SPATIAL-TEMPORAL MONITORING OF GROUNDWATER USING MULTIVARIATE STATISTICAL TECHNIQUES IN BAREILLY DISTRICT OF UTTAR PRADESH, INDIA

SUDHIR KUMAR SINGH[1)*], CHANDER KUMAR SINGH[2)], KEWAT SANJAY KUMAR[1)], RAMVTAR GUPTA[2)], SAUMITRA MUKHERJEE[2)]

[1) *]Corresponding author, Department of Forestry, Mizoram University, Aizwal-796009, India; mailto: sudhirinjnu@gmail.com
[2)]School of Environmental Sciences, Jawaharlal Nehru University, New Delhi-67, India.

Monitoring of groundwater quality in Bareilly district, Uttar Pradesh, India, was performed at 10 different sites during the years 2005–2006. Obtained quality parameters were treated using principal component analysis (PCA) and cluster analysis (CA). The study shows usefulness of multivariate statistical techniques for evaluation and interpretation of groundwater quality data sets.

KEY WORDS: Cluster Analysis, Principal Component Analysis, Groundwater Quality.

Sudhir Kumar Singh, Chander Kumar Singh, Kewat Sanjay Kumar, Ramvtar Gupta, Saumitra Mukherjee: MONITOROVÁNÍ PODZEMNÍ VODY V ČASE A V PROSTORU POMOCÍ MULTIVARIAČNÍCH STATISTICKÝCH METOD V OBLASTI BAREILLY, UTTAR PRADESH, INDIE. J. Hydrol. Hydromech., *57*, 2009, 1; 36 lit., 2 obr., 4 tab.

V letech 2005–2006 byla na deseti odběrných místech v regionu Bareilly, Uttar Pradesh, Indie sledována kvalita podzemní vody. Zjištěné kvalitativní parametry byly zpracovány pomocí analýzy hlavního prvku (PCA) a pomocí shlukové analýzy (CA). Studie dokládá vhodnost multivariačních statistických metod pro vyhodnocení a interpretaci změřených výsledků.

KLÍČOVÁ SLOVA: shluková analýza, analýza pomocí hlavního prvku, kvalita podzemní vody.

1. Introduction:

The quality of surface and groundwater is very serious concern today. The World Health Organization reports that ~80% of diseases are waterborne. The percentage mostly lies in the developing nations and underdeveloped nations. Hence for safe drinking water a regular assessment and monitoring of water resources is required. Water contamination and its sustainable management need our attention because of far reaching impact on human health. The natural processes (changes in precipitation inputs, soil erosion, weathering of crustal rocks) degrade surface and groundwater and impair their use of water resources for various purposes viz. drinking, industrial, agricultural, recreation or other. The quality of groundwater resources at any point reflects several major influences including lithology of the area, atmospheric inputs, climatic conditions and anthropogenic inputs. Municipal and industrial waste water discharge constitutes a constant polluting source, whereas surface runoff is a

seasonal phenomenon, largely affected by climate within the basin (*Singh* et al., 2004). The water chemistry varies at the spatial and temporal scale. Seasonal variations in the form and rate of precipitation, surface runoff, rate of groundwater flow and water interception and abstraction have strong effect on river discharge and subsequently on the concentration of the pollutants in surface and groundwater (*Vega* et al., 1998). So monitoring and assessment program will provide a representative and reliable estimation of the quantity and quality of groundwater resources. The regular monitoring program, results in huge and complex data matrix comprised of a large number of physico-chemical parameters, which are often difficult to interpret and draw a meaningful conclusions. However, for the effective pollution measure to control the water resource management it is required to identify pollution sources and their quantitative contribution. The application of different statistical approaches (CA and PCA) for the interpretation of these data offers a proper understanding of water quality and

