CHAPTER 5 GROUNDWATER CONTAMINATION SOURCE IDENTIFICATION

5.1 General

Principal Component Analysis (PCA), Factor Analysis (FA) and Hierarchical Cluster Analysis (HCA) are three distinct statistical techniques commonly employed in data analysis. PCA is utilized to reduce the dimensionality of a dataset by transforming the original variables into a new set of uncorrelated variables known as principal components. This technique is particularly useful for data compression, visualization, and noise reduction. Factor Analysis (FA) aims to uncover underlying hidden factors that explain the observed correlations among variables. On the other hand, Hierarchical Cluster Analysis (HCA) focuses on grouping data points or objects based on their similarities or dissimilarities, creating a hierarchical tree-like structure (dendrogram) that helps identify natural clusters within the data. HCA is often used in unsupervised learning to uncover patterns and relationships within datasets. These three techniques serve different purposes and are applied in various fields depending on the specific goals of data analysis.

5.1.1 Multivariate Statistical Analysis

The multivariate statistical analysis is used for the purpose of simplifying groundwater quality data containing a large set of geochemical parameters for a better understanding of regional hydrogeology. Such method provides strong basis for the problem of groundwater contamination influenced by either natural or anthropogenic activities when combined with spatial and temporal distribution of obtained key factors in GIS environment. This technique narrows down on the factors that are responsible for the groundwater system governance and draws attention towards tools that help in solving problems related to groundwater contamination.

5.1.2 Z-Score for Data standardization

Data standardization or normalization helps in bringing all the physico-chemical parameters of groundwater quality analysis on the same scale as the range of contaminant concentration differs from each other. For example, as per Indian drinking water standards IS-10500 (2012) pH ranges from 6.5 to 8.5, the permissible limit of groundwater Nitrate is 45 mg/l whereas TDS ranges from 500 to 2000 mg/l. Such large scale variation in values created the necessity of data standardization to reduce errors. The groundwater quality data were standardized using Z-score method in SPSS software using following equation.

$$\mathbf{Z = (X_{mn} - \mu_m)/\sigma_m} \qquad \textbf{(1)}$$

Where, X_{mn} - concentration of m quality parameter at nth site, μ_m – mean concentration of m quality parameter of all n sites, σ_m – standard deviation of m quality parameter

5.1.3 Principal Component Analysis (PCA)

The z-scale transformation was performed on the observed ground water quality data before the application of PCA for standardization. The basic function of PCA is dimension reduction of a large dataset in terms of Principle Components (PCs) so that a smaller number of factors can account for maximum variation of the given dataset. The PCA is done by computing eigen values and eigen vectors from the correlation matrix of groundwater quality dataset where PCs having eigen value greater than 1 show the best variance for the whole groundwater quality dataset. The PC loadings highlight the strong positive or negative correlation among the groundwater quality parameters and their respective PCs. The numbers of PCs to be retained were determined using Kaiser Normalization criterion (Kaiser H. 1958).

5.1.4 Hierarchical Cluster Analysis (HCA)

The HCA approach organizes the large number of sampling locations in terms of clusters with high homogeneity explaining qualitative similarity in groundwater dataset. Such segregation is helpful to understand the grouping of each well locations based on the similarity of each physico-chemical parameter to identify the sources of contamination. HCA approach is widely used where clusters are formed in sequence, begins with the most analogous pair of objects. The HCA outcome can be presented in the form of dendrogram which is a tree like structure showing the classification of clusters. Such dendrogram clears the pictures of various sampling sites falling in a specific cluster. The Ward-Linkage method was used for the HCA.

Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA) are two fundamental techniques in data analysis with distinct purposes. HCA is primarily employed for clustering and categorizing data points or objects based on their degrees of similarity or dissimilarity. In contrast, PCA serves as a powerful method for dimensionality reduction and feature extraction. Its main objective is to reconfigure the original variables into a fresh set of uncorrelated variables, known as principal components, while retaining as much of the data's variance as possible. While HCA is focused on identifying natural clusters within data, PCA is instrumental in simplifying data complexity and aiding subsequent analysis and interpretation.

