IOSR Journal of Applied Geology and Geophysics (IOSR-JAGG)
e-ISSN: 2321–0990, p-ISSN: 2321–0982.Volume 11, Issue 2 Ser. I (Mar. –Apr. 2023), 23-28
www.iosrjournals.org

A Systematic Approach to Approximate the Sources of Contaminants in the Groundwater Studies.

Priyadarshan S. Ganvir[1] and Gaurav R. Nimbarte[2]
[1]*Assistant Professor, Department of Geology, M. G. Arts, Science and Late N. P. Commerce College Armori, Dist. - Gadchiroli. Maharashtra, India.*
[2]*Assistant Professor,Department of Chemistry, MohsinbhaiZaweriMahavidyalaya, Desaiganj (Wadsa), Dist. - Gadchiroli. Maharashtra, India.*

Abstract
Thepollution in the groundwater system hasbeen a typical issue in the contemporary world.An evaluation and mitigationinspirations are requisites to take control over the pollution issues. The pollutants in the groundwater are clearlypersistent and pose a serious hazard for considerablelength to the depending biotain broad term and to the human beings in precise. The examinations of the contaminated groundwater systems or vulnerable ones arethe act of must. The most crucial phase among it is the source approximation with the latent strength of closingremarks of the study. The complications in the hydro-geochemistry andmanifoldoperating dynamics in the groundwater system,forcesthe utilization of aneffectualapproach to plot the sources of contaminations on the ground. The principal component analysis is the widelyassuredtechnique for thehunting down the sources ofevaluated variables in groundwater studies. The evaluated hydro-geochemical variables are subjected to the principal component analysis for raising the associations among them; to ordered branching of the principal components; to sort out the variance of the variables' data set in the principal components; to assessthe factor loadings of the variables in the principal components; to calculate the factor scores of the observations in the principal components. The association itself explains the togetherness of the assessed variables and hence the entanglements of the sources too. The forkedcomponents from the apex variance to the least one and the associated variance in the variable which isforeseeable from the factor loadings,elucidates the ascendancy of them in the groundwater system. The factor scores of the observations in the principal components arearithmetically potential to lead the way to the sources of the dominating variables. An assistance of the ordinary kriging technique for interpolating the factor scores on the study area map, promisinglydefines the spatial sources of the variables.
Keywords *- Principal Component Analysis, Groundwater Contamination, Source Approximation, Interpolation, ArcGIS.*

Date of Submission: 15-03-2023 Date of Acceptance: 31-03-2023

I. INTRODUCTION

There are many such studies, which have systematically publicized the confirmations of contaminations in the groundwater systems; especially in the localities of industrial and populace ascendancy. The term industrial and populace implicates the setups like of power plants, factories, active and dormant mining sites, high density population sites like slums, etc. Being the informal source of fresh water, the groundwater regimes are the most exploited on one flank and utmost neglected on another (Satapathy*et al.*, 2009; Mufid al-hadithi, 2012). Such misbalanced attitude towards it leads into many such issue, which can neither be resolved with ease nor can be avoided with poise. The groundwater can be at risk to versatile set of contaminations including; hydro-geochemical misbalance; trace metal contaminations; biotic infections and physical adulterations (Eckenfelder, 2000).

The hydro-geochemical misbalance implicates the concentration of certain set of ions among major anions viz., HCO_3^-, CO_3^{2-}, SO_4^{2-}, Cl^- and NO_3^- and cations viz., Ca^{2+}, Mg^{2+}, Na^+ and K^+in the groundwater leading into its unsuitability for the domestic purpose (Ganvir and Guhey, 2022). The trace metal contamination (also includes heavy metals) in the groundwater is the exceedingof metals' (As, Al, Cd, Cr, Hg, Mn, Ni, Pb, Zn, and many more) respective acceptable/permissible concentration (Ganvir and Papadkar, 2022). The contamination of such kind is considered to be most harmful in accordance with the effects it produces through bioaccumulation. The biotic infections fingers towards the any sort of microbiological adulteration which may cause a serious threats to biotic health like Cholera, Diarrhoea, Hepatitis A, typhoid, etc. The physical adulteration includes the turbidity, coarse suspended solid, raised temperature, etc. which makes the water look undrinkable at its very first site and it means too.

