Application of Time-series Model to Predict Groundwater Quality Parameters for Agriculture: (Plain Mehran Case Study)

Mir Mehrdad Mirmansari¹, Fatemeh Mohammadyari²

¹Department Environmental Sciences, Faculty of Natural Resources and Environment, Malayer University, Malayer, Iran
²PhD Student of Evaluation and land use Planning, Department Environmental Sciences, Faculty of Natural Resources and Environment, Malayer University, Malayer, Iran

corresponding author’s e-mail: mehrdadmirmansari@yahoo.com

ABSTRACT. Underground water is regarded as considerable water source which is mainly available in arid and semi-arid with deficient surface water source. Forecasting of hydrological variables are suitable tools in water resources management. On the other hand, time series concepts is considered efficient means in forecasting process of water management. In this study the data including qualitative parameters (electrical conductivity and sodium adsorption ratio) of 17 underground water wells in Mehran Plain has been used to model the trend of parameters change over time. Using determined model, the qualitative parameters of groundwater is predicted for the next seven years. Data from 2003 to 2016 has been collected and were fitted by AR, MA, ARMA, ARIMA and SARIMA models. Afterward, the best model is determined using information criterion or Akaike (AIC) and correlation coefficient. After modeling parameters, the map of agricultural land use in 2016 and 2023 were generated and the changes between these years were studied. Based on the results, the average of predicted SAR (Sodium Adsorption Rate) in all wells in the year 2023 will increase compared to 2016. EC (Electrical Conductivity) average in the ninth and fifteenth holes and decreases in other wells will be increased. The results indicate that the quality of groundwater for Agriculture Plain Mehran will decline in seven years.

1. Introduction
In terms of water resources and reduced water quality, water challenge with effects on agricultural production, energy conservation, as well as environmental protection has become a global issue [1]. Groundwater which is an important part of the reproducible water ecosystems most often existed in arid and semi-arid areas where surface water are very low [2]. Ground-water management is more difficult than surface water resource management; therefore, there is a need to seek for reasonable and cost-effective procedures to determine the status of these waters [3]. As a result, planning for the optimal use of water resources with the aim of achieving sustainable development is of utmost importance. A time series is defined as a sequence of observations ordered in time. The basic premise in the analysis of time series forecasting future values based on historical observations studied variables [4]. These models are widely used in hydrology and water quality parameters [5]. There are
some researchers who have conducted different predictions by using one time-series model \[6, 7\] and other scholars who have employed several time-series models \[8, 9\]. Time series models in different fields have gained the attention of researchers, such as air pollution \[5\], changes in groundwater levels \[10\] and water quality \[11\]. The combination of time-series models including AR, MA, ARMA, ARIMA and SARIMA to forecast ground-water quality parameters was provided by \[12\] (2014); but the combination of these models to predict water quality parameters has not been reported yet. Therefore, with respect to the conducted research studies and the ability of time-series techniques, this study is to examine the status of ground-water quality through different models of time series design by using 15 years of time series data of ground-water wells for agriculture in Mehran Plain, and eventually predict the status of ground-water quality parameters for the next 7 years (2021) via the best model obtained.

2. Materials and methods

2.1 Area of the Study
The study area (longitude: 46°05’ to 46°15’E, latitude: 32°59’ to 33°37’N) is located in Mehran plain, Ilam province, Iran. The Mehran plain has an area of 2391.9 km2. (Figure 1).