5.2 Methodology for Identification of Groundwater Contamination Sources

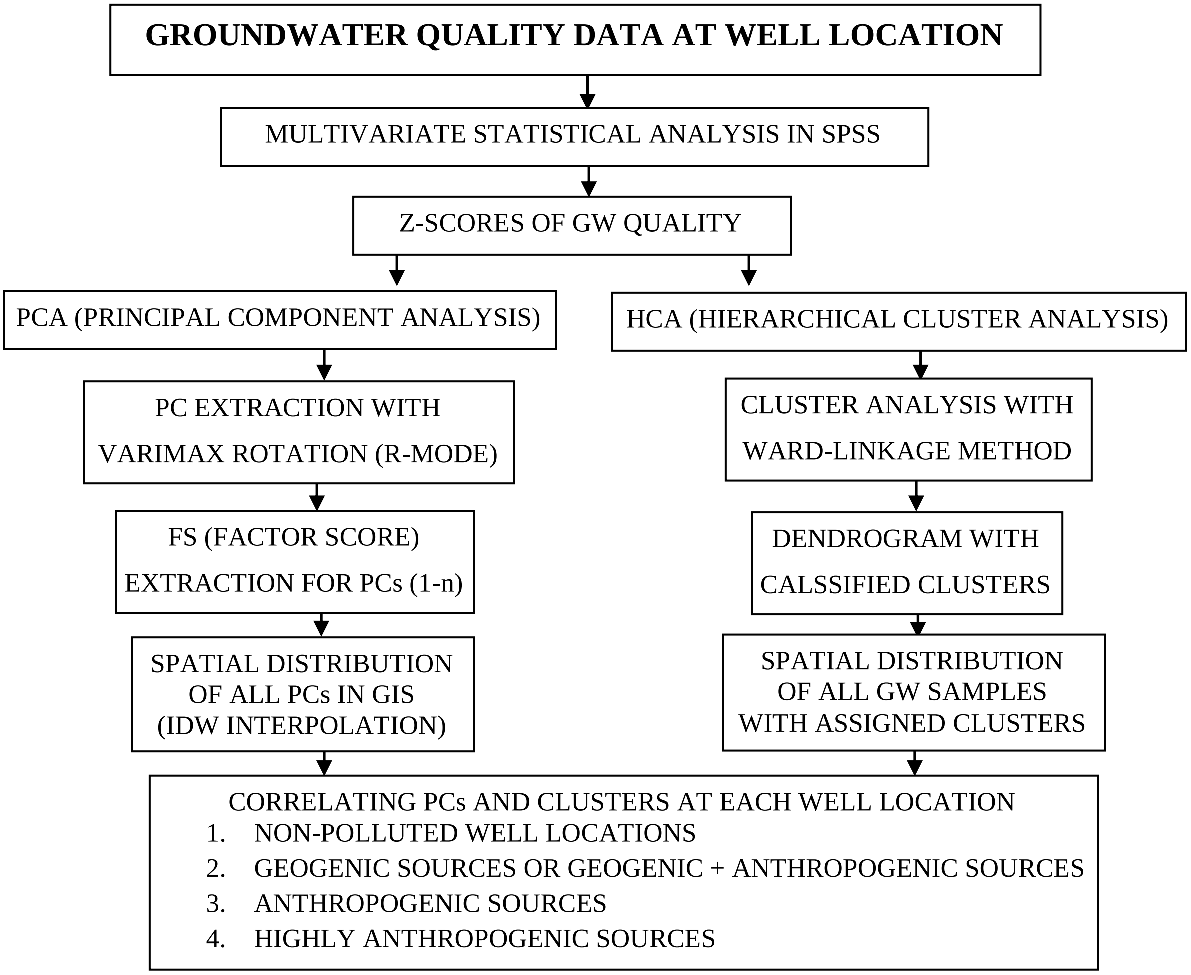


Fig 5.1 Methodology-Flow chart of Contaminant Source Identification

The above figure 5.1 explains step wise methodology for the identification of groundwater contaminant sources. First step is to perform multivariate statistical analysis on groundwater quality data collected from MOJS-DDWAS portal for 2018 pre monsoon season. At each well location 10 common groundwater quality parameters (pH, TDS, NO_3, F, Cl, SO_4, Ca, Mg, TH and ALK) data are available which were imported in SPSS software and all the parameter values were transformed into Z-scores. The PCA was performed using these Z-scores and PCs (1-n) as well as FS were extracted at each well location. These FS values at each location were input in GIS environment to obtain their spatial distribution. The HCA was performed on all the well location which resulted in a dendrogram that explained distribution of wells falling in the same clusters.

Basic statistics of groundwater quality data from MOJS-DDWAS is as below. The average values of groundwater Nitrate in both 2018-pre monsoon (table 5.1) for the entire alluvial region is above the permissible limit (45 mg/l) which created the necessity of identifying the sources of contamination. The high values of CV in Nitrate and Sulphate especially in the pre monsoon season. The parameters NO_3, TDS, Mg, TH and ALK showed highest values in the pre monsoon season.

Table 5.1 Statistical description of MOJS-DDWAS 2018-Pre monsoon data

PARA.	pH	TDS	NO_3	F	Cl	SO_4	Ca	Mg	TH	ALK
Avg.	7.80	1098	50	0.77	234	46	78	71	490	459
Min	6.89	382	2	0.07	36	6	14	15	104	228
Max	8.41	4120	284	2.15	848	473	239	424	1810	828
St.Dev.	0.26	614.61	53.21	0.43	188.34	69.39	51.30	60.82	313.24	146.52
CV	3.31	55.97	107.37	56.32	80.41	151.91	65.53	86.07	63.92	31.96

5.3 Principal Components Analysis Outcomes

The variance values of groundwater quality parameters are shown in table 5.2 as the initial eigen values for 2018. The Extraction Sums of Squared Loadings of variables includes only Eigen value > 1 which means more total variation in the data than individual parameter, and factor with Eigen value < 1 explains less total variation. Now, only factors with eigen value >1 are retained for further Rotation Sums of Squared Loadings which reduces the number of variables.

Table 5.2 Principal Component Analysis (MOJS-DDWAS-2018-Pre)

Total Variance Explained									
PC	Initial Eigenvalues			Extraction Sums Squared Loadings			Rotation Sums Squared Loadings		
	Total	Variance%	Cumulative%	Total	Variance%	Cumulative%	Total	% Variance	Cumulative %
1	5.144	51.441	51.441	5.144	51.441	51.441	4.963	49.634	49.634
2	1.502	15.018	66.459	1.502	15.018	66.459	1.523	15.226	64.860
3	1.223	12.231	78.691	1.223	12.231	78.691	1.383	13.830	78.691
4	0.827	8.273	86.963						
5	0.512	5.120	92.083						
6	0.468	4.684	96.767						
7	0.248	2.476	99.243						
8	0.063	0.632	99.875						
9	0.012	0.125	100.000						
10	0.00002	0.000	100.000						

In this research, three principal components with Varimax rotation having eigen values greater than 1 were extracted from SPSS environment which explained 78.69 % of cumulative variance (table 5.3) for 2018-pre. For the 2018-pre monsoon seasons (table 5.2), the factor loadings for PC-1 (TDS, NO_3, Cl, SO_4, Mg) showed 51.44% of total variance, the PC-2 highlighted 15.01% of variance with high positive loadings on (F, ALK) parameters and the PC-3 observed 12.23% variability containing (pH, Ca, TH) parameters.

