitted for the modified series .

ACF and PACF plots were used to specify ARIMA(p, d, q)(P, D, Q)s values. The observed mean monthly TWC per capita was shown to be stationary by the Mann-Kendall test, so there is no need for non-seasonal differencing (d = 0), but seasonal differencing is necessary for seasonal stationarity because the significant spikes were seen in ACF plots every 12 months as sine wave, which approved that the observed mean monthly TWC per capita was seasonally nonstationary. Seasonal differencing (D = 1) of the observed TWC per capita data was performed to make time series seasonal stationary.



(a) **(b)**

Fig. 2: a) ACF and b) PACF for observed data of TWC per capita



(a) **(b)**

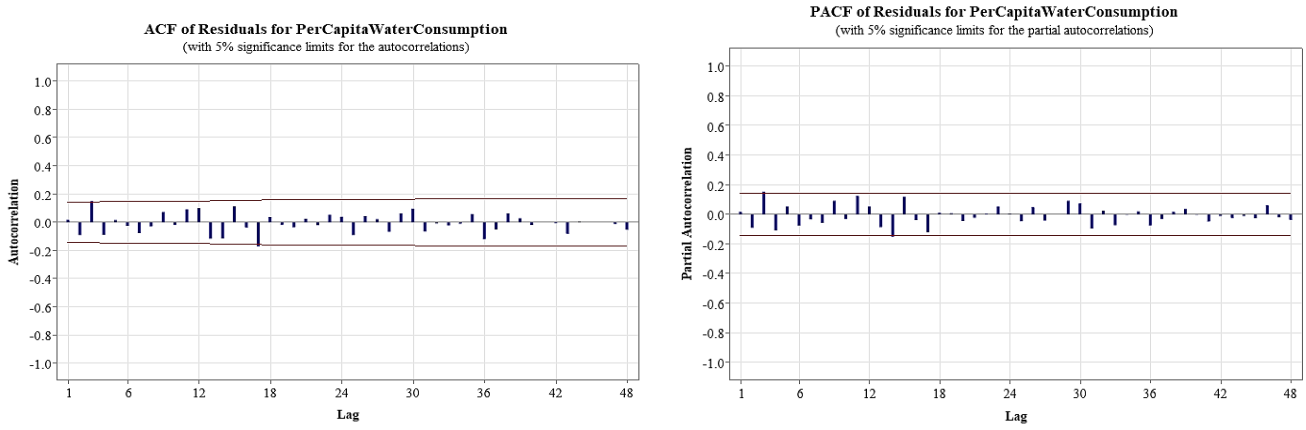
Fig. 3: a) ACF, and b) PACF for first order seasonal differencing and de-seasonal observed data of TWC per capita

For more consideration, SARIMA (p, 0, q) (P, 1, Q)12 models were proposed. The parameters of these models (p, q, P, and Q) were determined based on the characteristics of de-seasonalized ACF and PACF plots (Figure 3).

Based on Figure 3a the ACF plot of the seasonally differentiated time series is geometric decayed and has significant spikes at 1st, 2nd, 3rd, and 4th lags, and after the 4th lag, the confidence interval of the ACF plot cuts off and placed in the ACF confidence interval. So for the moving average parameter (MA) (non-seasonal component), the order was proposed, q=0-4. As well as, in Figure 3a, the ACF plot has also a negative significant spike at 12th lags and cuts off the ACF confidence interval in 12th lag. Then the ACF spikes were decayed after 24th lag, and they were within ACF confidence interval, so for seasonal moving average (SMA) values, the order was proposed Q=1. In the same way, for the non-seasonal parameter, the PACF plot of (Figure 3b) the seasonally differentiated time series has significant spikes at 1st lag, and the confidence interval of the PACF plot cuts off, therefore the value of autoregressive (AR) parameter was suggested p=1. As well as in Figure 3b, the PACF plot has significant spikes at 12st lag, and then it was placed in the confidence interval after the 12th lags, so the seasonal autoregressive (SAR) parameters were proposed P=1. Thereupon, based on the given conditions 9 SARIMA models were identified for further evaluation as represented in Table 6.

The models shown in Table 6 were fitted to the time series of TWC per capita, and the SARIMA(1,0,0)(0,1,1)12 model was chosen as the best model based on the AIC's (Akaike Information Criterion) lowest value.

The diagnostic checking is obtained by the selected model's residual analysis. The modified Ljung-Box test was used for goodness of fit test investigating and checking the residuals' independence. The independence of the residuals is confirmed by the results of the modified Ljung-Box test in



(a) **(b)**

Fig. 4: a) ACF, and b) PACF of residuals for TWC per capita model

Table 6: Optional SARIMA models for TWC per capita

Model Selection	
Model (d = 0, D = 1)	**AIC**
p = 1, q = 0, P = 0, Q = 1*	273.8
p = 1, q = 1, P = 0, Q = 1	285.7
p = 1, q = 2, P = 0, Q = 1	301.9
p = 1, q = 0, P = 1, Q = 1	304
p = 1, q = 1, P = 1, Q = 1	315.4
p = 1, q = 4, P = 0, Q = 1	325.3
p = 1, q = 3, P = 0, Q = 1	336.1
p = 1, q = 4, P = 1, Q = 1	339.7
p = 1, q = 3, P = 1, Q = 1	351.5

Table 7: The results of the modified Ljung-Box test of TWC per capita model

Lag	12	24	36	48
Chi-Square	13.99	30.40	42.15	46.39
P-Value	0.17	0.1	0.16	0.45

Table 7, by considering that the p-value is larger than 0.05 for all delays. Based on the ACF and PACF residual plots which are shown in Figure 5, all residual values, have no significant autocorrelation .

The diagnostic checking is obtained by the selected model's residual analysis. The modified Ljung-Box test was used for goodness of fit test investigating and checking the residuals' independence. The independence of the residuals is confirmed by the results of the modified Ljung-Box test in Table 7, by considering that the p-value is larger than 0.05 for all delays. Based on the ACF and PACF residual plots which are shown in Figure 5, all residual values, have no significant autocorrelation.

The p-value of Anderson-Darling test for TWC per capita residuals was obtained 0.1, which approves that the residuals are normally distributed at $\alpha = 0.05$. The normal probability and histogram plots in Figure 4a-b are a confirmation of the residuals normality distribution too.

Figure 5c shows the residuals versus fitted value plot, which approve the assumption that the residuals have constant variance. The points in Figure 5c randomly fall on both sides of zero and lack any discernible patterns, demonstrating the

constant variance of residuals. Also, the randomness of the residuals is confirmed by Run Test approval due to the p-value larger than 0.05 .

