Original Research Paper

Analyzing the Groundwater Quality Parameters Using Frequency Analysis

[1]Maryam Ghashghaie, [2] Kaveh Ostad-Ali-Askari, [3]Saeid Eslamian and [4]Vijay P. Singh

[1]Department of Water Resources Engineering, Faculty of Agriculture,
Bu-Ali Sina University, Hamedan, 6517833131, Iran
[2]Department of Civil Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran
[3]Department of Water Engineering, Isfahan University of Technology, Isfahan, Iran
[4]Department of Biological and Agricultural Engineering and Zachry Department of Civil Engineering,
Texas A and M University, 321 Scoates Hall, 2117 TAMU, College Station, Texas 77843-2117, U.S.A

Abstract: Water quality problems are world spread as a result of land use changes and industrial development. Water deficit problems become more severe in result of water quality ignorance. Determining the probability values of water quality parameters for specific return periods could be very helpful in water resources management and conservation in terms of quality. In this research some water quality parameters such as: SO_4^{2-}, Cl^-, HCO_3^-, TDS, TH, SAR, EC, Mg^{2+}, Ca^{2+}, K^+ and Na+ were selected from 7 wells to study through which the most convenient statistical distribution function was determined for each parameter using EASYFIT software. To reach this aim 9 functions were used and the most suitable function was identified through Kolmogorov-Smirnov test. At the second stage the value of each parameter was determined for 5, 10, 20, 50 and 100 year return periods. Results show that Wake by fits the best for most of parameters although GEV and Gen. Logistic are the second and third functions which best fit to data. Also return period graphs of each parameter in 7 sites show that SO_4^{2-} follows the steepest slope as well as SAR and Cl^-.

Keywords: Water Quality Parameters, Frequency Analysis, Wake By, Return Period

Introduction

Healthy and safe water is a necessary need for human beings. Our healthy life is highly dependent on the quality of water both for domestic and agricultural use. As water passes through different layers and constructions of soil in a vertical movement, the quality of groundwater decreases. Constituent minerals of soil solve in groundwater pass way and deteriorate the quality (Todd and Mays, 2005).

Nowadays water resources are exposed to more resources of pollution and less safe water is available in comparison with past decades. Water quality management is highly dependent on controlling the pollution originating from human activities. Industrial improvements usually decrease natural healthy resources especially water bodies as it is ignored in many cases unfortunately. Therefore an efficient water resources management system is required to predict and control the quality of natural water resources for future. There are many methods to predict the quality of water and in this study the methodology of frequency analysis was adopted to find a convenient function which best fits on quality data. Identifying the best frequency function which fits a special parameter and forecasting the probability of parameter occurrence the most important parameter in terms of quality deterioration is determined. In this research 11 water quality parameters were studied such as: $SO4^{2-}$, Cl^-, HCO_3^-, TDS, TH, SAR, EC, Mg^{2+}, Ca^{2+}, K^+ and Na^+.

SO_4^{2-} which is one of Onions originates from different sources. Its contribution from air is nearly 2ppm although it varies highly through movement in groundwater. These variations originate from permutation, sedimentation, solution and condensation. Any increase in groundwater SO_4^{2-} could be in result of pyrite oxidation, sulfide minerals Pyrite shills, lignite, coal, gypsum sediments and their oxidation.

The second important element is Cl- which could be increased in result of soil water drainage by desalinated

water in watersheds. Also high values of Cl⁻ in groundwater could be in result of pollution by waste water. Some cultivars such as coconut cause more salinity in soil and consequently the remained salt of soil is drained towards water bodies. Cl⁻ is considered as a salinity parameter in some researches (Hajrasuliha *et al*., 1991). Also Total Dissolved Solid (TDS), Total Suspended Solids (TSS) and Electrical Conductivity (EC) are considered as salinity parameters by some researchers (Tanji, 1990).

HCO_3^- is another important element and its initial origin comes from solute CO_2 in the rain, snow and soil. Bicarbonate deposits in soil and stone pores due to little variations of relative pressure on CO_2. TDS Solids is known as a parameter of water quality which shows total concentration of dissolved solids. Ca^{2+} is one of the main cations in groundwater, found in majority of igneous and vicissitudinou stones. The ratio of Ca^{2+} to Mg in sea water is equal to 1 to 5. Consequently high Mg in near sea groundwater could be in result of aquifer pollution by sea water.

The amount of Na or alkali risk is defined by Na absorption ratio (Gholami and Srikantaswamy, 2009). The percent of solution for Na is actually the percent of Na ratio to total kations and this ratio will be higher than 60 percent if soil structure is destroyed and Na is highly filled in result of weathering and infiltration (Hakim *et al*., 2009).

Na^+ concentration is very important in terms of irrigation water quality as high values of Na reduce soil permeability (Todd and Mays, 2005). The amount of soluble Na and EC are highly important in groundwater quality in terms of agricultural use (Khodapanah *et al*., 2009). Sodium bicarbonate is important in accepting the quality of water for irrigation (Bokhari and Kan, 1992). Permeability index is also an important factor in water quality evaluation for irrigation (Doneen, 1962). Total ratio is a parameter which evaluates the quality of water based on the value of Na to Ca^{2+} and Mg^{2+} (Kelley, 1940).

