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), 11-22
www.iosrjournals.org

Multivariate Geostatistical Modeling of Fluid Geochemistry in Peninsular and Extra-Peninsular India

Amitabha Roy
Ex-Senior Director, Geological Survey of India

Abstracts
Two sets of spatially dependent multivariate geothermal data representing two spatially distinctive regions of diverse geologic-tectonic settings – one from 2400 km long arcuate belt of tectonically active Extra-Peninsular Himalayan region and the other from Late-Precambrian or Proterozoic mobile belts in the Central Highland in otherwise a stable landmass or shield of Peninisular India, were subjected to robust statistical techniques of Exploratory Factor Analysis followed by multiple regression analyses to find out the genesis of geothermal hot springs spread over these areas conspicuously associating with the respective tectonic zones of different degrees of severity. The objective of Exploratory Factor analysis is to reduce a large number of variables into fewer number of factors or in other words to separate significant few from insignificant many variables.
The multiple linear regression analysis of multivariate data is aimed at explaining variability in dependent variable by means of one or more of independent or control variables. A multiple regression model is used when there is more than one independent variable affecting a dependent variable. While predicting the outcome variable, it is important to measure how each of the independent variables moves in their environment and how their changes will affect the output or target variable. Here in multiple regression analysis too choice of relevant variable (IV) out of many is an issue. One should never enter all the available variables at the same time. Carefully consider which independent variable is distinct or whether relevant to the problem. The first and the most reliable option adopted in the present study is to use factor analysis which creates a small number of factors that account for most of the original variables' information in them but which are mutually uncorrelated.

Both these Exploratory Factor and Multiple Regression analysises corroborate each other in deciphering the origin of these two suites of fluid geochemistry.The model study distinguishes non-magmatic thermal sources as K-Na-HCO3 for Peninsular springs as against the magmatic thermal sources as Cl-HCO3-SO4-Na type of Extra-Peninsular springs. The regression analysis revealed two statistically significant suites of fluid geochemistry – 1. The overall salt assemblage and concentration of Cl-HCO3-SO4-Na-F or chloride rich water suggest the existence of hydrothermal magmatic system operating in geotherms of Extra-Peninsular India and ***2.*** *Peninsular springs of K-Na-HCO3 bicarbonate rich waters with low SO_4^- content and relatively higher contents of HCO_3 compared to other anions SO_4, Cl and F suggestive of a non-magmatic origin*

Keywords: *Geostatistics, Geothermal, Geochemical, Factor Analysis, Multiple Linear Regression, Geologic-tectonic, Peninsular-Extra Peninsular India, Himalayan mountains, Proterozoic mobile belt, spatial dependency*

Date of Submission: 12-03-2023 Date of Acceptance: 26-03-2023

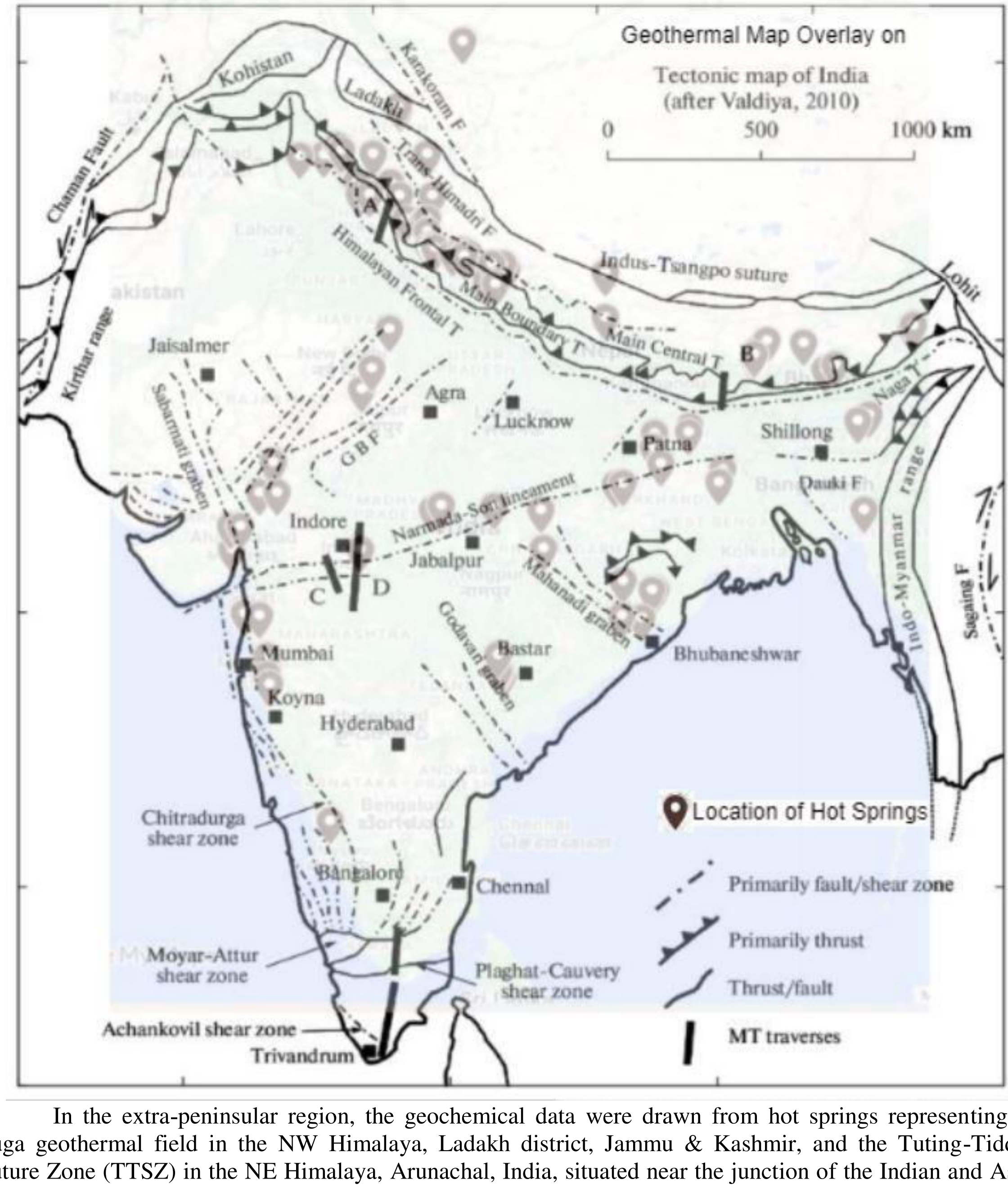
Introduction

In view of the ever-increasing demand as well as the dwindling resources of conventional energy resources like coal, oil, gas, etc., an urgent need was felt all over the world for all forms of alternative renewable energy resources. This enthused high interest in the exploration and exploitation of geothermal resources and gained momentum the world over. India was also not far behind in harnessing geothermal energy.