According to the investigations carried out, the precision of the SARIMA(1,0,0)(0,1,1)12 model was confirmed and the R2, E%, and NS efficiency coefficient were used to check the precision of the model (Table 8) .

Table 8: Precision criteria for TWC per capita model

R^2	0.98
E%	1.2
NS	0.8

Finally, to predict the future data of TWC per capita for 12 months in 2023, SARIMA(1,0,0)(0,1,1)12 model was used. Figure 6 is shown the comparison of the observed and modeled data.

4.2 Karaj Dam Inflow Prediction

The results of necessary statistical tests for the Karaj Dam inflow time series are represented in Table 9. Based on the p-values > 0.05 of Pettitt's and Mann-Kendall's tests, the homogeneity and trendlessness of the Karaj Dam inflow time series are confirmed. As well as, the p-value of

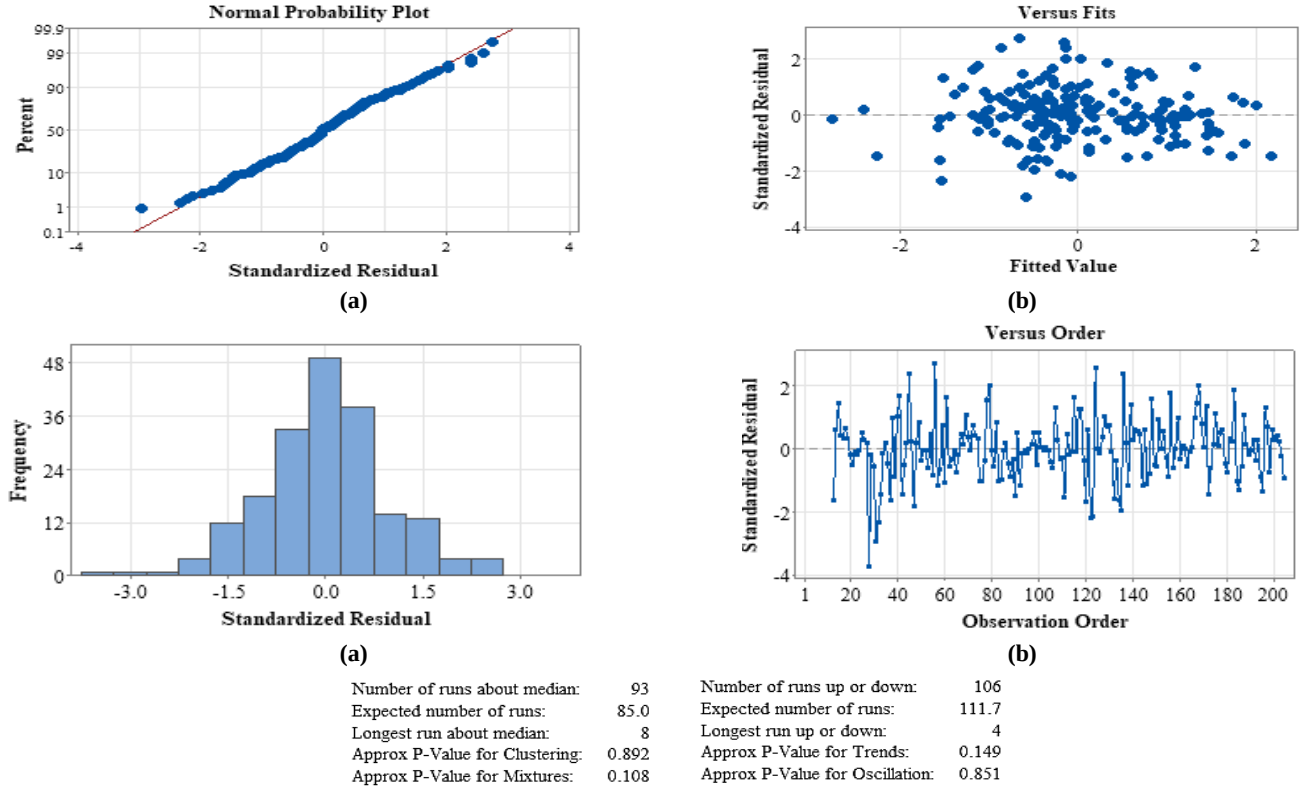


Fig. 5: a)Normal probability plot, b) Histogram plot, c) Residuals versus fitted value plot, d) RUN-test plot of residuals for TWC per capita model

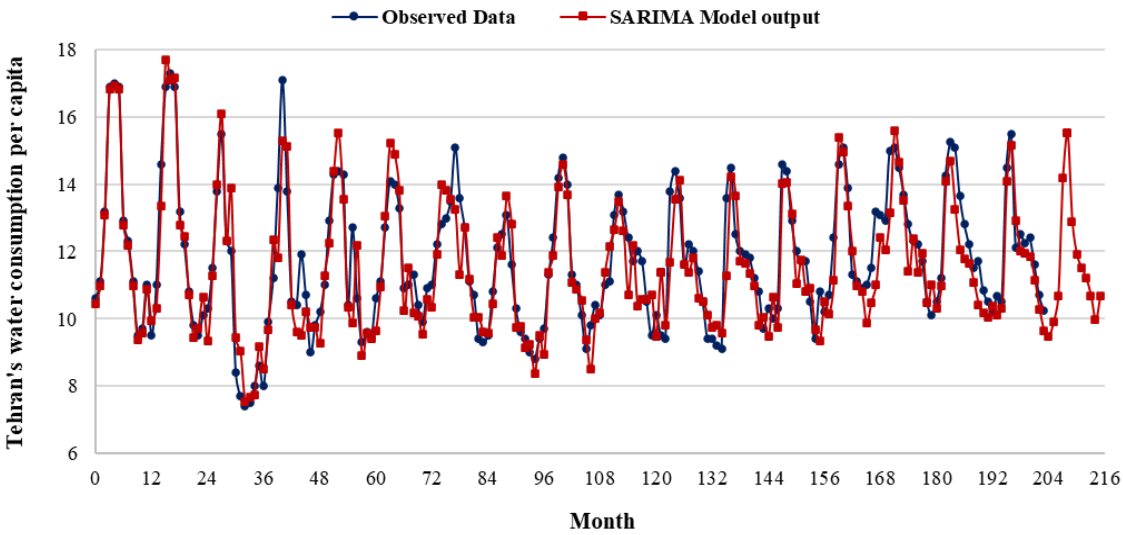


Fig. 6: Observed and SARIMA model output of TWC per capita

the ADF test is 0.007, which rejects the null hypothesis of the test that the data is non-stationary. Moreover, the result of the Anderson-Darling test shows that the Karaj dam inflow time series does not follow the normal distribution.