Once the sort of the groundwater contamination and its nature is confirmed, an endeavour for outlining the contaminants' source becomes an obligation for an ideal study. The source approximation is a process, where the plausible sources of the contaminants are demarcated in terms of time and space (Hwang *et al.*, 2001). Such demarcations are done in a very tactical way to keep the error at low. The source approximation unwaveringly provides a ground to initiate the all feasibleremediation (Ganvir and Guhey, 2020). The Principal Component Analysis (PCA)is the one of the usual statistical tool employed for the source approximations in the groundwater contamination studies (Facchinelli*et al.*, 2001). The present article deals with the procedures and elucidations of PCA in concern to groundwater contaminations. This attempt is exclusively motivated to edify the young scholars regarding the appropriate engagement of the PCA. The objectives can be enlisted as below;

- To introduce the PCA in general.
- To elucidate the procedures to be executed during PCA.
- To reveal the implications of PCA in groundwater studies.
- To enlighten the approach of getting best possible interpretations from PCA through the aid of other compatible tools.

II. METHODOLOGY

To illustrate the PCA significantly and to put a definite systematics for the potential researchers engaged in groundwater contamination studies, a wide-ranging works were engaged to sort the best valued articles dealing with PCA. Thecoreobtained from the survey done over the works has beenassociated with the existingdevelopments in the field of groundwater contaminations, entangled with the PCA or similar kind of tools. All the interpretationswere converged to deduce aninclusiverepresentationsignifying the practicability,efficacy, appropriateness, merits and demerits of the PCA. The endeavour has been so designed to make sure the attainment of the objectives.

III. PCA: GENERAL

The opening idea leading to the development of PCA was first proposed by Karl Pearson in 1901 and later advanced by Harold Hotelling in 1930s to the existing one. Technically, the PCA is a statistical tool to augment the interpretability of the massive information per observation. The existing facets of the derived information are put in a summarized dimension by PCA to get a clearer elucidation of variance as was earlier. Apart from the general utilization, it is largely designated as a pathfinder in source approximations. The usual engagement of PCA as a multivariate statistical tool to site and categorise the pollution sources can be justly illustrated by the significance it has stood in various contamination studies. It has been observed to be best in the spatial spotting of point sources.

Derivation of Principal Components

The PCA correlates the data by converging them into the merged variables of high implications. The operation includes the conversion of observed variables into the components signifying the variance in descending order. The component with apex variance is sorted as PC1, followed by the successor components viz., PC2, PC3, PC4 so on. These components are technically obtained by the process that essentially includes covariance matrix and it initiates form defining dataset proceeded by calculation of covariance matrix. Once the step of covariance matrix is done, the calculations of eigenvectors and eigenvalues from covariance matrix should be executed. The eigenvector with highest eigenvalue is the PC1 and with the dropping of eigenvalues respective components comes out. The foremost principal component i.e. PC1 is a variable resulting from the linear combination of the observed variables demonstrating apex variance throughout the dataset. The PC2 elucidates the apex variance left in the dataset on the omission of PC1's influence and the same goes on for further components. The Scree Plots are used to demonstrate the distribution of variance over principal components (Figure 1). The entire variance in the obtained dataset and eigenvalues are represented on the vertical axis; whereas, the horizontal axis represents the principal components.

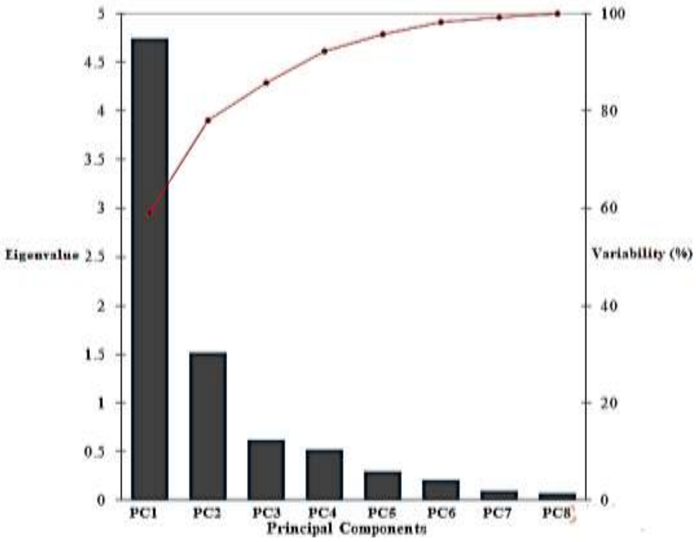

Figure 1: A sample scree plot representing derived principal components for a random dataset.

