Using Arima Time Series Model in Forecasting the Trend of Changes in Qualitative Parameters of Sefidrud River

Mina Ranjbar¹, Mohammadreza Khaledian²

1. MA Graduate in the field of Environmental Pollutions, Islamic Azad University, Tonekabon Branch,
2. Faculty Member of Irregation Education Dept., University of Guilan, E-mail: mkh572000@yahoo.com

Corresponding Author email: RanjbarM1982@yahoo.com

ABSTRACT: Sefidrud River is considered as the biggest river in northern Iran which is undergoing many changes due to increased water needs in agricultural, industrial and other sectors as well as increased pollution compiled by such consumptions. In this paper, ARIMA Model (2.0.0) was used to forecast the trend of changes in TSS, DO and NO₃ parameters of two stations of Sefidrud River in Astaneh and Manjil statistical year of 2010 separately at each station. The regression coefficient of actual and fitted values for each parameter were 85%, 74% and 75% in Astaneeh station and 78%, 80% and 82% in Manjil station, respectively. To test the forecasting conducted, we also used the data of year 2011 for Astaneh station and data of year 2012 for Manjil station, which emphasized on good and relatively good performance of the presented model.

Keywords: Sefidrud, Time Series, Qualitative Parameters, ARIMA.

INTRODUCTION

Rivers are among the important water ecosystems that have a great importance in terms of agriculture, environment, fisheries, economics and so on. Because, originating from the highlands and crossing through plains area and finally flowing into the wetlands, lakes, seas and oceans, these ecosystems have affected a wide area, whilst they are affected as well. Today, one of the existing ravages in the world is pollution of these valuable resources and, necessity of considering them and controlling pollution in water resources management has become particularly important, so that in applied planning, these important economic resources have a great importance in the future. Evaluation of temporal variations of water quality in rivers and amount of influence due to entry of pollutants resulting from human and natural activities are among the purposes of evaluating process of quality of rivers water.

Sefidrud River with an area of about 67,000 square kilometers is one of the largest and the most riviers in the country and is considered as one of the most important rivers in Caspian Basin. It is also considered as one the most important renewable sources of fresh water in order to be used in agricultural, industrial and drinking affairs. The study area is located in coastal plain of south of Caspian and on the northern slopes of the Alborz and Talesh mountains, from Sefidrud location (Manjil) to the Caspian Sea, between 49 degrees and 15 minutes to 50 degrees 15 minutes of eastern longitude and 36 degrees 45 minutes to 37 degrees 30 minutes north latitude. Sefidrud River, from GhezelOzan source to its mouth (the Caspian Sea), is about 786 km. This basin includes some parts of Guilan, Ardebil, Eastern Azarbaijan, Zanjan, Ghazvin, Kordestan, Tehran, Hamedan and Western Azarbaijan provinces. Necessity for taking into consideration and control of the pollutions entering the River has a great importance in its management.

The main objective of time series analysis is to find the changes model and forecasting. Forecasting time series means to estimate some values of data set which is unknown at the time of analysis. One of the techniques to forecast time series behavior is the Box-Jenkins method. The Box-Jenkins approach is based on examining a wide range of models for forecasting a time series. In the hydrology as rainfall and flowing water we have to deal with random processes. The random process variables set may be dependent or independent. If only non-zero
values are considered, the series would be from non-periodic type and otherwise it would be periodic. Qualitative parameters of river in daily, monthly and annual aspects are of this type of series. On the other hand, if the probability rules governing the process do not change over time, it is a static series and this statics may be defined in mean, variance and covariance. For example, in topics of hydrology, climate changes of land in an unknown period as rainy years followed by successive droughts make time series of quality parameters to be non-stationary. One of the basic terms to use data in time-series topics, is that they are static, otherwise non-stationary should be removed (6).