The geothermal energy is commonly manifested on the terrestrial surface in the form of fumaroles, hot springs, geysers, steaming grounds, and altered grounds. There are about 340 hot springs spread over different parts of India, covering the peninsular and extrapeninsular regions. The GSI initiated its geothermal exploration with the launch of the "Puga Project" in Jammu and Kashmir. Mainly due to the ever-increasing interest shown by various national agencies in geothermal energy, the growth rate of geothermal data has been constantly accumulated over the last few decades. The Geological Survey of India, being the repository of most of the information concerning geological and other related data in the country, has brought out a special publication titled 'Geothermal Atlas of India' (Ravi Shankar, 1991) based on the data compiled from all sources of information, both published and unpublished, on geothermal activities in India. Lack of uniformity and manual handling made data storage, retrieval, and analysis difficult. GTHERMIS is a computerized geothermal database system that offers both speed and storage capabilities (A. Roy, 1994).

Spatial distribution of geothermal fields coinciding with tectonic zones

In the present study, 62 geothermal fields or hot springs spread over both the extra-peninsular and peninsular regions of India have been considered. The GPS coordinates of each of the hot springs have been plotted on Google My Map. This geothermal map has been overlaid on the geologic-tectonic base map.



In the extra-peninsular region, the geochemical data were drawn from hot springs representing the Puga geothermal field in the NW Himalaya, Ladakh district, Jammu & Kashmir, and the Tuting-Tidding Suture Zone (TTSZ) in the NE Himalaya, Arunachal, India, situated near the junction of the Indian and Asian plates and characterized by volcanic sedimentary assemblages of rocks. The Uttarakhand and Himachal Pradesh geothermal fields are situated within the Middle or Lesser Himalayan Crystalline (LHC) Zone,

In the Peninsular region, the geochemical data were drawn from the hot springs spatially distributed in the ancient late Pre-Cambrian or Proterozoic (Aravalli-Delhi, Satpura, and Eastern Ghats) mobile belts in the Central Highland regions of the Peninsular India. The major prominent rifts that separate the southern and northern blocks of the Peninsular Shield are the Central India Tectonic Zone (CITZ), the Narmada-Son–Tapi (SONATA), or the Saihadri–Satpura Lineament.

Literature review

According to previous research (D. Rouwet, 2022; F. Tassi et al., 2010; Mohammad Noor et al., 2021), there is no clear direction about the origin of deep-seated acidic chloride-rich hot springs of the extra-peninsular region in contrast to the shallower alkaline bicarbonate-rich hot springs of peninsular India.

While the goal may be the same, the gaps in knowledge and unresolved problems that are lacking in their studies have been addressed in my research by adopting a definitive approach of statistical and mathematical modeling that gives an insight into arriving at the distinction of two distinct geothermal systems. The geothermal data were subjected to the combined techniques of exploratory factor analysis and multiple regression analysis, which made comparison simpler and easier to follow between the fluid geochemistry inherent in the two distinct geologic-tectonic environs with two very different tectonic histories of varying severity.

