Effect of Dam Construction on Lake Urmia: Time Series Analysis of Water Level via ARIMA

M. Ghashghaie¹, and H. Nozari¹*

ABSTRACT

Lake Urmia is one of the water bodies that face severe drought conditions nowadays. Therefore, the present study aimed to study monthly time series of the lake water level. Modeling the series was accomplished in two ways. First, all data were used to analyze the water level time series of the lake. Although the results of generating were quite well, the results of validity test were not satisfying. Second, only water level data after the year 1995 were used, which showed a continuously decreasing trend. These data start from the year when dam constructions and operations in Urmia Watershed were increased one by one. The values of $R^2$ and RMSE were 0.99, 0.001 m, respectively, for generating the data at this stage. These values were 0.93 and 0.03 m for the validity test of the model (from 2005 to 2009). The results of our study show that the lake water level behavior changed after 1995 due to constructing many dams in Urmia Lake Watershed.

Keywords: Drought condition, Modeling water level, Stochastic hydrology, Urmia Watershed.

INTRODUCTION

Time series analysis has been applied in different researches (Jadhav et al., 2017; Zhang et al., 2016). Also, it has been widely used in studies of water resources management. These researches demonstrate the efficiency and relevance of this kind of modeling due to considering the stochastic nature of hydrological processes such as discharge, temperature, precipitation, and water quality parameters.

Also, many studies have been conducted based on lake water level using time series analysis. Lakes are known as very important environments and natural ecosystems. The study of water level of lakes is important from two aspects. First, we can present a model capable of forecasting the water level, which increases our knowledge of its stochastic behavior. Secondly, time series analysis of water level identifies its reactions to anthropogenic and natural phenomena.

Barry et al. (1976) determined short and long term periodicity of great lakes water level such as: Ontario, Erie, Huron-Michigan and Superior. Guganesharaja and Shaw (1984) applied two models for predicting low levels of Chad Lake to support the planning of agricultural operation and describe the probability of future variations. The probabilistic model is based on the Markov chain process that was used to measure the risk involved in selecting various abstraction levels of water from the lake during the proposed life span of the project. Khavich and Ben-Zvi (1995) used a model based on water budget equation to forecast daily changes of water levels in Lake Kinneret during flood periods. Privalsky (1996) studied monthly and yearly water levels of Lake Erie at Cleveland, Ohio, from 1860 to 1988. He tried to estimate the attainable quality of least square predictions of water level using the

¹ Department of Water Science and Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Islamic Republic of Iran.
*Corresponding author; e-mail: hanozari@yahoo.com
AutoRegressive (AR) method and concepts of non-stationary random processes. Sen et al. (2000) used simple linear and periodic nonlinear models for modeling deterministic part of lake level time series and a second-order Markov model for the stochastic part. They developed a cluster regression model to deal with non-stationarity, especially in shifting means.

Also, LaValle et al. (2001) used first-order AutoRegressive process [AR (1)] to model short term fluctuations of Lake Erie to prove that Erie’s fluctuations occur in response to two ENSO phases such as El Nino and La Nina. Altunkaynak et al. (2003) introduced a Triple Diagram Model (TDM) based on the Kriging technique to forecast lake levels. Li et al. (2007) examined four models to forecast Great Salt Lake level: ARMA (AutoRegressive Moving Average), ARFIMA (Auto-Regressive Fractional Integral and Moving Average), GARCH (Generalized Auto-Regressive Conditional Heteroscedasticity) and FIGARCH (Fractional Integral Generalized Auto-Regressive Conditional Heteroskedasticity). They found that FIGARCH indicated the best performance offering that conditional heteroscedasticity should be included in time series with high inconsistency.

Sellinger et al. (2008) studied water level data (1860–2006) representing Lakes Michigan and Huron to evaluate changes in both long-term and seasonal patterns over time. They used Seasonal Trend decomposition using Loess (STL), and Dynamic Linear Models (DLM). Also Talebizadeh and Moridnejad (2011) developed ANN and ANFIS models to forecast Lake Urmia fluctuations. Rainfall, evaporation, and inflow time series were used as input variables to the models. Aksoy et al. (2013) studied water level of Lake Van using ARIMA modeling. They first investigated the series considering one trend for all data, and then multiple trends of time series were taken into account for modeling the series. Domenico et al. (2013) compared chaos theory and ARIMA techniques to model sea level for daily, weekly, 10- day, and monthly time scales. ARIMA technique was quite capable of modeling the levels, except for monthly data.

