


Statistical analysis of groundwater quality at Wonji Shoa Sugar Estate

Megersa Olumana Dinka 

University of Johannesburg, Faculty of Engineering and the Built Environment, Department of Civil Engineering Sciences,
PO Box 524, Auckland Park, 2006 Johannesburg, South Africa

RECEIVED 09.04.2021

ACCEPTED 06.04.2022

AVAILABLE ONLINE 08.06.2022

Abstract: Quality assessment of shallow groundwater table is very important because it is the water that recharges deeper aquifers and constantly feeds the water levels of our surface streams and wetlands. Continuous monitoring of large number of quality parameters is essential for effective maintenance of water quality through appropriate control measures. However, it is very difficult and laborious task for regular monitoring of all the parameters even if adequate manpower and laboratory facilities are available. Therefore, this study presents the statistical analysis of physico-chemical parameters (pH, EC, TDS, Na, K, Ca, Mg, HCO₃, Cl, CO₃, SO₄, TH, B, F) using correlation and Principal Component Analysis. The statistical analysis of the groundwater quality variables indicated that most of the variables are highly correlated. The strong correlation is an opportunity to develop a regression equation and monitor using few parameters. This provides an easy and rapid method of continuous groundwater quality monitoring. Moreover, groundwater of the area showed significant compositional variation. The compositional variability has implications for the source and origin of groundwater quality in the study area.

Keywords: correlation, groundwater quality, principal component, quality parameters, statistical analysis

INTRODUCTION

The groundwater at Wonji-Shoa Sugar Estate (WSSE) is very shallow (<1 m below ground) in most of the plantation fields. A large proportion of irrigated land in WSSE is affected by waterlogging problems [DINKA 2020; DINKA *et al.* 2013; DINKA, NDAMBUKI 2014]. In such very shallow groundwater (GW) table areas, soil salinisation and other water and soil quality deteriorations are also expected. Groundwater, once polluted, could remain for years, decades or even for centuries because of the slow movement (resident time) through the aquifers [MACDONALD, FOSTER 2016; ZAPOROZEC 1981]. It is difficult to remediate when GW is polluted persistently as the result of their large storage, long residence time and physical inaccessibility [DINKA 2017; FOSTER 2006].

The quality assessment of shallow groundwater is very important because it is the water that recharges in to deeper aquifers and constantly feeds the water levels of our surface streams and wetlands [FOSTER *et al.* 2013]. Continuous monitoring of large number of quality parameters is essential for effective

maintenance of groundwater quality through appropriate control measures. However, it is very difficult and laborious task for regular monitoring of all the parameters even if adequate facilities (manpower and laboratory) are available [JENAL *et al.* 2012]. Therefore, an alternative approach has been adopted by different researchers: statistical analysis of physico-chemical parameters and then developing regression equations between the highly correlated quality variables [CELESTINO *et al.* 2018; EL BAGHDADI *et al.* 2019; JENAL *et al.* 2012]. Statistical analysis techniques can be effective indicators of groundwater constituents and geochemistry [BELKHIRI *et al.* 2010].

No research has been done so far in WSSE regarding the groundwater quality (physico-chemistry) assessment using statistical method. Therefore, this study was initiated with the main objective to analyse the statistical correlation between the physico-chemical parameters. The highly correlated parameters are further verified by Principal Component Analysis (PCA). The significantly good correlation between the quality parameters provides an easy and rapid method of monitoring groundwater quality in the study area.

METHODS

BRIEF DESCRIPTION OF THE STUDY AREA

Wonji-Shoa Sugar Estate (WSSE) is located in uppermost part of Awash River basin, Central Oromia Region at a distance of about 110 km south east of the capital city of Ethiopia. The establishment of the sugar estate in 1954 marked the first era of large-scale irrigation development and era of domestic sugar production in the Ethiopian history. It has a total cultivated/irrigated area of about 8,000 ha (excluding the recent expansion). The area is characterised by erratic and bimodal rainfall distribution: major rainfall (67%) is concentrated in summer season (June–September), and minor, occasional rainfall (10%) in winter season (March). The mean annual rainfall of the area is about 704 mm. The mean average minimum and maximum temperatures of the region are 15.2 and 27.6°C, respectively. The average pan evaporation is 6.8 mm·day⁻¹. The climate of the area, in general, is classified to be semi-arid. The predominant soils in the plantation area are clay soil (clay black type). The sugar estate uses furrow (mostly) and sprinkler irrigation system. The water source for irrigation is Awash River, the most utilised river in Ethiopia. Furrow irrigation has been in use in the estate proper and some of the outgrower areas. Sprinkler irrigation is in use in the recent expansion areas (Wake Tio and Dodota). The excess water is drained from the field to Awash River through the network of surface drains. The details of the study area can be obtained from other publications [DINKA 2020; DINKA *et al.* 2013; DINKA, NDAMBUKI 2014; DINKA, TEWELDEBRIHAN 2019].