Two sets of multivariate fluid geochemical Data

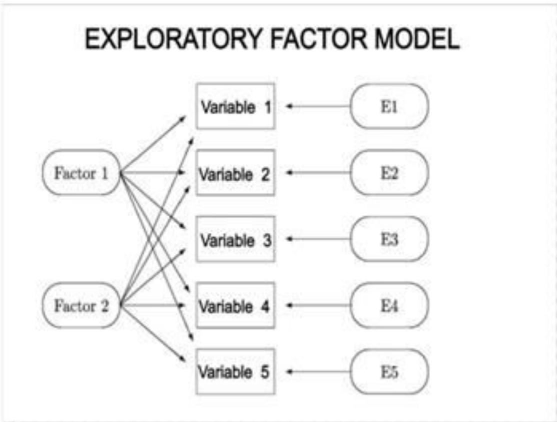
Data Set - I : Extra-Peninsular India

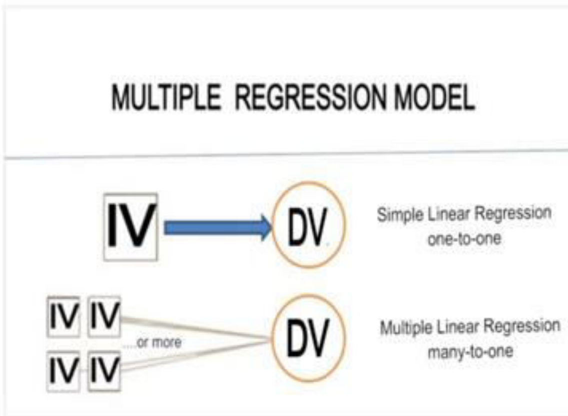
NUM	TEMPC	pH	SPCMHO/cm	HCO3 mg/L	Cl mg/L	SO4 mg/L	TotHard	Ca mg/L	Mg mg/L	Na mg/L	K mg/L	F mg/L	B mg/L	SiO2 mg/L	TDS mg/L
1	59	8.1	1271	300	163	62	0	14	5	210	13	12	5	80	800
2	96	7.7	827	170	133	36	131	44	15	88	19	0.8	33	60	514
3	59	7.1	5260	490	855	1244	1214	342	87	600	109	3.6	138	30	4072
4	24	8.2	795	210	102	83	136	30	15	110	19	1.2	25	25	488
5	56	7.6	1280	342	232	26	72	26	1	260	16	10	10	107	870
6	44	7.6	2015	303	200	340	302	103	11	260	45	6	13	87	1280
7	50	8.3	525	173	45	28	40	13	2	103	5	10	3	23	363
8	90	7.9	1045	276	170	33	0	52	12	135	27	3	10	83	874
9	73	7.5	410	145	30	55	0	38	13	30	7	1	3	50	378
10	52	7.8	25	15	2	0	12	3	1	1	0	0.2	0	2	20
11	50	8.2	700	248	72	48	40	13	2	140	6	5	3	65	480
12	55	7.7	400	272	10	14	0	56	24	8	5	0.4	0	68	366
13	54	7.6	845	445	35	0	0	50	52	50	10	1.2	0	41	536
14	55	6.8	2630	112	1485	22	230	70	13	490	37	1.6	19	115	1630
15	25	7.7	139	103	8	29	0	45	44	24	10	0.7	2	30	442
16	81	8.1	315	117	15	30	0	34	3	30	5	1.6	0	22	245
17	62	8	1460	861	48	14	72	14	99	290	43	3	5	91	1015
18	59	7.8	465	278	12	27	214	42	26	15	8	0.5	1	69	360
19	32	6.7	95	38	5	0	0	6	7	2	1	0.4	0	11	42
20	32	8.3	2000	953	86	0	0	0	47	80	83	0	0	18	0
21	68	6.4	1239	734	12	5	200	64	10	180	38	2	1	130	860
22	56	6.9	770	439	41	21	215	40	23	163	15	4	2	34	510
23	76	7.7	720	254	13	99	32	13	0	135	6	12.5	2.8	101	570
24	28	7.1	770	363	17	66	132	40	8	120	7	10	2	53	575
25	66	7.7	2030	1610	85	57	34	10	2	580	48	10	8	130	840
26	40	7.2	3641	259	11	1484	1352	504	22	200	6	2.5	1	35	2557
27	12	7	1060	233	58	383	536	169	28	10	2	0.2	0	18	834
28	12	7.1	63	32	3	0	24	9	0	2	0	0.4	0	6	42
29	68	6.9	386	112	30	72	40	14	1	56	4	6	1	35	235
30	9	8	178	0	6	12	52	15	3	9	3	1	0.9	9	114
31	18	8.3	205	0	7	2	88	27	5	6	2	0.2	0.9	9	123
32	34	7.4	2668	415	596	16	190	41	21	370	30	3	8	20	1399
33	40	7.6	469	264	13	10	186	44	18	19	10	0.3	0	28	279
34	35	6.6	446	49	104	6	20	7	1	75	3	7	1	28	260
35	52	7.2	8160	435	10	28	113	27	11	133	10	2.1	0	80	565
36	38	7.4	16630	362	154	370	396	127	19	150	17	1	0	60	1017
37	38	7.7	9800	353	35	36	158	54	5	86	9	0	0	40	638

Data Set-II : Peninsular India

NUM	Temp_C	pH	SPCMHO/	HCO3 mg/	Cl mg/L	SO4 mg/L	TotHard	Ca mg/L	Mg mg/L	Na mg/L	K mg/L	F mg/L	Bmg/L	SiO2 mg/L	TDS mg/
1	32	7.5	5090	154	1375	210	872	204	88	660	18	0.7	0	18	298
2	44	7.3	1115	339	165	24	270	82	16	110	6	0.4	0.9	46	50
3	40	7.7	1010	315	130	33	0	110	12	70	25	0	0	58	63
4	0	7.3	1410	390	195	75	0	65	40	210	5	0	0.1	40	83
5	26	7.8	1050	500	140	5	340	70	40	130	2	1	1.2	30	65
6	35.5	7.1	550	290	50	5	230	60	20	30	1	0.3	1.2	30	32
7	40	7.1	28980	190	1347	5	0	390	250	6810	55	0	0	16	[illegible]
8	0	7.2	960	410	110	25	0	45	15	95	2	0	0	4.5	[illegible]
9	43.5	7.4	8350	150	2725	10	0	105	40	1900	30	0.2	3	31	479
10	99	0	0	1534	2428	672	0	9	8	1167	145	0	0	0	574
11	40	7.4	4190	195	1485	0	0	90	40	875	14	0	0	25	[illegible]
12	39	7.6	550	183	71	33	160	40	21	40	2	0	0	22	32
13	54	7.5	13620	13	4800	185	4680	186	10	955	13	0	0.4	87	161
14	43	9	2950	11	850	130	432	170	0.1	368	7	2	0.4	50	186
15	64	8.6	4950	14	1210	144	890	348	0.2	391	8.5	7.2	0	65	270
16	35	8.3	883	18	78	242	109	40	15	155	2	2.5	0	60	56
17	35	8	1917	71	426	107	100	32	6	292	4	1.5	1	5	96
18	61	7.6	1457	30	375	100	147	56	1.8	231	7.8	4	0.4	122	95
19	0	8	0	63	265	108	210	80	44	148	6	0.1	0	60	18
20	91	0	0	177	67	70	0	3	1	133	0	3	0.5	96	51
21	0	7.5	0	364	30	8	100	35	3	110	16	0.3	0	57	[illegible]
22	0	7.4	0	99	457	128	100	42	2	360	19	0.5	0	70	[illegible]
23	0	8	0	366	257	55	530	96	70	98	15	0.2	0	45	85
24	33	7.8	765	171	50	120.6	0	50	7.9	95	7.4	4	0	35	484.
25	29	7.6	1077	128.6	166	182	0	20	13.4	208	4	5	0	28	75

<u>Theoretical framework</u> and <u>methodology</u>





Exploratory Factor Analysis

Basic concept of factor analysis and its significance in interpretation

$$X_j = \sum_{r=1}^{p} C_{jr} f_r$$

where f_r (r=1,2,3…….p) represents common underlying factors and C_{jr} indicates the factor loadings of variable X_i on factor f_r. Theoretical unknown factors can then be expressed in terms of distinct groups of variables (in the present case fluid geochemical elements), which when correlated with the observed features geothermal geochemistry of the area of investigation, provide significant insight into the causal factors.