Lake Urmia is the second largest hypersaline lake in the world (Shadkam et al., 2016; Sima et al., 2012). Both natural and anthropogenic factors have affected the lake level. Although many studies have been accomplished on the meteorological parameters of the lake such as temperature and precipitation (Mahsafar et al., 2011; Delju et al., 2012; Fathian and Morid, 2012; Fathian et al., 2013), few studies have investigated human driven factors of the lake level decline.

Fluctuations time series of a lake water level imply many facts about its watershed. It shows how human beings activities have affected this water body during the past. Also, it informs us about what is going on in a very important natural ecosystem. Different methods have been used to study and model Lake Urmia fluctuations (Talebzadeh and Moridnejad, 2011; Mahsafar et al., 2011; Jalili et al., 2011), and this study adopted time series analysis method.

The study aimed to apply time series modeling and use ARIMA or Box-Jenkins modeling to analyze the behavior of the lake water level. It also aimed to show the efficiency of this method in capturing deterministic component on a hydrologic time series to provide information on the impact of managerial strategy on this water resource.

MATERIALS AND METHODS

Study Area

Lake Urmia is located in northwest of Iran, 45º E–46º E and 37º 4´ N–38º 17´ N, at 1250 m asl (Figure 1). Different species benefit from this ecosystem as their territory. This lake also plays an important role in various socio-economic activities besides its ecological importance. However, it has faced many environmental problems,
especially during the recent years. Due to several factors such as droughts, overuse of surface water resources, and dam constructions, water level has decreased such that one quarter of the lake has converted to a saline land during the past 10 years. Also, as a result of drought and increased demands for agricultural water in its basin, the salinity of the lake has risen to more than 300 g L\(^{-1}\) in recent years, and large areas of the lake bed have been dried out (Eimanifar and Mohebbi, 2007).

Lake Urmia with an area of 5,000–6,000 km\(^2\) is the largest lake in Iran. In recent decade, the lake’s water level has been decreased up to 6 m (Water Research Institute, 2006). Precipitation, surface water, and groundwater are inflows to the lake. Nearly 20 perennial and seasonal rivers and springs supply the lake (Rasuly, 2005). Because of the special topography and poor quality of the Urmia Lake’s water, it is not suitable to convey and use for other purposes. In addition, by studying the groundwater gradient maps in the region’s plains, no considerable amount of groundwater flows from the lake to the nearby aquifers in the region. Therefore, the only outflow from the lake is supposed to be evaporation (Water Research Institute, 2003; Delavar, 2006). Figure 2 shows the lakes variations from 1990 to 2015. It could imply the outcome of dam constructions and mismanagement of water resources in Urmia Lake Watershed.

Figure 3 also supports this assumption that operating too many dams improperly has endangered the life of the lake. During the last 18 years, there has been a severe decline in water level of the lake. The starting point of this period is coincident with increasing the number of dams in the watershed. Figure 3 shows the variations of the lake water level during a century, adopted from different sources (Pengra, 2012)

Figure 4 shows the trends of precipitation and other inflows to the lake and its water level, where \(\frac{Q}{Q_{\text{mean}}}\) and \(\frac{R}{R_{\text{mean}}}\) represent the ratio of inflow to monthly mean inflow and precipitation to monthly mean precipitation, respectively. \(Y_L\) shows the trend line of the lake level. Also, \(Y_Q\) and \(Y_R\) show the equations for \(\frac{Q}{Q_{\text{mean}}}\) and \(\frac{R}{R_{\text{mean}}}\), respectively. The trend of precipitation is slightly increasing with a positive gradient (\(Y_R\) in Figure 4), while inflow and water level show a decreasing trend.

Also Fathian et al. (2013) studied the trends of hydro-climatic time series variables of the Lake Urmia Basin as well as changes in land use of the basin, using satellite images. They showed that the correlation between stream flow changes with simultaneous changes in climatic variables and land use was significant in
Figure 2. Urmia Lake changes over time Landsat (Google earth).

Figure 3. Water level variations of Urmia Lake during a century (Pengra, 2012).

Figure 4. Precipitation, inflow, and water level trends of Lake Urmia.
the case of temperature and land use; but insignificant in the case of precipitation. Also, the determination coefficient of land use was higher than temperature.

Generally, the reasons of any decrease in the amount of inflow to the lake can be investigated considering three factors:
1) Climate change
2) Water resource development projects
3) Water overuse in upstream (Hassanzadeh et al., 2012).