The more amount of Na^+ and Cl⁻ in groundwater the more dissolved substances in water. The amount of Na in is hundreds times more than Ca^{2+} and Mg in saline water.

K^+ varies between 1 and 10-15 ppm in potable water and between 100 and thousands ppm in some saline waters.

Many studies have been accomplished on rainfall and discharge for frequency analysis as well as flood frequency analysis round the world (Parent and Bernier, 2003; Reis and Stedinger, 2005; Payrastre *et al*., 2011; 2013;; Strupczewski *et al*., 2014; Lázaro *et al*., 2016; Jun *et al*., 2017; Gado *et al*., 2017; Strupczewski *et al*., 2017; Chen and Singh, 2018).

The estimation of return periods of hydrological events and the corresponding risks of failure in estimating such events are important aspects in many water resources studies in terms of quantity and quality (Fernandez and Salas, 1999). Frequency analysis has been used in hydrologic works. Rezaee (2001) showed

the application of common distributions functions in water resources such as daily, monthly and yearly precipitation, Rainfall, temperature, discharge and peak flows. Kroll and Vogel (2002) used it for low flows and reported that the distribution of low flows is unknown. However they implied that the United Stated Geological Survey uses LP3 distribution for frequency analysis of low flow as the best function. Waltemeyer (2002) used $Q_{4,3}$, which is the lowest 4-conccecutive-day discharge having reoccurrence interval of three years to design and administrate water quality standards in New Mexico, USA.

Kadri *et al*. (2005) derived appropriate probability distributions for frequency analysis of 7-day annual low flows at three gauging stations of the Çekerek Stream. Also there are many works on rainfall frequency analysis. Bhakar *et al*. (2005) used frequency analysis of maximum rainfall on consecutive days in Banswara in India to estimate the maximum value of rainfall for 1, 2 to 5 consecutive days in a year for different return periods.

The present study aims to apply this methodology for water quality parameters. Determining the value of each parameter helps to understand the behavior of groundwater resources in terms of quality and threatens human being for preventing from any hazardous treatment with water resources. In this study Easy Fit software was used to analyze the frequency of 11 groundwater quality parameters of Alashtar plane.

The Study Area and Methodology

The Study Area

The study area is called Alashtar plane which is located in Kashkan watershed. Alashtar plane lies between 33° 43"- 34° 5" N and 48° 2"- 48° 31" E and is shown in Fig. 1.

The aim of the frequency analysis for events in hydrology is to achieve the probability of an event occurrence such as: Maximum 24-h rainfall. Maximum series or partial series of data are needed for this aim and theoretical probability functions are fitted on these series. For the frequency analysis at first the data are sorted descending and the probability of occurrence is calculated using probability functions such as weibul (Alizade, 2006).

EASYFIT software was used in this study to identify the best fit to apply as a prediction tool for water quality parameters. This software supports over 50 continuous and discrete probability distribution functions. There are a number of well-known methods which can be used to estimate distribution parameters based on available sample data. For every supported distribution, EASYFIT implements one of the following parameter estimation methods: Method Of Moments (MOM); Maximum Likelihood Estimates (MLE); Least Squares Estimates (LSE); and method of L-moments.
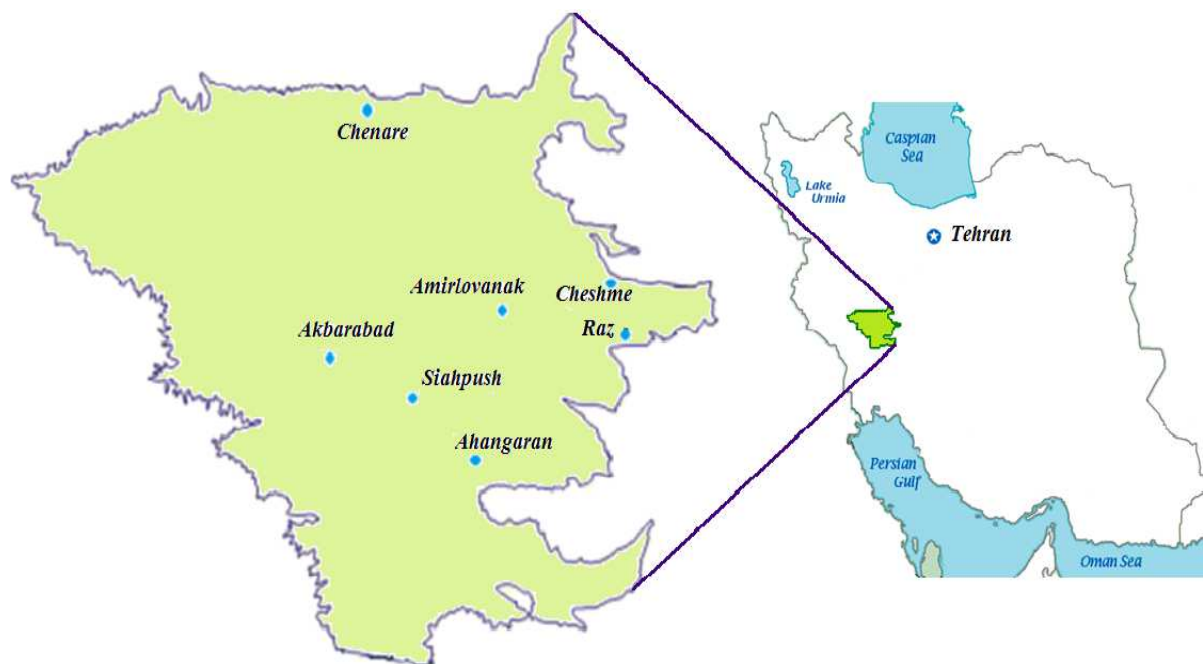
Fig. 1: The study area of Alashtar