The objective of the exploratory factor model study is to reduce a large number of variables into fewer factors, or, in other words, to separate the significant few from the insignificant many variables with a primary aim to differentiate distinctive assemblages of geochemical elements in an attempt to gain insight into the genesis of hot springs spread over these two regions of diverse geologic and tectonic settings.

Factor load, eigenvalue, communalities

Correlation Matrix

The first step in factor analysis is to calculate the correlation matrix. In correlation matrix we can see how strongly the variables correlate with each other. That is, we can overview of the correlations between the traits. Starting from the correlation matrix, the so-called eigenvalue problem is solved, which is used to calculate the factors (A.Roy, 2023).

The important terms or characteristic values for a factor analysis are factor loadings, eigenvalue and communalities. With their help, it is possible to see how strong the correlation between the individual variables and the factors is.

Factor load

- How high is the correlation between a variable and a factor

Eigenvalue

- How much variance explained by a factor of all variables
- Sum of the squared factor loadings

Communalities

- How much Variance of the variables, which is explained by all factors
- Sum of the squared factor loadings of a variable

Determination of number of factors to retain

● In factor analysis choice of number of factors out of as many as variables was a baffling issue (A.Roy, 2023). This issue is resolved by the three terms -1) factor loading indicates how high the correlation between the variables (recommended ≥ 0.6 (Awang, Z, 2014), 2) eigenvalue indicates how much variant can be explained by the factors (recommended >1) and 3) communality indicates how much of the variant of the variables can be explained by the factors (recommended at least 0.5).

- Furthermore, in the table "Explained total variance" the variance can be read, which explains each individual factor and the cumulative variance.
- Communalities - Once the number of factors is determined, the communalities can be calculated. How much of the variance of the variables can be explained by the chosen factors
- Component matrix

The component matrix indicates the factor loads of the factors on the variables. Since the first factor explains most of the variance, the values of the first component or factor are the largest. With this form of representation it is however difficult to make a meaningful interpretation about the factors, therefore this matrix is still rotated.

- Varimax Rotation

With the help of the Varimax rotation it should be analytically ensured that per factor certain variables load as high as possible and the other variables load as low as possible. This is obtained when the variance of the factor loadings per factor should be as high as possible.

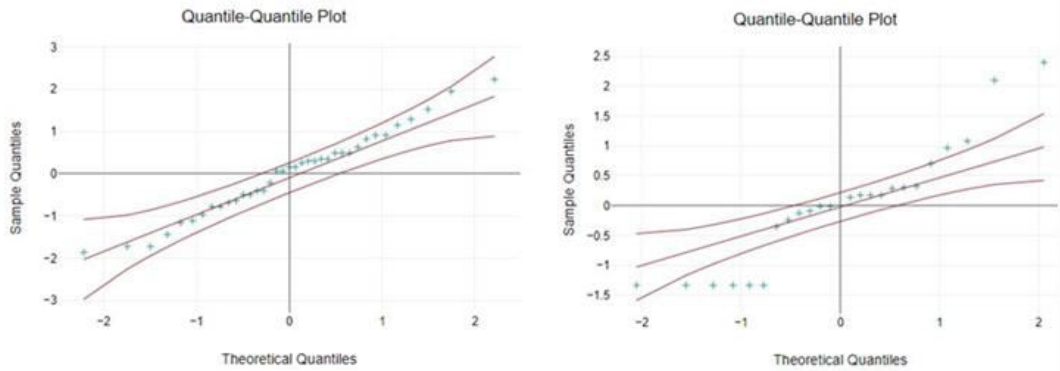
Two sets of geothermal data, one from the Exra-Peninsula and the other from the Peninsula, were subjected to robust statistical analysis, first by exploratory factor analysis followed by multiple regression analysis. The homogeneity and heterogeneity are very clearly established by quantile-quantile analysis. While a normal, uniform sample population distribution was visible in Extra-Peninsular geotherm data, there is some skewness at the two ends of the curve in the case of Peninsular data, which could be due to the interfering influence of mixing diverse geologic and tectonic data.

Presenting the results

Descriptive Basic Statistics (Upper:Extra-Peninsula, Lower:Peninsula)

	TEMPC	pH	SPCMHO/cm	HCO3 mg/L	Ca mg/L	TotHard	SO4 mg/L	Cl mg/L	Mg mg/L	Na mg/L	K mg/L	F mg/L	TDS mg/L	SiO2 mg/L	B mg/L
Mean	48.59	7.54	1,938.84	317.97	59.46	168.41	128.59	132.51	17.73	141.08	18.32	3.36	707.92	51.16	8.07
Std. Deviation	21.26	0.51	3,258.28	309.16	96.54	296.42	315.61	283.27	22.52	155.34	23.27	3.77	761.64	36.04	23.16
Minimum	9	6.4	25	0	0	0	0	2	0	1	0	0	0	2	0
Maximum	96	8.3	16,630	1,610	504	1,352	1,484	1,485	99	600	109	12.5	4,072	130	138

	Temp_C	pH	SPCMHO/cm	HCO3 mg/L	Ca mg/L	TotHard	SO4 mg/L	Cl mg/L	Mg mg/L	Na mg/L	K mg/L	F mg/L	TDS mg/L	SiO2 mg/L	Bmg/L
Mean	35.36	7.07	3,234.96	247.02	97.12	366.8	107.06	770.08	30.58	625.64	16.59	1.32	1,129.46	44.02	0.36
Std. Deviation	26.53	2.17	6,209.25	303.73	96.64	933.93	137.52	1,126.61	50.93	1,361.05	29.3	1.93	1,476.2	29.52	0.68
Minimum	0	0	0	11	3	0	0	30	0.1	30	0	0	0	0	0
Maximum	99	9	28,980	1,534	390	4,680	672	4,800	250	6,810	145	7.2	5,744	122	3



Exploratory Factor Analysis

Taking into account all of the above criteria for determining the number of factors, the researcher arrived at a compromise conclusion that in the current study, four factors that account for 75±2% of the total variance of the original variables "load" on a factor of Principal component are retained for rotation to avoid both overextraction and underextraction of factors, which may have negative effects on the results. This also corroborates the criteria for communalities for each of the variables closer to one (0.9) and eigenvalues. ≥ 1..