Typically for data analysis, observations are supposed to be independent, but in time series studies, temporal correlation is accounted for statistical analyses. Time series analysis helps to understand the structure of the series through finding the following characteristics:
- Any serial correlation.
- Any trend in the data over time
- Any seasonal variation in the data over time
- Suitability of the data to be used for forecasting the future observations

In this study, at first the data were plotted against time. The main goal of a time series analysis might be to understand seasonal changes and trends over time. Trends and seasonal variations are often evident in time plots. Another goal that is often of primary importance is to find out and model the correlational structure in the time series for generating new datasets. A monthly time series of the lake water level including 522 months (1965-2009) was used in this study for modeling and forecasting.

Although time series modeling originated from different scientific fields, it has demonstrated its capability and reliability in stochastic hydrology. Applications of time series analysis in water resources management are important.

Box and Jenkins (1976) innovated the basis of modern hydrologic stochastic modeling Equation (1):
\[ y_{t} = f(x_{t}, x_{t-1}, ..., y_{t-1}, y_{t-2}, ..., 1, 2, ...)+t \]  \hspace{1cm} (1)

Where, \( f \) shows the obtained mathematical function; \( y_{t} \) is the predicted output at time \( t \); \( y_{t-1}, y_{t-2}, \) and ..., denote the successive members of the output time series recorded at corresponding time intervals \( t, t-1, t-2; \) \( x_{t}, x_{t-1}, x_{t-2}, ... \), are the successive members of the input time series recorded at time intervals \( t, t-1, t-2; \) \( 1, 2, ... \), show the model parameters found by mathematically minimizing the differences between estimated (calculated) and observed \( y_{t} \) values; \( t \) is the model error (residual) given as the difference between the calculated and the recorded value of the output series at time \( t \).

Box and Jenkins (1976), also proposed ARMA models. The mathematical formulation of ARMA models is written as Equation (2):
\[ Z_{t} = \sum_{j=1}^{p} \phi_{j}Z_{t-j} - \sum_{j=0}^{q} \theta_{j}t - f + t \]  \hspace{1cm} (2)

Where, \( Z \) represents the time dependent variable with mean zero and variance one; \( 1, ..., p \) are time varying autoregressive coefficients; \( 0, ..., q \) denote time varying moving average coefficients and \( t \) is an independent normal variable.

Time series models are classified into AutoRegressive models [AR (p)], Moving Average models [MA (q)], and their combination ARMA (p, q) with variations, such as ARIMA models (p, d, q) and others, where p and q are the orders of autoregressive and moving average terms, respectively, and "d" denotes an order of differencing.

An AutoRegressive (AR) model estimates values for the dependent variable, \( Z_{t} \), as a regression function of previous values \( Z_{t-1}, Z_{t-2}, ..., Z_{t-n} \). A moving average model is conceptually a linear regression of the current value of the series against the white noise or random shocks of one or more prior values of the series. AR model, which is called a Thomas-Fiering model, has been applied extensively in hydrology for modeling annual and periodic hydrologic time series. AR models basically estimate values for the dependent variable, \( Z_{t} \), as regression function of previous values, \( Z_{t-1}, Z_{t-2}, ..., Z_{t-n} \). The first-order AR model [AR (1)] can be expressed as Equation (3):
\[ Z_{t} = \phi_{1}Z_{t-1} + a_{t} \]  \hspace{1cm} (3)
Where, $Z_t$ and $Z_{t-1}$ are the deviations from the mean of the time series, $\Phi_1$ shows the first-order AR coefficient describing the effect of a unit change in $Z_{t-1}$ on $Z_t$, and $\alpha_t$ represent random shock errors or white noise. Model stationarity requires that the variance of $Z_t$ be non-negative and finite (Vandaele, 1983).

Moving Average (MA) models incorporate past random fluctuations to represent the time series and the first order MA model [MA (1)] can be expressed as Eq. (4):

$$Z_t = \alpha_t - \theta_1 \alpha_{t-1}$$

Where, $\theta_1$ denotes the MA coefficient to be estimated and the random shocks ($\alpha_t$) are assumed normally and independently distributed with mean 0 and constant variance. The model structure requires the condition of reversibility to be met and $|\theta_1|$, therefore, must be less than 1.

In the modeling stage, at first, time series of the lake water level are plotted against time to find out any possible trend line of data series. Presence of a deterministic component like trend or jump in a natural time series reveals many facts about possible factors that influence the behavior of a system. Hence, the deterministic component must be treated accurately. Trend is one of the main non-stationary factors which affect time series gradually. Many factors inside and outside of the system could create trends. Sometimes current strategies must be changed to diminish the stress on natural resources like water. Results of modeling of the series are presented in the next section.

**RESULTS AND DISCUSSION**

At first, time series of monthly water levels were plotted. Figure 5 shows the time series of lake water level from 1965 to 2009. At the modeling stage, time series data were examined for normality of monthly data. All series showed a normal distribution at $\alpha = 0.05$. Then, all series were standardized by subtracting the mean and dividing by the standard deviation. The data from 1965 to 1997 were used for generating and from 1998 to 2009 were used for forecasting.