ecological status of the systems allows the identification of the possible factors/sources that influence the water systems and offers a valuable tool for reliable and effective management of water resources as well as rapid solutions on pollution problems (*Vega* et al., 1998; *Lee* et al., 2000; *Adams* et al., 2001; *Wunderlin* et al., 2001; *Reghunath* et al., 2002; *Simeonova* et al., 2003; 2004). There have been many studies confirming the usefulness of those techniques in the analysis of large, complex sets of measurement data and in the planning of measuring networks for efficient control of sources of surface water pollution (*Simeonov* et al., 2003). The literature provides many examples of the application of chemometric techniques in the interpretation of physico-chemical results of river water (*Einax* et al. 1998; *Vega* et al., 1998; *Simeonov* et al., 2000, 2001, 2002; *Marques da Silva* and *Sacomani*, 2001; *Brodnjak-Voncina* et al., 2002; *Mendiguchia* et al., 2004; *Singh* et al., 2004, 2005a,b; *Ouyang*, 2005; *Ouyang* et al., 2006; *Charkhabi* and *Sakizadeh*, 2006; *Kowalkowski* et al., 2006; *Sundaray* et al., 2006; *Tarrado* et al., 2006; *Zeilhofer* et al., 2006; *Kannel* et al., 2007; *Shrestha* and *Kazama*, 2007; *Sotelo* et al., 2007). Multivariate statistical technique(s) has been applied to characterize and evaluate surface and groundwater quality and it is useful in verifying spatial and temporal variations caused by natural and anthropogenic factors linked to seasonality (*Helena* et al., 2000; *Singh* et al., 2004, 2005).

2. Study area

Bareilly district is located in north western segment of Uttar Pradesh and lies between latitude 28° 1' to 28° 54' North and longitude 78° 58' to 78° 47' East. The study area encompasses a geographical area around 4,120 Sq. Km. For administrative convenience the Bareilly district has been divided into six tehsil (sub-division of district). The main economic activity is agriculture but some industries, such as urea plant Indian Farmers Fertilizer Cooperative Limited (IFFCO) in Aonla tehsil, rubber industry and other small scale industries such as paper and flour and rice mills etc. The study area having the groundwater potential, net groundwater availability 139714.95 hectare meter (Ham), minus all uses of groundwater is around 95524.76 hectare meter and divided by the net groundwater availability and multiply by 100, this represents the level of development of groundwater is 68.37%. The rate of decrease in groundwater table is around 0.10

m/year in most part of the study area. Fig. 1 is showing the study area.

2.1 Geographical features

2.1.1 Climate

The climate is semi-arid in nature. The summer temperature varies from 37 °C to 45 °C. Nights are relatively cooler with minimum temperature falling to 27–28 °C. The average annual rainfall is in the range of 5000–7500 millimeters that is precisely contributed by the Southwest jet of monsoon. Winters temperature varies between 4–5 °C.

2.1.2 Physiography

The study area lies in the part of the Ramganga and Ganga basins. The physiography of the study is mainly alluvium. A number of elevated structures found in and around the study area, are known as mounds. In general the area shows the following four distinctive geomorphic units.
1. Lower piedmont plain or tarai (area which is saturated with water)
2. Older alluvial plain or upland
3. Younger alluvial plain or lowland
4. Meander flood plain

2.1.3 Drainage

The study area occupies a part of the Ramganga sub basin of Ganga basin. The rivers in the study area are flowing throughout the year. The overall drainage system is controlled by the river Ramganga and its tributaries like the Siddha, the Dojora, the Bahgaul (west), Kicha, Sankha, the Deoranian, the Nakatia, the Bahagaul (east), Deoha and Aril etc. Except the river Aril all the rivers join the Ramganga River on its left bank. The river Ramganga is the great tributary of the river Ganga, rising in the mountains of Garhwal. The river Aril is an important right bank tributary of the river Ramganga.