Correlation Matrix of Extra-Peninsula

Correlation matrix - Extra-Peninsula

	TEMPC	pH	SPCMHO/cm	HCO3 mg/L	Cl mg/L	SO4 mg/L	TotHard	Ca mg/L	Mg mg/L	Na mg/L	K mg/L	F mg/L	B mg/L	SiO2 mg/L	TDS mg/L
TEMPC	1	0.05	-0.06	0.22	0.09	-0.05	-0.1	-0.06	0.03	0.25	0.17	0.26	0.18	0.55	0.14
pH	0.05	1	-0.11	0.05	-0.25	-0.2	-0.3	-0.26	0.04	-0.16	-0	0.04	-0.1	-0.12	-0.25
SPCMHO/cm	-0.06	-0.11	1	0.19	0.17	0.31	0.34	0.3	0.11	0.25	0.19	-0.13	0.13	0.13	0.34
HCO3 mg/L	0.22	0.05	0.19	1	0	0.03	0.05	0.01	0.35	0.57	0.6	0.23	0.09	0.51	0.21
Cl mg/L	0.09	-0.25	0.17	0	1	0.22	0.31	0.23	0.21	0.7	0.51	0.01	0.54	0.24	0.6
SO4 mg/L	-0.05	-0.2	0.31	0.03	0.22	1	0.96	0.97	0.34	0.37	0.35	-0	0.55	-0.09	0.83
TotHard	-0.1	-0.3	0.34	0.05	0.31	0.96	1	0.97	0.37	0.41	0.37	-0.1	0.56	-0.08	0.85
Ca mg/L	-0.06	-0.26	0.3	0.01	0.23	0.97	0.97	1	0.34	0.32	0.28	-0.12	0.46	-0.07	0.81
Mg mg/L	0.03	0.04	0.11	0.35	0.21	0.34	0.37	0.34	1	0.3	0.6	-0.26	0.48	-0.02	0.49
Na mg/L	0.25	-0.16	0.25	0.57	0.7	0.37	0.41	0.32	0.3	1	0.71	0.38	0.58	0.53	0.75
K mg/L	0.17	-0	0.19	0.6	0.51	0.35	0.37	0.28	0.6	0.71	1	0.03	0.7	0.25	0.62
F mg/L	0.26	0.04	-0.13	0.23	0.01	-0	-0.1	-0.12	-0.26	0.38	0.03	1	0.03	0.44	0.11
B mg/L	0.18	-0.1	0.13	0.09	0.54	0.55	0.56	0.46	0.48	0.58	0.7	0.03	1	-0	0.76
SiO2 mg/L	0.55	-0.12	0.13	0.51	0.24	-0.09	-0.08	-0.07	-0.02	0.53	0.25	0.44	-0	1	0.21
TDS mg/L	0.14	-0.25	0.34	0.21	0.6	0.83	0.85	0.81	0.49	0.75	0.62	0.11	0.76	0.21	1

Correlation Matrix of Peninsula

Correlation matrix - Peninsula

	Temp_C	pH	SPCMHO/cm	HCO3 mg/L	Cl mg/L	SO4 mg/L	TotHard	Ca mg/L	Mg mg/L	Na mg/L	K mg/L	F mg/L	Bmg/L	SiO2 mg/L	TDS mg/L
Temp_C	1	-0.63	0.15	0.26	0.39	0.48	0.14	0.11	-0.13	0.15	0.45	0.33	0.15	0.17	0.53
pH	-0.63	1	0.1	-0.64	-0.14	-0.52	0.12	0.31	0.08	-0.07	-0.59	0.08	0.02	0	-0.36
SPCMHO/cm	0.15	0.1	1	-0.19	0.53	-0.13	0.33	0.78	0.8	0.93	0.22	-0.13	0.08	-0.1	0.06
HCO3 mg/L	0.26	-0.64	-0.19	1	0.1	0.59	-0.21	-0.28	-0.02	0.03	0.8	-0.34	-0.07	-0.44	0.46
Cl mg/L	0.39	-0.14	0.53	0.1	1	0.36	0.73	0.39	0.12	0.36	0.43	-0.18	0.24	0.02	0.59
SO4 mg/L	0.48	-0.52	-0.13	0.59	0.36	1	0.12	-0.12	-0.19	-0.04	0.73	0.13	-0.26	-0.13	0.65
TotHard	0.14	0.12	0.33	-0.21	0.73	0.12	1	0.32	-0.07	0	-0.07	-0.06	-0.01	0.31	0.13
Ca mg/L	0.11	0.31	0.78	-0.28	0.39	-0.12	0.32	1	0.62	0.65	0.08	0.15	-0.09	-0.01	0.14
Mg mg/L	-0.13	0.08	0.8	-0.02	0.12	-0.19	-0.07	0.62	1	0.87	0.22	-0.3	-0.07	-0.32	-0.09
Na mg/L	0.15	-0.07	0.93	0.03	0.36	-0.04	0	0.65	0.87	1	0.42	-0.19	0.06	-0.24	0.1
K mg/L	0.45	-0.59	0.22	0.8	0.43	0.73	-0.07	0.08	0.22	0.42	1	-0.25	-0.08	-0.33	0.65
F mg/L	0.33	0.08	-0.13	-0.34	-0.18	0.13	-0.06	0.15	-0.3	-0.19	-0.25	1	-0.13	0.35	0.06
Bmg/L	0.15	0.02	0.08	-0.07	0.24	-0.26	-0.01	-0.09	-0.07	0.06	-0.08	-0.13	1	-0.07	0.36
SiO2 mg/L	0.17	0	-0.1	-0.44	0.02	-0.13	0.31	-0.01	-0.32	-0.24	-0.33	0.35	-0.07	1	-0.2
TDS mg/L	0.53	-0.36	0.06	0.46	0.59	0.65	0.13	0.14	-0.09	0.1	0.65	0.06	0.36	-0.2	1