![Figure 1. Location of the study region.](image)

2.2 Methodology
Using data extracted from ground-water quality parameters for 17 wells in 2001 to 2015 in Mehran Plain, the variation process in these parameters were modeled and then the State of agriculture Plain were predicted for the next seven years via the selected model. Effective parameters in agriculture according to classification by Wilcox (1995), shown in Table 1 and Table 2 is used (electrical conductivity and sodium adsorption ratio). The data used were monthly data and continuous measurements; therefore, the time series under investigation was of continuous type. Drawing time series data is the first step in the analysis of time series design. The goal is to determine the presence or absence of the process in the data. In the next step, the components of the process in time series are determined and then they are removed in order to make the data static. After that the stationarity of the data is examined, the appropriate model is fitted to the data to identify the best model and accordingly make the prediction. The third step in the analysis of time series is the investigation into the normality of the data obtained from the prediction of the model; so that the Kolomogrov-Smirnov test was used to assess the normalization of data in this study. The models used in this study included AR, MA, ARMA, ARIMA and SARIMA. based on the models used in this study, there is no need to determine the equation of the line of goodness of fit for the data and the elimination of the process, and the models used remove the deletion process on the data by themselves. In the ARIMA and SARIMA models, the seasonal data status is also removed through differentiation. Therefore, in this study, R software was employed to determine the best time series model and ultimately forecast the data required by the selected model. Out of the 64 data quality parameters, 28 were simulated and set aside for the calibration of the model. To describe time dependence in the structure of a time series model,
autocorrelation functions (ACF) and partial autocorrelation (PACF) were investigated. After the determination of goodness of fit for the model, detection of the accuracy of the selected pattern is essential. According to the charts of ACF and PACF functions, the best models were determined via information criterion such as Akaike (AIC) and coefficient of determination (R2). AIC test is used to compare different models of ARMA (p, q) and it is calculated as follows (Mirzavand and Ghazavi, 2014):

\[
\text{AIC (k)} = n \ln (\text{MSE}) + 2K
\]  

(1)

Where n is the number of data points (for calibration), and k is the number of free parameters used in models. MSE stands for mean square error. Usually, the preferred model gives a higher R2 or the smallest value of AIC. After validation of the best model fitted on time series, to prediction future of observations are done. In the forecast process, the current time period is shown by t and t+\tau represents the prediction for the time period of t+\tau at the time of t. The forecast is made through considering the mean at the origin of t from the model written at the time of t+τ. In general, prediction is provided for the time period of t+\tau-1, ..., t+2+\tau+1. In this method, \(x_t+j\) which has occurred at the time of t is replaced with the predictions of \(x_t+j\) (t), and \(e_t+j\) that have not occurred at the time of t are substituted with zero. As well as, \(e_{t+j}\) that have not occurred are replaced with a single-period prediction error of e1(t-j)=\(x_{t-j}-\hat{x}_{t-j}(t-j-1)\) (Erdem and Shi, 2011). After selecting the best model and predicting values of quality parameters in 2021, the changes between actual (2015) and modeled year (2021) was. For this purpose, the experimental variogram of SAR and EC parameters in the first two years, 2015 and 2021 are calculated separately and the best model was selected for each parameter. After drawing and modeling variograms, we used interpolation technique with kriging in Arc GIS 10. Kriging interpolation is considered as an advanced method for the topical trends [13]. Finally, after obtaining the distribution of EC and SAR maps every two years, agricultural zoning map were generated and possible changes in the year 2021 were discussed.

2.3 Time Series Models
The auto-regressive model (AR) (p) as Eq. (2) can be expressed:

\[ z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \ldots + \varphi_p z_{t-p} + a_t \]  

(2)

Where \(\varphi_1, \varphi_2 \) and \(\varphi_p\) are coefficients and model parameters and The random term of the data follows normal distribution with a zero mean. The model MA can be expressed as Eq. (3):

\[ z_t = \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q a_{t-q} + a_t \]  

(3)

Where \(\theta_1, \theta_2 \) and \(\theta_q\) are coefficient and model parameters and at is a random term of the data that follows by normal distribution with a zero mean [14]. ARMA (p, q) model also can be obtained from the following equation (4):

\[ Y_t = \delta + \sum_{i=1}^{p} \theta_i y_{t-1} + \sum_{j=1}^{q} \varphi_j e_{t-j} + e_t \]  

