**Poll Res. 43 (3-4) : 235-239 (2024)**

ISSN 0257–8050
**DOI No.: http://doi.org/10.53550/PR.2024.v43i03-04.008**

# ASSESSMENT OF GROUNDWATER QUALITY OF WINTER SEASON BY APPLYING MULTIVARIATE STATISTICS AS PRINCIPAL COMPONENT ANALYSIS FOR RURAL AREA OF RAISEN DISTRICT, MADHYA PRADESH

**MEENU SHARMA[1] AND VIPIN VYAS[2]**

[1]*Department of Environmental Science and Limnology,*
[2]*Department of Bioscience,*
*Barkatullah University, Bhopal 462 026, M.P., India*

(Received 13 January, 2024; Accepted 27 April, 2024)

## ABSTRACT

**Principal Component Analysis (PCA) multivariate statistical tool can be used to reduce the dimensionality of data set and maintain the characteristics of variables which contribute mostly to this variation. The aim of the present study is to assess the groundwater quality of twenty-four sampling sites which surrounds Mandideep Industrial Estate, Raisen district in winter season. Fourteen physico-chemical parameters such as pH, TDS, DO, conductivity, chloride, alkalinity, total hardness, sulphate, nitrate, fluoride, iron etc. have been studied to determine the groundwater quality. PCA was applied on experimental data using SPSS, Version 20 software. The outcome of PCA revealed that TDS, total hardness, chloride, sulphate, pH and DO affect the groundwater quality. Further, it also revealed that most of the samples are not suitable to use for drinking purpose without any treatment.**

**KEY WORDS**: Principal component analysis, physico-chemical parameters, groundwater, multivariate statistics

## INTRODUCTION

In both rural and urban areas, groundwater is a primary and ultimate source of fresh water that fulfils the basic requirements of the inhabitants. In recent times, with the development of industry, agriculture, urbanization, infrastructure developments, improvements in living standards and eco-environment construction, water demand has increased tremendously. In daily life, evaluation of quality and suitability of groundwater for various useful purposes are appraising extra concern (Ravikumar and Somashekar, 2017). To protect the ground water quality for drinking purpose in rural and urban areas, periodical monitoring for ascertaining the quality of groundwater is essential. However, the regular monitoring and testing of quality parameters of drinking water takes lot of time and involves high cost which is not viable for the rural areas in practical sense (Lusiana *et al.*, 2022). But with the advancement of technology and various methods, this can be resolved by using strategic technique as application of multivariate statistical technique.

Multivariate statistical techniques, including cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA) can interpret complex datamatrices of water quality and other environmental systems by allowing the identification of possible factors/ sources. This application serves as a worthy tool for quickly resolving pollution problems (Esdras *et al.*, 2017; Ravikumar and Somashekar, 2017).

The technique of Principal component analysis (PCA) has been utilized to separate the noise from huge data matrix and classify the variables into measurable components. Characterization and evaluation of surface and fresh water quality performed by multivariate statistical techniques has proved useful in verifying spatial and temporal variations (Gulgundi and Shetty, 2018).

Further, to aid this (PCA), a data reduction technique was used to find out the major contributing parameters which were involved in deciding the geochemistry and quality of ground water samples (Elemile *et al.*, 2021). The objective of this study was the application of multivariate statistics Principal Component Analysis (PCA) to assess the groundwater quality to find out major contributing groundwater quality parameters.

## MATERIALS AND METHODS

### Study area

Raisendistrict is situated in the central part of Madhya Pradesh. Twelve villages were selected from Goharganz tehsil that surrounds the Mandideep Industrial Estate for sample collection as shown in Figure 1. Twenty-Four sampling sites, two from each village were chosen for this study. Mandideep Industrial Estate has large number of small- and large-scale industries, which use groundwater as a primary source of water and release effluents and contaminate the groundwater.