Biplot: Introduction and Implication

A systematic illustration of the biplot was first given by K. R. Gabriel in 1971. It is a graphical representation of the observations and variables under observation in lone scatterplot. Abiplot in the PCA indicates the principal component scores of the observations and factor loadings of the variables in form of dots and vectors respectively (figure 2). The PC scores represent the extent of data variance; whereas, the factor loading represents the relationship coefficients between the PC scores and the observed variables. The magnitude of the loading value indicates the strength of relationship; whereas, the sign stands for positive and negative association.

The specificvector's dimensionimpliesfor thatspecific observed variable.The longer the vector of that variable, higher will be its credibility. The angle between any two variable vectors impliestheir bond. When theangle between any two variables is below than 90^o will representaffirmativebond. The word affirmative stands for positive covariance. The angle of 90^o signifies no bond between the variable; whereas the vectors directingin opposite ways i.e. more than 90^{o}are taken as negatively bonded. The dots represent the observations done for the variables. The biplot also sorts the observations into groups of likeness. The observation having analogous tendencies groups together representing similar kind of constitutions implying concomitant roots (Einax*et al.*, 1997).

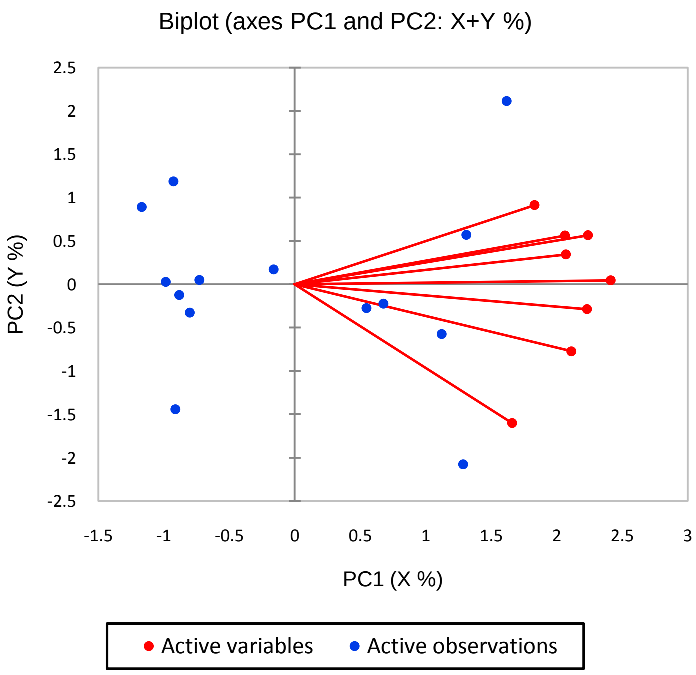

Figure 2: A sample PC biplot of observations and variables.

IV. PCA: SPECIFIC TO GROUNDWATER CONTAMINATION

The practicality of the PCA in the groundwater contamination studies have been proved through many endeavours. The most important aspect that constitutes the groundwater contamination studies is the assessed set of hydro-geochemical variables in an individual observation done on groundwater sample (Ganvir, 2023). The interpretation complexities upsurge with the higher number of observations taken. The underneath hydro-geochemical conditions are sometimes so fragile and varying, that the assessed set of data contrasts from place to place in the natural heterogeneous geologic condition; whereas, when the hydro-geochemical setup is disturbed due to some exploratory cum exploiting drive like that of mining, the hydro-geochemical conditions are far away from the stability, irrespective to the geology (Ganvir and Guhey, 2021). Hence, a tool like PCA is quite efficient in dealing with the variance in the data usual in endeavours of hydro-geochemistry.

To understand the exact course of PCA over the data generated in the assessment of the groundwater sample, a sample data can be taken into consideration (Table 1). Let's consider that in a groundwater assessment endeavour, all total 04 samples viz., 1, 2, 3 and 4 were taken. The assessed hydro-geochemical variables viz., A, B, C and D, are explained by the values in terms of W, X, Y and Z.

Table 1: Sample result table for the variables assessed in groundwater samples.

Observations	Hydro-geochemical Variables			
	A	B	C	D
1	W	W1	W2	W3
2	X	X1	X2	X3
3	Y	Y1	Y2	Y3
4	Z	Z1	Z2	Z3

Principal Components and Hydro-geochemical Variables

On an application of procedure cited in the section of *Derivation of Principal Components* the scree plot can be drawn considering all possible principal components in order of significance. The scree plot clearly indicates the percentage contribution and the eigenvalues of the respective components. In usual case, the first three principal components hold the maximum variance over the data set. Hence, tabulation can be made representing only first three principal components with the respective factor loadings and percentage contribution of the assessed variables (Table 2). The table manifests the individual contribution of the variable in the principal components and hence, the component wise dominant ones can be deduced with ease. Such tabulation elucidates the exact correlation among the assessed variables. The variables varying together i.e. having dominance together in the principal component implicates the common or associated source. With the decadency, the influence of the principal component over the subjected data set goes on waning, hence not been considered significantly.

