Journal of Applied Research in Water and Wastewater 1 (2014) 43-52

Razi University

Journal of
Applied Research in Water & Wastewater

Journal homepage: www.arww.razi.ac.ir

Original paper

Time series analysis of water quality parameters

AbdollahTaheri Tizro[1,*], Maryam Ghashghaie[1], Pantazis Georgiou[2], Konstantinos Voudouris[3]

[1]*Department of Water Engineering, College of Agriculture, Bu-Ali Sina University, Iran.*
[2]*Deptartment of Hydraulics, Soil Science & Agriculture Engineering, School of Agriculture, Faculty of Agriculture, Forestry and Natural Environment, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece.*
[3]*Laboratory of Engineering Geology & Hydrogeology, Department of Geology, Aristotle University, Egnatia St., 54124 Thessaloniki, Greece.*

ARTICLE INFO

Article history:
Received 18 December 2013
Received in revised form 13 February 2014
Accepted 12 March 2014
Available online 23 March 2014

Keywords:
Water quality
Time series
ARIMA
ACF
PACF

ABSTRACT

Water quality is a worldwide problem which affects human beings lives fundamentally. Water scarcity is intensified in result of quality deterioration. Different factors such as population increase, economic development and water pollution could be considered as the origins of the problem. The study and forecasting of water quality is necessary to prevent serious water quality deteriorations in future. Different methodologies have been used to predict and estimate the quality of water. In present study using time series modeling, the quality of Hor Rood River is studied at Kakareza station using time series analysis. 9 parameters of water quality are studied such as: TDS, EC, HCO_3^-, SO_4^{2-}, Mg^{2+}, Ca^{2+}, Na^+, pH and SAR. Investigation of observed time series show that there is an increasing trend for all parameters unless Na^+, pH and SAR. The order of model for each parameter was determined using auto correlation function (ACF) and partial auto correlation function (PACF) of time series. ARIMA (autoregressive, integrated, moving average) model was found suitable to generate and forecast the quality of river water. AIC, R^2, RMSE and VE % criteria were used for evaluating the generation and forecasting results. Results show that time series modeling is quite capable of water quality forecasting. For all generated and forecasted parameters the value of R^2 was greater than 0.66 Except for SO_4^{2-}. The value of R^2 for generated $SO4^{2-}$ was 0.48 and this value was 0.43 for forecasting this parameter. Also the study show that the quality of water is deteriorating based on an increasing trend for the majority of parameters and needs serious managerial actions.

1. Introduction

Water pollution is one of the most important environmental issues which the world faces. Universal problem of water deficiency and shortage of safe and healthy water require the investigation of problem. Water pollution, from a local river and basin to regional water pollution, from single to complex pollution, from surface water to groundwater, has been a serious restraint to sustainable economic development. Water quality could be affected by salinity, overdraw of ground water, urban and domestic wastewater entrance into surface streams as well as agricultural drainage. The purpose of most water quality and stream flow studies is to point out the information and necessary knowledge to manage water resources as well as their use, control and development. Time and money are saved through these studies and future development of water resources becomes inexpensive. The main objectives of water quality modeling could be to: (i) imply cause and effect relationships, (ii) identify impacts of pollutant sources, (iii) assess necessary levels of monitoring, (iv) evaluate planning and management alternatives, (v) focus on additional monitoring and management objectives and (vi) assess and evaluate future water quality conditions. Time series analysis is one of the useful methods which are applied in water quality modeling and forecasting.

Corresponding author E-mail: ttizro@yahoo.com

Nowadays time series analyses are used in different aspects of science such as physics, economy and engineering. Water resources engineering lies within this category as there are many characteristics of water bodies, streams and groundwater resources as well as lakes and seas which are defined using time series of data. This procedure is useful in understanding and modeling the process of a phenomenon through which the past observations are generated.

It is also helpful in forecasting the future values based on the past memory. Time series is a string of data over time and there is an equal interval between all data. The interval can be defined as daily, weekly, monthly as well as yearly time steps. Time series analyzing is used in decision making in many hydrological processes and operation systems. The aim of time series analysis in hydrology can be defined as two main goals: at first it is used to understand and model the stochastic mechanism of hydrologic phenomena and at the second stage it is used to forecast the future values of the phenomena.

Many works have been accomplished on hydrological components modeling using time series analysis. The application of this method for water quality forecasting is possible as well. Also evaluation of existing water resources, determining the quality of discharge as well as its quantity, identifying its variation on a watershed scale and forecasting these variables, could be a main step in integrated water resources management. Also stochastic characteristics of hydrological phenomena lead the hydrologists and water resources engineers

Please cite this article as: A. Taheri Tizro, M. Ghashghaie, P. Georgiou, K. Voudouris, Time series analysis of water quality parameters, Journal of Applied Research in Water and Wastewater, 1(1), 2014, 43-52.

towards benefiting from time series concepts in modeling and forecasting the future of water resources.

1.1. Applied time-series analysis

Time-series analysis using ARIMA approaches have been used to examine runoff and river discharge (Rao et al. 1982, Papamichail and Georgiou 2001; Yurekli et al. 2005), water levels in lakes (Irvine and Eberhardt 1992; Sheng and Chen 2011), sediment yield and erosion (Hanh et al. 2010), and water quality (Papamichail et al. 2000; Lehmann and Rode 2001; Faruk 2010; Hanh et al. 2010; Voudouris et al. 2010). ARIMA models are capable of reproducing the main statistical characteristics of a hydrologic or environmental time series. These models also provide information about system dynamics and could be used to forecast a time series for the future. Thomas and Fiering, (1962) used auto correlated models in their studies on stream analyzing. Chow and Kareliotis (1970) analyzed univarite time series of rainfall and temperature. They discovered yearly strict and 6 months slight periodic components in time series. The main step of time series application in hydrology was performed by McKerchar and Delleur (1974) as they used ARMA (Autoregressive Integrated Moving Average) and seasonal modeling in analyzing seasonal characteristics of stream parameters.

Zhang (2003) applied a hybrid of ARIMA and ANN model to take the advantages of the ARIMA and ANN models in linear and nonlinear modeling. Results showed that the combined model was capable of forecasting the real data sets more accurately in comparison with the separately applied methods of ANN and ARIMA.

Jalal Kamali (2006) also used time series models for monthly inflows to Jiroft dam. The results of this study showed that time series modeling is capable of identifying and forecasting monthly stream pattern and integrated water resources management. Also Komornk et al. (2006) studied hydrological time series in Czech through which high capability of forecasting by this kind of modeling was proved. Dalme and Yalcin (2007) applied time series analyzing in Mississippi River to forecast the values of flood. The results of their study showed the capability of time series modeling application in generating daily discharge as well as validity of forecasting. There are many studies which have been focused on water quality parameters mentioned as follow. Khashei and Bijari (2010) applied an artificial neural network (pdq) model to estimate time series forecasting. In this paper, a new hybrid model of ANN was introduced using ARIMA models in order to achieve a more accurate forecasting model than artificial neural networks. The empirical results with three well-known real data sets showed that the proposed model can be an effective way to improve forecasting accuracy achieved by artificial neural networks. The research proposed the application of model as a convenient alternative method to forecast accurately thanks to times series capabilities.

1.2. Applied time series analysis on surface water quality

Hirsch et al. (1982) introduced techniques to analyze monthly water quality data for monotonic trends. The first procedure is a non-parametric test to detect trend, which is used for seasonal time series. The second method of seasonal Kendall estimator estimates the magnitude of trends. The third procedure provides a tool to test temporal changes in correlation of constituent concentration and stream flow. Also El-Shaarawi et al. (1983) studied temporal changes in water quality parameters such as PH, Alkalinity, total Phosphorous and Nitrate concentrations using a 5 year data series of Niagara (on Ontario lake). Results showed that PH and Alkalinity were decreasing while Nitrate was increasing. Yu et al., (1993) examined surface water quality data of the Arkansas, Verdigris, and Neosho as well as Walnut river basin to study trends in 17 major constituents using 4 different nonparametric methods. Robson and Neal (1996) studied the trend of ten year upland stream and bulk deposition water quality data from Plynlimon, mid wales through the seasonal Kendall test, the stream water dissolved organic carbon was increasing over time. However any increase for PH was not found. It was suggested that long term monitoring programs should be applied for several decades.