Next, ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function) of time series were calculated. Figure 6 shows the ACF and PACF of monthly time series. This figure shows that the ACF is infinite,
which suggests the AR (p) as a good model for the series. The order of AR is determined through PACF diagram. Based on the ACF and PACF of data p=2, q=0 and d=1 is suggested to generate data. Table 1 demonstrates the result of data generation using 3 models.

Akaike Information Criterion (AIC), Determination Coefficient ($R^2$), Root Mean Square Error (RMSE), and MAE (Mean Absolute Error) are four factors which were used to evaluate generation and forecast results. The values of AIC and MAE were calculated by Equation (5).

$$AIC = n \times \ln(\sigma^2) + 2 \times (p + q)$$  \hspace{1cm} (5)

Where, $n$ denotes the number of data; $\sigma$ is the standard deviation of residuals; $p$ and $q$ show the order of the model.

The value of RMSE is calculated by Equation (6).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}{n}}$$  \hspace{1cm} (6)

Also MAE is calculated through Equation (7).

$$MAE = \frac{\sum_{i=1}^{n}|y_t - \hat{y}_t|}{n}$$  \hspace{1cm} (7)

Table 1 shows that the AR (2, 1, 0) model is a convenient model to forecast water level of the lake based on the AIC criteria. The normality of residuals was controlled as well as their independence. The ACF and PACF of residuals for (2, 1, 0) model is shown in Figure 7, which demonstrates that the residuals are independent.

The last 140 data points of total time series were used to control validity of the model through forecasting step. At this step, the results of forecasting were not convenient in spite of the generating step. The values of $R^2$ and RMSE were, respectively, 0.61 and 0.35 m.

Then, we decided to analyze only the last 166 data. This part of monthly data shows an absolutely decreasing trend. It starts with the year 1995, from which the number of operating dams was increased gradually. Figure 8 demonstrates two segments of water level. This figure shows that the second part has a significant decreasing trend.

Figure 9 shows the selected segment (1971-2006) for precipitation and inflow as

![Figure 6](image_url)  
**Figure 6.** The ACF and PACF of monthly time series.

<table>
<thead>
<tr>
<th>$(p, d, q)$</th>
<th>AIC</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2,1,0)</td>
<td>-2159</td>
<td>0.99</td>
<td>0.004</td>
<td>2.1</td>
</tr>
<tr>
<td>(2,0,0)</td>
<td>-2140</td>
<td>0.99</td>
<td>0.004</td>
<td>1.7</td>
</tr>
<tr>
<td>(3,0,0)</td>
<td>-2157</td>
<td>0.99</td>
<td>0.004</td>
<td>1.9</td>
</tr>
</tbody>
</table>

In both equations, $y_t$ is the observed data; $\hat{y}_t$ shows the estimated values, and $n$ denotes the number of data.
well as lake water level variations. Recall that $Q/Q_{\text{Mean1}}$ and $R/R_{\text{Mean1}}$ denote the ratios of inflow to monthly mean inflow and precipitation to monthly mean precipitation for the first part, respectively. These ratios for the second part are shown by $Q/Q_{\text{Mean2}}$ and $R/R_{\text{Mean2}}$. Also, Lake Level$_1$ and lake Level$_2$ show variations of the lake water level for the first and the second segment, respectively.

In Figure 9, the trends of the second part of these series are shown by $Y_L2$, $Y_Q2$, and $Y_R2$, which denote the trend equation for the lake water level, inflow to the lake, and precipitation, respectively.

The second segment was examined and modeled separately. The data from 1995 to 2005 were used for generating and from 2005 to 2009 were used for forecasting.

The ACF and PACF of the series (1971-2006) were calculated to suggest the suitable order of $p$ and $q$. Figure 10 shows the ACF and PACF of the second series.

Using one difference, the trend of data was removed for the second segment. The best model for generating this part of data was adopted based on the results shown in Table 2.

As clearly seen in the Table, (2, 1, 0) model is the best choice. Using these orders of $p$, $d$ and $q$, validity of the model was investigated using the last 48 data.

Normality of residuals was examined as well as their independence. The ACF and PACF of residuals for the selected model (2, 1, 0) are presented in Figure 11.
Figure 9. The trends of precipitation, inflow, and lake level for the second segment (1971-2006).

Table 2. The results of the second part generation.