Factor loadings for PCs are described in table 5.3 for 2018 pre monsoon season. The factor scores at each well location were used for spatial distribution of individual PCs as well as a composite PC in GIS considering IDW interpolation method. These maps highlighted the areas contaminated from anthropogenic sources having FS>2 as shown in the figures 5.1 to 5.4 for 2018-pre monsoon.

Table 5.3 Factor loadings for PCs after Varimax Rotation (MOJS-DDWAS-2018-Pre)

Rotated Component Matrix			
GWQ Parameters	**Component**		
	PC-1	**PC-2**	**PC-3**
TDS	0.919		
NO_3	0.783		
Cl	0.705		
SO_4	0.943		
Mg	0.954		
TH			0.511
ALK		0.818	
F		0.793	
pH			0.350
Ca			0.939
Extraction Method: Principal Component Analysis, Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 4 iterations.			

The following figures 5.2 to 5.5 show individual PCs as well as a composite PC map developed in GIS environment using IDW interpolation method for 2018-pre monsoon season. The PC-1 map (fig. 5.2), associated with high positive loadings of TDS, NO_3, Cl, SO_4 and Mg groundwater quality parameters which can be attributed as anthropogenic factor occurring from sources such as extensive use of fertilizers in agriculture, wastes from animal husbandry and improper arrangement of septic tanks and soak pits. The agricultural activities of rural area, petro-chemicals based industries and shallow depth of groundwater (around 35 fbgl) are the prominent causes for high factor scores of PC-1 in the Northern parts of the alluvial region. Wells 24 and 50 showing higher factor scores are located near a small river stream and an industrial pocket respectively. The dumping of ill-treated industrial waste water into small streams that eventually percolates towards groundwater is another significant factor which explains high loadings of Mg and TDS parameters. The gradual decrease in factors scores is in alignment with the movement of groundwater from North-East to South-West direction.

The PC-2 map (fig. 5.3) is associated with high positive loadings of F and ALK groundwater quality parameters. The higher factor scores are observed in the Northern and Central parts of the alluvial region where the depth of groundwater is comparatively higher (around 70 mbgl). Over exploitation of groundwater, extensive use of phosphate based fertilizers and industrial wastes are the key reasons for such high factor scores. The well 26 in the Northern zone, wells 42 and 43 in the central zone are located in the vicinity of small rivers stretches whereas well 44 is located around small scale industries in the central zone.

The PC-3 map (fig. 5.4) shows high positive loadings of both TH, Ca and pH parameters observed mostly in Northern parts of the alluvial region where the depth of groundwater is comparatively shallow (around 40 bgl). The higher factor scores from PC-3 also pointed towards the anthropogenic category of contamination source such as an industrial area that causes mixing of waste water into small drains which infiltrates into the groundwater. The well 8 located in the Northern zone is falling in such an industrial pocket where majority of the medium scale materials manufacturing units exist.

The composite map (fig. 5.5) obtained from overlay analysis in GIS environment helped in identification of the contamination sources. The majority of the contaminated pockets were identified in the Northern (wells 6, 8, 15 and 24) and central parts (43 and 44) of the alluvial region showing factor score greater than 2 being considered as the anthropogenic contamination sources. The conventional agriculture practices, rapid growth of minerals and manufacturing industries as well as shallow depth of groundwater are the key anthropogenic sources responsible for contamination.

Table 5.4 Source Identification by PCA

Sr. No.	Factor Score Range	No. of Wells	Remarks
1	< 1	1, 2, 3, 4, 5, 8, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 22, 25, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 45, 46, 47, 48, 49	Non polluted
2	1 to 1.5	15, 28, 41	Geogenic + Anthropogenic or Geogenic
3	1.5 to 3	6, 8, 9, 21, 23, 26, 39, 42, 43, 44, 50	Anthropogenic Source
4	> 3.0	24	Highly Anthropogenic Source

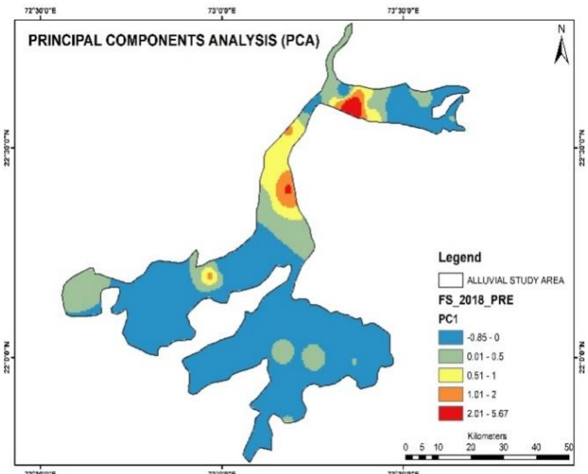


Fig 5.2 PC1 (TDS, NO_3, Cl, SO_4, Mg) (2018-pre)

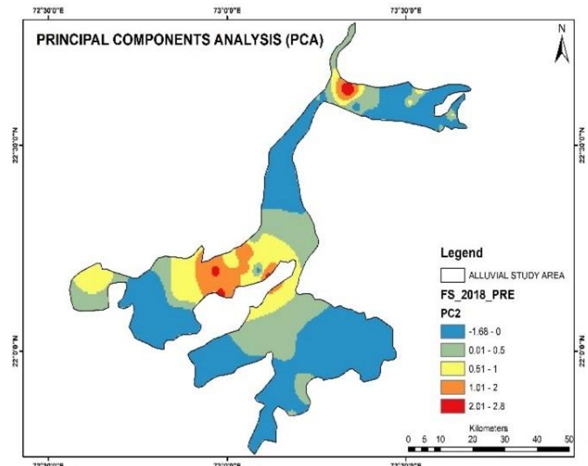


Fig 5.3 PC2 (F and ALK) (2018-pre)