In a study accomplished by Turner et al. 1995, Long-term simulations results of Lake Bosumtwi in Ghana, showed that stochastic climatic variations very similar to those observed in this century could produce the full range of lake levels observed in terrace deposits. The low salinity of about 1% suggests that dissolved solutes were removed by Lake Overflow in the recent geological past.

Primarily Graphical and statistical time series techniques have been used to analyze the trends and specified time changes, in river water quality data. The information obtained may be associated with some socio-economic variables, such as industrial or agricultural development, urban increase and wastewater discharge around or upstream of the measure station. Such a study may now be applied to more rural stations in order to compare the evolution of water quality and perhaps, historical monthly average values to evaluate the seasonality effect on annual trends (Gun and Vilagines 1997).

Papamichail et al. (2000) examined stochastic models to improve understanding and forecasting of monthly flow and some water quality parameters of Strymon River (Greek part) in an effort to reduce the negative impacts of incurred by interests using the river. Especially, they developed seasonal and nonseasonal ARIMA models for Strymon River using the time series of monthly measurements of flow and some water quality parameters. The selected models for each parameter data set can be used to forecast monthly values of one or more time periods ahead.

Antonopoulos et al. (2001) analyzed the time series of water quality parameters and the discharge of Strymon River in Greece from the 1980 to1997. The nonparametric Spearman's criterion was used to detect the trends for: discharge, EC_w, DO, SO_4^{2-}, Na^+, K and NO_3^{3-}. The Verification of the best fitted models was performed using χ^2 and Kolmogorov-Smirnov tests. The relationships between concentration and loads of constituents of both with the discharge were investigated as well. In spite of the relation between loads and discharge ($r > 0.9$), the correlation between concentrations and discharge is not good ($r < 0.59$). Ahmad et al., 2001, accomplished a study to analyze water quality data collected from Ganges River in India. Three approaches of stochastic modeling such as: multiplicative ARIMA model, deseasonalised model and Thomas–Fiering model were applied to model the observed time series of water quality. The multiplicative ARIMA model having non- seasonal and seasonal components we identified as a convenient model. The de seasonalised modeling approach was recommended to forecast water quality parameters of the river.

Through a water quality monitoring program (New Zeland) Stansfield (2001) illustrated the importance of considering detection limits of variables and sampling frequencies through analyzing the trends in water quality time series using nonparametric seasonal Kendall test and Sen Slope test. Result showed that if the sampling frequency was changed from monthly to quarterly fewer trends were detected. What is more results showed that the quarterly data present with a different magnitude in terms of slope in comparison with monthly data. Gangyan et al. (2002) investigated the temporal sediment load characteristics of the Yangtze River using the turning point test, Kendall's rank correlation test and the Anderson correlogram test to prove randomness and determine the trend. The annual sediment load data from 1950 to 1990 and the monthly sediment load data from 1950 to 1969 were used. The stochastic component was modeled using autoregressive model. Using the AR (1) model for the dependent stochastic component, 100 years of monthly sediment data were generated and the observed and generated data matched well.

Jassby et al. (2003) developed a time series model for Secchi depth in Lake Tahoe, USA considering an understanding of inter annual variability. The Secchi depth was found sometimes over 40 m. however the mean annual Secchi disk depth has declined about 10 m since 1967 inspiring a large scale restoration program. Yearly variability was extremely high, obscuring restoration actions and conformance with water quality standards. The model suggested a tool to determine the compliance with water quality standards when precipitation anomalies may persist for years. Also some studies have focused on water temperature time series such as Webb et al. (2003) who showed that when discharge is below the annual median, correlations between air and water temperature is high. Kurunc et al. (2005) applied time series analysis for water quality constituents and stream flow of the Yesxilirmak River at Durucasu which is a monitoring station. Two modeling approaches, ARIMA and Thomas–Fiering were evaluated in this study. A 13 year monthly time series records were used to obtain the best model of each water quality constituent and stream flow from both modeling approaches. The results of study showed that that between two approaches, for Yesxilirmak River Thomas–Fiering model presents more reliable forecasting of water quality constituents and stream flow than ARIMA model.

Panda et al. (2011) studied trends in sediment load of the tropical river basin of India and explored the influence of climate and human forcing mechanisms on the land ocean fluvial system. Sediment time series of different timescales during the period 1986- 87 to 2005- 06 from 133 gauging station were analyzed. Results showed significant

Please cite this article as: A. Taheri Tizro, M. Ghashghaie, P. Georgiou, K. Voudouris, Time series analysis of water quality parameters, Journal of Applied Research in Water and Wastewater, 1(1), 2014, 43-52.

diminishing in sediment load in the tropical river basins. The rainfall characterized by the non-significant decreasing trends and frequent drought years was found to be the reason of sediment load reductions for most of the river basins. Also results showed that the maximum reduction in sediment loads was referred to Narmada River among the tropical rivers (2.07×106 t / yr) because of construction of the dam. Also Irvine et al. (2011) accomplished a study on temporal variability of turbidity, dissolved oxygen, conductivity, temperature, and fluorescence in the lower Mekong River. Results showed that A strongly developed vertical variation of turbidity, DO, and conductivity in the flooded forest fringe may be related to a combination of factors, including dissolved material release from bed sediment and a floating organic-rich particulate layer near the bottom of the lake.

Halliday et al. (2012) studied two hydrochemical time-series derived from stream samples taken in the Upper Hafrencatchment, Plynlimon, Wales. A subset of determinants such as: aluminium, calcium, chloride, conductivity, dissolved organic carbon, iron; nitrate, pH, silicon and sulphate were examined within a framework of non-stationary time-series analysis to identify determinant trends, seasonality and short-term dynamics. The results demonstrate that both long-term and high frequency monitoring provide valuable and unique insights into the hydrochemistry of a catchment. Such studies moving forward demonstrate the need for both long-term and high-frequency monitoring to facilitate a thorough understanding of catchment hydro chemical dynamics.

Different studies on time series analyzing in water resources management demonstrate the efficiency and necessity of this kind of modeling as it takes into account the stochastic nature of hydrological processes such as discharge and climatology. Present study aims to apply this methodology on chemical water quality properties. In present study water quality parameters of Hor Rood is investigated at Kakareza station. The methodology and the study area are presented at the following stages.

The objectives of this study are to:

(i) plot time series of data to find any possible trend,
(ii) Omit the trend which demonstrates a deterministic nature of data,
(iii) Obtain the best model fits for each time series of parameters using a stochastic modeling approach including ARIMA, and
(iv) Evaluate the performance of modeling approach using five-year observed data vs. forecasted data

2. Methodology and the study area

2.1. The study area

The study area is located in the west of Iran in Kuhdasht region which is shown in Fig. 1. Kakareza station of Hor Rood River lies at 48o 15' E and 33o 43' N. The upstream area of the station is about 1148 Km2 located at Kashkan sub-basin of Karkheh watershed. The topography in the Karkheh watershed varies spatially with elevation ranging from 3 to more than 3500 masl. The average value of temperature varies from -3 to 21 ^{0}C across Kashkan basin. The mean annual precipitation of the basin varies between 345 mm/yr and 849 mm/yr. Also evaporation of the basin varies from 25 to 2922 based on a yearly average. The contribution of rainfall, snow and Karstic springs is significant to discharge of Kashkan River. Also Kakareza River is the main tributary of this river.

The land use of the study area is composed of nearly 26% agriculture, 16.9% pasture, 39.33 % forest, 0.12 % residential area and about 2.5 % of the area is composed of other kinds (Jamab Consulting Engineers 2005). Based on field studies agricultural drained waters and industrial waste waters are either directed or flow through surface streams as well as residential litters and swages. The time series of 9 water quality parameters such as TDS, EC, HCO_3^-, SO_4^{2-}, Mg^{2+}, Ca^{2+}, Na^+, pH and SAR of Kakareza station at Hor Rood River were studied in this research.