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>R²</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,1,0)</td>
<td>-679.8</td>
<td>0.997</td>
<td>0.005</td>
<td>0.64</td>
</tr>
<tr>
<td>(2,1,0)</td>
<td>-684.4</td>
<td>0.997</td>
<td>0.005</td>
<td>0.54</td>
</tr>
<tr>
<td>(3,1,0)</td>
<td>-682.9</td>
<td>0.997</td>
<td>0.005</td>
<td>0.58</td>
</tr>
</tbody>
</table>

The results of forecasting show that the model is quite capable of forecasting the series. The values of $R^2$ and RMSE are 0.93 and 0.03 m. Figure 12 shows the results of modeling for two stages. Also Figure 13 compares the results of modeling using all data and the second segment in a plot. This figure shows clearly the accuracy of estimating for both approaches.

CONCLUSIONS

Almost 70 percent of global net water loss occurs in the Middle East and central Asia, which are related to drought and human activities such as dam construction and excess withdrawal (Pekel et al., 2016). Furthermore, water level decline of Urmia Lake cannot be attributed merely to climate
change as watershed management played a crucial role in this problem (Shadkam et al., 2016; Sima et al., 2012).

This research tried to find the most suitable stochastic model for time series of Urmia Lake water level, which shows a normal and comprehensible behavior until 1995. After this time, a continuous sharp drop is observed in water level for 18 years. The study was accomplished in two ways. At first, the whole series was modeled. Although it demonstrated a suitable model for generating data, the results were not satisfying for the validity test of the model. Then, the series was divided into two segments. The first part showed a slightly increasing trend, while the second segment showed a significant decreasing trend. This segment started at 1995 when the number of operating dams in Urmia Lake Watershed was rapidly increased. The outcome of generating was quite well. The values of $R^2$ and RMSE were 0.99 and 0.001 m, respectively. Moreover, the values of $R^2$ and RMSE for the validity test of the model were 0.93 and 0.03 m, respectively. The results show that the second segment can better imply the behavior of time series.

Based on the information from Water Resources Management Company, 79 dams have been built in Urmia Lake Watershed since 1995, while the number of dams in this watershed was 34 before this year. Furthermore, in Figure 4, it is shown that the value of discharge has a decreasing trend, while the trend of precipitation is increasing in this watershed. This could be related to constructing too many dams in watershed.

---

**Figure 12.** The results of modeling the second segment.

**Figure 13.** The comparison between the results of the first and second approaches.
such that they negatively affected the value of discharge to the lake.

The results draw the attention to the point that, due to human activities, the variations in the Urmia Lake water level does not follow a normal behavior. Consequently, the study of Urmia Lake water level should focus on water management that has a greater effect on the behavior of the lake water level (Zafarnejad, 2009; Pekel et al., 2016; Shadkam et al., 2016; Sima et al., 2012). Therefore, an efficient management policy is necessary to operate all dams in the watershed and save the lakes environment.

ACKNOWLEDGEMENTS

We would like to express our sincere thanks to Dr. Hafzullah Aksoy and Dr. Ebru Eris for their valuable comments.

REFERENCES

تاثیر ساخت سد بر دریاچه ارومیه: تجزیه و تحلیل سری زمانی تراز سطح آب به
وسیله مدل ARIMA

م. قشقایی، و. ح. نوذری

چکیده
دریاچه ارومیه یکی از پیکرهای آبی است که امروزه با شرایط خشکسالی شدید مواجه است. بنابراین، هدف از مطالعه حاضر بررسی سری زمانی سطح آب ماهانه دریاچه می‌باشد. مدل سازی سری به دو صورت انجام شد. ابتدا تمام داده‌ها موجود در سری زمانی برای تجزیه و تحلیل سطح آب دریاچه مورد استفاده قرار گرفت. اگرچه نتایج حاصل از مرحله تولید بسیار خوب بود اما نتایج آزمون اعتبارسنجی رضایت بخش نبود. در مرحله دوم داده‌های سطح آب فقط پس از سال 1391 استفاده شد که یک روند کاملاً کاهشی را نشان می‌دهد. این داده‌ها از سالی که احداث و بهره‌برداری از سده‌ها در حوضه آبریز ارومیه، یکی پس از دیگری افزایش می‌یابد، شروع می‌شود. مقادیر RMSE و R² در مرحله تولید داده‌ها به ترتیب 99/00 و 0/0001 متر بود. برای اعتبارسنجی مدل (از 0/5 تا 0/9 تا 0/1 متر)، مقادیر به ترتیب برابر با 99/00 و 0/003 متر بود. نتایج مطالعه نشان می‌دهد که رفتار سطح آب دریاچه پس از سال 1395 به دلیل ساخت سد‌های متعدد در حوضه دریاچه ارومیه تغییر می‌کند.