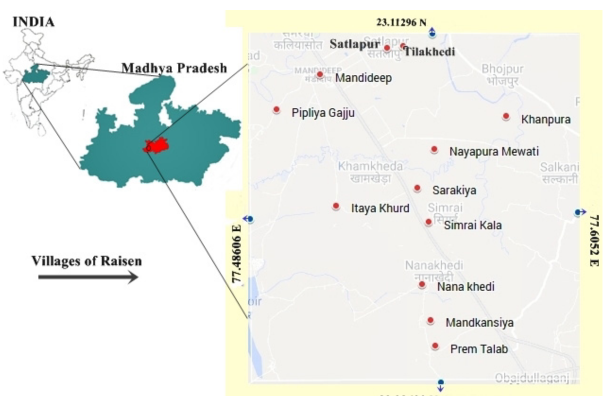


**Fig. 1.** Sampling sites of Industrial area

Thus, 24 groundwater samples were collected from selected sampling sites in 2018-2020 in winter season in the month of December and January. Samples were collected in pre-cleaned polyethylene plastic bottles for the physico-chemical analysis.

The 14 physico-chemical parameters were selected to analyse the ground water quality. These were pH, turbidity, total dissolved solids (TDS), conductivity (Conduct), Total Hardness (TH), Calcium Hardness (CaH), Magnesium Hardness, Alkalinity, Chloride (Cl), Sulphate (SO4), Nitrate (NO3), Fluoride (F), Iron, Dissolved Oxygen (DO). All the tests were conducted in accordance with the collection, preservation of samples, precautions, methods, and techniques described by APHA, 2017. PCA of the experimental data were performed by using SPSS (Statistical Package for Social Science), version 20 software.

## RESULTS AND DISCUSSION

Principal component analysis (PCA) is a statistical technique which reduces a larger set of variables into a smaller set of 'artificial' variables, called 'principal components' (PC). These PCs account for most of the variance in the original variables. PCA basically models the inter-relationships between items with fewer latent variables. Correlation among variables was found enough to go for PCA analysis.

Firstly, suitability of samples was checked for PCA analysis by KMO and Bartlett's Test. In this data analysis, the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy was found to be > 0.6. Thus, the test results were found to be useful on this criterion. The Bartlett's Test of Sphericity gave values of Sig. $p<0.05$, so the data was considered suitable for this type of investigation (Arumugam *et. al.*, 2010). The results of these tests are given below in Table 1.

**Table 1.** KMO and Bartlett's Test of samples of sites

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .673 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 690.910 |
| | df | 91 |
| | Sig. | .000 |

Communalities represent the proportions of variance in original variables explained by extracted components. Almost all the values in PCA analysis came out to be >0.50 (>50%) so all variables were considered for analysis. The communalities for selected sites are given below in Table 2.

The maximum number of components extracted always equals the number of variables. The plot is known as a Scree plot that shows the relation between Eigen Value and Component number as given in Figure 2. Only the first few components were taken for further analysis depending on the

**Table 2.** Communalities, showing proportions of variance explained by extractedcomponents.

| Parameter | Initial | Extraction |
|---|---|---|
| pH | 1.000 | .695 |
| Turbidity | 1.000 | .594 |
| TDS | 1.000 | .989 |
| Conduct. | 1.000 | .990 |
| TH | 1.000 | .986 |
| Ca H | 1.000 | .970 |
| Mg | 1.000 | .877 |
| Alkalinity | 1.000 | .824 |
| Cl | 1.000 | .924 |
| SO4 | 1.000 | .851 |
| NO3 | 1.000 | .621 |
| F | 1.000 | .756 |
| Iron | 1.000 | .764 |
| DO | 1.000 | .912 |

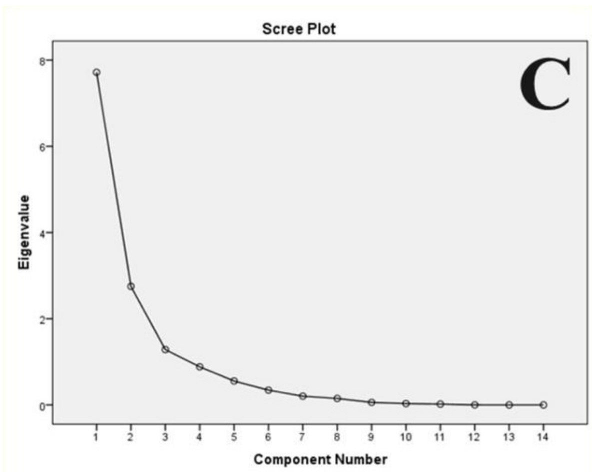


**Fig. 2.** Scree Plot for winter season

Eigen values and % of the variance explained by each PC. Eigen values represent the total amount of variance that can be explained by a given principal component. Components were retained where Eigen values exceeded 1 and where the % of the variance accommodated by each component both before and after (axis) rotation about the fixed origin was adequate. The data for total variance is explained herein below in Table 3.