Extra-Peninsula

Explained Total Variance

Component	Total	% of variance	Accumulated %
1	5.74	38.26	38.26
2	2.74	18.24	56.5
3	1.49	9.94	66.45
4	1.18	7.84	74.29
5	1.07	7.15	81.44
6	0.83	5.55	86.99
7	0.78	5.17	92.16
8	0.43	2.9	95.06
9	0.34	2.28	97.34
10	0.19	1.27	98.61
11	0.13	0.84	99.45
12	0.04	0.25	99.7
13	0.02	0.15	99.85
14	0.02	0.13	99.98
15	0	0.02	100

Communality

	Extraction
TEMPC	0.39
pH	0.38
SPCMHO/cm	0.31
HCO3 mg/L	0.83
Cl mg/L	0.82
SO4 mg/L	0.94
TotHard	0.97
Ca mg/L	0.94
Mg mg/L	0.74
Na mg/L	0.88
K mg/L	0.88
F mg/L	0.56
B mg/L	0.77
SiO2 mg/L	0.79
TDS mg/L	0.97

Peninsula

Explained Total Variance

Component	Total	% of variance	Accumulated %
1	4.29	28.6	28.6
2	3.61	24.09	52.7
3	2.29	15.27	67.97
4	1.44	9.63	77.6
5	1.21	8.08	85.68
6	0.93	6.2	91.88
7	0.39	2.61	94.49
8	0.27	1.82	96.3
9	0.17	1.16	97.47
10	0.17	1.14	98.61
11	0.12	0.78	99.38
12	0.05	0.36	99.75
13	0.02	0.14	99.89
14	0.02	0.1	99.99
15	0	0.01	100

Communality

	Extraction
Temp_C	0.7
pH	0.64
SPCMHO/cm	0.96
HCO3 mg/L	0.87
Cl mg/L	0.92
SO4 mg/L	0.78
TotHard	0.67
Ca mg/L	0.81
Mg mg/L	0.92
Na mg/L	0.92
K mg/L	0.92
F mg/L	0.7
Bmg/L	0.49
SiO2 mg/L	0.6
TDS mg/L	0.74

Unrotated Component Matrix

Extra-Peninsula

	Component			
	1	2	3	4
TEMPC	-0.14	0.58	0.16	0.02
pH	0.26	0.14	-0.49	0.23
SPCMHO/cm	-0.38	-0.09	0.08	0.38
HCO3 mg/L	-0.35	0.62*	-0.28	0.49
Cl mg/L	-0.6*	0.19	0.05	-0.65*
SO4 mg/L	-0.8*	-0.45	0.19	0.23
TotHard	-0.83*	-0.46	0.18	0.16
Ca mg/L	-0.77*	-0.49	0.22	0.23
Mg mg/L	-0.56	-0.04	-0.65*	0.08
Na mg/L	-0.78*	0.51	0.07	-0.12
K mg/L	-0.74*	0.34	-0.45	-0.07
F mg/L	-0.06	0.56	0.48	0.08
B mg/L	-0.78*	-0.01	-0.19	-0.35
SiO2 mg/L	-0.24	0.78*	0.32	0.17
TDS mg/L	-0.97*	-0.06	0.11	-0.05

Peninsula

	Component			
	1	2	3	4
Temp_C	0.59	-0.29	0.45	-0.26
pH	-0.54	0.55	0.07	0.2
SPCMHO/cm	0.52	0.83*	0.04	-0.07
HCO3 mg/L	0.6*	-0.56	-0.44	0.04
Cl mg/L	0.69*	0.21	0.49	0.38
SO4 mg/L	0.64*	-0.55	0.14	-0.21
TotHard	0.21	0.26	0.68*	0.31
Ca mg/L	0.35	0.77*	0.21	-0.24
Mg mg/L	0.38	0.75*	-0.43	-0.17
Na mg/L	0.6*	0.69*	-0.24	-0.17
K mg/L	0.89*	-0.27	-0.24	-0.09
F mg/L	-0.2	-0.11	0.53	-0.6*
Bmg/L	0.09	0.04	0.13	0.68*
SiO2 mg/L	-0.31	-0.01	0.67*	-0.22
TDS mg/L	0.73*	-0.32	0.25	0.22

Rotated (VARIMAX) Matrix

Extra-Peninsula

	Component			
	1	2	3	4
TEMPC	0.15	0.57	-0.07	-0.18
pH	0.33	-0.11	-0.42	0.29
SPCMHO/cm	-0.46	0.2	-0.16	0.17
HCO3 mg/L	0.01	0.61*	-0.67*	0.12
Cl mg/L	-0.19	0.11	-0.05	-0.87*
SO4 mg/L	-0.96*	0.03	-0.09	-0.08
TotHard	-0.97*	-0	-0.09	-0.15
Ca mg/L	-0.97*	0	-0.04	-0.05
Mg mg/L	-0.29	-0.16	-0.77*	-0.19
Na mg/L	-0.3	0.59	-0.33	-0.58
K mg/L	-0.24	0.22	-0.73*	-0.48
F mg/L	0.1	0.7*	0.22	-0.08
B mg/L	-0.45	-0.01	-0.36	-0.66*
SiO2 mg/L	0.11	0.87*	-0.07	-0.14
TDS mg/L	-0.78*	0.24	-0.24	-0.48

Peninsula

	Component			
	1	2	3	4
Temp_C	0.61	0.09	0.36	-0.43
pH	-0.78*	0.15	0.01	0.08
SPCMHO/cm	-0.02	0.97*	0.12	0.03
HCO3 mg/L	0.84*	-0.16	-0.14	0.34
Cl mg/L	0.3	0.45	0.79*	0.07
SO4 mg/L	0.83*	-0.11	0.16	-0.21
TotHard	-0.14	0.25	0.75*	-0.15
Ca mg/L	-0.12	0.86*	0.11	-0.24
Mg mg/L	0	0.87*	-0.34	0.22
Na mg/L	0.18	0.93*	-0.13	0.12
K mg/L	0.91*	0.25	0.01	0.16
F mg/L	-0.06	-0.12	0	-0.83*
Bmg/L	-0.1	-0.04	0.52	0.46
SiO2 mg/L	-0.3	-0.16	0.3	-0.63*
TDS mg/L	0.68*	0.06	0.52	0.08

Multiple Regression Analysis

The objective of regression analysis is to explain variability in a dependent variable by means of one or more independent or control variables. A multiple regression model is used when there is more than one independent variable affecting a dependent variable. While predicting the outcome variable, it is important to measure how each of the independent variables moves in their environment and how their changes will affect the output or target variable.