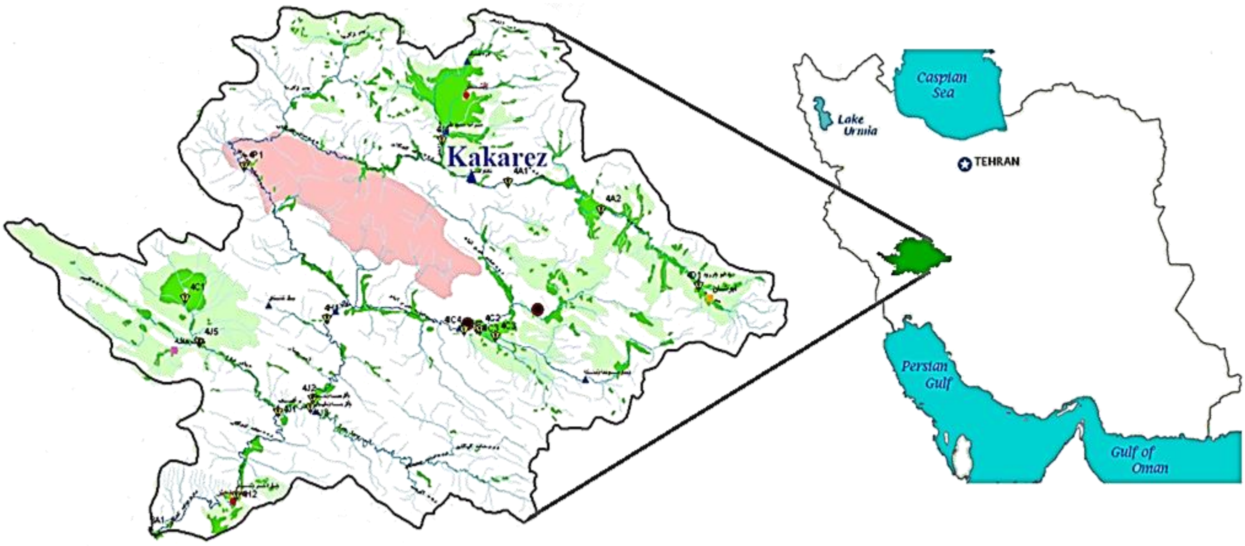

Fig.1. The location of the study area.

2.2. Methodology

Stochastic models reveal time series characteristics in terms of correlation as well as consider the randomness of phenomenon although they do not consider their physical nature. This method analyzes the past of the time series with respect to successive correlation which is used as system input in other words. The present or the future is then predicted as the system output. Two main applications of the time series models are generating simulated samples and forecasting hydrologic events. Forecasted time series are used as input for analyzing complex water resources systems. Generated series could show many possible hydrologic conditions that do not appear in the historic series explicitly. Consequently, using simulated time series, different designs and operational strategies can be tested under various conditions. Forecasted data from known historic observations can help to assess and evaluate options for a real system operation.

Although time series modeling originated from different scientific fields, it has demonstrated its capability and reliability in stochastic hydrology, and the applications of time series analysis in water resources management are important. Development of stochastic modeling in hydrology began at the beginning of the 1960s, when time series analyses of hydrologic phenomena were extended to the synthetic generation of stream flow using a table of normal random numbers. Thomas and Fiering (1962) were the first to propose a first-order Markov model to generate stream-flow data. The classic book on time series analysis by Box and Jenkins (1976) presents the foundation of modern hydrologic stochastic modeling eq. (1):

$$y_t = f(x_t, x_{t-1}, \ldots; y_{t-1}, y_{t-2}, \ldots; 1, 2, \ldots) + \varepsilon \quad (1)$$

where f is the selected mathematical function; y_t is the predicted output at time t; y_{t-1}, y_{t-2}, ... , are the successive members of the output time series recorded at corresponding time intervals 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, ... , are the model parameters found by mathematically minimizing the differences between estimated (calculated) and observed y_t values; ε is the model

Please cite this article as: A. Taheri Tizro, M. Ghashghaie, P. Georgiou, K. Voudouris, Time series analysis of water quality parameters, Journal of Applied Research in Water and Wastewater, 1(1), 2014, 43-52.

error (residual) given as the difference between the calculated and the recorded value of the output series at time t.

Stochastic modeling generally follows the approach proposed by Box and Jenkins, (1976), who introduced autoregressive moving average (ARMA) models. The mathematical formulation of ARMA models is written as eq. (2):

$$Z_t=\varphi_1 Z_{t-1}+\ldots+\varphi_p Z_{t-p}+\alpha_t-\theta_1\alpha_{t-1}-\ldots-\theta_q\alpha_{t-q} \quad (2)$$

Where Z_t represents the time dependent series, φ_i, i=1,2,..., p are nonseasonal AR parameters, θ_i, i=1,2,..., q are the nonseasonal MA parameters,Time series models used to generate synthetic time series can be classified into autoregressive models (AR (p)), moving average models (MA (q)), and their combination, autoregressive moving average (ARMA (p, q)) with variations, such as autoregressive, integrated moving average (ARIMA) models (p, d, q) and others, where p and q are the orders of autoregressive and moving average terms, respectively, and "d " is an order of differencing.

An autoregressive 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.

An autoregressive (AR) model, which is called a Thomas-Fiering model, has been applied extensively in hydrology for modeling annual and periodic hydrologic time series. Autoregressive (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}. An AR model of order 1 (i.e. an AR (1) model) can be expressed as eq. (3):

$$Z_t = \varphi_1 Z_{t-1} + \alpha_t \quad (3)$$

where, Z_t and Z_{t-1} are the deviations from the mean of the time series, φ_1 is the first-order AR coefficient describing the effect of a unit change in Z_{t-1} on Z_t, and α_t represent random shock errors or white noise. Values for α_t are assumed normally and independently distributed with zero mean and constant variance. Model stationarity requires that the variance of Z_t be non-negative and finite (Vandaele 1983) and for these conditions to be met, $|\varphi_1|$ must be less than 1. Higher order AR models are possible, much like a multiple regressions, and in this case, the absolute value of each AR coefficient should be less than 1.

Moving average (MA) models incorporate past random fluctuations to represent the time series and an MA model of order 1 (i.e. an MA (1) model) can be expressed as eq. (4):

$$Z_t = \alpha_t - \theta_1 \alpha_{t-1} \quad (4)$$

where, θ_1 is the MA coefficient to be estimated and the random shocks (αt) are assumed normally and independently distributed with mean 0 and constant variance. The model structure requires the condition of reversibility to be met and | θ1 | therefore must be less than 1. Values greater than 1 indicate that observations further in the past have a greater influence on Z_t than more recent observations which is unlikely in hydrologic time series. Higher order of MA models is possible, and like the AR model coefficients, the absolute value of each MA coefficient should be less than 1.

A parsimonious model can be achieved using a mixed ARMA model as a combination of a moving average process and an autoregressive process rather than a merely AR or MA model. Therefore, low-order of ARIMA models has been widely used in hydrological practice (Salas et al., 1982; Weeks and Boughton 1987; Padilla et al., 1996; Montanari et al. 2000).

The statistical structure of a time series should be represented by a parsimonious model, and in some cases, parsimony can be achieved using a mixed (ARMA) model rather than a pure AR or MA model. As such, it would be more parsimonious to represent a time series with an ARMA (1, 1) model than an AR (3) model because fewer model parameters need to be estimated. It is possible to mix models because these models theoretically can be rewritten as pure AR or MA models of infinite order (Vandaele 1983). Furthermore, a hydrologic time series is the result of several interactive processes that may have both a seasonal and a random fluctuation component. The mixed model structure can provide additional flexibility in describing the result of the interaction between the processes (Salas et al. 1980).

2.3. Modeling water quality time series of Hor Rood River

The main goal of a time series analysis may be to understand seasonal changes and/or trends over time. Plotting the data against time was accomplished as the first step of analyzing time series. Time plot show a lot of information about the time series. Trends and seasonal variations are often evident in time plots. Hydrologic time series frequently exhibit a regular seasonal pattern that can be removed by standardizing the data for the seasonal mean and standard deviation and then retrending the forecasts using the inverse of the deseasonalizing transformation. Also time plots indicate the presence of outliers in the time series which are observations that are not consistent with the rest of the data.

However, another goal that is often of primary importance is to understand and model the correlational structure in the time series. This type of analysis is generally done on stationary processes. A stationary process is one that looks basically the same at any given time point. That is, a stationary time series is one without any systematic change in its mean and variance and does not have periodic variations.

Many studies have been written about the theory of ARIMA modeling as well as its applications (Pankratz, 1983; Vandaele, 1983; Nelson, 1973; Box and Jenkins, 1976). Here a brief description is presented to point out main stages accomplished in this study.