EASYFIT uses the least computationally intensive methods. Thus, it employs the method of moments for those distributions whose moment estimates are available for all possible parameter values and do not involve the use of iterative numerical methods.

For many distributions, EASYFIT uses the MLE method involving the maximization of the log-likelihood function. For some distributions, such as the 2-parameter Exponential and the 2-parameter Weibull, a closed form solution of this problem exists. For other distributions, EASYFIT implements the numerical method for multi-dimensional function minimization. Given the initial parameter estimates vector, this method tries to improve it for subsequent iteration. The algorithm terminates when the stopping criteria is met (the specified accuracy of the estimation is reached, or the number of iterations reaches the specified maximum).

Totally to fit a probability function on a series of observed data at first a model is selected. This selection could be either based on the experience of observed data analysis or different tests. After the model selection the parameters of distribution function are identified and a new series is generated using those parameters. Then the verification of generated data is evaluated through goodness of fit tests (Sharifi and Shahidi, 2001). Clearly there is no especial distribution function which fits on observed data thoroughly however a function is selected through comparison with other functions (Shahabfar and Ghiami, 2004).

The present study is composed of two main stages; at first the data are fitted using 9 theoretical distribution functions. Then the best fit is adopted using goodness of fit test of Kolmogorov- Smirnov. Finally the value of each parameter is determined based on the selected distribution function.

At the second stage a diagram is provided for each well which demonstrates the estimated values of parameters for different return periods. These diagrams show that which parameter of water quality in a well has a sharper slope and consequently is riskier. Also a diagram is prepared for each parameter which shows estimated values of a parameter for different return periods in 7 wells. Through these diagrams it is easily found that which well is more exposed to the risk of abnormal excess.

The template is used to format your paper and style the text. All margins, column widths, line spaces and text fonts are prescribed; please do not alter them. Your paper is one part of the entire proceedings, not an independent document. Please do not revise any of the current designations.

Text font. The entire document should be in times New Roman font size 10. Paper title must be Left size, bold, regular font size 18 and the first letter of each word capitalized. Author names must be regular font size 10, bold. Author affiliation must be regular font size 9. Email addresses font size 9. Level 1 main headings must be Left size, bold, regular font size 12 and first word capitalized. Level 2 Sub headings must be left-justified, bold, italic, font size 11 and the first letter of each word capitalized.

Goodness of Fit Tests

Goodness of fit tests is used to verify the generated data by distribution functions. In EASYFIT software there are 3 tests for goodness of fit which are mentioned in the following steps briefly.

Kolmogorov-Smirnov Test

This test is used to decide if a sample comes from a hypothesized continuous distribution. It is based on the Empirical Cumulative Distribution Function (ECDF). Assuming that a random sample $x_1,..., x_n$ from some distribution with CDF $F(x)$, the empirical CDF is denoted by Equation (1):

$$F(x) = \frac{1}{n}[Number\ of\ Observation \leq x] \tag{1}$$

The Kolmogorov-Smirnov statistic (D) is based on the largest vertical difference between the theoretical and the empirical cumulative distribution function Equation (2):

$$D = \max_{1 \leq i \leq n}\left(F(xi) - \frac{i-1}{n}, \frac{i}{n} - F(xi)\right) \tag{2}$$

Anderson-Darling Test

The Anderson-Darling procedure is a general test to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. This test gives more weight to the tails than the Kolmogorov-Smirnov test. The Anderson-Darling statistic (A2) is defined as Equation (3):

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F(X_i) + \ln(1 - F(X_{n-i+1}))\right] \tag{3}$$