2.1.4 Geology

The Gangetic alluvial plain occupies a structural trough in front of Himalayan ranges. It was interpreted to be a fore deep (Sues, 1904–24) or Great Rift Valley filled with alluvium of enormous thickness varying from 4500 meters (*Oldham*, 1917) to as much as 25 km (*Pascoe*, 1968). Aeromagnetic,
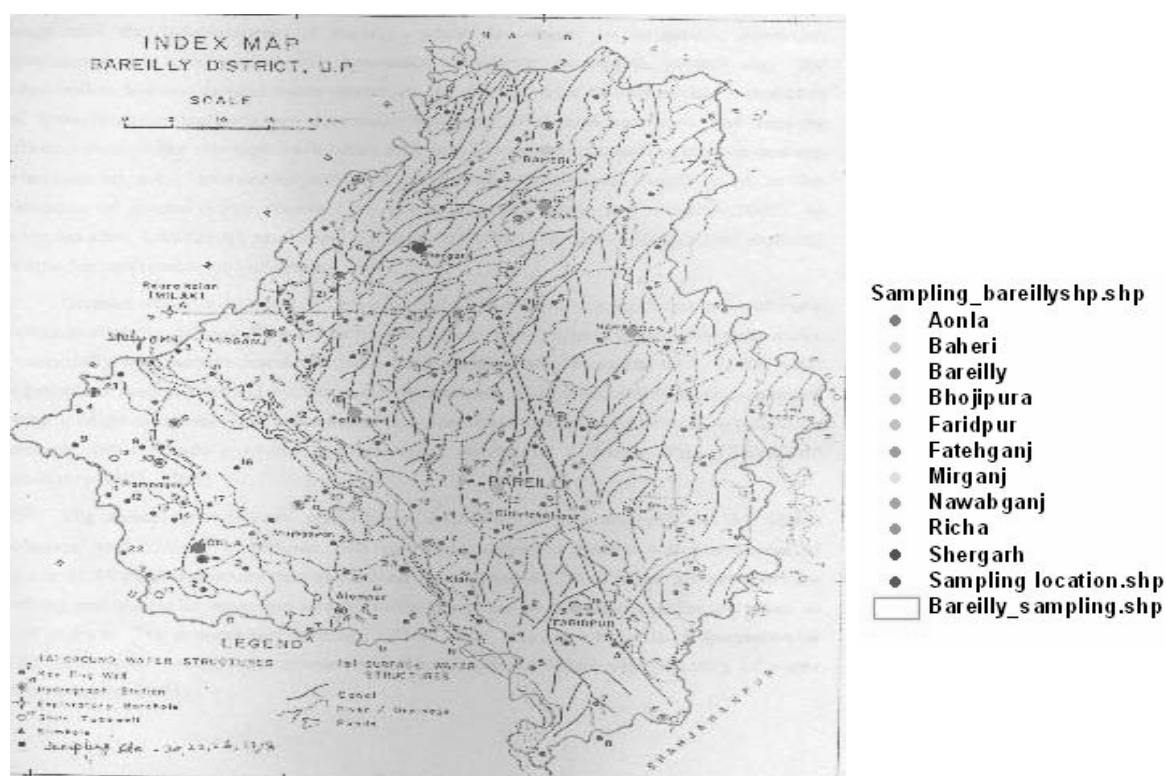
Fig. 1. Showing water quality monitoring sites in the study region.
Obr. 1. Odběrná místa ve sledované oblasti.

T a b l e 1. Stratigraphy of Bareilly.
T a b u l k a 1. Geologická stratigrafie regionu Bareilly.

System	Age	Formation	Lithology
Quaternary	Recent to upper	Lower piedmont	Clays, sands
	pleistocene	Plain (tarai = swampy area)	Sand, clay, and gravel
	Recent to upper pleistocene	Younger alluvium	Fine sand, silt, clay, mixed with gravel
	Upper pleistocene	Older alluvium	Clay wise kankar and sands of different grades

seismic and magnetic surveys conducted by ONGC have indicated that Ganga basin consists of several depressions separated by sub surface ridge. The study area falls in such one depression called Ramganga depression.

The general geological sequence of the formation in the district is as follows in the Tab. 1.

2.1.5 Hydrogeology

This hydrological (Central Ganga Plain) unit is confined between tarai zone in the north and marginal alluvial plain in the south. Several extensive exploration studies have indicated the presence of four aquifer groups within a depth of 700 meter below ground. The first shallow phreatic aquifer occurs within the 50 meter depth. The groundwater is present under confined to semi-confined state in shallow aquifer while in the deeper aquifer, it occurs under the confined state. The second aquifer, which occurs between the depths of 50 to 200 meter, is being extensively exploited to fulfill the irrigational need. The hydro-geological cross sections show(s) that there occurs a single tier (layer) system down to 124 mbgl (meter below ground level), however in some places it is observed occasionally when the single tier aquifer system interleaved with the two clay beds showing three tier aquifer system. These aquifers appear to merge with each other and behave(s) as a single bodied aquifer system. The depth to water level in the area varies between 8 and more than 30 mbgl. The regional groundwater flow is from north-west to north south direction.