The basic stages in ARIMA modeling are composed of: (1) identifying the autocorrelation and partial autocorrelation of time series, (2) estimating the orders of the identified model, and (3) verify the model through standard tests. The results of time series analyzing of this study are explained in the following stage.

In this study 8 river quality parameters were studied. All series showed trend line. MINITAB 14 was used to analyze these 9 time series. Also Normality test of series was investigated using Easy fit. All series were normal. Then ACF and PACF of time series were plotted at next stage. Fig. 2 shows the ACF and PACF for TDS time series. Based on this Figure p= 1 and q= 3 is suggested for TDS time series. Fig. 3 shows the ACF and PACF of EC time series. For EC p= 1 and q=2 is offered. ACF and PACF of HCO_3^- time series are shown in Fig. 4 and p=1 and q=2 is suggested for HCO_3^-.

Also ACF and PACF of SO_4^{2-} are demonstrated in Fig. 5. P=1 and q=2 is proposed for the series. Fig 6 presents the ACF and PACF of Ca^{2+} and p=1 and q=2 is suggested for the series. Fig 7 shows the ACF and PACF of Mg^{2+}. As it is clear in the Fig. 7 there is not any autocorrelation between data. ACF and PACF of Na^+ are shown in Fig. 8. P=1 and q=2 is proposed for the series. Fig. 9 shows the ACF and PACF for pH series. P= 1 and q= 1 is suggested for this series. And finally ACF and PACF of SAR are shown in Fig. 10. P=1 and q=2 is offered for SAR series.

All parameters suggest the value of p= 1 except Mg^{2+}. At the following steps generation of data are accomplished and the results of it are demonstrated for TDS, EC, HCO_3^-, $SO4^{2-}$, Ca^{2+}, Na^+, pH and SAR respectively. The standard (Z) time series of all parameters were plotted at the second stage, which are shown in Fig. 11. Using one difference the data were transformed to make a yearly stationary time series.

After identifying the ACF and PACF and removing the trend of each time series first the order of model was determined and then 4 criteria such as Akaike Information Criterion (AIC), Determination Coefficient (R^2), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE %) (Karamouz and Araghinejad 2005) were used to compare the results of series generation through suggested models. The value of AIC is estimated through eq. (5):

$$AIC= n\times \ln(\sigma^2) + 2 \times (p+ q) \quad (5)$$

where, σ denotes the standard error of residuals; n shows the sample size; p and q show the order of AR and MA, respectively.

Also the value of VE % is calculated using eq. (6):

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{y_t-\hat{y}_t}{y_t}\right|100}{n} \quad (6)$$

where, y_t and $\hat{y}_t$ show the observed and estimated values respectively and n is the sample size. A thirty five (35) year time series were generated for each parameter.

Please cite this article as: A. Taheri Tizro, M. Ghashghaie, P. Georgiou, K. Voudouris, Time series analysis of water quality parameters, Journal of Applied Research in Water and Wastewater, 1(1), 2014, 43-52.

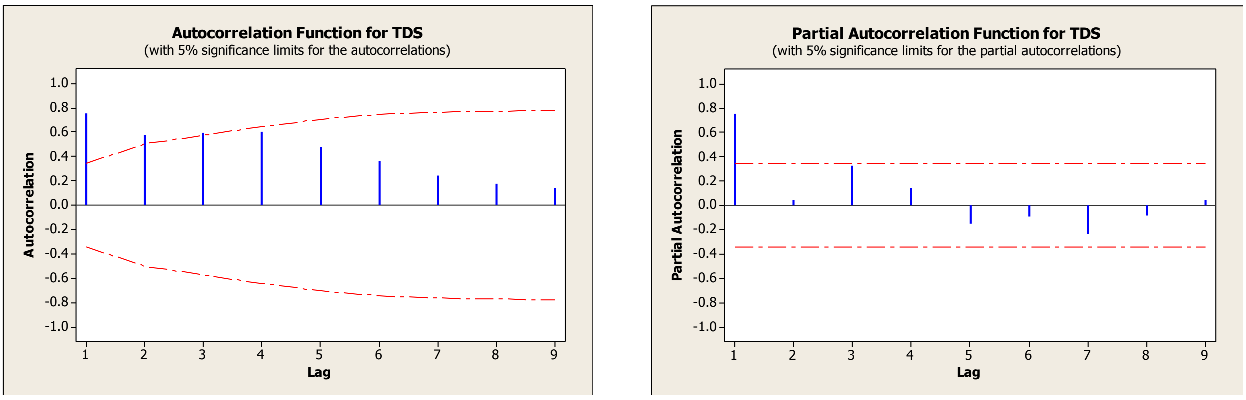

Fig. 2. ACF and PAC of TDS time series.

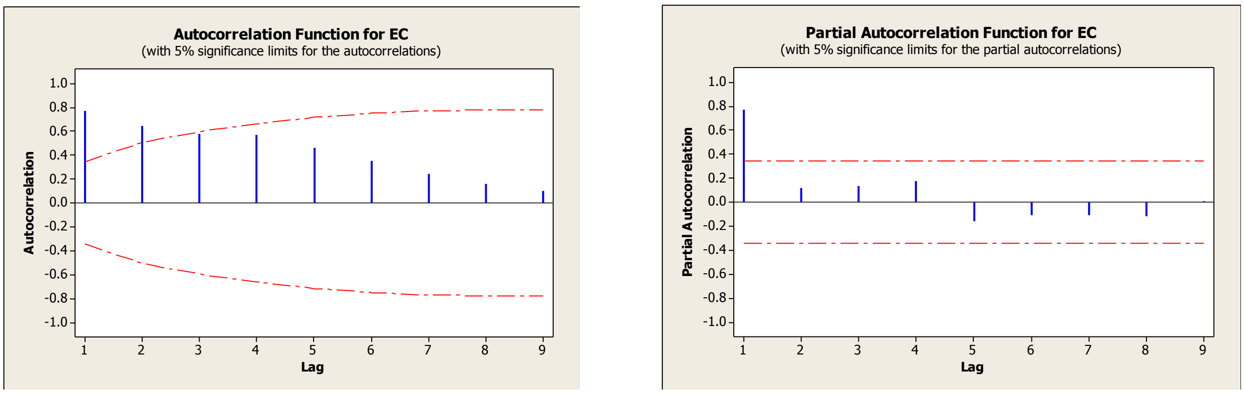

Fig. 3. ACF and PAC of EC time series.

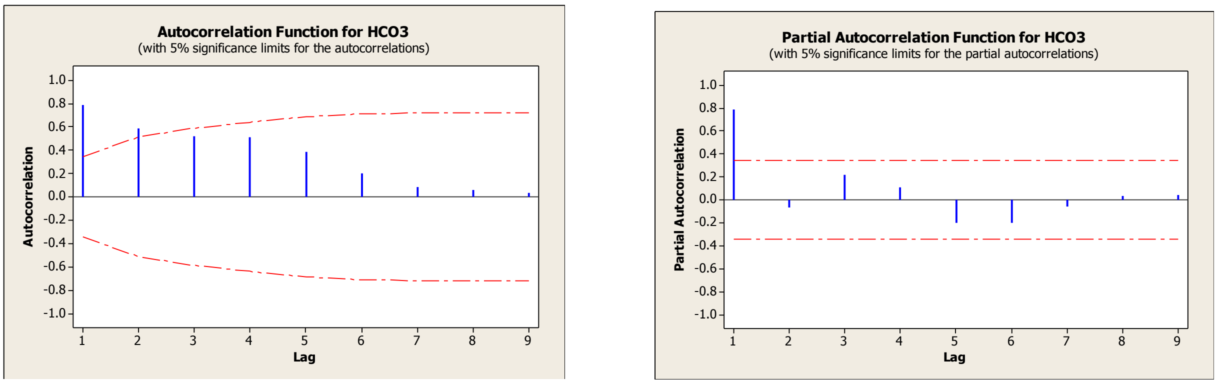

Fig. 4. ACF and PAC of HCO_3^- time series.