Chi- Squared Test

The Chi-Squared test is used to determine if a sample comes from a population with a specific distribution. This test is applied to binned data, so the value of the test statistic depends on how the data is binned. This test is only available for continuous sample data. Although there is no optimal choice for the number of bins (k), there are several formulas which can be used to calculate this number based on the sample size (N). For example, EASYFIT employs the following empirical formula Equation (4):

$$k = 1 + \log_2 N \tag{4}$$

The data can be grouped into intervals of equal probability or equal width. The first approach is generally more acceptable since it handles peaked data much better (changing the binning method in the Fitting Options dialog is possible). Each bin should contain at least 5 or more data points, so certain adjacent bins sometimes need to be joined together for this condition to be satisfied. The Chi-Squared statistic is defined as Equation (5):

$$X_2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{5}$$

where, O_i is the observed frequency for bin i and E_i is the expected frequency for bin i calculated by Equation (6):

$$E_i = F(x_2) - F(x_1) \tag{6}$$

where, F is the CDF of the probability distribution being tested and x_1, x_2 are the limits for bin i. However this test is reliable were the number of samples are enough large so that at least 5 data is categorized in each Class, unless Kolmogorov-Smirnov test is advisable.

In present study Kolmogorov-Smirnov test was used and the results of goodness of fit were based on this test.

Quality parameters of 7 wells in the study area were used and the best fit function for each parameter was found for each well. A 25 yearly data is available for each parameter. About 9 functions were examined in this study to fit data including: GEV, Generalized Logistic, Gumbel Max, Gumbel Min, Log Pearson 3, Logistic, Normal, Pearson 5 and Wake by. The best function is determined among these fits to use for predicting future amounts of each parameter. Probability function of each parameter was determined through EASYFIT at the first stage and for return periods of 2, 5, 10, 50 and 100 the value of each parameter was predicted. Finding the Value of each parameter for different return periods has two advantages: 1- The most possible (probable) value of a parameter is known for the future 2- the parameter which varies faster and has a sharper slope is determined.

Results

Totally 7 wells were examined in this study the first of which is called Ahangaran. For this sampling well, the result of parameter prediction for 2, 5, 10, 50 and 100 return periods are shown in Fig. 2a. It is obvious that SO_4^{2-}, Cl^- and EC show a sharper slope while other parameters vary gradually.

Figure 2b shows the values of parameters for Amirlovank well in determined return periods.

As it is seen in Fig. 2b SO_4^{2-}, EC and TDS show a sharper slope in comparison with other parameters for this well.

The results for Akbarabad well are shown in Fig. 2c as well.

For Akbarabad well SO_4 varies with a sharper slope among other parameters.

Probability values for groundwater quality parameters of Chenare well are shown in Fig. 2d. SAR, SO_4^{2-} and EC demonstrate sharper slopes in comparison with the others.

Also probability values for defined return periods of water quality parameters in Cheshme, Raz and Siahpush wells are presented in Fig. 2e, 2f and 2g respectively.

In Cheshme well SAR, Na^+, K^+ and SO_4^- show a sharper slope in comparison with other parameters.

Figure 2g show that only SO_4^- has a sharp slope which means there is a danger of $SO4^{2-}$ exceeding in groundwater resources.

The frequency analysis for quality parameters of last well which is called Siahpush, show that SO_4^{2-} values follow a sharper slope.

Also a diagram was prepared for each well based on a parameter to show that which site is more sensitive in terms of groundwater quality. Figure 3 shows the results of this section.

