ANOVA-based analysis using MATLAB for Groundwater quality assessment

Shankar B S[1] , Usha A[2] and Sreevidya Raman S[3]

1. Department of Civil Engineering, Cambridge Institute of Technology, Bengaluru
Email: shanky5525@gmail.com
2. Department of Mathematics, Alliance School of Applied Mathematics, Alliance University, Bengaluru
Email: ushmath@gmail.com
3. Department of Civil Engineering, Cambridge Institute of Technology, Bengaluru
Email: sreevidya.raman@gmail.com

Corresponding author: Shankar B S ; shanky5525@gmail.com

ABSTRACT

The study involves an ANOVA-based analysis using MATLAB to detect the quality of groundwater in Bellandur area in Bengaluru, Karnataka. The main purpose of this study is to ascertain the suitability of water for drinking and domestic uses. In this work, water samples were collected from thirty groundwater sampling locations of Bellandur, in two (pre-monsoon and post-monsoon seasons) of 2020. Student's t-test and one way ANOVA are performed using MATLAB version 2015a for key physico-chemical characteristics of water. Six major parameters of water are considered, for which the testing of hypothesis has been carried out employing t-test and analysis of variance. In both the cases, after the hypothesis test was conducted and it was observed that the calculated values were not within the acceptance region. The p-value of the anova table shows that it is above the significance level which hypothize that the differences between the means are not statistically significant. This infers that the water is not potable as it falls outside the permissible limits as prescribed by Bureau of Indian Standards (BIS). Karl Pearson's coefficient of correlation was calculated for the six parameters of the collected samples, and a positive correlation was observed, ranging from 0.7 to 0.9 for total hardness and chloride, total dissolved solids, and nitrate, chloride and TDS as well as nitrate and TDS.

Keywords: Statistical analysis, Student's t-test, One-way ANOVA, groundwater, correlation coefficient

1. Introduction

Water is a very precious resource for all the lives on Earth and all the activities of living beings mainly depend on water. Groundwater is a critical fresh water resource and the quality of groundwater is extremely significant, as the towns and villages completely depend on it for domestic and agricultural activities [1, 2]. The groundwater is rich in minerals compared to the surface water as the water underground is less polluted. Earth's surface is 71% covered with water and only 29% with land. Almost 70% of fresh ground water is used by us for irrigation which supply 45% of global food [3]. Saltwater bodies also play a vital role as it is a reservoir of many fishes. Indirectly, humans depend on salt water for their living/food. Seas, rivers, and canals are a major source of trade of oil, LPG, and other commodities.

The cause of chronic diseases is mainly from the polluted water [4,5]. So, it is very important to test the quality of groundwater to assure whether the groundwater can be utilized for drinking and various purposes. Main reasons of water getting polluted is because of the industrial waste being disposed into water bodies and poor maintenance of drain water getting mixed in the surrounding wells, lakes, and ponds [6]. It is in this regard; the present study has been undertaken for the said study area to critically evaluate the quality of groundwater.

Statistics provide many tools to test the data from which an analysis can be made [7, 8]. It helps in condensing complex data into important figures that are understandable. There are many techniques available in statistics to test the hypothesis and forecast the happenings with the available data [9-12]. In statistics, population refers to larger population and sample refers to a small quantity of population. To test the population, samples are tested, and the results are generalised for the population. In this study, the water quality is tested for Bellandur area as population for which the samples are collected at 30 different places of the study area.

To ascertain the plausibility of a hypothesis using sample data, hypothesis testing is done. Such data may come from a larger population, or from a data-generating process. The word "population" will be used for both cases in the following descriptions.

The hypothesis that is being tested for possible rejection is the null hypothesis while the hypothesis that is accepted when the null hypothesis is rejected is termed as alternative

hypothesis. While testing the parameters, the null and alternative hypotheses are assumed accordingly for the permissible range of each property. Each parameter test conducted is verified with the tabulated values and an inference is drawn regarding the same. It is observed that for both the seasons, the water samples showed that the six physico-chemical properties not being within the range of permissible values. This declares that the water quality is not safe for consumption in the study area where the samples are collected.