WATER SAMPLING AND ANALYSIS

A total of 60 groundwater samples were collected from piezometers (50) and hand-dug wells (10). Water sampling was done in the month of May (just before the end of irrigation season) for three years (2009, 2010 and 2014). The procedures and protocols required for sample collection, preservation and transport were

done as per the standard guidelines [APHA, AWWA, WPCF 2005; WHO 2011]. The water sampling was done early in the morning using acid-washed clean polyethylene bottles (0.5 dm³). The sampling bottles were labelled with permanent ink and recorded in the datasheet immediately after sampling. The bottles were rinsed with distilled water and preserved airtight to avoid evaporation. For quality control and reliability, all the water samples were collected in duplicate on the same day (morning). The samples were preserved in an ice bucket container and immediately delivered to the laboratory (Wonji Research Station) for physico-chemical analysis. Sampling period was almost at the end of the irrigation season. In the meantime, during the sample collection, the location of the area is measured using GPS. Figure 1 presents the distribution of the sugar estate (estate proper and outgrowers), network of irrigation and drainage systems, Awash River flow and the location of sampling sites.

The collected water samples were analysed in the laboratory of the Research Directorate, on the same day, for selected physico-chemical parameters following standard procedures [APHA, AWWA, WPCF 2005; WHO 2011]. The pH measured by pH meter, electrical conductivity (EC) and total dissolved solids (TDS) by conductivity meter, Ca and Mg by EDTA titrimetric, Na and K by flame photometric, Cl, CO₃, HCO₃ by titration, SO₄ by gravimetric, F and B by potentiometric methods. The methods used by the Central Laboratory of the Research Directorate are presented by DINKA *et al.* [2015].

DATA ANALYSIS

The statistical parameters and correlation between the quality parameters were conducted using SPSS (ver. 16.0) statistical software package. A correlation matrix was performed using Pearson correlation coefficient (*r*) to identify the highly correlated water quality parameters. The Pearson correlation (1 tailed) was done at two confidence levels: 99% (*p* = 0.01) and 95% (*p* = 0.05). The significant correlation has been further verified by Principal Component Analysis (PCA). Factorial analysis (PCA) was per-

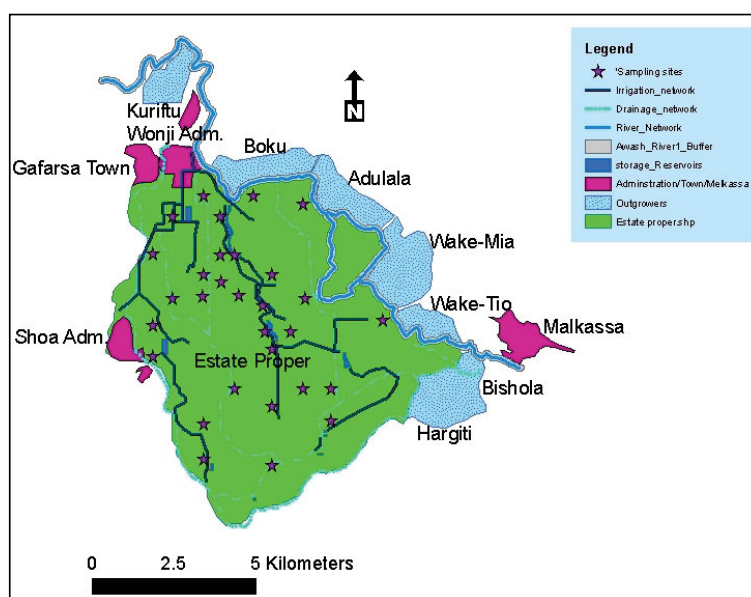


Fig. 1. Map of the sugar estate (estate proper and outgrowers), network of irrigation and drainage systems, Awash River flow and the location of sampling sites; source: own elaboration

formed in SPSS in order to further identify the contributing sources on the bases of their physico-chemical signature. PCA are used to reduce the effect of outliers (large dataset) in order to improve the clustering [CELESTINO *et al.* 2018]. PCA are based on the eigenvectors of a variance-covariance or a correlation matrix of the original data matrix [IRANMANESH *et al.* 2014]. In this study, the PCA was analysed using the eigenvalue greater than one criteria and the varimax with Kaiser normalisation rotation method.