486
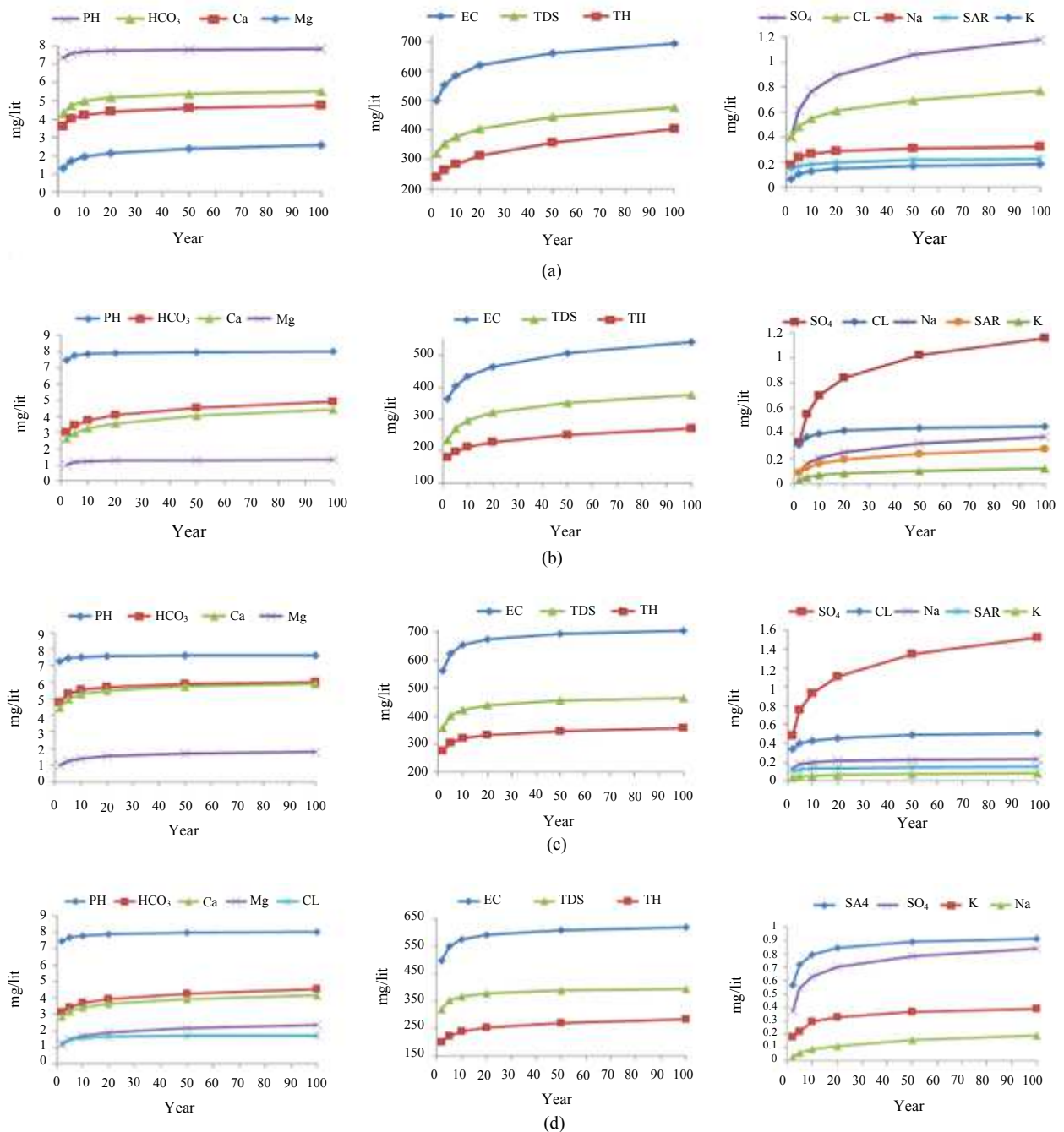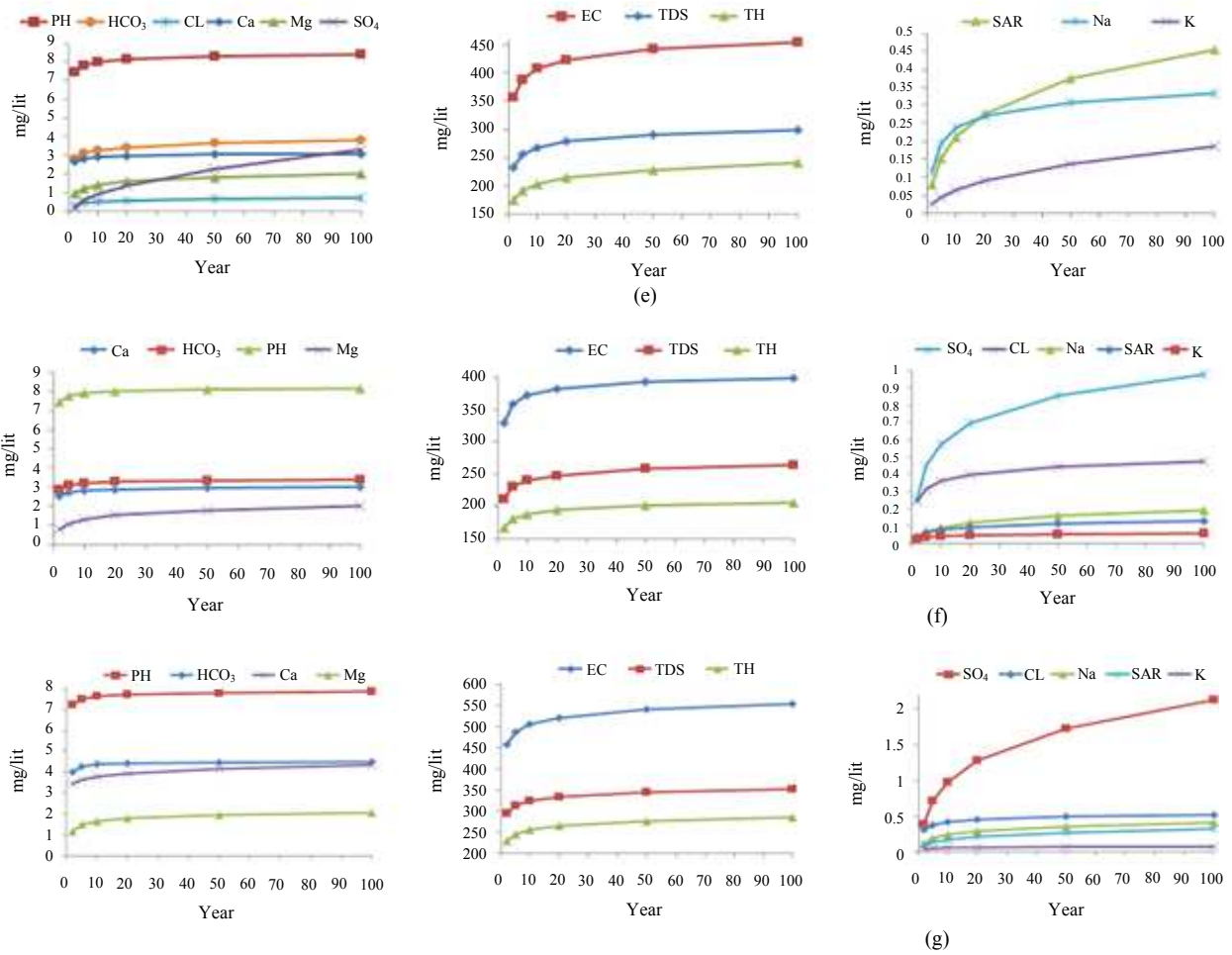
Fig. 2: a,b,c,d,e,f,g- probability values of SO_4^{2-}, Cl^-, HCO_3^-, TDS, TH, SAR, EC, Mg^{2+}, Ca^{2+}, K^+ and Na^+ for return periods; Ahangaran, Amirlovank, Akbarabad, Chenare, Cheshme, Raz