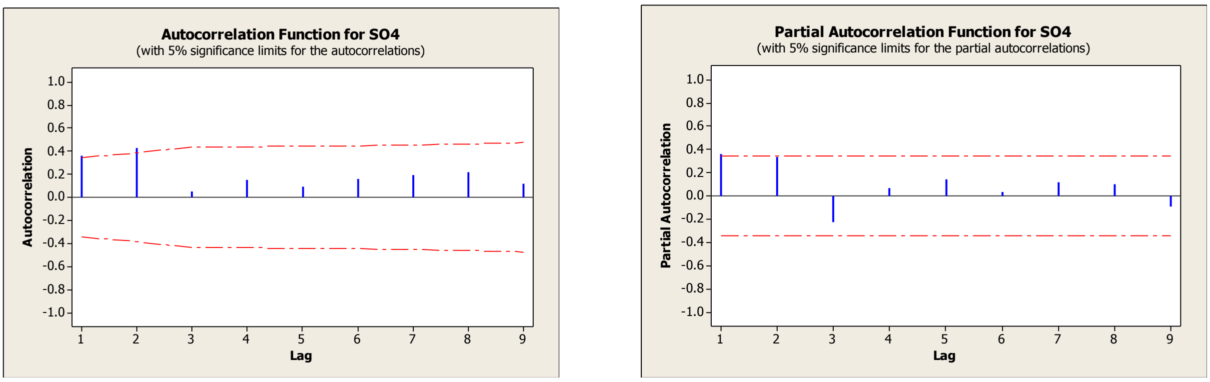

Fig. 5. ACF and PAC of SO_4^{2-} time series.

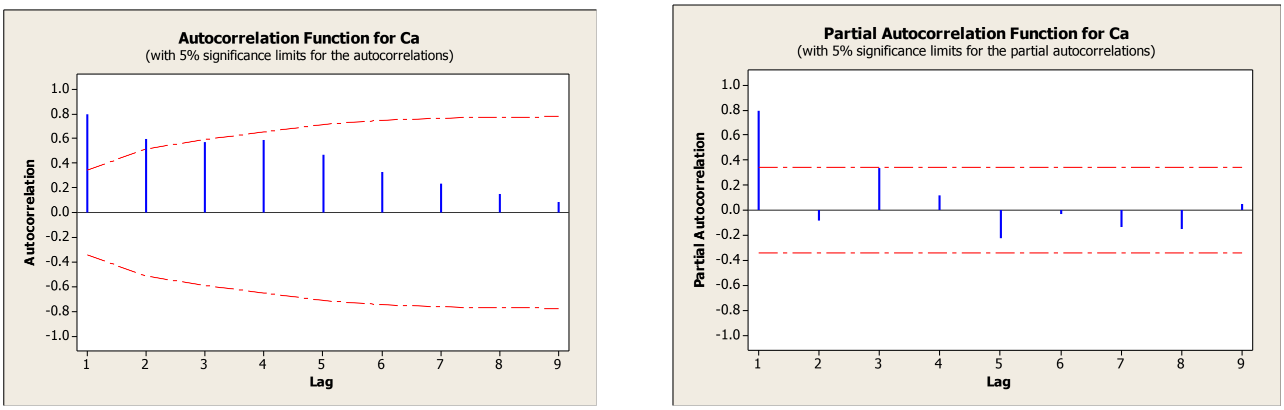

Fig. 6. ACF and PAC of Ca^{2+} time series.

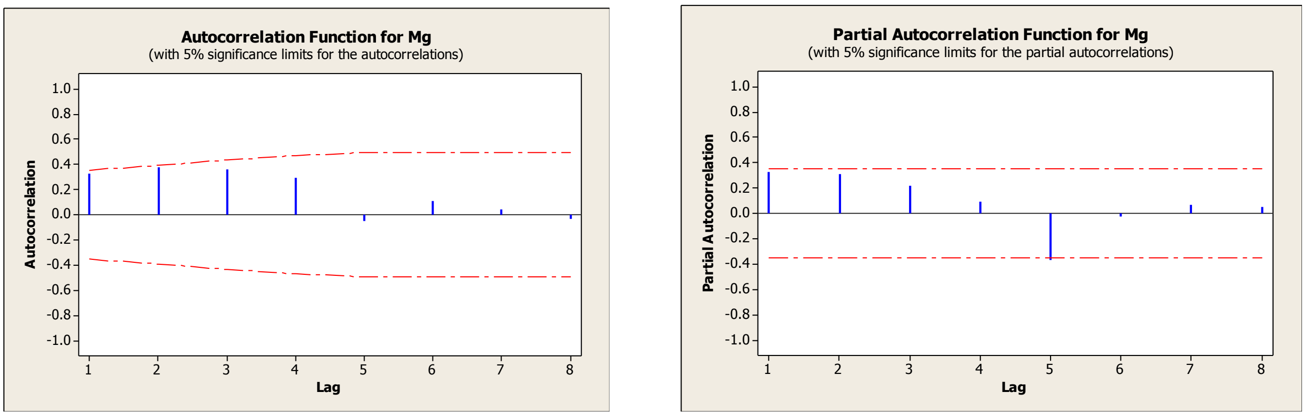

Fig. 7. ACF and PAC of Mg^{2+} time series.

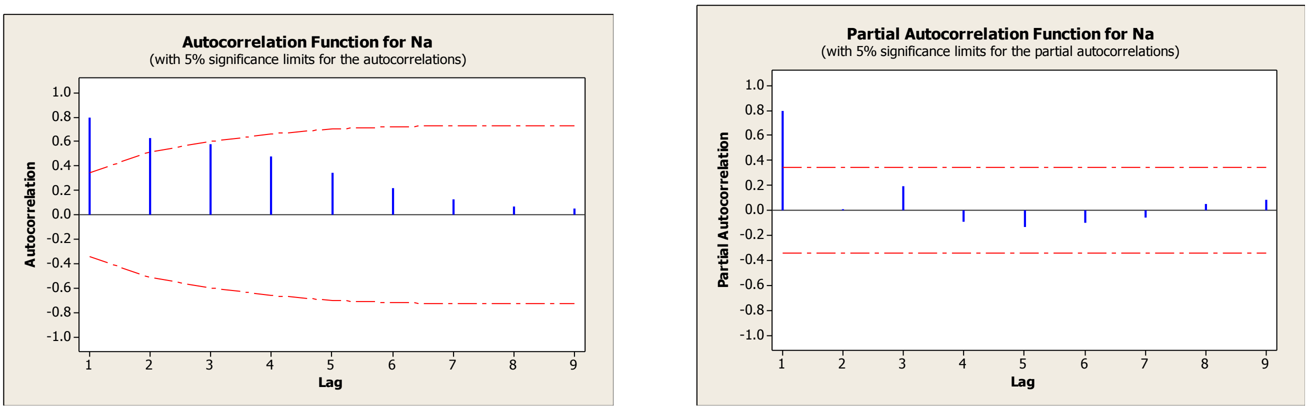

Fig. 8. ACF and PAC of Na^{+} time series.

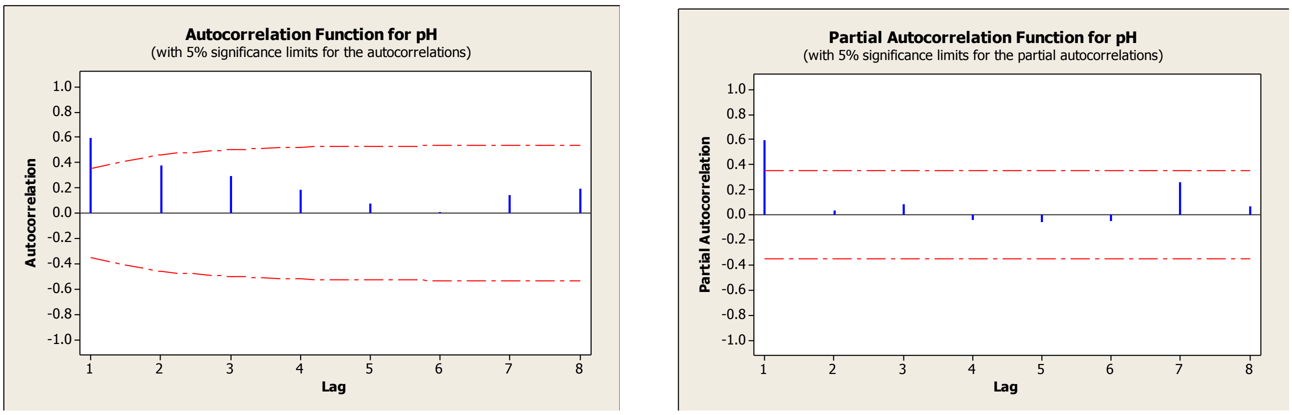

Fig. 9. ACF and PAC of pH time series.