RESULTS AND DISCUSSION

STATISTICAL SUMMARY

The statistical summary of the GW quality parameters are presented in Table 1. Note that the results for each of the 30 samples are published by DINKA [2019]. Majority of the considered parameters have higher skewness, except pH. All the quality parameters have a positive skewness, indicating that the distribution is right-skewed (i.e., mean > median > mode). In such type, the right tail is more pronounced than that the left tail and the distribution is not normal. It is interesting to see that groundwater quality of the study area showed large compositional variability (except pH, EC, B and F). The compositional variability has implications for the source and origin of groundwater quality in the study area. As indicated by DINKA [2019; 2020], the compositional variability is indicator for anthropogenic activities and geochemical processes prevailing in the region. Based on concentration, the cations and anions chemical abundance is in the order of Na > Ca > Mg > K and HCO₃ > SO₄ > Cl > CO₃, respectively. The order of abundance based on the CV (coefficient of variation) do not follow the order of relative abundance based on concentration: K > Mg > Na > Ca and Cl > HCO₃ > CO₃ > SO₄, respectively. From the measured

values, K is the highly variable parameters (CV = 140%) and followed by Cl (84%), Mg (70%), EC (59.4%), TDS (59.5%), HCO₃ (58.3%), Na (55.3), TH (55.1%), Ca (51%), etc.

CORRELATION ANALYSIS

Table 2 presents the correlation matrix for the GW quality parameters. All of them have positive correlation, except pH indicating the good water-rock interaction, which can be revealed from the significant correlation of EC and TDS with each other and other parameters. EC showed a significant positive (direct) correlation with all the parameters (except pH and F) at $p < 0.01$ (99% confidence level), indicating the possibility of mineral or gypsum or halite dissolution, silicate weathering processes, etc. [DINKA 2020] in the region. EC correlated with TDS, HCO₃, Ca, Na, Mg, Cl, with a correlation coefficient of 0.999, 0.868, 0.828, 0.806, 0.775, and 0.557, respectively. EC showed a significant positive correlation with K ($r = 0.456$) and a significant negative correlation with pH ($r = -0.376$) at $p < 0.05$ (95% confidence level). TDS also showed a significant positive correlation with EC ($r = 0.999$), Ca ($r = 0.817$), Na ($r = 0.809$), Mg ($r = 0.760$) and K ($r = 0.454$) at $p < 0.01$ and showed a significant negative (inverse) correlation with pH ($r = -0.363$) at $p < 0.05$ significant level. Chloride ion bears significant positive correlation ($p < 0.01$) with Na ($r = 0.678$), Mg ($r = 0.590$), EC ($r = 0.557$) and TDS ($r = 0.565$). Bicarbonate ions (HCO₃) bear a significant positive correlation with EC ($r = 0.868$), TDS ($r = 0.865$), Ca ($r = 0.782$), Na ($r = 0.607$) and K ($r = 0.466$) and has a significant negative (inverse) correlation with pH ($r = -0.423$) at $p < 0.05$ significant level. Cl ion do not have a significant correlation with pH, K, Ca and HCO₃. The strong correlation of Cl with Mg and Na reveals that Mg mainly remains present as MgCl₂ and Na as NaCl. The lack of correlation between Cl and K ions indicates that KCl in the water may be very low or absent in the sample [BELKHIRI *et al.* 2010].

Table 1. Statistical summary of groundwater quality variables

Quality parameter ¹⁾	Range	Min.	Max.	Mean	CV (%)	Skewness
pH	1.04	6.90	7.94	7.47	3.21	0.02
EC	2.803	0.187	3.00	1.038	59.44	1.62
Na	30.00	31.96	488.53	171.94	55.34	1.60
K	68.03	1.17	69.20	9.72	144.03	3.43
Ca	136.27	24.05	160.32	58.25	50.95	1.71
Mg	68.66	6.08	74.74	23.32	69.98	1.84
HCO ₃	30.00	219.66	1928.11	617.48	58.31	1.92
Cl	271.57	28.36	299.93	65.62	84.04	3.39
TDS	30.00	114.98	1846.50	641.25	59.49	1.66
TH	499.00	105.00	604.00	241.27	55.10	1.59
B	0.38	0.20	0.58	0.31	29.03	1.12
F	4.75	1.51	6.26	3.68	31.79	0.30
SO ₄	174.59	29.95	204.55	94.66	40.76	1.23
CO ₃	42.00	0.00	42.00	23.50	45.02	-0.77

¹⁾ All units are in mg·dm⁻³, except EC (dS·m⁻¹), pH (-), TDS (ppm).