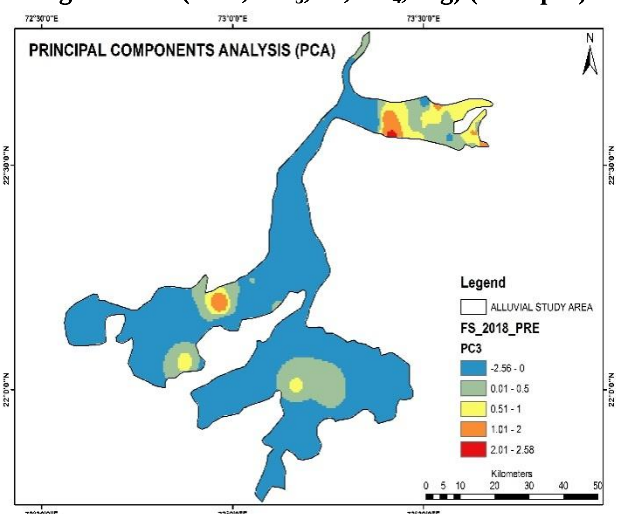


Fig 5.4 PC3 (pH, Ca, TH) (2018-pre)

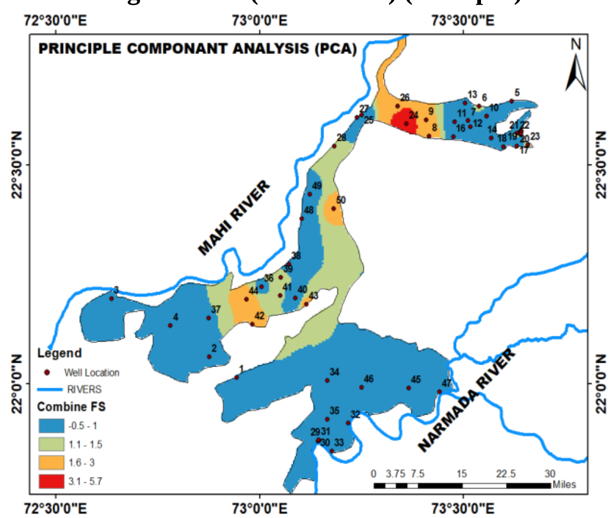


Fig 5.5 Composite PC (2018-pre)

Negative and extremely low factor scores in specific areas suggest that the groundwater quality is not polluted. Conversely, higher factor scores indicate varying degrees of contamination, in ascending order from Geogenic sources, a mix of Geogenic and Anthropogenic sources and anthropogenic sources of contamination (Deepash Machiwal, 2015; N. Suresh Nathan, 2017).

Table 5.5 Factor Score for 2018 – Pre monsoon

SR	District Name	Village Name	ZPH	ZTDS	ZNO3	ZF	ZCL	ZSO4	ZCA	ZMG	ZTH	ZALK	FAC1_1	FAC2_1	FAC3_1	Max FS
1	Bharuch	Vedcha	-0.55	-1.15	-0.36	-1.28	-1.04	-0.57	-1.09	-0.55	-0.87	-1.37	-0.53	-1.68	-1.09	-0.53
2	Bharuch	Jafarpara	0.60	-0.15	-0.66	-1.32	0.30	0.00	0.81	-0.69	-0.22	-1.10	-0.39	-1.09	0.78	0.78
3	Bharuch	Kavi	-3.51	-0.16	-0.79	1.41	-0.68	0.26	-1.20	-0.45	-0.85	1.42	0.36	0.81	-2.56	0.81
4	Bharuch	Nondhana	1.06	-0.82	-0.75	-1.60	-0.81	-0.32	-0.80	-0.58	-0.79	-0.53	-0.73	-0.96	-0.45	-0.45
5	Panch Mahals	Adadara	0.45	-0.25	-0.10	0.08	-0.35	-0.31	0.79	0.02	0.33	-0.18	-0.28	-0.08	0.89	0.89
6	Panch Mahals	Alali	-0.02	0.41	1.48	1.60	-0.05	-0.15	1.69	0.46	1.07	0.74	0.30	1.02	1.60	1.60
7	Panch Mahals	Alva	0.60	-0.39	-0.27	1.23	-0.54	-0.30	0.69	-0.03	0.26	0.15	-0.45	0.72	0.86	0.86
8	Panch Mahals	Bakrol	-0.55	1.15	1.01	0.03	-0.01	0.95	3.10	1.19	2.23	-0.21	0.95	-0.43	2.60	2.60
9	Panch Mahals	Bedhiya	0.29	0.07	1.53	-0.33	-0.83	-0.07	1.69	0.46	1.08	1.15	0.24	0.13	1.53	1.53
10	Panch Mahals	Devpura	0.25	0.30	0.86	-1.07	1.13	-0.32	0.28	-0.24	-0.07	-0.31	0.19	-0.57	0.38	0.38
11	Panch Mahals	Jetpur	0.37	-0.78	-0.14	-0.79	-0.89	-0.35	0.15	-0.30	-0.18	-0.31	-0.48	-0.70	0.21	0.21
12	Panch Mahals	Medapur	-0.25	-0.58	-0.14	-0.72	-0.92	-0.27	0.61	-0.08	0.20	0.21	-0.24	-0.54	0.37	0.37
13	Panch Mahals	Shaktipura	-0.40	-0.50	1.63	0.36	-0.12	-0.51	-0.41	-0.58	-0.65	-1.10	-0.11	-0.65	-0.36	-0.11
14	Panch Mahals	Arad	-0.40	-0.83	-0.22	-0.38	-0.79	-0.39	-0.06	-0.42	-0.36	-0.94	-0.43	-0.95	-0.17	-0.17
15	Panch Mahals	Govindpuri	0.37	0.61	-0.40	1.23	1.12	-0.21	1.27	0.25	0.73	-0.10	0.04	0.84	1.41	1.41
16	Panch Mahals	Navaria	-0.09	-0.58	-0.49	0.08	-0.79	-0.34	0.59	-0.08	0.17	-0.31	-0.39	-0.31	0.51	0.51
17	Panch Mahals	Dhaneshwar	0.33	-0.48	-0.14	-0.52	-0.22	-0.42	0.15	-0.30	-0.20	-0.34	-0.37	-0.46	0.25	0.25
18	Panch Mahals	Jitpura	0.33	0.34	-0.35	0.08	1.61	-0.41	0.15	-0.30	-0.20	-0.58	-0.06	0.09	0.36	0.36
19	Panch Mahals	Kanbipalli	-0.02	-0.70	-0.10	0.40	-0.81	-0.45	-0.16	-0.47	-0.44	-0.02	-0.50	0.05	-0.18	0.05
20	Panch Mahals	Kumbhar Palli	-0.02	-0.93	-0.16	0.08	-0.75	-0.51	-0.33	-0.56	-0.59	-1.23	-0.61	-0.77	-0.22	-0.22
21	Panch Mahals	Ranipura (dam	-0.09	1.10	-0.38	-0.10	1.44	-0.07	2.04	0.64	1.36	0.66	0.50	0.47	1.83	1.83
22	Panch Mahals	Rayan Muvada	-0.32	-0.07	-0.10	-0.20	0.32	-0.37	0.48	-0.14	0.08	-0.88	-0.08	-0.63	0.43	0.43
23	Panch Mahals	Savapura	0.06	0.01	0.62	-0.15	-0.56	-0.20	1.46	0.35	0.88	-1.56	-0.01	-1.14	1.55	1.55
24	Vadodara	Dhantej	-3.09	4.87	4.36	-0.52	3.23	6.10	-1.05	5.75	4.17	0.97	5.69	-0.09	-1.62	5.69
25	Vadodara	Kalupura	0.71	-0.44	1.01	0.26	-0.66	-0.12	-1.14	-0.40	-0.80	0.04	-0.28	0.18	-0.78	0.18