The components (accounted for Eigen values <1) were dropped because of their smaller contributions for variance. In this case, for selected sites in winter season, the first 3 principal components had Eigen values >1. After rotation 1st, 2nd, 3rd components could explain 51.5%, 17.1%, and 15.3% of variance respectively. And collectively those 3 components explained 83.9% of the variation in the data. The interpreted results are shown in Table 3.

The scree plot displayed the number of principal components versus corresponding Eigen value (Vega *et al.*, 1998). The plots showed Eigen values from largest to smallest, which started to form about straight line after the 3rd or 4th principal component.

The next output Rotated Component Matrix showed the respective loadings of each of the variable onto the extracted components. Those loadings represented correlation of components with each variable. The interpretation of the principal components is based on finding for which variables are most strongly correlated with each component, i.e., which of these numbers are large in magnitude, the farthest from zero in either direction in rotated component matrix (Osborne, 2015). According to Liu *et al.* (2003), loadings with > 0.75 loading value was considered as strong, 0.75- 0.50 as moderate and 0.5-0.3 as weakly correlated loadings with corresponding components.

The rotated component matrix was used to understand the correlation between water quality parameters and corresponding principal components. The most favoured quality parameters were chosen by corresponding components.

**Principal Component Analysis of selected sites**

In this study, the first principal component (PC1) was strongly correlated with eight of the original variables as shown in Table 4 and Fig. 3. The first principal component had strong positive loadings of

**Table 3.** Total Variance Explained

| Total Variance Explained | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Component | Initial Eigen values | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| | Total | % of Variance | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| **1** | 7.718 | 55.125 | 7.718 | 55.125 | 55.125 | 7.218 | 51.557 | 51.557 |
| **2** | 2.751 | 19.647 | 2.751 | 19.647 | 74.772 | 2.389 | 17.063 | 68.620 |
| **3** | 1.284 | 9.171 | 1.284 | 9.171 | 83.944 | 2.145 | 15.324 | 83.944 |

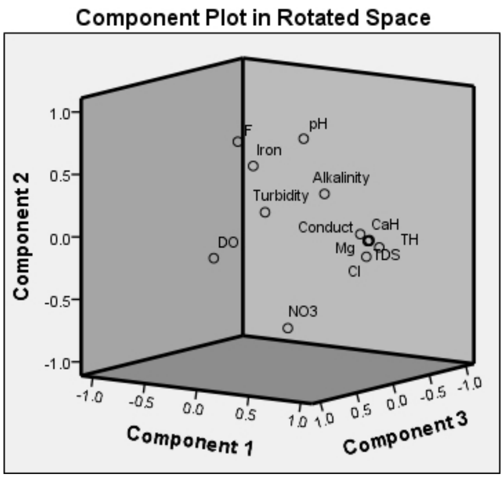


**Fig. 3.** Component Plot in Rotated Space

TDS (.981), Conductivity (.980), TH (.982), CaH (.975), Mg (.916), Cl (.947), SO4 (.917) and alkalinity (.712). This suggested that these eight criteria varied together. If one increased, then the remaining ones tended to increase as well. This component can be viewed as a measure of natural and anthropogenic sources as discussed earlier. The second principal component (PC2) had high positive loadings of pH (.777), F (.763) and iron (.622), but high negative loading of nitrate (-.727). So PC2 had positive correlation with pH, Fluoride, and iron, but negative correlation with nitrate. The third principal component (PC3) was associated with high ratings of DO (0.950) and moderately associated with Turbidity (.641). It also moderately correlated with iron (0.588) and weakly with Fluoride (0.404).

This method helped in reduction of number of dependent variables and to determine the independent variables as water quality parameters that affect the groundwater quality (Zuska *et al.*, 2019).