and Siahpush respectively
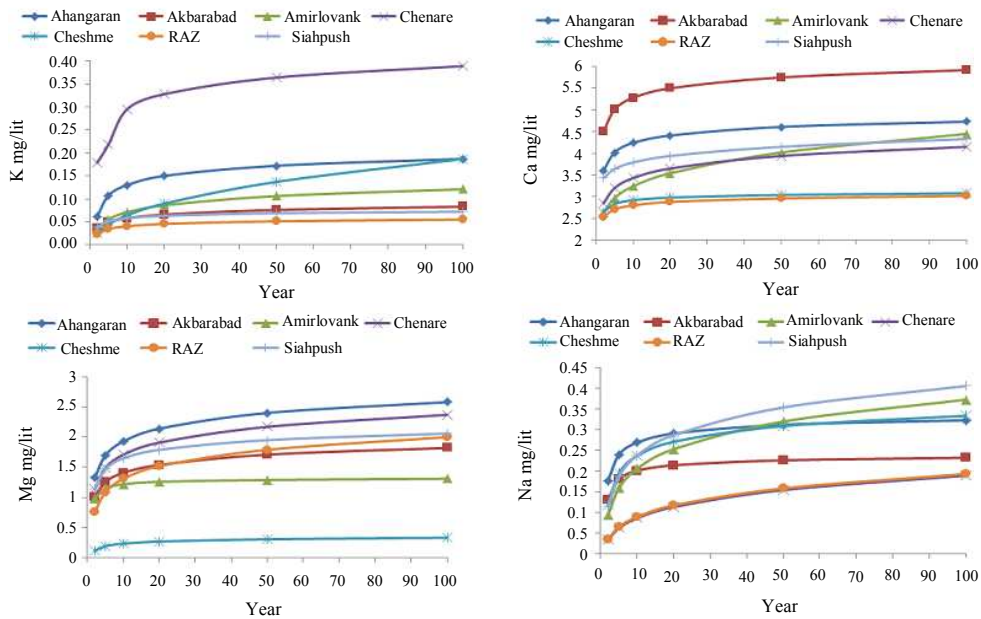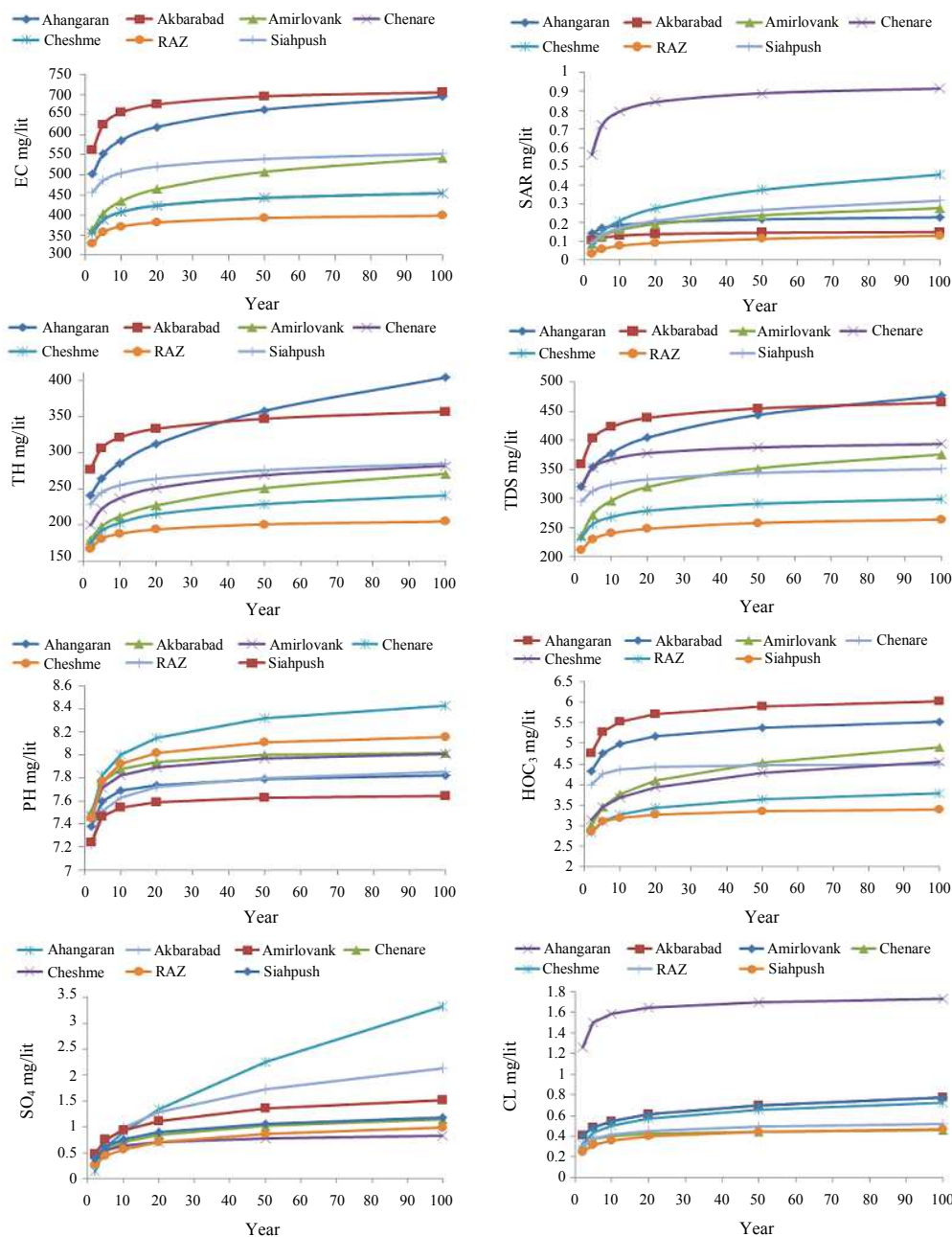
Fig. 3: Probability values of SO_4^{2-}, Cl-, HCO_3-, TDS, TH, SAR, EC, Mg^{2+}, Ca^{2+}, K+ and Na+ for return periods; all sites