Explanations: EC = electrical conductivity, TDS = total dissolved solids, TH = total hardness.

Source: own study.

Table 2. Pearson correlation matrix for groundwater quality parameters

Parameter	pH	EC	Na	K	Ca	Mg	HCO ₃	Cl	TDS	TH	B	F	SO ₄	CO ₃
pH	1													
EC	-0.376*	1												
Na	-0.119	0.806**	1											
K	-0.398*	0.456**	0.005	1										
Ca	-0.376*	0.828**	0.434**	0.522**	1									
Mg	-0.358*	0.775**	0.592**	0.15	0.772**	1								
HCO ₃	-0.423**	0.868**	0.607**	0.466**	0.782**	0.695**	1							
Cl	0.054	0.557**	0.678**	-0.07	0.360*	0.590**	0.302	1						
TDS	-0.363*	0.999**	0.809**	0.454**	0.817**	0.760**	0.865**	0.565**	1					
TH	-0.390*	0.852**	0.541**	0.367*	0.947**	0.935**	0.787**	0.499**	0.839**	1				
B	-0.184	0.595**	0.399*	0.082	0.693**	0.659**	0.535**	0.310*	0.584**	0.719**	1			
F	0.227	-0.706**	-0.467**	-0.314*	-0.887**	-0.676**	-0.591**	-0.329*	-0.699**	-0.835**	-0.712**	1		
SO ₄	0.015	0.523**	0.875**	-0.12	0.039	0.194	0.348*	0.401*	0.530**	0.119	0.104	-0.103	1	
CO ₃	-0.139	0.464**	0.252	0.063	0.634**	0.551**	0.361*	0.383*	0.469**	0.631**	0.461**	-0.625**	-0.08	1

Explanations: * correlation is significant at the 0.05 level (1-tailed); ** correlation is significant at the 0.01 level (1-tailed), other explanations as in Table 1. The bold values indicate the highest correlation from all the parameters considered.

Source: own study.

PRINCIPAL COMPONENT ANALYSIS (PCA)

As indicated in Table 1, most of the parameters are positively skewed. Thus, the values were normalised for the highly skewed parameters before conducting the PCA analysis. After the PCA analysis, four principal components were extracted, depending the eigenvalue greater than one criteria, using the varimax with Kaiser normalisation rotation method as shown in Table 3. The rotated component converged in four iterations. The four components explain about 85% of the total variance in the variables which are included in the component analysis. The percent of variances explained by PC1, PC2, PC3 and PC4 are about 37, 23, 15 and 10%, respectively. Most of the parameters were grouped under PC1, which significantly correlated with Ca, Mg, HCO₃, TH, B, F, CO₃. All of the parameters, except F, were positively correlated with PC1. F showed negative loading under PC1, in both cases: with- and without outliers. This reveals that F concentration in the groundwater of the study area is favoured by low values of Ca, Mg, HCO₃, TH, B, CO₃, respectively according to their correlation. The pH, K, and SO₄ have no significant impact on F. The significant negative correlation of F with those parameters indicates that the increment of these parameters, especially Ca, EC, TDS, Mg, Na, TH, will decrease the F concentration in the area. Groundwater of the study area are characterised by high F content due to the fact that the area is found within the geologically unstable rift valley region. Also, pH has a negative loading under PC3, indicating pH values are influenced by EC, K, Ca, HCO₃, and TDS. Na, Cl, TDS, EC, and SO₄ were grouped under PC2, where all of these parameters have significant positive correlation.