So to normalize the time series some transformation functions were checked as shown in Table 10. Due to the higher p-value of Anderson-Darling test (p-value>0.05), Box-Cox Transformation function was selected as the best transformation functions to normalize data.

Table 9: P-values of statistical tests of Karaj Dam inflow time series

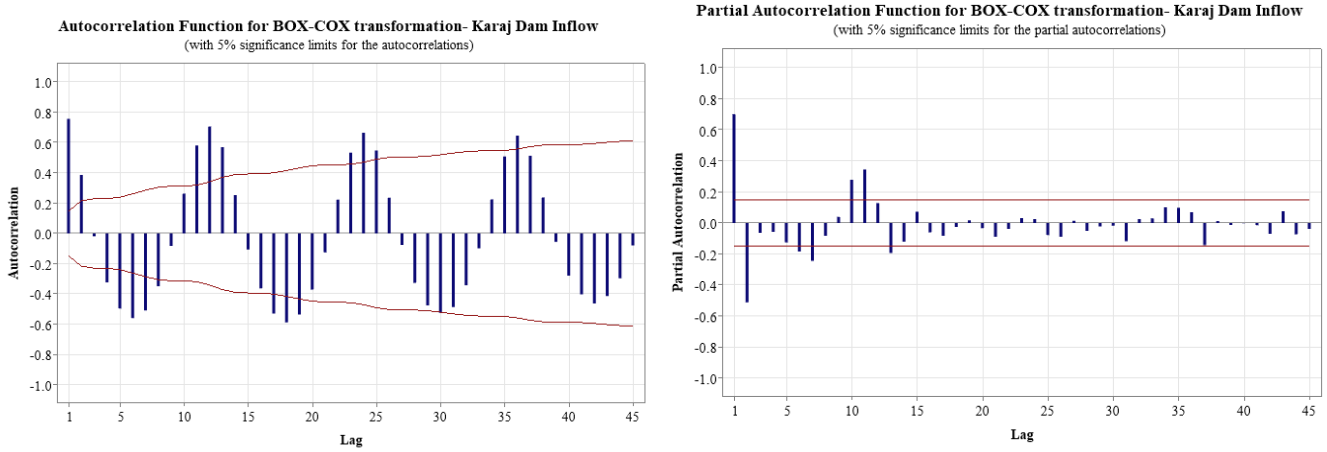
Time series	Pettitt's Test	Mann-Kendall Tests	ADF Test	Anderson-Darling test
Original	0.63	0.3	0.007	0.0001

Table 10. Transformation identification for Karaj Dam inflow

Transformation	P-Value
Normal	<0.005
Box-Cox Transformation	0.155
Lognormal	0.002
Exponential	0.001
Gamma	0.03
Johnson Transformation	0.01

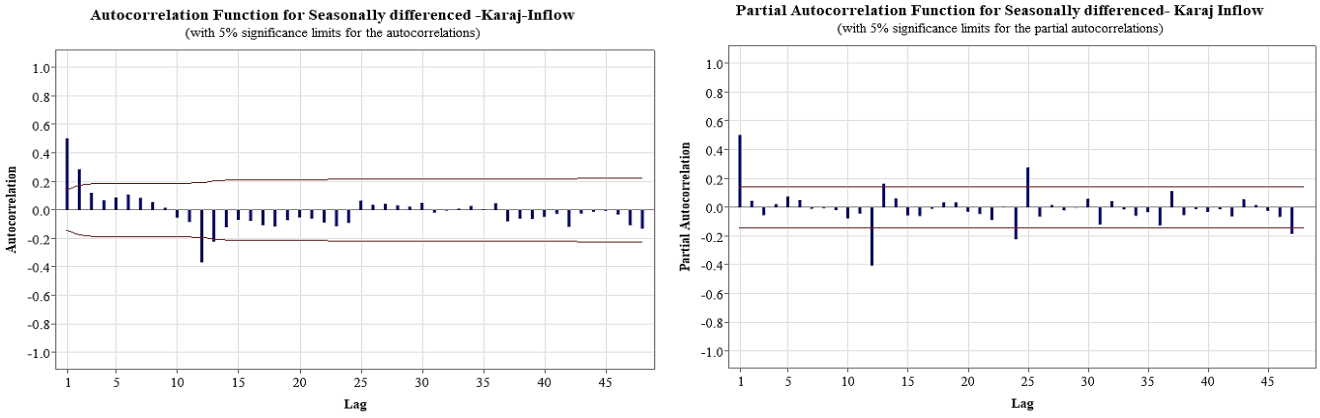
After Box-Cox Transformation, the results show that, the p-value of Anderson-Darling test is 0.115, which shows the transformed data follows the normal distribution.

The ACF and PACF plots of the observed data of Karaj Dam inflow are represented in Figure 7. The ACF plot shows a seasonal oscillation occurs every 12 months. By focusing on the ACF of observed data, a slow decrescent trend is seen in the ACF spikes at seasonal lags (12^{th}, 24^{th}, 36^{th}) where s = 12. It displays a nonstationary characteristic and proposes fist-order seasonal difference. The ACF and PACF plots of differentiated time series are represented in Figure 8. According to Figure 8a, the ACF plot of the seasonally differentiated Karaj Dam inflow time series has significant spikes at 1^{st}, and 2^{nd} lags, and after 2^{nd} lag the spiked placed in the ACF confidence interval. So the moving average parameter order (MA) (non-seasonal component), was proposed q=0-2. As well as, in Figure 8a, the ACF plot has also a negative significant spike at 12^{th} lags and cuts off the ACF confidence interval in 12^{th} lag. Then the ACF spikes were decayed after 24^{th} lag, and they were within ACF confidence interval, so for seasonal moving average (SMA) values, the order was proposed Q=1.



(a) **(b)**

Fig. 7: a) ACF, and b) PACF observed data of Karaj dam inflow



(a) **(b)**

Fig. 8: a) ACF, and b) PACF for first order seasonal differencing and de-seasonal observed data of Karaj dam inflow

In the same way, for the non-seasonal parameter, the PACF plot of (Figure 8b) the seasonally differentiated time series has significant spikes at 1st lag, and the confidence interval of the PACF plot cuts off, therefore the value of autoregressive (AR) parameter was suggested p=1. As well as in Figure 8b, the PACF plot has significant spikes at 12th and 24the lag, and then it was placed in the confidence interval after the 36th lags, so the seasonal autoregressive (SAR) parameters were proposed P=0-2. Thereupon, based on the given conditions 9 SARIMA models were identified for further evaluation as represented in Table 11.