In this work, various significant physico-chemical water quality parameters namely total hardness, chloride, nitrate, iron, total dissolved solids, and fluoride are analysed. The test results were compared with the stipulated permissible limits as per the Bureau of Indian Standards (BIS) to assess the quality of water. In this connection, an effort has been made to carry out statistical analysis using MATLAB to establish the relationship amongst the evaluated parameters.

1.1 Details of study area

Bellandur, the study area lies at the southern part of Bangalore covering 24 square kilometres, situated between latitude $12^{0}53^{1}30^{11}$ to $12^{0}56^{1}50^{11}$ East and Longitude $77^{0}34^{1}$ to $77^{0}41^{1}40^{11}$ North. The area falls under 57H/9 Topographical sheet of Survey of India. The study area is encompassed by about 158 industries, most of which have contributed to the contamination of groundwaters in Bellandur.

2. MATERIALS AND METHODOLOGY

2.1 Sampling and methods of analysis

From the groundwaters of Bellandur, 30 samples were drawn in two litre cans, during the study period of April and November, the respective pre and post-monsoon seasons of 2020. Subsequently, these samples were stored and preserved in line with the protocols of the American Public and Health Association. The global positioning system (GPS) technique was used for locating the sampling stations by recording their longitude and latitude and of these stations.

The samples were analysed for the important quality parameters in the environmental engineering laboratory. The determination of pH was done in the field while collecting the samples. The analysis for chemical parameters was carried out strictly adhering to the protocols [13]. The results of the physico-chemical analyses in accordance with the quality

standards of 'IS 10500' of Bureau of Indian Standards [14] were interpreted. Figure 1 depicts the location map of study area indicating the location of stations.

Figure 1. Location map of study area depicting the sampling stations

*The red dots denote the sample stations

2.2 *Parametric evaluation Methodology*

The collected 30 groundwater samples are analysed for 6 physico-chemical parameters namely total hardness (TH), iron (Fe), chloride (Cl), nitrate (NO_3), total dissolved solids (TDS), and fluoride (F) as per the Standard methods protocol [13]. Table 1 shows the analysis techniques used for evaluation of physico-chemical parameters.

Table 1. Techniques adopted for physico-chemical parametric evaluation

Water Quality Parameter	**Analysis Methods**
Chloride (Cl)	Argentometric titration method
Total hardness (TH)	EDTA titrimetric analysis
Fluoride (F)	UV-Visible spectrophotometry
Nitrate (NO_3)	UV-Visible spectrophotometry

Iron (Fe)	Atomic absorption spectrophotometry
Total Dissolved Solids (TDS)	Gravimetric analysis

2.3 *Statistical Methodology*

The statistical study has been carried out using MATLAB version R2015a software. The one-way analysis of variance study is performed using the annova1 () MATLAB function. QGIS, a free and open-source software is used to obtain the geographical data and map creation [15]. Toy data set is used to obtain the district boundaries, which is then imported into QGIS. The latitude and longitude of the sample locations are saved as a text file and added to QGIS as a delimited test layer. Using the 'new print layout' option, the final map has been created and saved as a jpeg.

3. RESULTS AND DISCUSSIONS

Statistical tests namely Student's t-test and one-way ANOVA were performed, and observations noted. The data set has samples collected during the two seasons namely, pre-monsoon and post-monsoons of the year 2020 [16-18]. It is observed from both the tests using the analysis that the water is not consumable and cannot be used for domestic purposes. The details of the test results have been discussed in the following section.

3.1 *Statistical Analysis*

To verify the quality of groundwater, statistical analysis has been carried out in which Student's t-test and the analysis of variance (one-way ANOVA) were performed for pre-monsoon and post-monsoon values of each parameter collected from the study area [19-21]. Appropriately, the null hypothesis is considered for permissible limits of the parameters. The observations of all the six significant parameters are as follows.

Total Hardness

The calculated values of t are found to be in the range of -27.2778 to 4.73542. The tabulated value of t for a state of level of confidence (95%) for degrees of freedom 29 is 1.70. The calculated values are obtained using MATLAB version 2019B. Hard waters usually originate in locations with thick top soils and limestone formations are present [22]. The permissible limit of total hardness as per BIS is considered as 200 to 600 mg/l as $CaCO_3$. The null hypothesis is considered as $\geq$200 and accordingly the alternative hypothesis is assumed. For both the cases, the test resulted in the rejected region of null hypothesis. Similarly, for both

the data of the seasons, the upper limit is considered and found that the water is not in with in the permissible limit, with a variation of values.