3. Material and methods

3.1 Sampling

Total 10 water samples collected from 10 different sampling sites, which contain groundwater, were collected in and around Bareilly district, State Uttar Pradesh, India. The sampling was manually performed and collected samples belong to grab samples. Water samples were collected before onset of monsoon (May-June) season and post monsoon (August-September) successively during year 2005 to 2006. Hand pumps and tub well were used to collect groundwater samples. All water samples were brought to lab and stored at 4 °C in order to avoid any chemical alteration. The sampling was done in 250 ml polypropylene bottle. The methods of sampling and collection are in accordance with Standard Methods for the Examination of Water and Wastewater (APHA – American Public Health Association 1985).

3.2 Laboratory analysis

The collected water samples were divided in two bottles. One bottle was acidified with nitric acid to pH < 2 and stored at 4 °C for the later analysis of the metals Fe, Zn, Mg and Na; where analysis of Fe and Mg were carried out according to the protocols of ultra-trace-metal analysis (*Bruland* et al., 1979) using AAS-Graphite furnace technique, analysis of Na was accomplished using flame photometer technique and Zn was determined using AAS method. Working standard solutions were prepared by dilution of stock solutions (1 mg metal/ml in 2% HNO_3). The other bottle was stored at 4 °C without the addition of preservatives for the analysis of major water parameters Cl^- (0.0141N mercuric nitrate), Ca, Mg and total hardness (0.01 M EDTA). The pH of the samples was determined by using digital pH meter, turbidity by Nephelo-turbidity meter, total hardness by titrimetric method, BOD as per standard method (*Neeri*, 1991), and COD by potassium dichromate open reflex method (*Apha*, 1985).

3.3 Data treatment and multivariate statistical methods

The Kolmogorov-Smirnov (K-S) statistics were used to test the goodness of fit of the data to log normal distribution. According to the K-S test, all the variables are log normally distributed with 95% or higher confidence. Similarly, to examine the suitability of the data for PCA, Kaiser-Meyer-Olkin (KMO) and Bartlett's test were performed. KMO is a measure of sampling adequacy that indicates the proportion of variance which is common variance, i.e., which might be caused by underlying factors. High value (close to 1) generally indicates that PCA analysis may be useful, which is the case in this study: KMO = 0.87. Bartlett's test of sphericity indicates whether correlation matrix is an identity matrix, which would indicate that variables are unrelated. The significance level which is 0 in this study (less than 0.05) indicates that there are significant relationships among variables.

Spearman rank-order correlations (Spearman R coefficient) were used to study the correlation structure between variables to account for non-normal distribution of water quality parameters (*Wunderlin* et al., 2001). In this study temporal variations of groundwater quality parameters were first evaluated through a season parameters correlation matrix, using Spearman non parametric correlation coefficients (Spearman's). The groundwater quality parameters were grouped into two parts: pre monsoon and post monsoon, and each assigned a numerical value in the data file (Pre monsoon = 1 and post monsoon = 2), where 1, as variables corresponding to the season, was correlated (pair by pair) with all the measured parameters. River water quality data sets were subjected to three multivariate techniques: CA, PCA and FA (*Wunderlin* et al., 2001; *Simeonov* et al., 2003; *Singh* et al., 2004, 2005). CA and PCA were applied to experimental data, standardized through z-scale transformation to avoid misclassification arising from the different orders of magnitude of both numerical values and variance of the parameters analyzed (*Liu* et al., 2003; *Simeonov* et al., 2003). All the mathematical and statistical computations were made using Microsoft Office Excel 2007 and SPSS 12.

3.3.1 Cluster analysis

CA groups the objects into the classes or clusters on the basis of similarities within a class and dissimilarities between different classes. The results of CA help in interpreting the data and indicate patterns. In hierarchical clustering, clusters are formed sequentially by starting with the most similar pair of objects and forming higher clusters step by step. Hierarchical agglomerative CA was performed on the normalized data set (mean observations over the whole period) by means of Ward's method using

T a b l e 2. Summary basic statistics (concentration units [mg l⁻¹] for metal ions; conductivity [μs cm⁻¹]).
T a b u l k a 2. Souhrn základních parametrů (koncentrace iontů kovů [mg l⁻¹], vodivost [μs cm⁻¹]).