Please cite this article as: A. Taheri Tizro, M. Ghashghaie, P. Georgiou, K. Voudouris, Time series analysis of water quality parameters, Journal of Applied Research in Water and Wastewater, 1(1), 2014, 43-52.

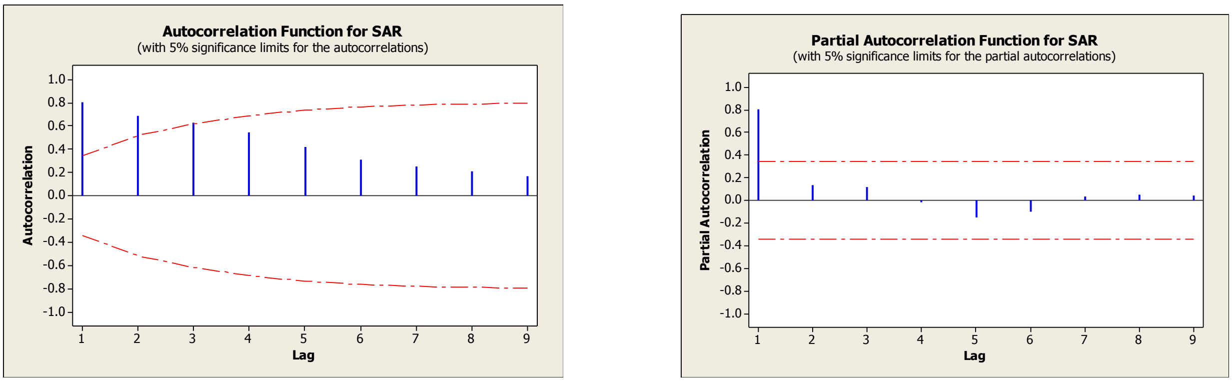

Fig. 10. ACF and PAC of SAR time series.

4. Results

4.1. Results of data generation

Time series of TDS values shows a positive trend which was removed through one difference. The best fitted model of the series based on ACF and PACF, AIC, R^2, RMSE and MAPE is demonstrated in Table 1. Results show that the model is capable of modeling the time series well.

Table 1. The results of TDS generation, order (1, 3).

MODEL	R^2	AIC	RMSE	MAPE %
(1,1,3)	0.73	-45.66	0.07	0.64
(2,1,3)	0.78	-49.13	0.07	0.86
(3,1,3)	0.76	-45.85	0.07	0.63

For the second parameter, as it is clear from Fig. 11, EC time series follows an increasing trend. Table 2 shows the results for generating and choosing the best fit after removing the trend.

Standardized time series of HCO_3^- is presented in Fig. 11. HCO_3^- follows an increasing slope. Modeling was accomplished after trend elimination. Table 3 shows the results of modeling time series for this parameter.

Table 2. The results of EC generation, order (1, 2).

MODEL	R^2	AIC	RMSE	MAPE %
(1,1,3)	0.72	-46.00	0.07	8.11
(2,1,3)	0.81	-54.11	0.06	3.90
(3,1,3)	0.79	-54.55	0.06	1.28

Table 3. The results of HCO_3^- generation, order (1, 2).

MODEL	R^2	AIC	RMSE	MAPE %
(1,1,2)	0.67	-29.22	0.10	4.16
(2,1,1)	0.61	-21.73	0.11	3.19
(2,1,2)				
(1,1,3)	0.75	-35.69	0.08	3.17
(2,1,3)	0.76	-35.50	0.08	3.20
(3,1,3)	0.76	-33.61	0.08	3.21

Also Z time series of SO_4^{2-} is shown in Fig. 11. The results of generation after trend elimination and the best fit are shown in table 4. Results show that ARIMA modeling for this parameter is rather capable of predicting as well as previous parameters.

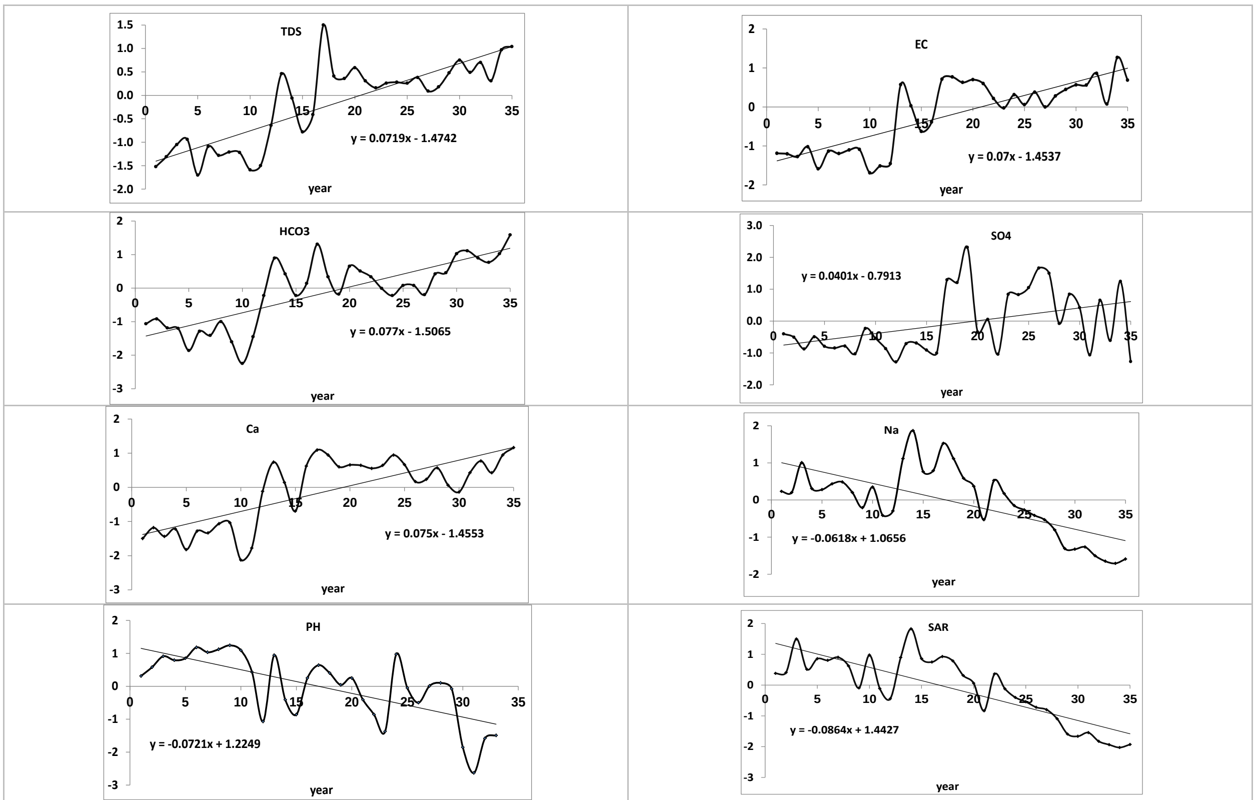

Fig. 11. The standard time series of TDS, EC, HCO_3^-, $SO4^{2-}$, Ca^{2+}, Na^+, pH and SAR.

Please cite this article as: A. Taheri Tizro, M. Ghashghaie, P. Georgiou, K. Voudouris, Time series analysis of water quality parameters, Journal of Applied Research in Water and Wastewater, 1(1), 2014, 43-52.

Table 4. The results of SO_4^{2-} generation, order (1, 1).

MODEL	R^2	AIC	RMSE	MAPE %
(1,1,2)	0.19	-1.01	0.15	1.08
(2,1,2)	0.35	-6.89	0.13	1.47
(1,1,3)	0.43	-11.37	0.12	1.11
(2,1,3)	0.39	-6.92	0.13	1.26
(1,1,1)	0.28	-1.95	0.15	2.04
(2,0,1)	0.29	-7.12	0.14	1.48
(2,1,1)	0.28	-0.01	0.15	2.04

Also time series of Z for Ca^{2+} is presented in Fig. 11. The results of generating presented in Table 5 show that the selected model, shown in Table 5, is capable of modeling the series quite well.