Iron

The permissible limit of Iron (Fe) as per BIS is considered as 0.3 to 1(maximum allowable limit). Accordingly, the data for both seasons has been checked for the t-test. The null hypothesis is considered for both the limits of Iron. In both the cases, the Iron level in which null hypothesis considered for values greater than 0.3 was accepted, while for less than 1, was rejected. This infers that the Iron content in the water is not with in the permissible limits, and well beyond the upper limit of the prescribed value.

Chloride

Chloride beyond the permissible concentrations generally imparts a very salty taste to water. People who are allergic to high chloride have experienced laxative effects [23].The permissible limit of chloride is considered as 250 to 1000 mg/l. The null hypothesis is greater than 250 and accordingly the alternative hypothesis is assumed. In both the cases of pre-monsoon and post-monsoon, the null hypothesis is rejected which means that the chloride content in water of the study area is not with in the permissible limits. It shows that the water is non-potable, as the calculated values are in the rejected region of the hypothesis.

Nitrate

Ingestion of nitrate in concentrations exceeding 45 mg/l may cause methemoglobinemia in infants, gastric cancer and thyroid issues [24]. As per the Bureau of Indian Standards, the maximum permissible limit of nitrate is 45mg/l and hence the null hypothesis is equal to 45. A two-tailed test is examined, and it was found that the calculated value does not fall under acceptance region rejecting the null hypothesis. This confirms that the nitrate content is beyond the permissible range and hence the water is not suitable for domestic purposes

Total dissolved solids

Total dissolved solids values are invariably the least at recharge zones and maximum at the points of discharge on travelling through the rock media and dissolving more materials along its path of travel [25]. The permissible limit range for TDS (Total Dissolved Solids) as per BIS is 500 to 2000mg/l. The pre-monsoon and the post-monsoon values of the parameter is checked for t-test by assuming the null hypothesis to be greater than 500 in first case and less

than or equal to in second case. Accordingly, the alternative hypothesis is assumed. t-test shows that the null hypotheses in both the cases for two seasons are rejected confirming the water contamination.

Fluoride

The permissible limit range for fluoride is 1 to 1.5mg/l. t-test is performed for both pre and post monsoon results and the null hypothesis is taken as greater than 1 and less than 1.5 which resulted in the acceptance of null hypothesis for greater than 1 only. The upper limit of the permissible limit when tested showed that the null hypothesis is rejected.

For each analysed parameter, the results of the statistical analysis, F test and significance are presented in Table.4. The p-value of the ANOVA table shows that it is above the significance level which infers that the differences between the means are not statistically significant. The values of correlation matrix have been presented in tables 2. and 3 which reveal a high positive correlation ranging between 0.7 and 0.9 between TH with Cl, TDS, and NO_3, positive correlation of Cl and TDS as well as NO_3 and TDS for pre-monsoon and post-monsoon samples. Thus it is clear that the water quality of the study area is not fit for domestic purposes.