Parameters	Site No. 1				Site No. 2				Site No. 3				Site No. 4			
	Min.	Max.	Mean	S.E	Min.	Max.	Mean	S.E	Min.	Max.	Mean	S.E	Min.	Max.	Mean	S.E
pH	7.38	7.90	7.62	0.13	7.17	8.00	7.59	0.2	7.64	7.9	7.81	0.06	7.55	7.80	7.66	0.06
Turbidity	2.47	2.6	2.51	0.02	3.0	3.72	3.36	0.19	5.11	5.74	5.36	0.013	2.80	3.12	2.94	0.07
Conductivity	0.18	0.82	0.65	0.15	0.99	1.50	1.20	0.12	1.52	1.68	1.60	0.03	0.51	0.71	0.56	0.04
TS	600	769	685.25	44.53	810	990	880.25	41.9	1100	1290	1227.5	43.27	390.00	484.00	433.5	19.3
TDS	398	582	488.75	51.83	660	690	673.75	6.56	850	988	936.5	30.46	297.00	404.00	349.25	28.3
DO	3.99	5.02	4.50	0.23	3.18	3.62	3.39	0.09	3.00	3.49	3.19	0.11	4.32	5.02	4.63	0.16
BOD	6.97	11.00	8.61	0.09	6.59	7.80	7.19	0.32	6.37	7.81	6.81	0.33	8.51	11.50	9.47	0.68
COD	24.00	24.7	24.17	0.17	12.00	21.00	15.5	2.17	29.00	58.00	38.50	6.66	15.00	20.00	17.00	1.08
TH	230	275	260.75	10.35	260	480.00	412.00	51.2	500	590	561.75	20.77	114.00	182.00	154.00	14.9
Ca	55.3	69.9	64.12	3.16	86.6	109.07	101.29	5.12	120.24	159.2	145.39	8.26	34.47	37.13	35.31	0.62
Mg	22.4	29.1	26.32	1.39	35.1	44.36	40.04	2.06	48.74	57.02	53.22	1.71	14.14	63.00	37.58	13.3
Cl	45.9	55.7	49.86	2.08	68.7	79.52	75.63	2.39	159.04	172.1	163.5	2.91	19.88	29.08	23.92	2.00
Zn	0.58	0.91	0.70	0.07	0.50	0.94	0.66	0.08	1.85	2.20	1.96	0.08	1.10	1.67	1.32	0.12
Fe	0.23	0.29	0.25	0.01	0.11	0.28	0.17	0.30	0.70	0.92	0.82	0.40	0.57	0.84	0.67	0.60
F	0.31	0.42	0.37	0.02	0.78	1.02	0.92	0.05	1.24	1.57	1.04	0.07	0.52	0.61	0.55	0.02

Table 2 continued

Parameters	Site No.5				Site No. 6				Site No. 7				Site No. 8			
	Min.	Max.	Mean	S.E	Min.	Max.	Mean	S.E	Min.	Max.	Mean	S.E	Min.	Max.	Mean	S.E
pH	7.60	7.80	7.69	0.05	7.50	7.65	7.58	0.03	7.52	8.30	7.73	0.18	7.12	8.2	7.64	0.29
Turbidity	2.94	3.25	3.09	0.08	2.23	3.59	3.07	0.29	3.12	3.71	3.34	0.12	4.87	5.21	5.05	0.08
Conductivity	0.57	0.59	0.58	0.01	0.49	0.62	0.56	0.02	0.66	0.72	0.67	0.01	1.45	1.64	1.55	0.04
TS	418	428.00	421.00	2.38	360.00	392.00	375.5	6.84	491.00	564.00	514	17.25	1181	1270	1235	19.00
TDS	290	393.00	315.75	25.75	248.00	310.00	277.75	13.6	381.00	483.00	430.75	28.74	991	1014	1002.8	5.12
DO	4.50	4.20	4.08	0.12	4.18	5.50	5.68	0.29	4.48	4.69	4.54	0.04	1.98	3.21	2.30	0.30
BOD	9.24	10.80	10.06	0.42	8.67	11.53	9.60	0.65	7.78	7.84	7.81	0.01	3.02	5.21	4.11	0.58
COD	23.00	25.00	24.00	0.57	14.00	27.00	17.75	3.09	30.00	43.00	33.75	3.09	50.00	68.00	56.50	3.94
T H	158.00	184.00	169.00	6.55	128.00	157.00	146.25	6.35	250.00	284	260.25	8.08	532	602	566.5	14.5
Ca	37.70	45.10	41.11	1.99	30.50	42.53	36.98	2.51	60.92	78.72	66.07	4.26	127	174	155	10.00
Mg	16.10	52.00	24.27	10.23	12.20	13.74	13.15	0.35	23.88	43.01	28.74	4.75	41.21	63.21	49.82	5.00
Cl	14.90	33.2	27.44	4.23	15.90	28.18	20.24	2.72	77.57	87.02	80.04	2.32	52	153	120.25	23.1
Zn	0.37	0.87	0.70	0.11	0.20	0.37	0.25	0.03	0.70	0.89	0.77	0.40	0.87	0.99	0.93	0.02
Fe	0.18	0.40	0.27	0.04	0.11	0.19	0.13	0.01	0.09	0.14	0.11	0.01	0.64	0.79	0.67	0.04
F	0.43	0.67	0.54	0.05	0.79	0.98	0.86	0.04	0.84	1.49	1.06	0.15	1.02	1.8	1.27	0.18