Table 3. The factor loading for factor analysis¹⁾ and rotated component analysis

Quality parameter	Principal component						
	including outliers				excluding outliers		
	1	2	3	4	1	2	3
pH	-	-	-0.541	-0.732	-	-0.735	-
EC	0.586	0.641	0.476	-	0.500	0.847	-
Na	0.313	0.939	-	-	-	0.975	-
K	-	-	0.860	-	-	-	0.868
Ca	0.831	-	0.489	-	0.917	-	-
Mg	0.794	0.369	-	-	0.794	0.426	-
HCO ₃	0.499	0.463	0.599	-	0.383	0.720	-
Cl	0.472	0.615	-	-	0.429	-	0.467
TDS	0.578	0.648	0.468	-	0.481	0.851	-
TH	0.864	-	0.374	-	0.907	0.371	-
B	0.775	-	-	-	0.680	-	-0.468
F	-0.833	-	-	-	-0.836	-	-
SO ₄	-	0.952	-	-	-	0.921	-
CO ₃	0.824	-	-	-	0.750	-	-
% of variance	31.2	23.03	15.2	10.1	38.08	25.03	18.68

¹⁾ Extraction method: Principal Component Analysis; rotation method: varimax with Kaiser normalisation. Source: own study.

The presence of large values indicating outliers is also checked by PCA analysis based on the factor score limit (± 3.0). Two samples fall outside of these acceptable range and hence the analysis was performed without these outliers (two samples) just to check whether it can change the interpretation of the analysis. The result indicates that the pattern of factor loading is not the same in the two split analysis (with and without outliers). Especially the loading of variables: pH, HCO_3 and Cl will be affected. The pH and HCO_3 are shifted from PC3 to PC2, whereas Cl is shifted from PC2 to PC3. Therefore, the final PC without the effect of outliers is presented as in Table 3. The three components explain about 82% of the total variance in the variables. The result indicates that the first component (PC1) explains about 38.1% of the variance and it includes Ca, Mg, TH, B, F and CO_3 . Variables within PC2 (pH, EC, Na, HCO_3 , TDS and SO_4) and within PC3 (K and Cl) explains about 25 and 18.7%, respectively of the total variance. Variables under PC1 represents high loading of salinity factor; PC2 represents oxidation-reduction factor; and PC3 represents acidity loading factor [IRANMANESH *et al.* 2014].

This study result (especially PCA categories) indicates that anthropogenic (industrial, domestic and agricultural activity) and natural (mineralisation, cation/ion exchange, reverse ion exchange, halite dissolution, silicate weathering, etc.) processes are controlling the hydro-chemistry of groundwater of the study area. The suggested idea also agrees with the hydro-chemical characterisation conducted by DINKA [2019] in the study area. Moreover, DINKA [2019] reported that the groundwater of the area were due to the mixing tendency between the deep-seated magmatic groundwater with the shallow groundwater, resulting in changing and mixing trend of hydrochemical-facies and compositions.

CONCLUSIONS

This study presents the statistical analysis of groundwater quality at Wonji-Shoa Sugar Estate. The correlation has been done and verified using Principal Component Analysis. The study result indicated that most of the considered quality parameters showed high compositional variability and correlation. The spatial variation could be the result of the respective groundwaters being in contact with different aquifer materials or minerals for significantly different periods as well as the anthropogenic inputs from industrial, domestic and agricultural effluents. With the exception of pH, all of the considered water quality parameters have positive correlation. Interestingly, EC showered a strong positive correlation with all the other parameters, except pH. The correlation analysis result indicated that groundwater of the area is mostly due to the good water-rock interaction. The strong correlation between Cl, Mg and Na reveals that Mg mainly remains present as MgCl_2 and Na as NaCl. The factorial (PCA) analysis result indicates all of the water quality parameters (except F) were positively correlated with PC1, indicating F concentration in the groundwater of the study area is favoured by low Ca, Mg, HCO_3 , TH, B, CO_3 . The strong negative correlation between Ca and F could point to CaF_2 solubility, which limits the concentrations of Ca. Moreover, groundwater quality of the area has a characteristics of deep and shallow groundwater. The

strong positive correlation of most of the parameters indicates the possibility for evaporation and mixing with surface water.

In general, the compositional variability, correlation and PCA analysis results indicated that the physicochemical sources of groundwater are related to anthropogenic and natural sources. The anthropogenic sources could be factory waste, domestic waste and agricultural effluents. The natural sources could be ion exchange, reverse ion exchange, halite dissolution, silicate weathering, etc. The relatively higher correlation found between the variables provides a good opportunity for the development of a regression equation for the different quality variables, which minimises/reduces the time, cost and labour in collecting, analysing and interpretation of large volumes of quality variables. Especially, since the groundwater in the area is very shallow (<1 m), the developed equation provides an easy and rapid method of continuous groundwater quality monitoring.