In order to model the time series, in most cases, the professionals use different kinds of mathematical and statistical models so that correlation between time and observations is considered (6). One of the techniques to forecast the time series behavior is the Box–Jenkins method. Box–Jenkins approach is based on an extensive review of a wide range of models for time series prediction. General group of models for a time series in Box–Jenkins methodology, are the combined model of autoregressive and moving which are known as ARIMA Model in statistics (2).

The model is written in the form of ARIMA (p,d,q). In this form, p which determines the autoregressive or AR rank, indicates dependence of a parameter to its previous effective amounts in current time. In this model, regression of each element is determined based on its previous values. Q also determines the average moving by which dependence of the series to the random elements of its present and past is defined. In the Box–Jenkins and ARIMA model, there are three major stages, each of them may be used multiple times: 1- recognition of model, 2- fitting of model, 3-diagnosis the model accuracy (2).

Several studies to predict quality of river water through random methods in time series analysis were conducted by different researchers. Dordoi (2009) predicted boron concentrations in hydrology stations of Buick River in Turkey using ARIMA and SARIMA models for a period of 36 months (11). Ben Yahya et al (2007) modeled weekly time series for maximum temperature of Descgutes River using the autocorrelated model (AR) and periodic autocorrelated model (PAR) (10). Krunch et al (2005) studied performance of two methods of SARIMA and Thomas-Faring in prediction of water quality and current of the river in Doroijaso monitoring stations located in Sirlemak River in Turkey, by measuring parameters of potassium, sodium, chloride, magnesium, calcium, temperature, sulfate, pH and SAR (13). Ragavan and Fernandez (2006) predicted long-term trend of water quality of some rivers that were selected randomly by using ARIMA seasonal adjustment parameters model in SAS software (14). Ahmed et al (2001) used ARIMA model to predict water quality of Gangs River in India in order to properly manage the basin of that river and they concluded that at the appropriate time, we can take necessary measures to reduce emissions in allowable standard area (9). Hassanazadeh et al (2008) studied the time series analysis, frequency distribution and prediction of SO2 concentration for 2000-2005 course in five hygrometry stations of Tehran City using the ARMA model (12). Also, Haltiner and Salas (1988) used a seasonal ARMA (1,1) model in bivariate modeling of monthly flows in Yampa and White 1 Rivers in northwestern California. They obtained the parameters of model from two maximum likelihood procedures and moments and compared to each other. Dodangeh and Koupaei (2011) examined two models using the time series of monthly flow of Poldokhtar River by using Box–Jenkins method and selected the ARIMA (2,0,2) (2,0,0) model as the best model for forecasting of flow of that river (3).

The purpose of this study is to use time series to forecast qualitative parameters (TSS, NO3 and DO) in two stations of Astaneh and Manjil of Sefidrud River and to present a forecasting model for the future.
Methodology of this study is statistical and is based on using time series models. The main goal of time series analysis is to find a model by which we could predict quantities in some future time series based on existing observations (1). A time series is an example of repeated measurements of a separate variable that is observed at a given time interval. Duration of the intervals can be on hourly, daily or monthly basis (in this paper, daily intervals are considered). What is important is that it is regular (8).

Since qualitative parameters such as TSS, NO$_3$ and DO are occurred considering the time and evidences show that there is a relationship (dependence) between the previous amount of data and future amount of data. Therefore, the best option to analyze the data is the time series methods. In this research, ARIMA Non-seasonal method was used (autocorrelated models, average combined moving, ARIMA (p,d,q), ARIMA model which is also called the Box–Jenkins method, refers to a model that obtains a time series from combination of one or more other time series. Box et al (1994) offered the ARIMA model while considering differential times for static conditions. In ARIMA (p,d,q) process, p, d, q, are respectively indicating number of autoregressive sentences, differential times and number of average moving sentences. If the ARIMA Model would be used for prediction, features of this model should be fixed during the times. Therefore, the reason of need for a static time-series is that any model which would be obtained out of these series can be assumed fixed and considered a valid basis for prediction (7).