As it is shown in Fig. 3 the estimated probability values of each parameter for different return periods were plotted in all sites. K^+ has the higher value in Chenare well and the values of Ca^{2+} are more significant in Akbarabad well. Mg^{2+} has shown to have the higher values in Ahangaran and Na^+ in Siahpush well is more in comparison with other sites. EC is relatively high in Akbarabad and SAR values of Chenare are absolutely significant. pH of Ahangaran and Akbarabad are considerable. TDS of Akbarabad is relatively more than

other sites. pH values of Cheshmeh lie higher than the other sites and HCO_3^- in Akbarabad is more than the others. SO_4^{2-} follows a severe slope in Cheshme and Cl- values are absolutely significant in Chenare Site.

Discussion

Frequency analysis of water quality parameters was accomplished in present study. The study area was located in the west of Iran where there are fresh resources of water.

Conclusion

Based on field studies existence of karstic and alluvial constructions has provided a good resources of safe water in high amounts which is very useful in drought periods.

However, land use plays an important role in groundwater of the region in terms of quantity and quality. High growth in the region as well as water overdrawn, inconvenient methods of burying litters, agricultural fertilizers and pesticides are considered as the major reasons of ground water quality deterioration. As it is demonstrated in Fig. 3, Abarabad site is exposed to serious water quality problems in near future. Also the danger of water quality aggravation is increasing in result of high population growth in the region and efficient actions are necessary in the region to prevent more deterioration of groundwater quality.

Author's Contributions

Maryam Ghashghaie: Write manuscript and analyzed data.

Saeid Eslamian and Vijay P. Singh: Design the study and revise manuscript.

Kaveh Ostad-Ali-Askari: Write manuscript and revise it.

Ethics

This study was approved by Bu-Ali Sina University, Hamedan, 6517833131, Iran.

References

Alizade, A., 2006. Principals of applied hydrology. Imam Reza University, Mashhad, Iran.

Bhakar, S.R., A.K. Bansal, N. Chhajed and R.C. Purohit, 2005. Department of soil and water engineering CTAE Udaipur, Rajasthan, India.

Bokhari, A.Y. and M.A. Kan. 1992. Deterministic modeling of al-Madinah groundwater quality using lumped parameter approach. King Abdul Aziz Uni. Earth Sci., 5: 89-107. DOI: 10.4197/Ear.5-1.5

Chen, L. and V.P. Singh, 2018. Entropy-based derivation of generalized distributions for hydrometeorological frequency analysis. J. Hydrol., 557: 699-712. DOI: 10.1016/j.jhydrol.2017.12.066

Doneen, L.D., 1962. The influence of crop and soil on percolating water. Proceeding of Biennial Conference on Groundwater Recharge, (CGR' 62).