(4)

Where \(\delta\) is the constant term of the ARMA model, \(\varphi_i\) indicates the ith auto-regressive coefficient, \(\varphi_j\) is the jth moving average coefficient, et shows the error term at time period t, and \(Y_t\) refers the value of groundwater level observed or forecasted at time period t [15]. ARIMA models originated from the combination of autoregressive models (AR) and moving average models (MA). ARIMA fits a Box-Jenkins ARIMA model to a time series [6]. In ARIMA, the future value of a variable is assumed to be a linear function of several past observations and random errors. A SARIMA model can be explained as ARIMA (p, d, q) (P, D, Q)s, where (p, d, q) is the non-seasonal part of the model and (P, D, Q)s is the seasonal part of the model in which p is the order of non-seasonal autoregression, d is the number of regular differencing, q is the order of non-seasonal MA, P is the order of seasonal auto-regression, D is the number of seasonal differencing, Q is the order of seasonal MA, and s is the length of the season [11].
12 models on groundwater parameters, the results, for four wells are shown in Table 3.

3. Result and Discussion

12 models on groundwater parameters, the results, for four wells are shown in Table 3.

Table 1. Classification criteria in terms of agricultural water quality

| Water quality | EC | Class | SAR | Class |
|---------------|----|-------|-----|-------|
| Excellent     | <250 | C_1   | <10 | S_1   |
| Good          | 250-750 | C_2   | 10-18 | S_2   |
| Intermediate  | 750-2250 | C_3   | 18-26 | S_3   |
| Unsuitable    | >2250 | C_4   | >26  | S_4   |

Table 2. Various categories and types of water quality classification based on Wilcox [16]

| Row | Water Category | Quality of water for agriculture |
|-----|----------------|----------------------------------|
| 1   | C1S1           | Sweet-for Agricultural quite harmless |
| 2   | C2S1,C2S2,C1S2 | Little salty-for Agricultural almost perfect |
| 3   | C3S3,C3S2,C3S1,C2S3,C1S3 | Salted-for agriculture by applying the necessary measures |
| 4   | C4S1,C4S2,C4S3,C4S4,C4S2,C4S4,C1S4 | Very Salt harmful for agriculture |

Table 3. Results of time series models in 4 Wells

| Used models | Well 1 | | Well 2 | | Well 3 | | Well 4 |
|-------------|--------|--------|--------|--------|--------|--------|--------|
| AR(1)       | AIC    | R2     | AIC    | R2     | AIC    | R2     | AIC    | R2     |
| SAR          | -0.1   | 509.0  | -0.1   | 415.6  | -0.1   | 342.7  | -0.1   | 479.5  |
| EC           | -0.3   | 22.5   | -0.3   | 9.77   | -0.3   | 9.76   | -0.3   | 9.37   |
| AR(2)       | AIC    | R2     | AIC    | R2     | AIC    | R2     | AIC    | R2     |
| SAR          | -0.3   | 799.0  | -0.3   | 9.77   | -0.3   | 9.76   | -0.3   | 9.37   |
| EC           | -0.5   | 9.79   | -0.5   | 9.76   | -0.5   | 9.76   | -0.5   | 9.37   |
| MA(1)       | AIC    | R2     | AIC    | R2     | AIC    | R2     | AIC    | R2     |
| SAR          | -0.3   | 799.0  | -0.3   | 9.77   | -0.3   | 9.76   | -0.3   | 9.37   |
| EC           | -0.1   | 0.844  | -0.1   | 0.672  | -0.1   | 0.844  | -0.1   | 0.672  |
| MA(2)       | AIC    | R2     | AIC    | R2     | AIC    | R2     | AIC    | R2     |
| SAR          | -0.1   | 0.843  | -0.1   | 0.650  | -0.1   | 0.843  | -0.1   | 0.650  |
| EC           | -0.1   | 0.843  | -0.1   | 0.650  | -0.1   | 0.843  | -0.1   | 0.650  |