Table 11: Optional SARIMA models for Karaj dam inflow

Model Selection	
Model (d = 0, D = 1)	**AIC**
p = 1, q = 0, P = 0, Q = 1*	210.701
p = 1, q = 0, P = 1, Q = 1	235.091
p = 1, q = 2, P = 0, Q = 1	239.336
p = 1, q = 0, P = 2, Q = 1	249.731
p = 1, q = 2, P = 1, Q = 1	277.813
p = 1, q = 1, P = 0, Q = 1	288.694
p = 1, q = 2, P = 2, Q = 1	307.221
p = 1, q = 1, P = 1, Q = 1	380.557
p = 1, q = 1, P = 2, Q = 1	421.22

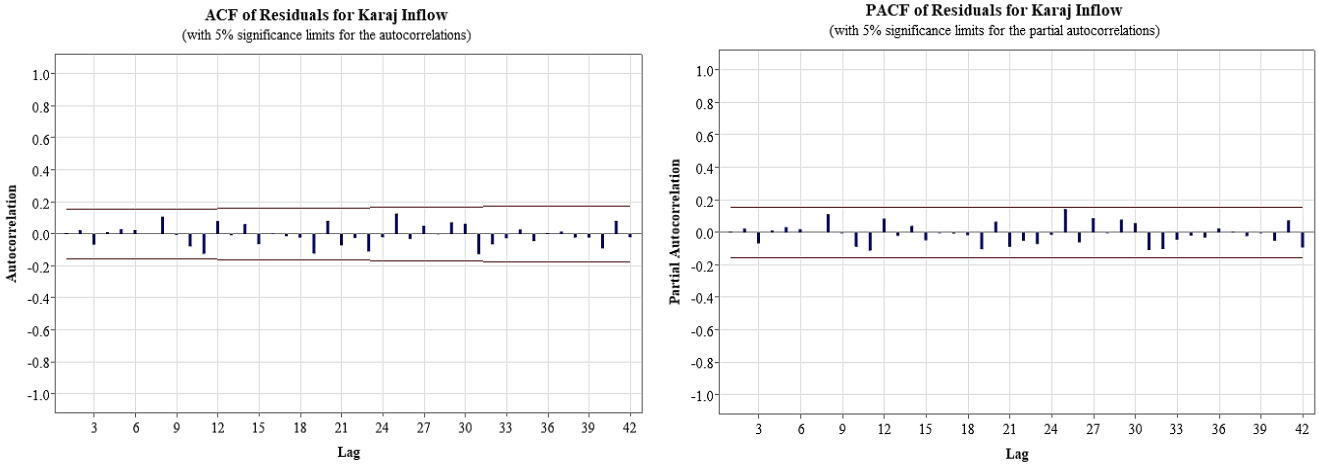
The SARIMA(1,0,0)(0,1,1)12 model has the lowest value of AIC, hence it was selected as the best model for Karaj Dam inflow. The modified Ljung-Box test was used for goodness of fit test investigating and checking the independence of the residuals due to all lags p-values>0.05. It means that the approval of the null hypothesis of no autocorrelation of the residuals was successful; Additionally, based on the ACF and PACF residual plots which are shown in Figure 9, all residual values, have no significant autocorrelation.

Table 12: The results of the modified Ljung-Box test of the Karaj Dam inflow

Lag	12	24	36	48
Chi-Square	20.76	30.59	42.32	56.03
P-Value	0.103	0.12	0.15	0.16

To test the normality of the residuals, Anderson Darling's test was used along with the normal probability plot and histogram of residuals. Anderson Darling's test p-value was obtained 0.12, which indicates the normality of the residual, as also Figure 10a-b is a confirmation of residual normality distribution. Moreover, the variance of the residuals is constant given the residuals versus fitted value plot (Figure 10c), and also the randomness of the residuals is confirmed by Run Test approval due to the p-value greater than 0.05 (Figure 10d).

According to the investigations carried out, the precision of the SARIMA(1,0,0)(0,1,1)12 model was confirmed and the R2, E%, and NS efficiency coefficient were used to check the precision of the model in Table 13.



(a) **(b)**

Fig. 9: a) ACF, and b) PACF of residuals for Karaj Dam inflow

distribution. Moreover, the variance of the residuals is constant given the residuals versus fitted value plot (Figure 10c), and also the randomness of the residuals is confirmed by Run Test approval due to the p-value greater than 0.05 (Figure 10d).

According to the investigations carried out, the precision of the SARIMA(1,0,0)(0,1,1)12 model was confirmed and the R^2, E%, and NS efficiency coefficient were used to check the precision of the model in Table 13.

Table 13. Precision criteria for Karaj Dam inflow

R^2	0.93
E%	0.6
NS	0.95

Finally, to predict the future data of Karaj Dam inflow for 12 months in 2023, SARIMA(1,0,0)(0,1,1)12 model was used. Figure 11 is shown the comparison of the observed and modeled data of Karaj Dam inflow