Table 2: Sample table for thefactor loading and the respective percentage contributions of the first 03 principal components for the assessed variables.

Variables	PC1	PC1's Contribution in %	PC2	PC2's Contribution in %	PC3	PC3's Contribution in %
A	×	×	×	×	×	×
B	×	×	×	×	×	×
C	×	×	×	×	×	×
D	×	×	×	×	×	×
Eigen Value	×	100.00	×	100.00	×	100.00
% of Variance	×		×		×	

Biplot: Hydro-geochemical Variables and Observations

The best upshot of the PCAis the biplot manifesting the positioning of assessed variables as vectors and observations as dots, denoting manifold associations (Figure 3). The vectors representing A, B and D are longer than the vector B, which implies the insufficient representation of variable B in the data set. The acute angles among the A, B and D, indicates positive relation; whereas the vector C is non-associated to the rest. An insignificant negative relation between can be predicted between A and C due to obtuse angularity. The dots demonstrating observations are positioned on the behalf on likeliness. The together the observation will be, the more will be the likeliness. The togetherness also provides a sound manifestation of similar kind of source they have.

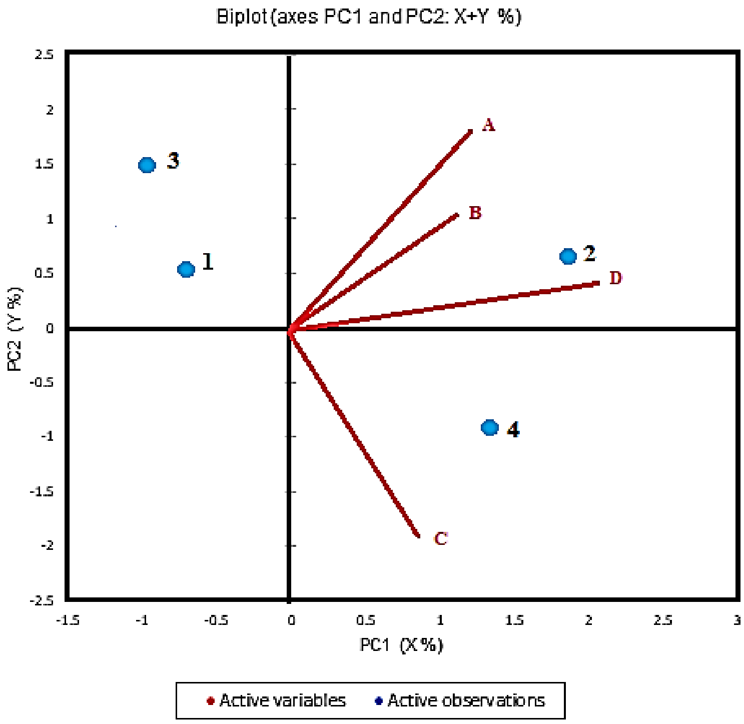

Figure 3: A sample PC biplot for assumed observations and variables.

Source Approximation by Interpolation

The factor scores of the observations in concern to the respective principal components can be enlisted (Table 3). This essentially shows the influence of the observation over the derived components. These factor scores of observations can be used very efficiently to plot them on the spatial maps of the study areaby the process of ordinary kriging method of interpolation. The best results are obtained on the base maps superimposed by geology of the study area. In geo-statistics the ordinary kriging technique produces the best and the unbiased linear approximations of the variables in the known and unknown sites of groundwater samplings. It has also been engaged for the finest kind simulations of vagueness around the study area by projected values (Deutsch and Journel, 1998). This technique can be implemented by the ArcGIS software. Once the interpolation of factor score is done, a systematic depiction of the hotspots implying sources of the variables on spatial region can be achieved. As the principal component is dominated by certain set of variables, therespective interpolation map only highlights the sources of the same set dominating variables. Hence, principal component specific maps are advised to be prepared; for at least first two components.

Table 3: Sample table for the factor scores and the respective percentage contributions of the first 02 principal components for the observations.