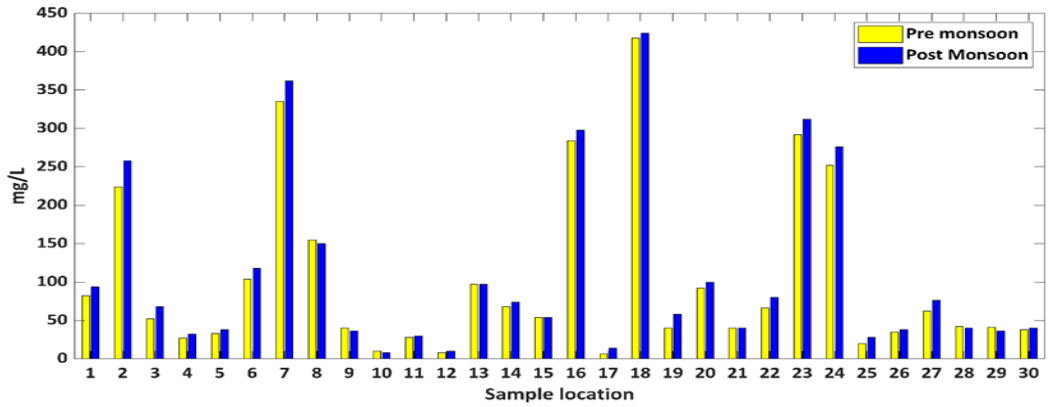

Figure 2. NO_3 Concentration in water at different sample locations

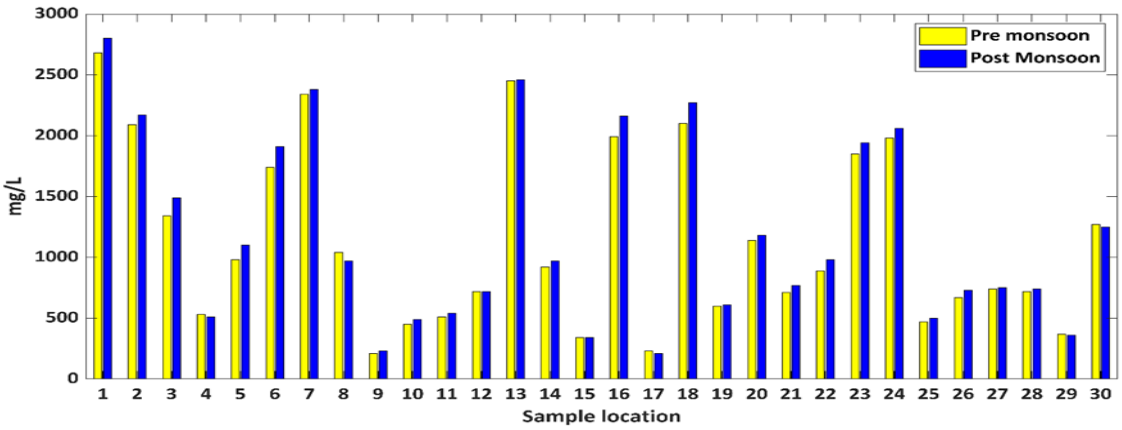

Figure 3. TDS Concentration in water at different sample locations

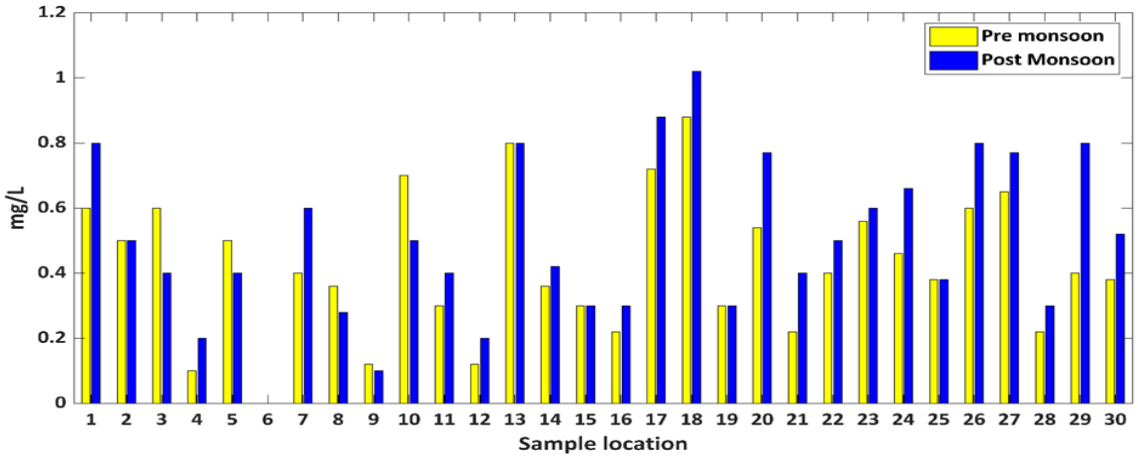

Figure 4. Fluoride Concentration in water at different sample locations

Table 2. Correlation matrix for various water quality parameters (Pre-monsoon)

Parameter	T.H	Fe	Cl	NO3	TDS	F
T.H	1	-0.0213	0.8936	0.7220	0.9410	0.2209
Fe		1	-0.0048	0.0586	0.0266	-0.0314
Cl			1	0.5251	0.9377	0.2553
NO_3				1	0.7420	0.2863
TDS					1	0.2966
F						1