Table 2 continued

Parameters	Site No. 9				Site No. 10			
	Min.	Max.	Mean	S.E	Min.	Max.	Mean	S.E
pH	7.30	8.12	7.72	0.22	7.2	7.78	7.75	0.12
Turbidity	4.17	4.75	4.45	0.15	3.27	3.77	3.55	0.11
Conductivity	1.39	1.64	1.5	0.05	0.61	0.63	0.62	0.01
TS	1118	1178	1147	12.5	700	743	721	9.23
TDS	914	979	937.5	14.49	592	598	594	1.41
DO	2.22	2.71	2.38	0.1	2.25	4.00	2.89	0.38
BOD	4.00	5.02	4.31	0.24	5.50	8.00	6.25	0.58
COD	35.00	49.00	41.75	3.40	28.00	39.00	32.75	2.32
TH	522	584	544	13.68	244	311	282.75	14.6
Ca	12.2	139	91.45	28.71	58.5	94.18	72.75	7.84
Mg	49.1	68.00	59.13	4.49	23.4	41.76	33.17	4.28
Cl	44.00	173	129.76	29.20	8.76	71.5	39.96	17.1
Zn	0.5	0.69	0.61	0.04	0.2	0.32	0.37	0.01
Fe	0.14	0.2	0.17	0.01	0.25	0.6	0.45	0.07
F	1.11	1.9	1.51	0.19	0.33	0.72	0.58	0.08

Units of parameters in SI:
NTU – turbidity, conductivity [µs cm^{-1}], TS – total solid [mg l^{-1}], TDS – total dissolved solids [mg l^{-1}], DO – dissolved oxygen [mg l^{-1}], BOD – biological oxygen demand [mg l^{-1}], COD – chemical oxygen demand [mg l^{-1}], TH – total hardness [mg l^{-1}], Ca – calcium [mg l^{-1}], Mg – magnesium [mg l^{-1}], Cl – chloride [mg l^{-1}], Zn – zinc [mg l^{-1}], Fe – iron [mg l^{-1}], F – fluoride [mg l^{-1}].

squared Euclidean distances as a measure of similarity. Cluster significance was determined using the criterion of 0.66 D_{max}.

Cluster analysis was applied to the river water quality data set with a view to group the similar sampling sites (spatial variability) spread over the river stretch and in the resulted agglomerative hierarchical cluster (dendrogram), the linkage distance is reported as D_{link}/D_{max}, which represents the quotient between the linkage distance for a particular case divided by maximal distance, multiplied by 100 as way to standardized the linkage distance represented on x-axis.

3.3.2 Principal component analysis

PCA is designed to transform the original variable into the new, uncorrelated variables called the principal components, which are linear combinations of the original variables. The new axes lie along the direction of the maximum variance. PCA provides an objective way of finding indices of this type so that the variation in the data can be accounted for as concisely as possible. PC provides information on the most meaningful parameters, which describes a whole data set affording data reduction with minimum loss of original information. Principal component is a linear combination of observable water quality variables. PCA of the normalized variables was performed to extract significant PCs.

4. Result and discussion

4.1. Site similarity

Cluster analysis was applied to detect the similarity among different sampling sites. The data was treated after data scaling by Z-transformation with the Ward's method of linkage with squared Euclidean distance as measure of similarity. The significance of the clusters obtained was tested by Sneath's Index disjunction. The clustering reveals the groups of similar site in a quite convincing way. These clusters include sampling sites with similar characteristic features and natural background that are affected by sources of similar type or strength.