Observations	PC1	PC1's Contribution in %	PC2	PC2's Contribution in %
1	×	×	×	×
2	×	×	×	×
3	×	×	×	×
4	×	×	×	×
Total		100 %		100 %

V. CONCLUSION

The groundwater studies involve massive data for interpretation. The assessment of the hydro-geochemical variables is not sufficient enough for a fruitful outcome. The source approximation has to be done for the comprehensive significance of the endeavoured study. The principal component analysis is the statistical tool which not only lowers the variance in the dataset, but also represents a scenario with high level of elucidation. The derived principal components along with the contribution of evaluated variables in them demonstrations all the facets perfectly. The biplot of two major principal components indicating the observations and variables discloses various associations, which were otherwise concealed in the dataset. The interpolation of the observations' factor scores on the base maps promisingly leads to the source approximations of the variables.

REFERENCES

[1]. Satapathy, D.R., Salve, P.R. and Katpatal, Y.B. Spatial distribution of metals in ground/surface water in the Chandrapur district (Central India) and their plausible sources. Environmental Geology. 2009;v.56:323-1352.
[2]. Mufid al-hadithi. Application of water quality index to assess suitability of groundwater quality for drinking purposes in Ratmao – PathriRao watershed, Haridwar District, India. American Journal of Scientific and Industrial Research. 2012;v.3(6):395-402.
[3]. Eckenfelder Jr., W.W. Kirk- OthmerEncyclopedia of Chemical Technology. John Wiley & Sons, 2000.
[4]. Ganvir, P.S. and Guhey, R. Hydro-geochemical elucidation and its implications in the Wardha valley coalfields of central India. IOP Conference Series: Earth and Environmental Science. 2022;v.1032:012015.
[5]. Ganvir, P.S. and Papadkar, J.N. Hydro-Geochemistry and Human Health: A Brief Review. International Journal of Food and Nutritional Sciences. 2022;v.11:223-227.
[6]. Hwang, C.K., Cha, J.M., Kim, K.W. and Lee, H.K. Application of multivariate statistical analysis and geographical information system to trace element contamination in the Chungnam coal mine area, Korea. Applied Geochemistry. 2001;v.16:1455–1464.
[7]. Ganvir, P.S. and Guhey, R. Assessment of Some Heavy Metals Toxicity and Its Probable Remediation in Groundwater around Telwasa and Ghugus Area of Wardha Valley Coalfields, Maharashtra. Journal of Indian Association of Sedimentologists. 2020;v.37(1):33-43.
[8]. Ganvir, P.S. and Guhey, R. Groundwater quality assessment with reference to some heavy metals toxicity and its probable remediation around Ballarpur area of Wardha valley coalfields, Maharashtra. IOP Conference Series: Earth and Environmental Science. 2020;v.597:012001.
[9]. Facchinelli, A., Sacchi, E. and Mallen, L. Multivariate statistical and GIS-based approach toidentify heavy metal sources in soils. Environmental Pollution. 2001;v.114:245–276.
[10]. Pearson, K. On Lines and Planes of Closest Fit to System of Points in Space. Philosophical Magazine. 1901;v.2(11):559-572.
[11]. Hotelling, H. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology. 1933;v.24:417–441 and 498–520.
[12]. Gabriel, K. R. The biplot graphic display of matrices with application to principal component analysis. Biometrika. 1971;v.58(3):453–467.
[13]. Einax, J. W., Zwanziger, H.W. and Geiss, S. Chemometrics in environmental analysis. Wiley-VCH, Weinheim, 1997.
[14]. Ganvir, P.S. Hydro-Geochemical Plots: An Efficient Tool for the Elucidation of Groundwater Chemistry. International Journal of Innovative Science and Research Technology. 2023;v.8(2):95-100.
[15]. Ganvir, P.S. and Guhey, R. Geochemical Studies of some Heavy Metals' Toxicity in Groundwater with their Plausible Sources around GondwanaSupergroup, Wardha valley Coalfields, Maharashtra. Journal Geological Society of India. 2021;v.97:1415-1421.
[16]. Deutsch, C. V. and Journel, A. G. GSLIB: Geostatistical software library and user's guide (2nd edn.). Oxford University Press, New York, 1998.

Priyadarshan S. Ganvir, et. al. "A Systematic Approach to Approximate the Sources of Contaminants in the Groundwater Studies."*IOSR Journal of Applied Geology andGeophysics (IOSR-JAGG)*,11(2), 2023, pp. 23-28.