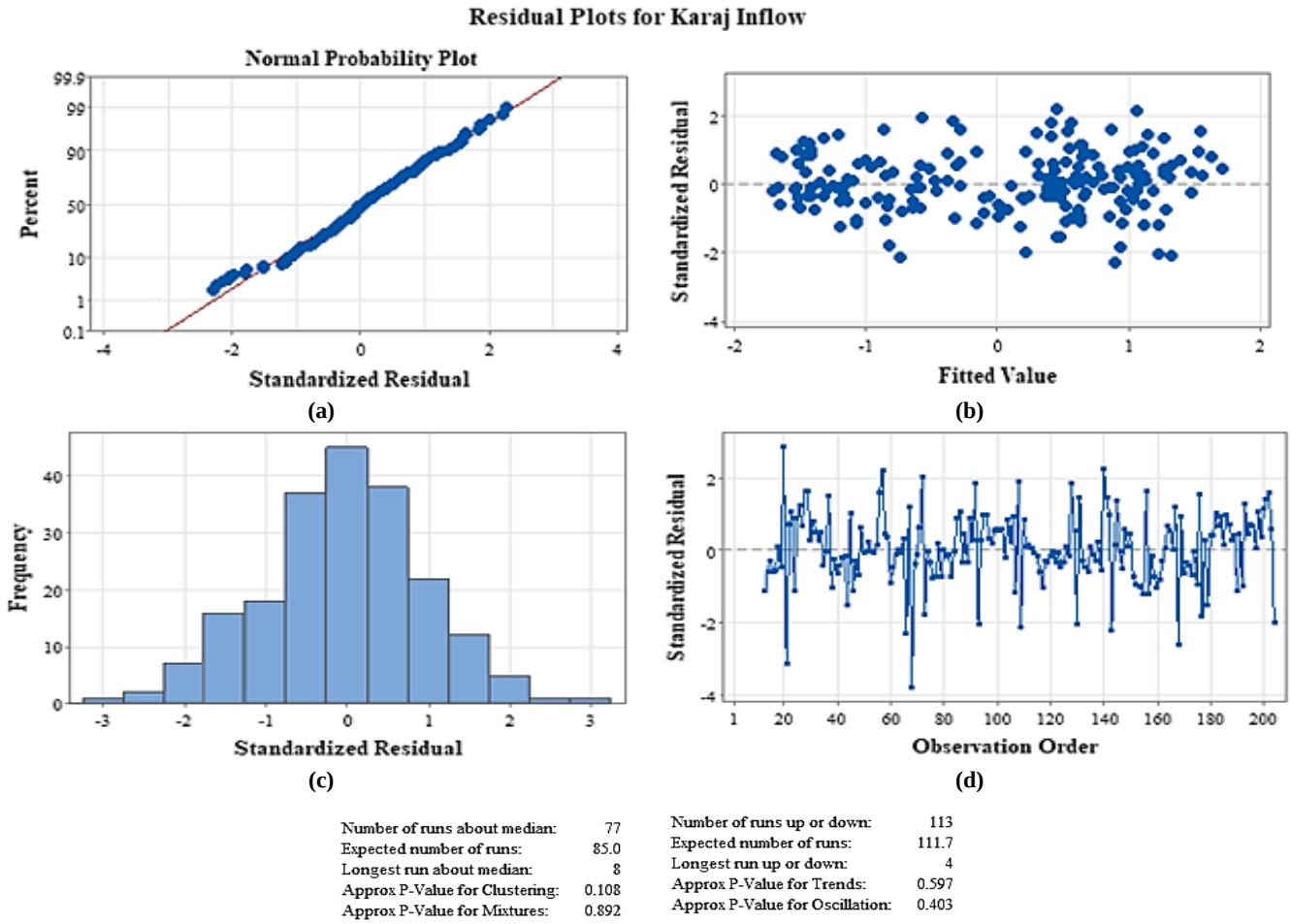


Fig. 10: a)Normal probability plot, b) Histogram plot, c) Residuals versus fitted value plot, d) RUN-test plot of residuals for Karaj Dam inflow

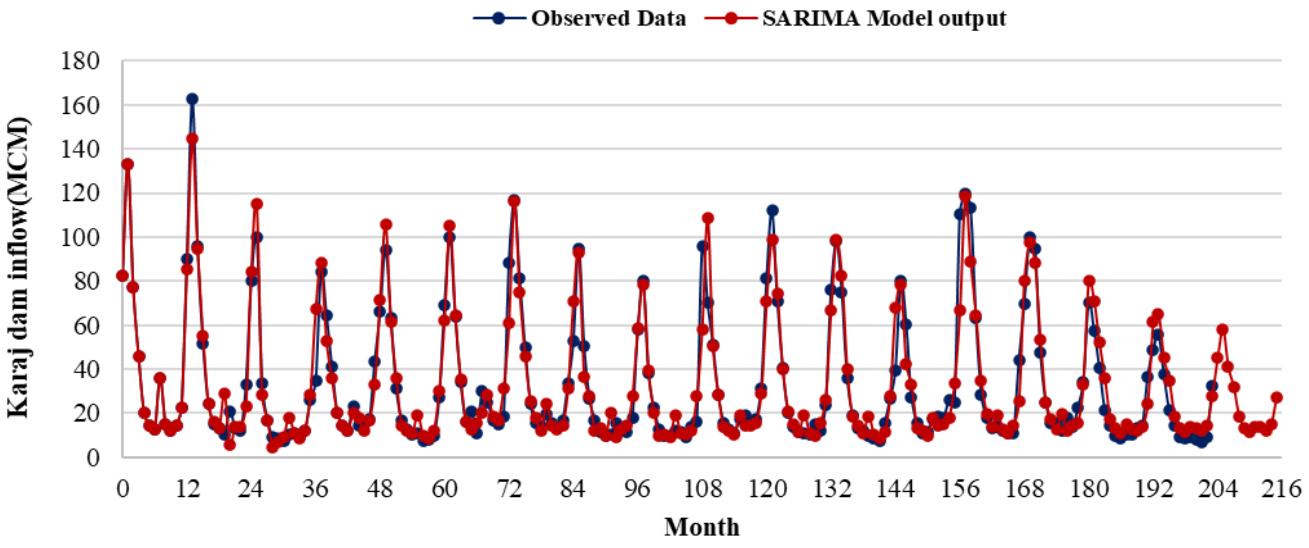


Fig. 11: Observed and SARIMA model output of Karaj Dam inflow

4.3 Model Calibration

Model calibration should be done before predictive analysis. The simulation results are compared with real historical data to verify their agreement so that the reliability of model parameters and model accuracy can be judged. The simulation results of Karaj Dam Reservoir Volume, Karaj Dam household water supply, Karaj Outflow are compared with the actual data from National Water and Wastewater Engineering Company of Iran. Present SD Model has several parameters to tune it makes sense to use the built-in

optimizer to search for the best combination. The objective in this case is to minimize the difference between the observed simulation output and historic data. The simulation length for this study is from 2006 to 2023, a total of 17 years. The total simulation period is divided into three stages: the model calibration stage, which runs from 2006 to 2016 (Time: 0-131), the model validation stage, which runs from 2017-2022 (Time: 132-203) and the model prediction stage, which runs from 2022 to 2023 (Time: 204-215). The goal of model calibration is to obtain reasonable parameter values by comparing model output to historical data. The simulation is based on a monthly time step (i.e. the water balance is computed monthly) and it is focused on the simulation of Karaj Dam water supply for Tehran. Table 13 lists the model parameters that are adjusted to match the simulation results to the historical data. Based on optimized parameters in Table 14 the model calibrated, the model validation was checked ,and the calibrated model predicted the Karaj Dam Reservoir Volume, Karaj Dam household water supply, Karaj Outflow in 2022-2023.