To predict future fluctuations in groundwater quality series validation models for each time series was used for a period of 7 years. Figure 2 an example of real and simulated of that SAR and EC of the well third shows.
Figure 2. Models prediction versus observed data in wells 3, Bottom: SAR Top: EC

Models were next fitted on parameters, the appropriate model to predict ground-water parameters were selected using Akaike and correlation coefficient indices. As shown in Figure 2 and Table 4, ARMA, AR and MA was selected as suitable model.

| Wells          | Parameters | Models                  |
|----------------|------------|-------------------------|
| 1, 5, 9, 13, 16| EC         | ARMA(2,1)               |
|                | SAR        | AR(2)                   |
| 2, 8, 11       | EC         | ARMA(2,2)               |
|                | SAR        | MA(2)                   |
| 3, 7, 12, 15, 17| EC       | MA(1)                   |
|                | SAR        | AR(2)                   |
| 4, 6, 10, 14   | EC         | AR(2)                   |
|                | SAR        | MA(2)                   |

According to the results, the mean SAR predicted in the years 2022 to 2015 all wells will increase. EC average decreases in the ninth and fifteenth wells and will be increased in other wells. with the best chosen models by using software ARC GIS10.3, EC and SAR parameters were interpolated by kriging(Figure 3).
Finally groundwater quality in terms of agricultural zoning map was produced according to the Wilcox [16] are shown in Figure 4 and Table 5.

| Year | Class     | Area (km²) | Area (%) | Area (km²) | Area (%) |
|-----|-----------|------------|----------|------------|----------|
| 2015| Excellent | 206.61     | 9        | 279.7      | 12       |
|     | Good      | 1938.3     | 80       | 1557.99    | 65       |
|     | Intermediate | 162.38   | 7        | 404.98     | 17       |
|     | Unsuitable | 84.63      | 4        | 149.13     | 6        |
|     | Total     | 2391.9     | 100      | 2391.9     | 100      |

4. Conclusions
To predict the parameters of ground-water quality using monthly data, the quality parameters were analyzed into four sections including random, seasonal, trend and observed; and the random parameter was selected as the model. Additionally, five models with 12 sub-models were evaluated. AIC and R2 criteria for model selection were used to make more accurate predictions. Based on the results
illustrated in Table 3 and according to AIC and R2 coefficient values for the different models examined, the final models of quality parameters were determined for each cluster (Table 4). In these models, R2 and AIC had respectively the maximum and the minimum values and the absolute value for the parameters of the selected models did not exceed 1. The results show that the time series models in the modeling of spatial and temporal variations of groundwater quality. The time-series models, four models of ARMA (1,1), ARMA (2,1), ARMA (2,2) and SARIMA (1,1,1) (1,1,1) [4] for ground-water quality parameters in mehran Plain were identified through the autocorrelation and partial correlation functions. The appropriateness of the model was confirmed by the analysis of the remaining fitted model. Normality test results showed normality of the data was predicted. From the results it is expected in 2022, the EC is higher than the standard for agriculture. The SAR value is higher than the standard in all wells for agriculture. The comparison of the data from model that is predicted by different methods and the original data are shown in Table 3 and Figure 2. These data demonstrate that the selected time series models have had a good performance in predicting the time series of ground-water quality parameters which are also consistent with the findings by [17, 9, 6]. The distribution maps of qualitative indicators modeled of the year 2022 reflects the deteriorating quality of groundwater resources in the future years. as indicated in Figure 3, EC value is appropriate in all parts of the plain in 2015 but it is gradually decreased and in the year 2022 a small part of the plain possesses relatively adequate value and the majority of it which lie on the floor would be undesirable. The SAR region in 2022 and a few changes in gradation, as compared to the year 2015 follow the trend of declining quality stems. The final map in 2015 due to higher agricultural area for the middle class and 4% of the area is unsuitable for agriculture. But in 2022 the extent of suitable area for agriculture will be reduced that reflects alarming circumstance of water quality for agricultural utilization (Table 5). Irregular exploitation of groundwater aquifers and unauthorized wells are the most important reasons of quality decrease in Mehran Plain. Protecting approach may include pressurized irrigation projects, recharge artificial wells and installing water meters on agriculture, unauthorized filling of wells and finally raising awareness. Furthermore, considering the studied area which is located in arid and semi-arid area and dealing with water scarcity, the results shows a clear picture of the groundwater quality parameters in future and can be used by professionals and planners.