SR	District Name	Village Name	ZPH	ZTDS	ZNO3	ZF	ZCL	ZSO4	ZCA	ZMG	ZTH	ZALK	FAC1_1	FAC2_1	FAC3_1	Max FS
26	Vadodara	Dipapura	0.79	-0.13	-0.29	3.18	-0.64	-0.01	-1.24	-0.76	-1.12	1.66	-0.63	2.81	-1.02	2.81
27	Vadodara	Rasawadi	0.52	-0.59	0.68	-0.95	-0.68	-0.18	-0.12	-0.24	-0.25	-1.23	-0.29	-1.25	0.14	0.14
28	Vadodara	Wankaner	-1.93	0.34	2.20	-0.98	-0.52	-0.08	0.42	1.13	1.07	0.19	1.09	-1.15	-0.07	1.09
29	Vadodara	Vadava	0.37	-0.53	-0.88	-1.09	-0.41	-0.22	-0.58	-0.48	-0.63	-0.15	-0.48	-0.55	-0.51	-0.48
30	Vadodara	Delvada	0.91	-0.83	-0.87	-0.56	-0.68	-0.45	-1.09	-0.71	-1.02	-0.67	-0.83	-0.46	-0.69	-0.46
31	Vadodara	Somaj	-0.25	-0.84	-0.74	-0.79	-0.85	-0.28	-0.72	-0.45	-0.65	-0.50	-0.48	-0.82	-0.78	-0.48
32	Vadodara	Fatepur	0.14	-0.45	-0.85	-0.06	-0.79	-0.34	-0.84	-0.53	-0.77	0.93	-0.52	0.50	-0.88	0.50
33	Vadodara	Pura	-0.86	0.52	-0.88	0.13	1.02	0.16	0.25	0.28	0.33	0.04	0.36	0.14	-0.04	0.36
34	Vadodara	Sandarna	0.91	0.74	-0.08	0.06	1.27	0.20	0.25	0.28	0.33	-0.67	0.20	0.10	0.69	0.69
35	Vadodara	Saniyad	0.21	-1.11	-0.87	1.27	-0.98	-0.35	-0.85	-0.56	-0.80	-1.21	-0.85	-0.03	-0.58	-0.03
36	Vadodara	Bhoj	0.91	-0.07	-0.88	0.81	0.39	-0.05	-0.22	-0.09	-0.16	0.01	-0.36	0.77	0.15	0.77
37	Vadodara	Brahmanvasi	1.17	-0.32	-0.79	0.22	-0.45	-0.34	-0.84	-0.48	-0.73	0.93	-0.64	0.93	-0.48	0.93
38	Vadodara	Chokari	0.56	-0.25	-0.74	-0.10	-0.01	-0.49	-1.18	-0.81	-1.13	0.20	-0.58	0.34	-0.95	0.34
39	Vadodara	Dabhasa	0.68	-0.34	-0.08	2.79	-0.31	-0.11	-0.33	-0.17	-0.27	-0.08	-0.48	1.53	0.07	1.53
40	Vadodara	Goriyad	0.29	-0.68	-0.85	0.63	-0.22	-0.28	-0.72	-0.47	-0.66	-1.10	-0.60	-0.16	-0.44	-0.16
41	Vadodara	Pipli	-0.25	0.01	-0.35	0.26	-0.33	-0.48	-1.20	-0.91	-1.22	1.63	-0.35	1.10	-1.43	1.10
42	Vadodara	Sadra	0.83	0.49	-0.49	0.95	0.20	-0.47	-1.12	-0.76	-1.07	2.48	-0.39	2.28	-1.00	2.28
43	Vadodara	Shihor	2.33	0.84	-0.75	0.13	1.75	-0.14	-0.41	-0.22	-0.35	2.12	-0.21	2.15	0.25	2.15
44	Vadodara	Vishrampura	0.68	1.83	0.98	1.37	1.08	0.97	2.00	1.60	2.12	2.50	1.14	2.27	1.99	2.27
45	Vadodara	Simli	-0.32	-0.13	-0.61	-1.00	0.01	-0.02	-0.10	0.02	-0.02	0.06	0.01	-0.52	-0.26	0.01
46	Vadodara	Tinglod	0.29	0.28	-0.10	-0.98	0.32	0.18	0.30	0.28	0.35	-0.61	0.17	-0.73	0.43	0.43
47	Vadodara	Bhimpura	-0.17	-1.00	-0.83	-0.93	-0.83	-0.35	-0.35	-0.56	-0.59	-1.48	-0.62	-1.43	-0.35	-0.35
48	Vadodara	Karchiya	-1.20	-0.32	-0.49	-0.63	-0.46	0.35	-0.57	0.01	-0.24	0.62	0.19	-0.29	-1.11	0.19
49	Vadodara	Sankarda	-0.82	0.26	0.79	-0.68	0.66	-0.21	-1.03	0.67	0.10	0.45	0.65	-0.22	-1.00	0.65
50	Vadodara	Sisva	-1.90	2.25	0.85	-0.89	3.14	2.34	0.59	0.98	1.02	-0.37	2.14	-0.57	-0.24	2.14