REFERENCES

- APHA, AWWA, WPCF 2005. Standard methods for the examination of water and wastewater. 21st ed. Washington DC. American Public Health Association, American Water Works Association, Water Environment Federation. ISBN 0875530478 pp. 2671.
- BELKHIRI L., BOUDOUKHA A., MOUNI L.A. 2010. Multivariate statistical analysis of groundwater chemistry. *International Journal of Environmental Research*. Vol. 5(2) p. 537–544. DOI 10.22059/IJER.2011.338.
- CELESTINO A.E.M., CRUZ D.A.M., SÁNCHEZ E.M.O., REYES F.G., SOTO D.V. 2018. Groundwater quality assessment: An improved approach to K-means clustering, principal component analysis and spatial analysis: A case study. *Water*. Vol. 10(4) 437. DOI 10.3390/w10040437.
- DINKA M.O. 2017. Hydrochemical composition and origin of surface water and groundwater in the Matahara area, Ethiopia. *Inland Waters*. Vol. 7. Iss. 3 p. 297–304. DOI 10.1080/20442041.2017.1329909.
- DINKA M.O. 2019. Groundwater property and composition variability under long-term irrigated area of Wonji Plain, Ethiopia. *Journal of Water and Land Development*. No. 41 (IV–VI) p. 37–46. DOI 10.2478/jwld-2019-0025.
- DINKA M.O. 2020. Groundwater quality composition and its suitability for drinking in long-term irrigated area. *Journal of Water and Land Development*. No. 44 (I–III) p. 43–54. DOI 10.24425/jwld.2019.127044.
- DINKA M.O., LOISKANDL W., NDAMBUKI J.M. 2013. Seasonal behaviour and spatial fluctuations of groundwater levels in long-term irrigated agriculture: The case of Wonji Shoa Sugar Estate (Ethiopia). *Polish Journal of Environmental Studies*. Vol. 22. Iss. 5 p. 1325–1334.
- DINKA M.O., LOISKANDL W., NDAMBUKI J.M. 2015. Hydrochemical characterization of various surface water and groundwater resources available in Matahara areas, Fantalle Woreda of Oromiya region. *Journal of Hydrology: Regional Studies*. Vol. 3 p. 444–456. DOI 10.1016/j.ejrh.2015.02.007.
- DINKA M.O., NDAMBUKI J.M. 2014. Identifying the potential causes of waterlogging under long-term irrigated agriculture: The case of Wonji-Shoa Sugarcane Plantation (Ethiopia). *Irrigation and Drainage*. Vol. 63 p. 80–92. DOI 10.1002/ird.1791.

- EL BAGHDADI M., MEDAH R., JOUIDER A. 2019. Using statistical analysis to assess urban groundwater in Beni Mellal City (Morocco). *International Journal of Agronomy*. Vol. 2019, 7469741 p. 1–22.
- FOSTER S. 2006. Groundwater-sustainability issues and governance needs. *Episodes*. Vol. 29(4) p. 238–243.
- FOSTER S., CHILTON J., NIJSTEN G.J., RICHTS A. 2013. Groundwater – a global focus on the ‘local resource’. *Current Opinion in Environmental Sustainability*. Vol. 5(6) p. 685–695. DOI 10.1016/j.cosust.2013.10.010.
- IRANMANESHA A., LOCKE R.A., WIMMER B.T. 2014. Multivariate statistical evaluation of groundwater compliance data from the Illinois Basin – Decatur Project. *Energy Procedia*. Vol. 63 p. 3182–3194. DOI 10.1016/j.egypro.2014.11.343.
- JENAL V., DIXIT. S., GUPTA S. 2012. Physico-chemical parameters assessment of ground water in different sites of Bhilai City, Chhattisgarh. *Rasayan Journal of Chemistry*. Vol. 5. Iss. 4 p. 506–509.
- MACDONALD A.M., FOSTER S.S.D. 2016. Groundwater systems. In: *Spring: managing groundwater sustainably*. Ed. M. Smith. Gland, Switzerland. IUCN p. 24–48.
- WHO 2011. *Guidelines for drinking-water quality*. 4th ed. Geneva, Switzerland. World Health Organization. ISBN 978 92 4 154815 1 pp. 541.
- ZAPOROZEC A. 1981. Ground-water pollution and its sources. *GeoJournal*. Vol. 5 p. 457–471. DOI 10.1007/BF02484718.