Since this study had used hierarchical agglomerative cluster analysis, the number of clusters was also decided practically from the results as there is required information available on the sampling site in the form of report.

The clustering procedure generates either cluster or groups on the basis of similar characteristic features and source type. The results indicate that CA techniques is useful in offering reliable classification of water resources in the study area and will make it possible to design future spatial sampling strategy in an optimal manner. Thus, the number of sampling sites in the monitoring network will be reduced hence lesser cost without losing any significance of the result. There are earlier studies (*Simeonov* et al., 2003; *Singh* et al., 2004, 2005; *Kim* et al., 2005) where similar approaches has

successfully been applied to water quality assessment and monitoring program.

It is illustrated in the form of dendrogram (Fig. 2), grouping all the ten sampling sites into three statistically significant clusters at $(D_{link}/D_{max})*100 < < 60$. Sampling sites 3, 8 and 9 were highly polluted (HP) due the anthropogenic activity (industrial and urban activity). Sampling sites 2, 4, 5, 6, and 7 were medium polluted (MP) because of less anthropogenic activity (agricultural activity). Sampling sites 1 and 10 were less polluted (LP) due to very less anthropogenic interference.

4.2. Data structure determination and source identification

Principal component analysis is an effective pattern recognition technique that attempts to explain the variance of a data set of inter-correlated variables with a smaller set of independent variables (PC). PCA was performed on the normalized data, to compare the compositional patterns between the examined water systems and to identify the factors that influence each other. Three principal components were identified with eigen value > 1 summing almost 90% of the total variance in water dataset, factor loading matrix calculated on the basis of groundwater quality parameters measured in the period from year 2005–2006 in the study area (Tab. 3). An eigen value gives a measure of the significance of the factor, the factors with highest eigen value are most significant, eigen value of 1.0 or greater are considered significant (*Kim* and *Muller*, 1987). PCA was performed on data to compare the compositional patterns between the examined water systems and to identify the factors that influence each other. The first PC, accounting for ~63.51% of total variance was correlated with representing influences from point sources such as municipal and industrial effluents and weathering processes from rocks, soil leaching. The second PC ~17.44%, this nutrient factor represents influences from non point sources such as agriculture runoff and atmospheric deposition. Third PC was weighted on pH and electrical conductivity (EC) and represents the physicochemical source of the variability. The third PC was loaded on ~8.35% probably represents anthropogenic toxic pollution from metal activities or industrial effluents. It shows the total variance explained in terms of eigen values, extraction sums of squared loadings and rotation sums of squared loadings (Tab. 4).
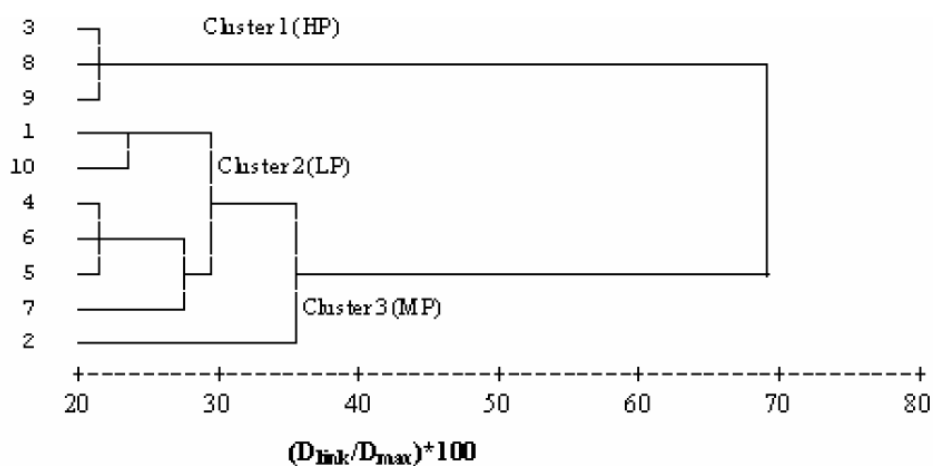
Sampling sites	Fatehganj (site 1)	Baheri (site 2)	Aonla (site 3)	Bhojipur (site 4)	Nawabganj (site 5)	Mirganj (site 6)	Richa (site 7)	Faridpur (site 8)	Bareilly (site 9)	Shergarh (site 10)
Longitude	79.3	79.5	79.16	79.44	79.63	79.23	79.5	79.54	79.46	79.37
Latitude	28.5	28.83	28.28	28.53	28.58	28.57	28.76	28.23	28.38	28.7