Table 5. The results of Ca^{2+} generation, order (1, 1).

MODEL	R^2	AIC	RMSE	MAPE %
(1,0,1)	0.59	-61.39	0.06	1.32
(1,0,2)	0.67	-66.36	0.06	1.33
(2,0,1)				
(2,0,2)				
(1,1,1)	0.60	-61.90	0.06	1.15
(1,1,2)	0.70	-69.94	0.05	1.16
(2,1,1)	0.76	-76.41	0.05	1.23
(2,1,2)	0.76	-74.52	0.05	1.22

Z time series of Na^+ show that series follow a decreasing trend. Modeling was done for Na^+ series after trend elimination. Also the results of modeling are presented in Table 6. The results show that the selected model is capable of modeling the series. Time series of pH values shows a decreasing trend which was removed through one difference. Table 7 shows the results of generating for the series. The results show that the model is quite capable of generating the data. Finally modeling the SAR series was done after trend elimination. Table 8 shows that selected model is capable of modeling the series properly.

4.2. The results of forecasting

Fig. 12 shows the results of forecasting 5 data for the 5 last years. to evaluate the selected models. These results are shown for time series of TDS, EC, HCO_3^-, $SO4^{2-}$, Ca^{2+}, Na^+, pH and SAR, respectively.

Table 6. The results of Na+ generation, order (1, 2).

MODEL	R^2	AIC	RMSE	MAPE %
(1,0,1)	0.69	-42.63	0.09	0.66
(1,1,1)	0.70	-42.62	0.08	0.61
(1,0,2)	0.71	-42.87	0.08	0.73
(2,0,1)	0.70	-41.23	0.09	0.68
(2,0,2)	0.72	-41.55	0.08	0.70
(1,1,2)	0.72	-43.14	0.08	0.64
(2,1,2)	0.74	-42.89	0.08	0.61
(2,1,1)	0.74	-44.75	0.08	0.66
(1,1,3)	0.72	-40.54	0.08	0.66

Table 7. The results of pH generation, order (1, 2).

MODEL	R^2	AIC	RMSE	MAPE %
(1,0,1)	0.40	-12.67	0.13	1.91
(1,0,2)	0.32	-5.99	0.14	1.60
(2,0,2)	0.45	-11.97	0.13	1.26
(1,1,1)	0.53	-17.38	0.12	3.12
(2,0,1)				
(2,1,1)	0.52	-14.94	0.12	3.18

Table 8. The results of SAR generation, order (2, 2).

MODEL	R2	AIC	RMSE	MAPE %
(1,0,1)				
(1,1,1)	0.76	-39.42	0.09	0.92
(1,0,2)				
(2,0,1)				
(2,0,2)				
(1,1,2)	0.77	-38.36	0.09	0.93
(2,1,2)	0.84	-50.05	0.07	0.80
(1,1,3)	0.77	-36.86	0.09	0.94
(2,1,3)	0.78	-36.13	0.08	0.86

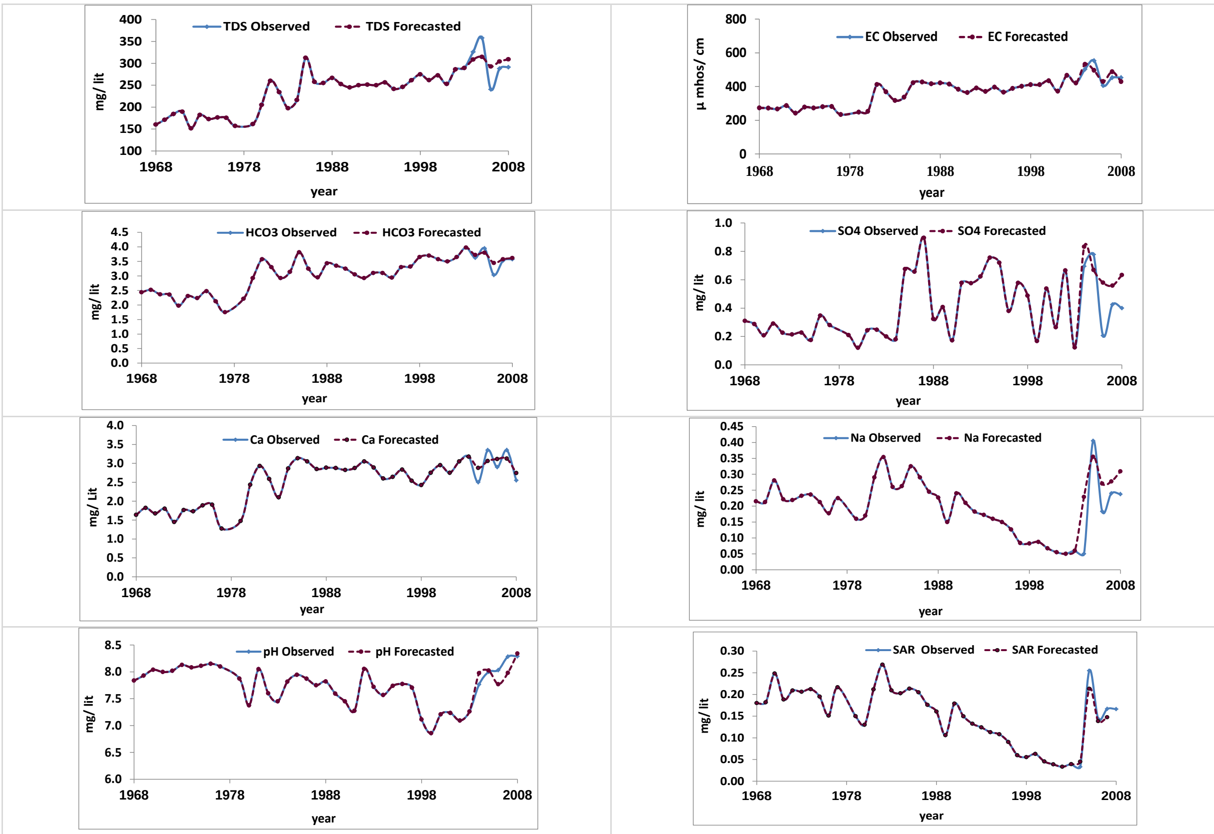

Fig. 12. Forecasted values of 5 last years for TDS, EC, HCO_3^-, SO_4^{2-}, Ca^{2+}, Na^+, pH and SAR.

Please cite this article as: A. Taheri Tizro, M. Ghashghaie, P. Georgiou, K. Voudouris, Time series analysis of water quality parameters, Journal of Applied Research in Water and Wastewater, 1(1), 2014, 43-52.

Table 9 shows the results of forecasting water quality parameters. AIC, RMSE and VE %, criteria were used at this stage to show the capability of each model to forecast the value of data. Results show that the selected model for each parameter is quite able to estimate the future values of water quality parameters.

5. Conclusions

In this study, nine water quality parameters of Hor Rood River were studied at Kakareza station. First the normality of series was examined. All parameters showed that they follow a normal distribution. Then the ACF and PACF of each series were estimated to guess the best model for generating the value of series. Also standard time series of all parameters were plotted. Na^+, pH and SAR show decreasing trend in spite of other elements of water quality which show an increasing trend.

Investigation of observed time series proves that EC, Ca^{2+}, SO_4^{2-} and HCO_3^- show a significant increasing trend which is a sign for water quality deterioration in the region. Before data generation using one difference to eliminate the trend, stationary time series were prepared to work on. The results of modeling show that ARIMA modeling process is suitable in generating and forecasting the parameters.

Table 9. Results of forecasting 5 years of parameters.

Parameter	RMSE	VE %	R2
TDS	18.87	0.1	0.87
EC	16.19	0.07	0.49
HCO_3^-	0.09	0.04	0.91
SO_4^{2-}	0.10	0.61	0.48
Ca^{2+}	0.12	0.09	0.66
Na^+	0.04	0.32	0.9
pH	2.08	1.45	0.92
SAR	0.02	0.23	0.7

Reference

Ahmad, S., Khan, I.H., and Parida, B.P., Performance of stochastic approaches for forecasting river water quality, Water Research 35 (2001) 4261–4266.