Regression line for a multivariate data

$Y = a + b_1 \times X_1 + b_2 \times X_2 + \ldots + b_n \times X_n$,

Where

Y = dependent variable

X_i = independent variables

a = constant (y-intersect)

b_i= regression coefficient of the variable X_i

Assumptions

1) Is the sample size sufficient or the chosen samples are representative of the population
2) Do the DV and IV show variation
3) Is the DV interval or ratio scaled
4) Is linearity or linear relationship between IV and DV exist
5) Multivariate normality i.e. approximately normally distributed (with a mean of zero)
6) No or little multicolinearity (occurs when the IVs are too highly correlated with each other)
6) No auto-correlation
7) Homoskedasticity vs Heteroskedasticity: The scatter plot is a good way to check whether the data are homoscedasticity meaning the residuals are equal around the regression line.

Here in multiple regression analysis, too, the choice of one relevant variable (IV) out of many is an issue. One should never enter all the available variables at the same time. Carefully consider which independent variable is distinct or whether it is relevant to the problem. The first and most reliable option is to use factor analysis, which creates a small number of factors that account for most of the original variables' information in them but are mutually uncorrelated.

Extra-Peninsular India

Dependant Variable = Cl Independent variables = HCO3, SO4, Na F

R	R^2	Adjusted R^2	Standard error of the estimate
0.92	0.84	0.82	120.07

ANOVA

Model	df	F	p
Regression	4	42.1	<.001

Coefficients

	Unstandardized Coefficients	Standardized Coefficients				95% confidence interval for B	
Model	B	Beta	Standard error	t	p	lower bound	uppe bour
(Constant)	103.2		32.45	3.18	.003	37.08	169.
HCO3 mg/L	-0.6	-0.66	0.08	-7.38	<.001	-0.77	-0.4

SO4 mg/L	-0.22	-0.25	0.07	-3.09	.004	-0.37	-0.08
Na mg/L	2.36	1.29	0.19	12.61	<.001	1.97	2.74
F mg/L	-24.81	-0.33	5.83	-4.25	<.001	-36.7	-12.92

B= This value represents the slope of the line between the predictor variable and the dependent variable; SE B= standard error for the unstandardized beta, similar to the standard deviation for a mean. The larger the number, the more spread out the; (β)= the standardized beta similar to a correlation coefficient, ranging between 0 and ±1; t= the test statistic calculated for the individual predictor variable and used to calculate the p value; p = the probability level to tell whether or not an individual variable significantly predicts the dependent variable points.

Peninsular India

Dependant variable = HCO3 Independent variable - Na , K

R	R^2	Adjusted R^2	Standard error of the estimate
0.87	0.76	0.74	156.16

ANOVA

Model	df	F	p
Regression	2	34.39	<.001

Coefficients

	Unstandardized Coefficients	Standardized Coefficients				95% confidence interval for B	
Model	B	Beta	Standard error	t	p	lower bound	uppe boun
(Constant)	134.62		36.9	3.65	.001	58.1	211.[illegible]
Na mg/L	-0.08	-0.38	0.03	3.25	.004	-0.14	-0.0[illegible]
K mg/L	9.94	0.96	1.2				

Interpretation of the results

Exploratory Factor Analysis

EXTRA-PENINSULA		PENINSULA	
Unrotated PCA	Rotated VARIMAX	Unrotated PCA	Rotated VARIMAX
F1 Cl-,SO4-, Ca-, Na-, K-,B-TDS-	F1 SO4, Ca, TDS	**F1** HCO3, Cl, SO4, Na, K, TDS	**F1** pH- , HCO3, SO4. K, TDS
F2 HCO3, SiO2	**F2** HCO3, F, SiO2	**F2** SPCMHO*, Ca, Mg, Na	**F2** SPCMHO*,Ca, Mg, Na
F3 Mg-	**F3** HCO3-, Mg-, K-	**F3** SiO2	**F3** Cl
F4 Cl -	**F4** Cl-, F-	**F4** B, F-	**F4** F-, SiO2-

SPCMHO/Cm* - is correctly defined as the electrical conductance of 1 cubic centimeter of a solution at 25 °C used to estimate the salinity , ionic strength and concentrations of major TDS solutes itrix n natural waters.

Multiple Regression Analysis

Extra-Peninsular India

A multiple linear regression analysis was performed to examine the influence of the variables HCO3 mg/L, SO4 mg/L, Na mg/L and F mg/L on the variable Cl mg/L.
The regression model showed that the variables HCO3 mg/L, SO4 mg/L, Na mg/L and F mg/L explained 84.03% of the variance from the variable CL mg/L. An ANOVA was used to test whether this value was significantly different from zero. Using the present example, it was found that the effect was significantly different from zero (meaning statistically significant), F= 42.1, p= <.001, R^2=0.84.
Regression Coefficients

The following regression model is obtained:
Cl mg/L = 103.2-0.6 - HCO3 mg/L -0.22 - SO4 mg/L +2.36 - Na mg/L -24.81 - F mg/L
When all independent variables are zero, the value of the variable Cl mg/L is 103.2
The standardized coefficients beta are independent of the measured variable and are between -1 and +1, with the larger the amount of beta, the greater the contribution of the respective independent variable to explain the dependent variable Cl mg/L. In this model, the variable Na mg/L has the greatest influence on the variable Cl mg/L. The calculated regression coefficients refer to the sample used for the calculation of the regression analysis, and the null hypothesis is made for each coefficient that it is equal to zero in the population. The test statistic t is then calculated from the standard error and the coefficient. The P-value is used to test whether individual coefficients deviate from zero in the population. It is calculated from the standard error and the coefficient. The coefficients of HCO3 mg/L, SO4 mg/L, Na mg/L, and F mg/L all have p values of.001.Thus, the p-value is smaller than the significance level of 0.05, and the null hypothesis of zero is rejected, so it is assured that the coefficient for the variable age in the population is different from zero.