**Table 4.** Component matrix of Industrial sites in winter season

| Rotated Component Matrix[a] | Component 1 | Component 2 | Component 3 |
|---|---|---|---|
| Ph | .295 | .777 | .063 |
| Turbidity | .329 | .275 | .641 |
| TDS | .981 | .040 | .156 |
| Conduct. | .980 | .042 | .164 |
| TH | .982 | .047 | .141 |
| Ca H | .975 | .037 | .132 |
| Mg | .916 | .092 | .172 |
| Alkalinity | .712 | .422 | .372 |
| Cl | .947 | -.092 | .132 |
| SO4 | .917 | -.052 | -.088 |
| NO3 | .234 | -.727 | .193 |
| F | -.099 | .763 | .404 |
| Iron | .177 | .622 | .588 |
| DO | .052 | -.076 | .950 |

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.[a]a. Rotation converged in 4 iterations

## CONCLUSION

Statistical methods are one of the important approaches to display the information related to groundwater quality variables. Principal Component Analysis (PCA), a multivariate statistic was successfully applied to evaluate the important parameters that strongly influenced the groundwater quality in this season.It gave more simpler and more easily interpretable results for the evaluation of observed data. The multivariate analysis of data simplified and summarized the huge data and interpret by means of relatively few parameters. The outcome of PCA revealed that total dissolved solids, conductivity, total hardness, calcium hardness, chloride, sulphate strongly corelated and besides these pH and dissolved oxygen also affected the groundwater quality.

## ACKNOWLEDGMENT

The first author is thankful to express her sincere gratitude to Barkatullah University, Bhopal for providing Institutional Research Fellowship for this Study.

## REFERENCES

APHA, 2017. *Standard Methods for the Examination of Water and Wastewater*, American Public Health Association, American water works Association, Water Environment Federation, Washington, DC, 23rd edition.

Arumugam, N., Fatimah M.A., Chiew, E.F.C. and Zainalabidin, M. 2010. Supply chain analysis of fresh fruits and vegetables (FFV): prospects of contracts farming. *Agr Econ.* 56(9): 435-442.

Elemile, O.O., Ibitogbe, O.P., Ejiboye, P.O. and Adewumi, J.R. 2021. Principal component analysis of groundwater sources pollutionin Omu Aran Community, Nigeria. *Environmental Earth Sciences.* 80: 690, https://doi.org/10.1007/s12665-021-09975-y

Esdras, J., Zandagba, B., Adandedji, F.M., Lokonon, B.E., Chabi, A., Dan, O. and Mama, D. 2017. Application Use of Water Quality Index (WQI) and Multivariate Analysis for Nokoué Lake Water Quality Assessment. *American Journal of Environmental Science and Engineering.* 1(4): 117-127.

Gulgandi, M.S. and Shetty, Amba, 2018. Groundwater quality assessment of urban Bengrluru using multivariate statistical techniques. *Applied Water Science.* 8(43): 1-15. DOI:10.1007/s13201-018-0684-z

Liu, C.W., Lin, K.H. and Kuo, Y.M. 2003. Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwen. *Sci Total Environ.* 333(1-3): 77-89.

Lusiana, E.D., Mahmudi, M., Hutahaean, S. M., Darmawan, A., Buwono, N.R., Sulastri Arsad, S. and Musa, M. 2022. *Journal of Ecological Engineering.* 23(2): 123-131. https://doi.org/10.12911/22998993/144420

Osborne, J.W. 2015. What is rotating in exploratory factor analysis? *Pract Assess Res Eval.* 20(2): 1-7.

Ravikumar, P. and Somashekhar, R. 2017. Principal component analysis and hydrochemical facies characterization to evaluate groundwater quality in Varahi River basin, Karnataka state, India. *Applied Water Science.* 7: 745-755. DOI:10.1007/s13201-015-0287-x

Vega, M., Pardo, R., Barrado, E. and Deban, L. 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res.* 32(12): 3581-3592.

Zuœka, Z., Kopciñska, J., Dacewicz, E., Skowera, B., Wojkowski, J. and Wojtaszek, A.Z. 2019. Application of the Principal Component Analysis (PCA) Method to Assess the Impact of Meteorological Elements on Concentrations of Particulate Matter (PM10): A Case Study of the Mountain Valley (the S¹cz Basin, Poland). *Sustainability.* 11(23): 6740. https://doi.org/10.3390/su11236740