Fernandez, B. and J.D. Salas, 1999. Return period and risk of hydrologic events: 1. Mathematical formulation. ASCE J. Hydrol. Eng., 4: 297-307. DOI: 10.1061/(ASCE)1084-0699(1999)4:4(297)

Gado, T.A., K. Hsu and S. Sorooshian, 2017. Rainfall frequency analysis for ungauged sites using satellite precipitation products. J. Hydrol., 554: 646-655. DOI: 10.1016/j.jhydrol.2017.09.043

Gholami, S. and S. Srikantaswamy, 2009. Analysis of agricultural impact on the Cauvery River water around KRS dam. World Applied Sci. J., 6: 1157-1169.

Hajrasuliha, S., D.K. Cassel and Y. Rezainejad, 1991. Estimation of chloride ion concentration in saline soils from measurement of electrical conductivity of saturated soil extracts. Geoderma, 49: 117-127. DOI: 10.1016/0016-7061(91)90095-B

Hakim, M.A., A.S. Juraimi, M. Begum, M. Hasanuzzaman and M.K. Uddin *et al*., 2009. Suitability evaluation of groundwater for irrigation drinking and industrial purposes. Am. J. Environ. Sci., 5: 413-419. DOI: 10.3844/ajessp.2009.413.419

Jun, C., X. Qin, T.Y. Gan, Y.K. Tung and C.D. Michele, 2017. Bivariate frequency analysis of rainfall intensity and duration for urban stormwater infrastructure design. J. Hydrol., 553: 374-383. DOI: 10.1016/j.jhydrol.2017.08.004

Kadri, Y., K. Ahmet and G. Selçuk. 2005. Frequency Analysis of Low Flow Series from ˡekerek Stream Basin. Tarim Bilimleri Dergisi, 11: 72-77

Kelley, W.P., 1940. Permissible composition and concentration of irrigation water. Am. Society Civil Eng., 66: 607-613.

Khodapanah, L., W.N. Sulaiman and M. Khodapanah, 2009. Ground water quality assessment for different purposes in Eshtehard district, Tehran, Iran. Eur. J. Sci. Res., 36: 543-553.

Kroll, C.N. and R.M. Vogel, 2002. Probability distribution of low stream flow series in the United States. J. Hydrol. Eng., 7: 137-146.

Lázaro, J.M., J.A.S. Navarro, A.G. Gil and V.E. Romero, 2016. Flood Frequency Analysis (FFA) in Spanish catchments. J. Hydrol., 538: 598-608. DOI: 10.1016/j.jhydrol.2016.04.058

Parent, E. and J. Bernier, 2003. Bayesian POT modeling for historical data. J. Hydrol., 274: 95-108. DOI: 10.1016/S0022-1394(02)00396-7.

Payrastre, O., E. Gaume and A. Herve, 2013. Historical information and flood frequency analysis:Which optimal features for historical floods inventories? Houille Blanche-Revue Int. De L Eau, 3: 5-11. DOI: 10.1051/1hb/2013019

Payrastre, O., E. Gaume and H. Andrieu, 2011. Usefulness of historical information for flood frequency analysis: Developments based on a case study. Water Resources Res., 47: 1-15. DOI: 10.1029/2010WR009812

Reis, D.S.J. and J.R. Stedinger, 2005. Bayesian MCMC flood frequency analysis with historical information. J. Hydrol., 313: 97-116. DOI: 10.1016/j.jhydrol.2005.02.028

Rezaee, P.H., 2001. Application of statistics and probability in water resources.

Shahabfar, A. and A. Ghiami, 2004. Evaluating the goodness of fit methods of statistical distributions and using time series to forecast yearly rainfall of Masshad City, Educational Workshop of study the strategies against water crisis, national committee of draining and irrigation, Iran.

Sharifi, M. and M.P. Shahidi, 2001. Water resources system analysis. Ferdowsi University, Mashhad.

Strupczewski, W.G., K. Kochanek and E. Bogdanowicz, 2014. Flood frequency analysis supported by the largest historical flood. Nat. Hazards Earth Syst. Sci., 14: 1543-1551. DOI: 10.5194/nhess-14-1543.

Strupczewski, W.G., K. Krzysztof and E. Bogdanowicz, 2017. Historical floods in flood frequency analysis: Is this game worth the candle? J. Hydrol., 554:800-816. DOI: 10.1016/j.jhydrol.2017.09.034

Tanji, K.K., 1990. Agricultural salinity assessment and management. Am. Society Civil Eng. New York.

Todd, K.D. and L.W. Mays, 2005. Groundwater Hydrology. 2nd Edn., John Wiley and Sons, New York, pp: 636.

Waltemeyer, S.D., 2002. Analysis of the magnitude and frequency of the 4-Day annual low flow and regression equations for estimating the 4-day, 3-year low-flow frequency at ungagged sites on unregulated streams in New Mexico. U.S. Department of the Interior, U.S. Geological Survey, Water Resources Investigation Report 01-4271, New Mexico, USA.