Peninsular India

A multiple linear regression analysis was performed to examine the influence of the variables Na mg/L and K mg/L on the variable HCO3 mg/L.
The regression model showed that the variables Na mg/L and K mg/ explained 75.77% of the variance from the variable HCO3 mg/L. An ANOVA was used to test whether this value was significantly different from zero. Using the present example, it was found that the effect was significantly different from zero (meaning statistically significant), F= 34.39, p= <.001, R^2=0.76.
Regression Coefficients
The following regression model is obtained:
HCO3 mg/L = 134.62 -0.08 - Na mg/L +9.94 - K mg/L
When all independent variables are zero, the value of the variable HCO3 mg/L is 134.62
The standardized coefficients beta are independent of the measured variable and are between -1 and +1, with the larger the amount of beta, the greater the contribution of the respective independent variable to explain the dependent variable HCO3 mg/L In this model, the variable K mg/L has the greatest influence on the variable HCO3 mg/L. The calculated regression coefficients refer to the sample used for the calculation of the regression analysis, and the null hypothesis is made for each coefficient that it is equal to zero in the population. The test statistic t is then calculated from the standard error and the coefficient. The P-value is used to test whether individual coefficients deviate from zero in the population. It is calculated from the standard error and the coefficient. The p value for the coefficient of Na mg/L is.004, which is smaller than the significance level of 0.05, and the null hypothesis that it is zero is rejected. For the coefficient of K mg/L, the p-value is .001, which is also smaller than the significance level of 0.05, and the null hypothesis that it is zero in the population is rejected. This indicates that the coefficient for the variable age in the population is different from zero.

Conclusion

Some researchers from their experience while doing factor analysis raised pertinent questions about the efficacy of rotated component matrix (VARIMAX) over unrotated matrix (A.Fog, 2020).

From the above Factor analysis Table, following points of Fluid geochemistry emerge:

1.The thermal springs of Extra-Peninsular region are deep-seated hot acidic in contrast to shallower relatively cold waters with high pH alkalinity of Peninsular springs.

2.The overall salt assemblage and concentration of F, Cl, SO4, Na, K, Mg, and Ca suggest the existence of hydrothermal system operating in geotherms of Extra-Peninsula. In contrast, Peninsular springs are K-Na-bicarbonate rich waters with lowSO_4^- content and relatively higher contents of HCO_3 compared to other anions SO_4, Cl, and F

3.Extra Peninsular springs are magmatic-hydrothermal manifestations, a phenomenon of magma progressively degassing in their decreasing order of solubility $CO_2 < SO_2 < HCl < HF$ i.e."CO_2-first till HF-last" (Giggenbach 1987).

4.In case of the non-magmatic thermal springs of Peninsular India, the water is heated by convective circulation: groundwater percolating downward through fracture, faults reaching great depths of a kilometre or more where the temperature of rocks is high because of the normal temperature gradient of the Earth's crust—about 30 °C (54 °F) per kilometre in the first 10 km.

5. The geochemical characteristic as established by the exploratory factor analysis as tabulated above distinguishes non-magmatic thermal sources as K-Na-HCO_3 of Peninsular springs as against the the magmatic thermal sources as Cl-HCO_3-SO_4-Na type of Extra-Peninsular springs.

.6. Unlike Extra-Peninsular Himalayan region which is an extension of active tectonic Alpine-Himalayan main thrust zone with a homogeneous lithology, there is a heterogeneity both in lithological as well as multi-directional tectonic settings. Naturally, these distinctive geologic-tectonic environs have definite bearing on their fluid geochemistry.

Both these exploratory factor and multiple regression analyses corroborate each other in deciphering the origin of these two suites of fluid geochemistry. The model studies distinguish two statistically significant suites of fluid geochemistry: 1. the overall salt assemblage and concentration of Cl-HCO3-SO4-Na-F or chloride rich water suggestive of the existence of a hydrothermal magmatic system operating in the geotherms of extra-Peninsular India; and 2. Peninsular springs of K-Na-HCO3 bicarbonate rich waters with low SO4-content and relatively higher contents of HCO_3 compared to other anions SO_4, Cl, and F suggestive of a non-magmatic origin.

References

[1]. Ravi Shankar et al., 1991. Geothermal Atlas of India, GSI Spec Publ,
[2]. A.Roy, 1994. GTHERMIS – An information management and analysis system in geothermal data of India. A field season 1993-94 program as an R&D Item No.7/WB-5
[3]. Amitabha Roy, 2023. Comparative Statistical Study of Geochemistry of Geothermal fields of Peninsular and Extra Peninsular India. J. Appl. Geol. & Geophys (ISOR-JAGG), V. 11, Issue I, Ser. II, pp. 32-44
[4]. Awang , Z. 2014. Research Methodology and data analysis…..
[5]. Miroslaw Grzesik, 2022. Table containing the values measured and calculated from the regression equation as well as the statistical tests on the regression equation and its coefficients, Institute of Chemical Engineering, Polish Academi of Sciences
[6]. WF Giggenbach et al, 1987. The effects of hydrothermal processes on the chemistry of some recent volcanic gas.
[7]. F.Tassi et al., 2010. Fluid geochemistry of hydrothermal systems in the Arica-Parinacota, Tarapaca and Antofagasta regions (northern Chile), J. Volcanol.Geotherm.Res.
[8]. Mohammad Noor et al.,2021. A geochemical comparison between volcanic and non-volcanic hot springs from East Malaysia: Implications for their origin and geothermometry, Journal of Asian Earth Sciences.
[9]. D Rouwet, 2017. Fluid Geochemistry and Volcanic Unrest: Dissolving the Haze ...Springer.
[10]. H. Baioumy, 2015. Geochemistry and geothermometry of non-volcanic hot …, J. Volcanol. Geotherm. Res.
[11]. A.Fog, 2014. Why is factor rotation always recommended, though it obscures general factors?, J. ross-Cultural Rese