5.4 Hierarchical Cluster Analysis (HCA) Outcomes

The Z-scores obtained for the groundwater quality dataset (MOJS-DDWAS-2018-pre) were used in Hierarchical Cluster analysis (Machiwal and Jha, 2015) with the Ward's linkage method and squared Euclidean distances as a measure of similarity (Loganathan and Ahmed, 2017). The Hierarchical Cluster analysis distributed the dataset into 4 significant clusters. The Dendrogram obtained from HCA shown in figure 5.7 helped in understanding the distribution of well location in each cluster Tables 5.6. Average values of each quality parameter is in tables 5.7 for 2018 pre monsoon season.

Table 5.6 Cluster Classification (MOJS-DDWAS-2018-pre)

Clusters	Well Locations
Cluster-1	1, 2, 4, 5, 7, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 22, 23, 25, 27, 29, 30, 31, 33, 34, 35, 36, 40, 45, 46, 47, 48, 49
Cluster-2	3, 26, 32, 37, 38, 39, 41, 42, 43
Cluster-3	6, 8, 9, 15, 21, 28, 44, 50
Cluster-4	24

Table 5.7 Average of GWQ parameters for each cluster (MOJS-DDWAS-2018-pre)

Cluster	pH	TDS	NO_3	F	Cl	SO_4	Ca	Mg	TH	ALK
1	7.83	859	39	0.63	191	29	73	57	418	381
2	7.88	1076	19	1.19	208	29	30	36	226	644
3	7.70	1700	98	0.87	362	78	161	122	912	543
4	7.00	4120	284	0.54	848	473	24	424	1810	602

For the 2018-pre-monsoon season, cluster 1 majorly located in Southern part showed 64% of the well samples with mean values of pH (7.83), Ca (73 mg/L) and TH (418 mg/L) groundwater quality parameters. Figure 5.5 highlights cluster 2 containing 18% of well samples in the central and western parts with high ALK (644 mg/L) and F (1.19 mg/L) parameters. Cluster 3, which included 12% of well locations with high mean values of NO_3 (98 mg/L), Mg (122 mg/L) and TH (912 mg/L) explaining high anthropogenic sources in shallow aquifers of alluvial region. The well number 24 under cluster-4 with the highest amount of NO_3 (284 mg/L), TDS (4120 mg/L), SO_4 (473 mg/L), Mg (424 mg/L) and TH (1810 mg/L) is pointing towards severe contamination from anthropogenic sources. Below fig. 5.6 is showing the cluster distribution of well locations obtained from HCA for 2018 pre monsoon season.