Table 3. Correlation matrix for various water quality parameters (post-monsoon)

Helpful tools in promoting research and widening the knowledge is possible by studying inter relationship between the parameters considered. The uncertainty [26] with decision making

Parameter	**T.H**	**Fe**	**Cl**	**NO3**	**TDS**	**F**
T.H	1	0.003	0.8957	0.6976	0.9385	0.2442
Fe		1	0.0037	0.0814	0.0393	-0.0383
Cl			1	0.5030	0.9335	0.2491
NO_3				1	0.7244	0.2509
TDS					1	0.3097
F						1

could be avoided using the study of correlation. Karl Pearson's coefficient of correlation 'r' is calculated using the formula

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

Tables 2 and 3 represents the correlation coefficient between the parameters considered for pre-monsoon and post-monsoon respectively.

The water quality parameters are graphically (figures 2, 3, 4) compared for the set of two values in two seasons. Even though the graphs show moderate variations in each parameter, the analysis has proved that the parameters considered are not within the permissible limits of the BIS standards.

The Sample code used for statistical analysis is mentioned below.

```
%% T-test
n=30; %sample stations
x1=xlsread('data_pre.xlsx','C:C') %read from database
sumx=sum(x1);
x2=x1.*x1;
sumx2=sum(x2);
first=sumx2/n;
second=(sumx/n)^2;
S=sqrt(first-second);
samplemean=sumx/n;
permissible=500;
t5=(samplemean-permissible)/(S/sqrt(n-1));
disp (['the T test value of TDS of samples =', num2str(t5,6)])
```

```
%% ANNOVA
x= xlsread('data_pre.xlsx','C:C')
y= xlsread('data_post.xlsx','C:C')
data= [x' y'];
[p,tbl,stats] = anova1(data);
```

Table 4. One-Way Analysis of Variance

Parameters	SS	df	MS	F	Significance
TH					
Between the Groups	13680.6	1	13680.6	0.1	0.7558
Within the Groups	8123608.1	58	140062.2		
Total	8137288.7	59			
Fe					
Between the Groups	0.356	1	0.3557	0.2	0.8843
Within the Groups	965.481	58	16.6462		
Total	965.836	59			
Cl					
Between the Groups	18270.1	1	18270.1	0.18	0.6707
Within the Groups	5803226.8	58	100055.6		
Total	5821497	59			
NO_3					
Between the Groups	992.3	1	992.3	0.08	0.7806
Within the Groups	735176.5	58	12675.5		
Total	736168.7	59			
TDS					
Between the Groups	38506.7	1	38506.7	0.07	0.7955
Within the Groups	32931633.3	58	567786.8		
Total	32970140	59			
F					
Between the Groups	0.0814	1	0.0814	1.48	0.2281
Within the Groups	3.1817	58	0.05486		
Total	3.2631	59			

4. Conclusions

This study is aimed at MATLAB based analysis using statistical tools for evaluating the quality of groundwater in 30 different locations of Bellandur area, Bengaluru, Karnataka, India during two different seasons and the analysis was carried out. Student's t-test and one-way ANOVA were the tools used in testing the six physico chemical properties of water. It is clear p-value of the anova table ranging from 0.22 to 0.89 shows that it is above the significance level which infers that the differences between the means are not statistically significant. On testing using above said tools, it is concluded that the groundwater quality is not within the permissible range according to t-test and the F test values. They are not significant at 95% level as per BIS standards. The results of the investigation clearly reveal that the quality of water in the study area is unfit for domestic purposes.

Acknowledgments

The researchers are extremely indebted to the managements of Cambridge Institute of Technology & Alliance University for their wholehearted support and facilities extended during the course of this work.

Abbreviation

ANOVA: Analysis of Variance

APHA: American Public Health Association

GPS: Global Positioning System

MATLAB – Matrix Laboratory Biological Oxygen Demand

QGIS: Quantum Geographic Information System

Acknowledgements

The authors wish to thank the Management and Principal of Cambridge Institute of Technology and Alliance University for the excellent laboratory and library facilities provided during the course of this research work.