Antonopoulos, V.Z., Papamichail, D.M., and Mitsiou, K.A., Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece, Hydrology and Earth System Sciences, 5 (2001) 679- 691.

Box, G.E.P. and Jenkins, G.M., Time Series Analysis, Forecasting and Control. Revised ed. Toronto: Holden-Day (1976).

Chow, W.T. and Kareliotis, S.J., Analysis of stochastic hydrologic systems. Water Resources Research, 16 (1970) 1569-1582.

Dalme, C. and Yalcin, A., Flood prediction using time series data mining, Journal of Hydrology, 333, (1970) 305-316.

El-Shaarawi, A.H., Esterby, S.R. and Kuntz, K.W., A statistical evaluation of trends in the water quality of the Niagara river, Journal of Great Lakes Research, 9 (1983) 234- 240.

Faruk, D.O., A hybrid neural network and ARIMA model for water quality time series prediction, Engineering Applications of Artificial Intelligence, 23 (2010) 586–594.

Gangyan, Z., Goel, N.K. and Bhatt, V.K., Stochastic modeling of the sediment load of the upper Yangtze river (Chaina), Hydrological Sciences Journal, 47 (2002) 93-105.

Gun, C. and Vilagines, R., Time series analysis on chlorides, nitrates, ammonium and dissolved oxygen concentrations in the Seine, The Science of the Total Environment 208 (1997) 59-69.

Halliday, S.J., Wade A.J., Skeffington, R.A., Neal, C., Reynolds, B., Rowland. P., Neal, M. and Norris, D., An analysis of long-term trends, seasonality and short-term dynamics in water quality data from Plynlimon, Wales, The Science of the Total Environment 434 (2012) 186–200.

Hanh, P.T.M., Analysis of variation and relation of climate, hydrology and water quality in the lower Mekong river, Water Science and Technology 62 (7) (2010), 1587–1594.

Hirsch, R.M., Slack J.R. and Smith, R.A., Techniques of Trend analysis for monthly water quality data. Water Resources Research, 18(1) (1982) 107- 121.

Irvine, K.N. Richey, J.E. Holtgrieve, G.W. Sarkkula, J. and Sampson, M., Spatial and temporal variability of turbidity, dissolved oxygen, conductivity, temperature, and fluorescence in the lower Mekong River–Tonle Sap system identified using continuous monitoring, International Journal of River Basin Management, 9:2, (2011) 151-168; http://dx.doi.org/10.1080/15715124.2011.621430.

Irvine, K.N. and Eberhardt, A.J., Multiplicative, seasonal ARIMA models for Lake Erie and Lake Ontario water levels, Water Resources Bulletin, 28 (2) (1992) 385–396.

Jalal Kamali, N., Forecasting the variations of inflow to Jiroft Dam using Time Series Theories, 6th international seminar on River Engineering, ShahidChamran University, Ahvaz, Iran, 2006.

Jamab Consulting Engineers, Integrated Program of Adaptation to Climate Study, Karkhe Watershed 1 (2005).

Jassby, A.D. Reuter, J.E. and Goldman, C.R., Determining long term water quality change in the presence of climate variability. Lake Tahoe (USA). Canadian Journal of Fisheries and Aquatic Sciences 60 (2003) 1452- 1461.

Karamouz, M. and Araghinejad, SH. Advanced Hydrology. Industrial University of Amir Kabir (Poly Technics), Tehran, Iran, Publication Centre of Amir Kabir University, (2005).

Khashei, M. and Bijari, M., An artificial neural network (p,d,q) model for time series forecasting, Expert Systems with Applications 37 (2010) 479–489.

Kim, J.-H., Lee, J., Cheong, T.-J., Kim, R.H. Koh, D.-C. Ryu, J.-S. and Chang,H.–W., Use of time series for theidentification of tidal effect on groundwater in the coastal area of Kimje, Korea, Journal of Hydrology 300 (2005) 188- 198.

Komornık, J., Komornıkova, M., Mesiar, R., Szokeova, D. and J. Szolgay. Comparison of forecasting performance of nonlinear models of hydrological time series. Physics and Chemistry of the Earth 31 (2006) 1127–1145.

Kurunc, A., Yurekli, K. and Cevik,O., Performance of two stochastic approaches for forecasting water quality and stream flow data from Yesilirmak River, Turkey, Environmental Modeling & Software 20 (2005) 1195–1200.

Lehmann, A. and Rode, M., Long-term behavior and cross-correlation water quality analysis of the River Elbe, Germany, Water Research 35 (2001) 2153–2160.

McKerchar, A.I. and Delleur, L.W.,. Application of seasonal parametric linear stochastic models to monthly flow data, Journal of Water Resource Reservoir 10 (1974) 246-255.

Montanari, A., Rosso, R., Taqqu, M.S., A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan. Journal of Water Resource Reservoir 36 (2000), 1249–1259.

Nelson, C. R. Applied Time Series Analysis for Managerial Forecasting. San Francisco: Holden-Day. 1973.

Padilla, A., Pulido-Bosch, A., Calvache, M.L. and Vallejos, A., The ARMA models applied to the flow of karstic springs, Journal of Water Resource Reservoir 32 (1996) 917–928.

Panda, D.K., Kumar, A. and Mohanty, S., Recent trends in sediment load of the tropical (Peninsular) river basins of India. Global and Planetary Change, 75 (2011) 108- 118.

Pankratz, A. Forecasting with Univariate Box-Jenkins Models. New York: John Wiley & Sons. (1983).

Papamichail, D.M., and Georgiou, P.E., Seasonal ARIMA inflow models for reservoir sizing. Journal of American Water Resources Association 37(2001) 877-885.

Papamichail, D.M., Antonopoulos, V.Z., and Georgiou, P.E., Stochastic models for Strymon river flow and water quality parameters. Proc. of International Conference “Protection and Restoration of Environment V”, I (2000) 219-226.

Rao, A.R., Kashyap, R.L., and Mao, L.-T, Optimal choice of type and order of river flow time series models. Water Resources Research, 18 (1982) 1097–1109.

Robson, A.J. and Neal, C., Water quality trends at an upland site in Wales, UK, (1983- 1993), Hydrological Processes, 10 (1996) 183-203.

Salas, J.D., Boes, D.C. and Smith, R.A., Estimation of ARMA models with seasonal parameters. Water Resources Research 18 (1982) 1006–1010.

Salas, J.D., Applied Modeling of Hydrologic Time Series, Littleton, CO: Water Resources Publications. 1980.

Sheng, H. and Y.Q. Chen., FARIMA with stable innovations model of Great Salt Lake elevation time series. Signal Processing, 91 (2011) 553–561.

Stansfield, B., Effects of sampling frequency and laboratory detection limits on the determination of time series water quality trends, New Zeland, Journal of Marine and Freshwater Research, 35 (2001).

Thomas, H.A. and Fiering, M.B., Mathematical synthesis of stream flow sequences for the analysis of river basin by simulation. Harward University Press. Cambridge, 1962.

Turner, B.F., Gardner, L.R. and Sharp, W.E., The hydrology of Lake Bosumtwi, a climate-sensitive lake in Ghana, West Africa, Journal of Hydrology 183 (1996) 243-261.

Vandaele, W. Applied Time Series and Box-Jenkins Models. New York: Academic Press, Inc. (1983).

Voudouris, K., Georgiou, P., Stiakakis, E. and Monopolis, D., Comparative analysis of stochastic models for simulation of discharge and chloride concentration in Almyroskartsic spring in Greece. e-Proceedings of the 14th Annual Conference of the International Association of Mathematical Geosciences, IAMG (2010), Budapest, Hungary, 1-15.

Webb, B.W., Clack, P.D. and Walling, D.E., Water- Air Temperature Relationships in a Devon River System and the Role of Flow. Hydrological processes, 17 (2003) 3069- 3084.

Weeks, W.D. and Boughton, W.C, Tests of ARMA model forms for rainfall-runoff modeling. Journal Hydrology 91(1987) 29–47.

Yu, Y-S, Zou, S. and Whittemore, D., Non parametric trend analysis of water quality data of rivers in Kansas. Journal of Hydrology, 260 (1993) 161-175.

Yurekli, K. and Kurunc, A., Performance of stochastic approaches in generating low streamflow data for drought analysis, Journal of Spatial Hydrology, 5 (2005) 20–32.

Zhang, G.P. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50 (2003)159–175.