Table 14: The Range of model parameter changes for calibration purposes

Model Parameter	Dimension	Min	Max	Optimized
KarajDam_ Minimum_Storage	MCM	54	66	61.5
KJK2_A1	-	8.4	10.2	9.5
KJK2_A2		0.08	0.10	0.09
KJK2_A3		0.021	0.025	0.022
KJK1_A1		41.8	51.1	44.5
KJK1_A2		0.16	0.20	0.175
KJK1_A3		0.0018	0.0022	0.002
KJK_Agri1		-24.6	-30.14	-26.5
KJK_Agri2		0.42	0.52	0.461
KJK_Agri3		0.13	0.154	0.143

4.4 SD Model Results

The model focuses on the systematic analysis of the Karaj Dam water supply system, and based on water balance principles and compares water demand and water supply. It runs over a time scale of 17 years, accounting for the evolution of the Karaj Dam water supply system from April 2006 to March 2023. The main objectives of Karaj Dam modeling through SD approach are analyzing and predicting, Karaj Dam share in Tehran's water supply as well as the volume and outflow of the Karaj Dam reservoir until 2023.

Figure 12 is shown the observed and calibrated model output of the Karaj Dam reservoir volume, which indicated that the Karaj Dam Reservoir volume will be reduced from April 2023 to March 2023, which the main reason is the reduction of Karaj Dam inflow besides the increment of TWC per capita. On the other hand, according to Figure 13, the Karaj Dam share in Tehran's water supply has not significantly increased, and it will be provided at the same level as it was in the years 2021 and 2022, thereupon the effective factor in the reduction of Karaj Dam Reservoir volume is Karaj Dam inflow. Figure 14 represents the comparison of the Observed and calibrated model output of Karaj Dam Outflow. It shows that the Karaj Dam Outflow will be reduced from April 2023 to March 2023 due to the reduction of agricultural usage. Based on Eq 5 agricultural usage of Karaj Dam has more dependence on Karaj Dam inflow due to its higher coefficient value. The precision criteria of the calibrated models is represented in Table 14.

Table 15. Precision criteria for Karaj Dam SD model (April 2006 to March 2022)

Model	R^2	E%	NS%
Reservoir volume	0.9	3.8	0.88
Karaj Dam share in Tehran's water supply	0.85	0.7	0.81
Karaj dam outflow	0.96	0.6	0.87

5. CONCLUSIONS

In this study, Karaj Dam reservoir as the main surface water resource of Tehran's water supply was modeled in AnyLogic software through the SD approach by using the historical data from April 2006 to March 2022 to predict the Karaj Dam share in Tehran's water supply for April 2023 to March 2023. At first, TWC was modeled by the SARIMA model. As a result, among the suggested models, SARIMA$(1,0,0)(0,1,1)_{12}$ was chosen for TWC prediction, which had the lowest value of AIC. Similar to TWC, Karaj Dam inflow was also modeled with SARIMA models, which the same model, SARIMA$(1,0,0)(0,1,1)_{12}$ was obtained for Karaj Dam inflow prediction. The diagnostic check was done for the residuals of the SARIMA models for TWC and Karaj Dam inflow. The results of the SARIMA models were applied as the input data of the Karaj Dam SD model for the simulation of the Karaj Dam water supply system from April 2006 to March 2023. Finally, to evaluate the precision of the results obtained from the SD model, the criteria of R^2, E%, and NS were calculated. The SD model results represented that the reservoir volume of Karaj Dam will be decreased from April 2023 to March 2023 because of the Karaj Dam inflow reduction, consequently the priority usage of Karaj Dam as the household water supply is provided as before, but the agricultural usage is reduced due to the dependence of agricultural water allocation to Karaj Dam inflow.

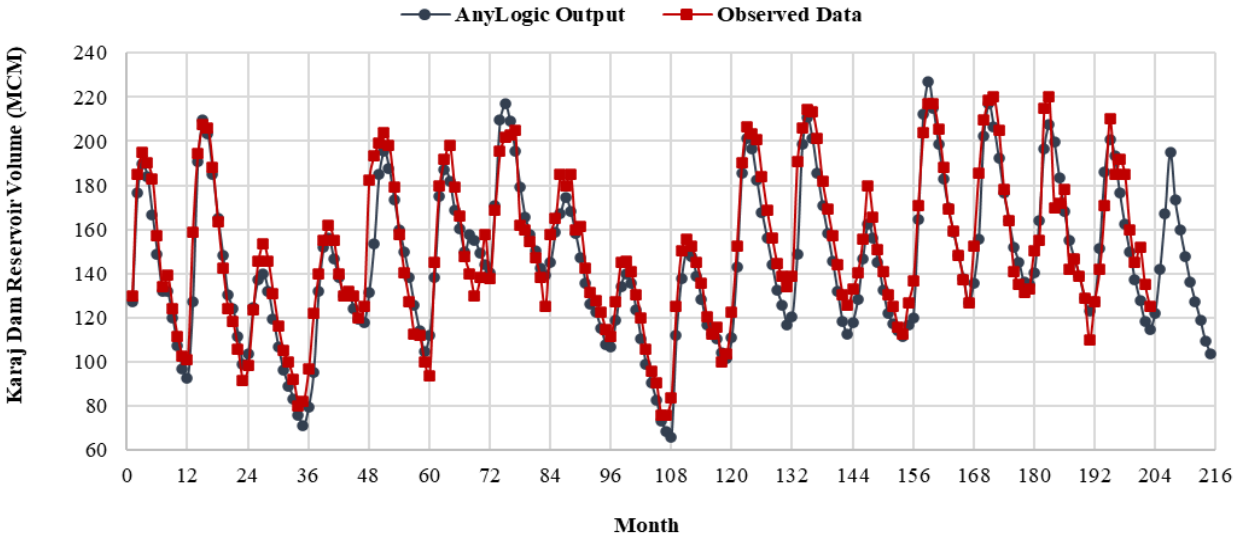


Figure 12. Observed and calibrated AnyLogic model output of Karaj Dam Reservoir Volume