References:

[1] S. Subramaniam, S. Abinandan and B. Anand, International Journal of Environmental Health Engineering 3, (2014). doi:10.4103/2277-9183.131809

[2] T. Sheikhy Narany, M. Ramli, A. Aris, W. Sulaiman and K. Fakharian, Environmental Monitoring and Assessment 186, (2014). doi:10.1007/s10661-014-3820-8

[3] P. Thenkabail, M. Hanjra, V. Dheeravath and M. Gumma, Remote Sensing 2, (2010). doi:10.3390/rs2010211

[4] K. Mohammed Abdul, S. Jayasinghe, E. Chandana, C. Jayasumana and P. De Silva, Environmental Toxicology and Pharmacology 40, (2015). doi: 10.1016/j.etap.2015.09.016

[5] R. Schwarzenbach, T. Egli, T. Hofstetter, U. Von Gunten and B. Wehrli, Annual Review of Environment And Resources 35, (2010). doi:10.1146/annurev-environ-100809-125342

[6] B.S Shankar and Usha Arcot, International Journal of Environmental Analytical Chemistry (2021). doi:10.1080/03067319.2021.2002311

[7] A. Octova, Afrinaldi, M. Gusman, Y. Anaperta and F. Ramadhan, Journal Of Physics: Conference Series 1387, (2019). doi:10.1088/1742-6596/1387/1/012083

[8] J. Agbasi and J. Egbueri, Geocarto International (2022). doi:10.1080/10106049.2022.2034990

[9] Singh, Yashbir, and Manish Kumar, Journal of Earth Sciences and Geotechnical Engineering (2011)

[10] A. Aydin Uncumusaoğlu and T. Akkan, Polish Journal of Environmental Studies 26, (2017). doi:10.15244/pjoes/68952

[11] P. Pujar, H. Kenchannavar, R. Kulkarni and U. Kulkarni, Applied Water Science 10, (2019). doi:10.1007/s13201-019-1111-9

[12] F. Wang, G. Huang, Y. Fan and Y. Li, Water Resources Management 34, (2020). doi:10.1007/s11269-020-02608-2

[13] A.E. Greenberg, Proc. - AWWA Water Qual. Technol. Conf. 1984, 179.

[14] I. Standards, Water Supply 25, 1 (2012).

[15] Docs.Qgis. Org (2022).

[16] I. Leščešen, M. Pantelić, D. Dolinaj, V. Stojanović and D. Milošević, Polish Journal of Environmental Studies 24, (2015). doi:10.15244/pjoes/29684

[17] R. Gyimah, C. Gyamfi, G. Anornu, A. Karikari and F. Tsyawo, International Journal of River Basin Management 19, (2020). doi:10.1080/15715124.2020.1803337

[18] M. Gradilla-Hernández, J. de Anda, A. Garcia-Gonzalez, D. Meza-Rodríguez, C. Yebra Montes and Y. Perfecto-Avalos, Environmental Monitoring and Assessment 192, (2019). doi:10.1007/s10661-019-7972-4

[19] I. Aboim, D. Gomes and P. Mafalda Junior, Environmental Monitoring and Assessment 192, (2019). doi:10.1007/s10661-019-7882-5

[20] J. Zhang, S. Li and C. Jiang, Ecological Indicators 113, (2020). doi:10.1016/j.ecolind.2020.106226

[21] Z. Luo, Q. Shao, Q. Zuo and Y. Cui, Journal of Hydrology 584, (2020). doi:10.1016/j.jhydrol.2020.124655

[22] Rubia Khan and D. C. Jhariya, Journal of Geological Society of India 90, (2017). doi: 10.1007/s12594-017-0665-0

[23] S.M Sadat-Noori, K. Ebrahimi and A.M Liaghat, Environ Earth Sci 71, (2014) doi: 10.1007/s12665-013-2770-8

[24] S.K. Kumar, A. Logeshkumaran, N.S. Magesh, P.S. Godson and N. Chandrasekar, Appld. Water Sci, (2014) doi: 10.1007/s13201-014-0196-4

[25] Larry Pax Chegbeleh, Bismark Awinbire Akurugu, Sandow Mark Yidana, The Scientific World Journal, 2020. https://doi.org/10.1155/2020/8450860

[26] S. Dandwate, E-Journal of Chemistry 9, (2012). doi:10.1155/2012/650212