References

[1] Wu D, Ekama GA, Wang HG, Wei L, Lu H, and Chui H 2014 Simultaneous nitrogen and phosphorus removal in the sulfur cycle-associated Enhanced Biological Phosphorus Removal (EBPR) process. Water Research 49:251-264.

[2] Nas B, and Berktay A 2008 Groundwater quality mapping in urban groundwater using GIS. Environ monit Assess. 160 (1–4): 215–227

[3] Hajihashemi Jazy MR, Atashgahi M, and Hamidian AH 2011 Spatial estimation of groundwater quality factors using geostatistical methods (case study: Golpayegan plain). Journal of Natural Environmental, Iranian Journal of Natural Resources, 4: 347-357

[4] Taneja K, Ahmad sh, Ahmad K, and Attri SD 2016 Time series analysis of aerosol optical depth over New Delhi using BoxeJenkins ARIMA modeling approach, Atmospheric Pollution Research, 3: 1-12

[5] Wang H R, Wang C, Lin X, and Kang J 2014 An improved ARIMA model for precipitation simulations. Nonlinear Process. Geophys. 21: 1159-1168.

[6] Shirmohammadi B, Vafakhah M, Moosavi V, and Moghaddamnia A 2013 Application of several data-driven techniques for predicting groundwater level. Water Resour Manag 27:419–432.

[7] Adamowski J, and Chan HF 2011 Awavelet neural network conjunc tion model for groundwater level forecasting. Hydrol 407(1–4):28–40. doi:10.1016/j.jhydrol.2011.06.013

[8] Parmar KS, and Bhardwaj R 2014 Water quality management using statistical analysis and time-series prediction model. Appl Water Sci. 4: 425–434.
[9] Malekian A, Chobin B, Sajedi Hossini F, and Rahmati A 2015. Prediction water table using time series and Fuzzy Inference System - Adaptive Neural, Journal of Soil and Water Research Iran. 45: 19-48.
[10] Panda D K, and Kumar A 2011 Evaluation of an over-used costal aquifer (Orissa, India) using statistical approaches. Hydrol. Sci. Jour., 56: 486-497.
[11] Faruk D 2010, A hybrid neural network and ARIMA model for water quality time series prediction. Eng Appl Artif Intell 23(4):586–594. doi:10.1016/j.engappai.2009.09.015
[12] Mirzavand M, and Ghazavi R 2014 Stochastic Modelling Technique for Groundwater Level Forecasting in an Arid Environment Using Time Series Methods, journal Water Resources Management. Doi: 10.1007/s11269-014-0875-9 Online ISSN 1573-1650.
[13] Gahrevaridi Tally M 2005 In the three-dimensional geographic information system ArcGIS, Publications Teacher Training University Jahad.
[14] Hannan EJ 1971 Multiple time series. Wiley, New York
[15] Erdem E, and Shi J 2011 ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 84(2):1405–1414. doi:10.1016/j.apenergy.2010.10.031
[16] Wilcox L V, 1955 Classification and use of irrigation
[17] Mirzaee S, Chitsazan M, Chinipardaz R, and Samady H 2010 Forecast plain SHAHREKORD using time-series models and examine ways to improve, The first regional conference on optimal utilization of water resources and river basins Karon. 46: 1-8.