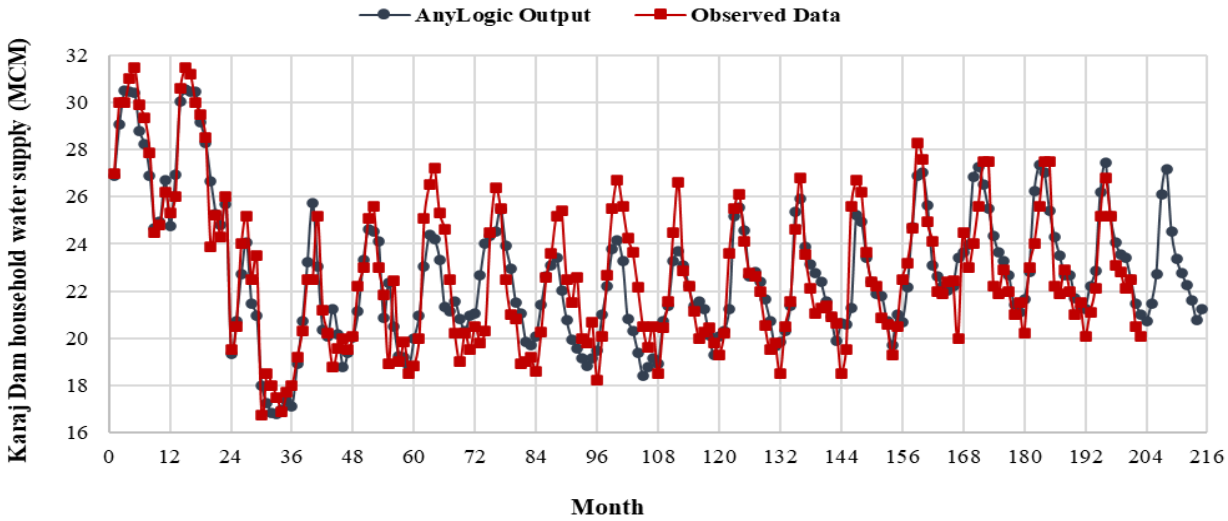


Figure 13. Observed and calibrated AnyLogic model output of Karaj Dam household water supply

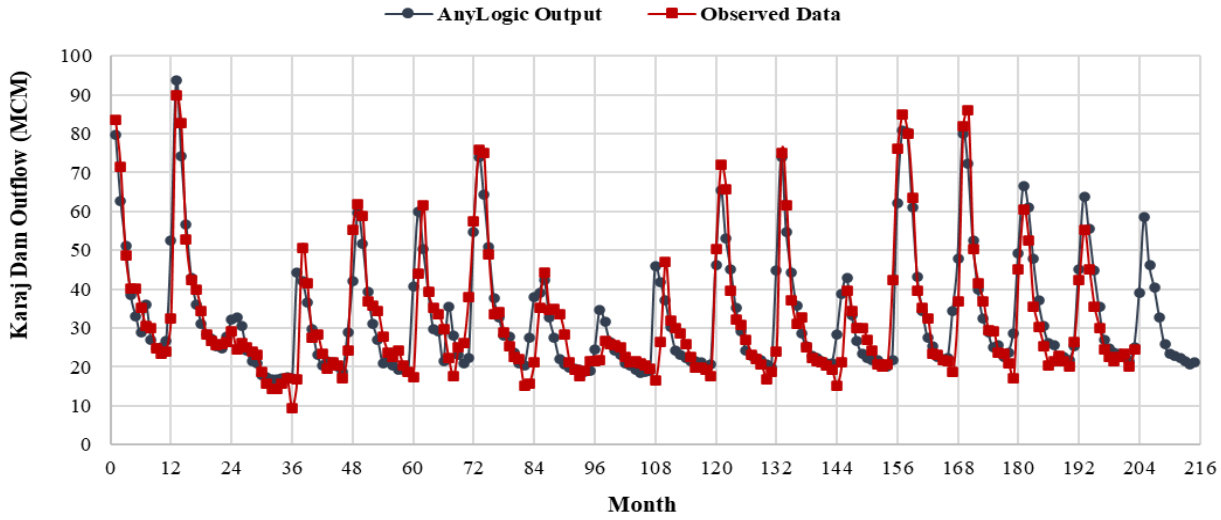


Figure 14. Observed and calibrated AnyLogic model output of Karaj Outflow

Reference

[1] B.A. Bryan, Future global urban water scarcity and potential solutions, Nat. Commun. (2021) 1–11. https://doi.org/10.1038/s41467-021-25026-3.

[2] O. Bozorg-Haddad, P. Dehghan, B. Zolghadr-Asli, V.P. Singh, X. Chu, H.A. Loáiciga, System dynamics modeling of lake water management under climate change, Sci. Rep. 12 (2022) 5828.

[3] H. Sarmadi, E. Salehi, N. Kusari, The mega city of Tehran water quantity assessment based on DPSIR model, J. Phys. Conf. Ser. 1834 (2021) 0–6. https://doi.org/10.1088/1742-6596/1834/1/012007.

[4] R. Modarres, V. de Paulo Rodrigues da Silva, Rainfall trends in arid and semi-arid regions of Iran, J. Arid Environ. 70 (2007) 344–355. https://doi.org/10.1016/j.jaridenv.2006.12.024.

[5] worldpopulationreview, No Title, (n.d.) https://worldpopulationreview.com/search?query=teh.

[6] M. Zomorodian, S.H. Lai, M. Homayounfar, S. Ibrahim, S.E. Fatemi, A. El-Shafie, The state-of-the-art system dynamics application in integrated water resources modeling, J. Environ. Manage. 227 (2018) 294–304. https://doi.org/10.1016/j.jenvman.2018.08.097.

[7] H. Nozari, P. Moradi, E. Godarzi, Simulation and optimization of control system operation and surface water allocation based on system dynamics modeling, J. Hydroinformatics. 23 (2021) 211–230. https://doi.org/10.2166/HYDRO.2020.294.

[8] J.M. García, Theory and practical exercises of system dynamics: modeling and simulation with Vensim PLE. Preface John Sterman, Juan Martin Garcia, 2023.

[9] T.D. Phan, E. Bertone, R.A. Stewart, Critical review of system dynamics modelling applications for water resources planning and management, Clean. Environ. Syst. 2 (2021) 100031. https://doi.org/10.1016/j.cesys.2021.100031.

[10] A. Bafkar, J. Mozafari, H. Alizadeh, Determination of Optimal Irrigation Water Supply Scenario for Karkheh Dam to Prevent Drainage Problems of Dashte Abbas Plain Using System Dynamics Approach, J. Agric. Sci. Technol. 24 (2022) 707–722.

[11] A.F. Abdolvandi, S.S. Eslamian, M. Heidarpour, H. Babazadeh, Armaghanparsamehr, Simultaneous simulation of both surface and groundwater resources using system dynamics approach (Case study: Taleghan dam), Adv. Environ. Biol. 7 (2013) 562–570.