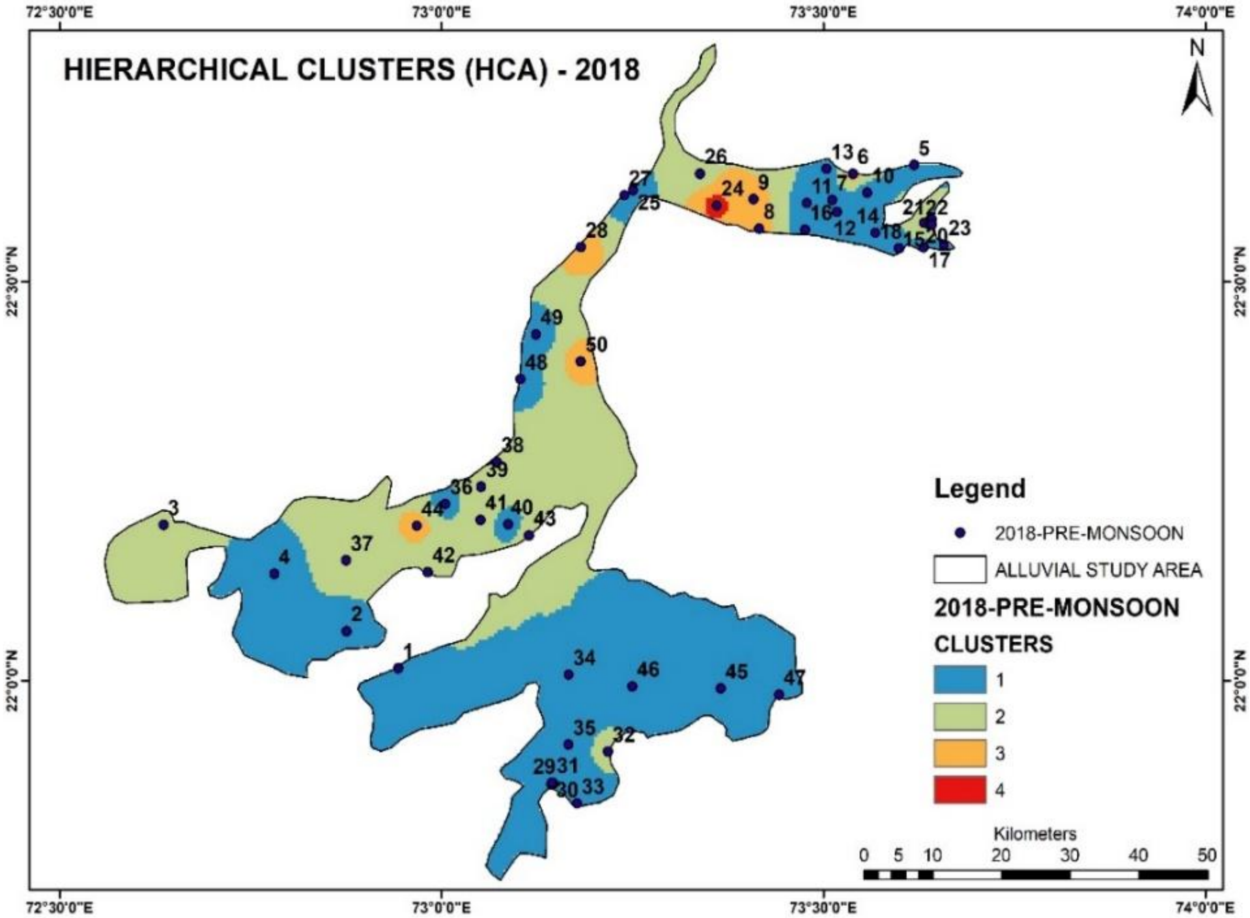


Fig 5.6 Spatial distribution of Clusters from HCA 2018 pre monsoon

The below fig. 5.7 are showing the dendrogram from HCA for 2018 pre monsoon seasons respectively.

5.3 Results and Discussion

The multivariate statistical analysis methods, including Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA) and Factor Score (FS), have proven their efficiency in identifying contamination sources within a large dataset of groundwater quality parameters in the alluvial region. The application of PCA revealed three significant principal components, PC-1 (comprising TDS, NO_3, Cl, SO_4, Mg), PC-2 (consisting of F, ALK), and PC-3 (including pH, TH, Ca), all of which consistently pointed towards the same regions - the northern and central parts of the area. These areas exhibited high contamination levels originating from anthropogenic sources.

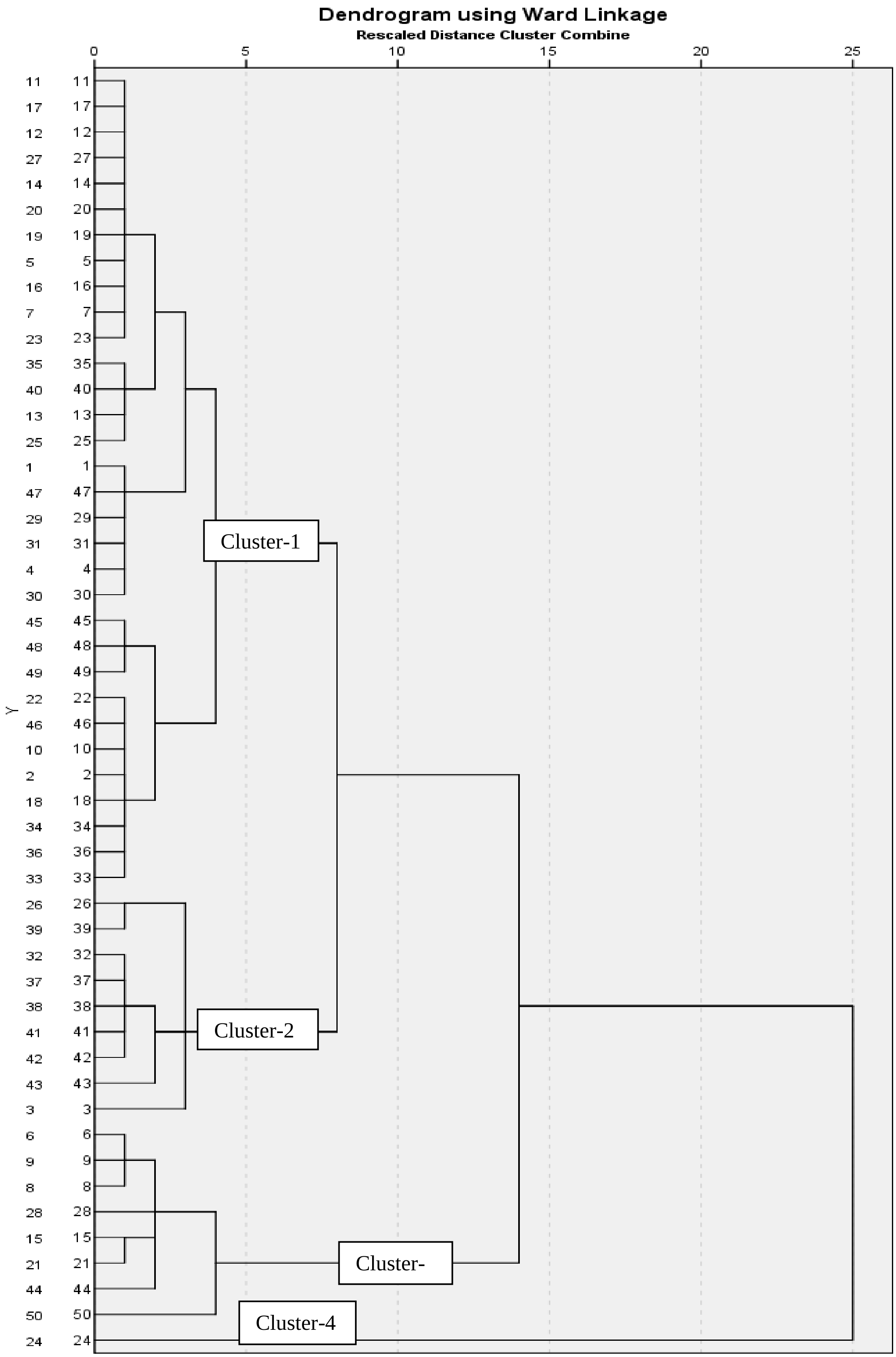
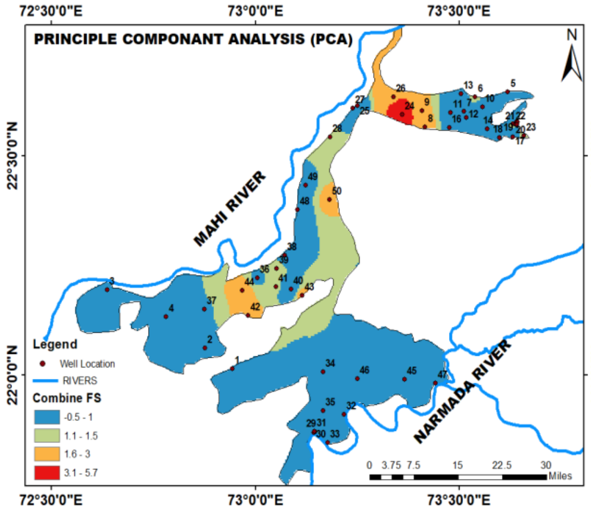