ARIMA models make an important part of Box-Jenkins approach for time series model. ARIMA models are applied to the data with non-stationary processes which have a quite distinguishable trend (4).

General equation of ARIMA (p,d,q) model can be presented as Equation 1 in which p, d and q are zero or positive values (Mishra, 2005) and φ(B) and θ(B) are polynomials from p and q times, respectively (11). Equations (2) and (3) indicate AR(p) and MP (q) models, respectively.

\[ \phi(B)V^dZ_t = \theta(B)a_t \]  
\[ \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \]  
\[ \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \]

Time modeling involves three stages: diagnosis, mode estimation and diagnostic control (11). In the model estimation stage, we assess the adequacy of the model by using coefficient of determination.

\[ R^2 = 1 - \frac{\sum(Y(t) - \bar{Y})^2}{\sum(Y(t) - \bar{Y})^2} \]  

In this paper, desired parameters were prepared from Astaneh station (utm$_x$=405433, utm$_y$=4126172) and Manjil station (utm$_x$=356885, utm$_y$=4070434), for water year 2010 which has been registered every 30 minutes in daily form by HACH JANGE machine in Guilan Province Regional Water Organization. Then, average of data was taken in daily mode by Excel Software and normality of data distribution was studied. Finally, with ARIMA (2,0) Model each parameter was modeled separately. Determination coefficient (R$^2$) was calculated for each model. In this case, the adequacy of models to predict the intended quality parameters should be determined. Then, to test models in Astaneh Station, average of daily data of 2 months of year 2011 and in Manjil Station average of daily data of 4 months of year 2012 were examined.

CONCLUSION AND SUGGESTIONS

Table 1 corresponds to the models predicted from the parameters mentioned in two stations of Astaneh and Manjil with ARIMA (2,0,0) Model. Determination coefficient (R$^2$) of paramaters TSS, NO$_3$, DO were calculated 85%, 74%, 90% and 75% in Astaneh Station and 78%, 80% and 82% in Manjil Station, respectively. In order to test the models obtained in Astaneh Station, prediction models were performed on daily averages of 2 months of year 2011 of the same station and the results for the said parameters was gained 91%, 70%, 43% and 86% respectively which indicate reliability of the model. And in order to test models in Manjil Station, the prediction models were performed on daily averages of 4 months of year 2012 at the same station. The results for TSS, DO and NO$_3$ parameters were obtained 95%, 81% and 76% respectively which indicate good performance of the model.
Thus, figures 2, 3 and 4 also indicate the prediction charts of the mentioned parameters in Astaneh Station and figures 5, 6, and 7 indicate the prediction charts of the mentioned parameters in Manjil Station. The horizontal axis on the charts show the studied days and vertical axis on the chart show the changes of each parameter. The red line is the data series on which modeling has been done. The blue line represents built the model. The purple chart and figures 5, 6, and 7 indicate the prediction of the mentioned parameters in astro stations of Asaneh and Manjil.

With regard to this fact that surface water pollution is one of the main problems of today’s societies, it is recommended that sampling work on water streams for qualitative analysis, would be done in shorter intervals and longer, continuous and more carefully time series. Avoid direct discharge of domestic sewage into the river and use domestic sewage treatment methods. Treatment system must be provided where wastewaters are discharged into rivers, lakes and in general water resources so that entry of harmful contaminants to the water supplies would be
prevented. Also it is recommended to use other ARIMA models in forecasting of studied factors and compare them with the present study as well as using more time series models for prediction of qualitative changes of rivers in Iran.

![Figure 2. NO3 Prediction Chart in Astaneh](image2)

![Figure 3. DO Prediction Chart in Astaneh](image3)

![Figure 4. TSS Prediction Chart in Astaneh](image4)

REFERENCES


Figure 5. NO₃ Prediction Chart in Manjil

Figure 6. DO Prediction Chart in Manjil

Figure 7. TSS Prediction Chart in Manjil