[12] Z. Li, C. Li, X. Wang, C. Peng, Y. Cai, W. Huang, A hybrid system dynamics and optimization approach for supporting sustainable water resources planning in Zhengzhou City, China, J. Hydrol. 556 (2018) 50–60. https://doi.org/10.1016/j.jhydrol.2017.11.007.

[13] E. Bakhshianlamouki, S. Masia, P. Karimi, P. van der Zaag, J. Sušnik, A system dynamics model to quantify the impacts of restoration measures on the water-energy-food nexus in the Urmia lake Basin, Iran, Sci. Total Environ. 708 (2020) 134874. https://doi.org/10.1016/j.scitotenv.2019.134874.

[14] A. Babolhakami, M.A. Gholami Sefidkouhi, A. Emadi, Application of system dynamics model for reservoir performance under future climatic scenarios in Gelevard Dam, Iran, AQUA — Water Infrastructure, Ecosyst. Soc. 00 (2023) 1–15. https://doi.org/10.2166/aqua.2023.193.

[15] K.B. Tadesse, M.O. Dinka, Application of SARIMA model to forecasting monthly flows in Waterval River, South Africa, J. Water L. Dev. 35 (2017) 229–236. https://doi.org/10.1515/jwld-2017-0088.

[16] M. Imran, M.D. Majeed, M. Zaman, M.A. Shahid, D. Zhang, S.M. Zahra, R.M. Sabir, M. Safdar, Z. Maqbool, Artificial Neural Networks and Regression Modeling for Water Resources Management in the Upper Indus Basin, (2023) 53. https://doi.org/10.3390/ecws-7-14199.

[17] H. Allah, M. Sadegh, Estimation of Water Demand in Iran Based on SARIMA Models, (2013) 559–565. https://doi.org/10.1007/s10666-013-9364-4.

[18] A.S. Azad, R. Sokkalingam, H. Daud, S.K. Adhikary, H. Khurshid, S.N.A. Mazlan, M.B.A. Rabbani, Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study, Sustain. 14 (2022). https://doi.org/10.3390/su14031843.

[19] M. Valipour, M.E. Banihabib, S.M.R. Behbahani, Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir, J. Hydrol. 476 (2013) 433–441. https://doi.org/10.1016/j.jhydrol.2012.11.017.

[20] W. chuan Wang, K. wing Chau, D. mei Xu, X.Y. Chen, Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition, Water Resour. Manag. 29 (2015) 2655–2675. https://doi.org/10.1007/s11269-015-0962-6.

[21] R.M. Adnan, X. Yuan, O. Kisi, V. Curtef, Application of time series models for streamflow forecasting, Civ. Environ. Res. 9 (2017) 56–63.

[22] I.M. Almanjahie, Z. Chikr-Elmezouar, A. Bachir, Modeling and forecasting the household water consumption in Saudi Arabia, Appl. Ecol. Environ. Res. 17 (2019) 1299–1309. https://doi.org/10.15666/aeer/1701_12991309.

[23] P. Aghelpour, V. Varshavian, Evaluation of stochastic and artificial intelligence models in modeling and predicting of river daily flow time series, Stoch. Environ. Res. Risk Assess. 34 (2020) 33–50. https://doi.org/10.1007/s00477-019-01761-4.

[24] X. Wang, W. Tian, Z. Liao, Statistical comparison between SARIMA and ANN's performance for surface water quality time series prediction, Environ. Sci. Pollut. Res. 28 (2021) 33531–33544. https://doi.org/10.1007/s11356-021-13086-3.

[25] M.M.R. Tabari, R. Safari, Development of water re-allocation policy under uncertainty conditions in the inflow to reservoir and demands parameters: a case study of Karaj AmirKabir dam, Soft Comput. 27 (2023) 6521–6547. https://doi.org/10.1007/s00500-023-07885-8.

[26] X. jun Wang, J. yun Zhang, J. fu Liu, G. qing Wang, R. min He, A. Elmahdi, S. Elsawah, Water resources planning and management based on system dynamics: A case study of Yulin city, Environ. Dev. Sustain.13(2011),331–351.

https://doi.org/10.1007/s10668-010-9264-6.

[27] F. Nasiri, T. Savage, R. Wang, N. Barawid, J.B. Zimmerman, A system dynamics approach for urban water reuse planning: A case study from the Great Lakes region, Stoch. Environ. Res. Risk Assess. 27 (2013) 675–691. https://doi.org/10.1007/s00477-012-0631-8.

[28] L. Willuweit, J.J. O'Sullivan, A decision support tool for sustainable planning of urban water systems: Presenting the dynamic urban water simulation model, Water Res. 47 (2013) 7206–7220. https://doi.org/10.1016/j.watres.2013.09.060.

[29] W. Zeng, B. Wu, Y. Chai, Dynamic simulation of urban water metabolism under water environmental carrying capacity restrictions, Front. Environ. Sci. Eng.10,(2016),114–128. https://doi.org/10.1007/s11783-014-0669-6.

[30] A. Kumar Dubey, A. Kumar, V. García-Díaz, A. Kumar Sharma, K. Kanhaiya, Study and analysis of SARIMA and LSTM in forecasting time series data, Sustain. Energy Technol. Assessments. 47 (2021) 101474. https://doi.org/10.1016/j.seta.2021.101474.

[31] W.M. Thupeng, K. Samakokore, Modelling Monthly water consumption for Gaborone using log-logistic and related distributions, (n.d.).

[32] P.P. Dabral, M.Z. Murry, Modelling and Forecasting of Rainfall Time Series Using SARIMA, Environ. Process. 4 (2017) 399–419. https://doi.org/10.1007/s40710-017-0226-y.

[33] A. Chakrabarti, J.K. Ghosh, AIC, BIC and Recent Advances in Model Selection, Elsevier B.V., 2011. https://doi.org/10.1016/B978-0-444-51862-0.50018-6.

[34] S.O. Adams, M. Bamanga, M. Ardo Bamanga, Modelling and Forecasting Seasonal Behavior of Rainfall in Abuja, Nigeria; A SARIMA Approach, Am. J. Math. Stat. 2020 (2020) 10–19. https://doi.org/10.5923/j.ajms.20201001.02.

[35] D.K.D. et. al. . D. K. Dwivedi et.al., Forecasting Mean Temperature using Sarima Model for Junagadh City of Gujarat, Int. J. Agric. Sci. Res. 7 (2017)183–194. https://doi.org/10.24247/ijasraug201723.