Fig 5.7 Dendrogram for 2018-Pre-Monsoon data (MOJS-DDWAS)



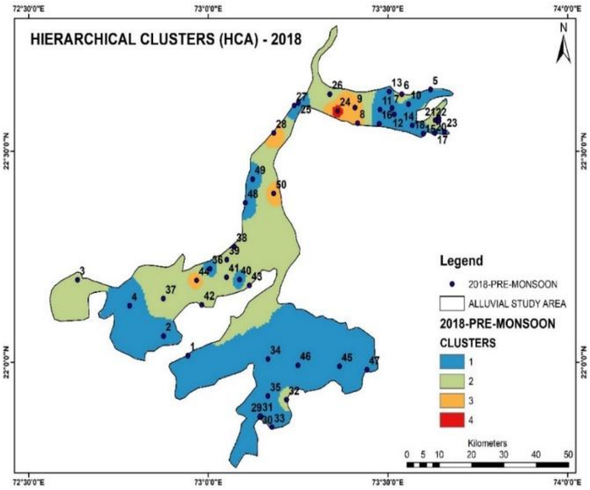


Fig 5.8 Composite PCA results **Fig 5.9 HCA Cluster map**

Further analysis of individual PC maps highlighted the influence of the two main perennial rivers, Mahi and Narmada, flowing in the northern and southern parts of the alluvial region. Shallow groundwater depths (5 – 20 mbgl) towards these rivers correlated with a gradual decrease in factor scores (fig 5.5), indicating the impact of these hydrological features on contamination patterns. Within the study area, there are a total of 50 well locations. Among these, 35 wells exhibit a factor score less than 1, indicating they are non-polluted. Additionally, there are 3 wells with factor scores ranging from 1 to 1.5, suggesting geogenic sources of contamination. Furthermore, there are 11 wells with factor scores between 1.5 and 3, primarily concentrated in the northern region of the study area. The highest factor score recorded is 5.68, predominantly found in the northern area of the study zone.

The categorization of well locations into clusters based on factor scores provided additional insights. Wells falling under cluster 3 and 4, characterized by factor scores higher than 1, indicated contamination from anthropogenic sources. Additionally, cluster 2 contained wells with positive factor scores greater than 1, signifying the need for ongoing monitoring due to combined geogenic and anthropogenic influences.

In the extensive analysis of 50 well locations, notable patterns have emerged across distinct clusters. Table no. 5.7 reveals that, Cluster 1 comprising 32 wells, predominantly situated in the southern and western regions of the alluvial area, represents negative factor scores and is deemed non-polluted.

Cluster 2, situated in central area, demonstrates a collective feature of high alkalinity across its 9 wells. The factor scores, surpassing a value of 1 in majority of wells, signify this pattern, with the origin of pollution linked to inadequate solid waste disposal practices. Cluster 3, which encompasses 8 wells in the northern upper zone, the examination common characteristics of Total Dissolved Solids (TDS) and Total Hardness (TH). Within this cluster, there is a notable presence of high Total Hardness, attributable to the proximity to industrial zones, consequently resulting in heightened Nitrate concentrations.

Cluster 4, specifically well number 24, manifests a highly contaminated zone in the northern part, validated by excessive Urea usage, animal husbandry practices, and improper domestic waste disposal near a pond. Lastly, a significant anthropogenic influence stems from the rapid expansion of Petroleum and Mineral-based industrial manufacturing units, coupled with improper waste disposal into small drains and the Mini river, notably in the northern reaches of the alluvial region. These findings emphasize the pressing need for targeted interventions and sustainable practices to preserve the groundwater quality and ecological integrity of the studied area. The combination of PCA, HCA and FS has effectively revealed contamination sources in the alluvial region's groundwater.