Fig. 2. Dendrogram showing sampling site clusters in the study region, Bareilly, Uttar Pradesh using single link, rescaled distance cluster combine.
Obr. 2. Dendrogram shluku odběrů v regionu Bareilly, Uttar Pradesh. Je použita jednoduchá vazba a vzdálenost shluků je zkreslená.

T a b l e 3. Factor loadings matrix for the whole dataset after varimax rotation.
T a b u l k a 3. Matice vstupních faktorů pro celý souhrn dat.

Component matrix (a)

Variable	Component		
	PC 1	PC 2	PC 3
pH	0.470	−0.689	0.295
Turbidity	**0.935**	0.028	0.167
Conductivity	**0.970**	−0.117	−0.055
TS	**0.969**	0.109	−0.037
TDS	**0.978**	0.143	−0.021
DO	−0.358	−0.321	**0.810**
BOD	−0.873	−0.353	0.096
COD	**0.807**	0.166	0.203
T H	**0.985**	0.033	−0.046
Ca	**0.924**	0.094	−0.041
Mg	**0.862**	−0.051	0.257
Cl	**0.961**	−0.148	0.002
Zn	−0.039	**0.932**	0.259
Fe	−0.096	**0.956**	0.225
F	**0.869**	−0.225	−0.097

Extraction method: Principal component analysis
    a) 3 components extracted;
    b) significant factor loadings are bold faced.

T a b l e 4. It shows the total variance explained in terms of eigen values, extraction sums of squared loadings and rotation sums of squared loadings.
T a b u l k a 4. Souhrnné rozdíly ukázané pomocí charakteristických hodnot, vybraných součtů čtverců zatížení a střídavých součtů čtverců zatížení.

Total variance explained

Component	Initial eigen values			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	[%] of variance	Cumulative [%]	Total	[%] of variance	Cumulative [%]	Total	[%] of variance	Cumulative %
1	9.730	64.866	64.866	9.730	64.866	64.866	9.527	63.511	63.511
2	2.642	17.616	82.482	2.642	17.616	82.482	2.617	17.447	80.958
3	1.024	6.826	89.308	1.024	6.826	89.308	1.253	8.351	89.308
4	0.540	3.597	92.905						
5	0.436	2.909	95.814						
6	0.289	1.925	97.739						
7	0.223	1.489	99.228						
8	0.085	0.569	99.796						
9	0.031	0.204	100.000						
10	0.000	0.000	100.000						
11	0.000	0.000	100.000						
12	0.000	0.000	100.000						
13	0.000	0.000	100.000						
14	0.000	0.000	100.000						
15	0.000	0.000	100.000						

Extraction method: Principal component analysis

5. Conclusion

In this case study, different multivariate statistical techniques were used to evaluate spatial variations in water quality. Hierarchical cluster analysis grouped 10 different sampling sites into three clusters of similar water quality characteristics. Based on obtained information it is possible to design a future, optimal sampling strategy, which could reduce the number of sampling stations and cost of sampling. PCA helped in identifying the factors and sources responsible for water quality variations in 10 different regions. Variables factors obtained from PC by varimax rotation indicate that the pa-

rameters responsible for water quality variations are mainly related to trace metals (leaching from soil and industrial waste sites), dissolved salts (natural), organic pollution and nutrients (anthropogenic). Thus this study rendered the usefulness of multivariate statistical techniques for analysis and interpretation of datasets, and in water quality assessment, identification of pollution source/factors and understanding spatial/temporal variations in groundwater quality for effective groundwater quality management.

REFERENCES

AMERICAN PUBLIC HEALTH ASSOCIATION, 1985: Standard methods for the examination of water and waste water (10th edn). Washington, DC, American Public Health Association.

BRODNJAK VONCINA D., DOBCNIK D., NOVIC M., ZUPAN J., 2002: Chemometrics characterisation of the quality of river water. Analytica Chimica Acta, 462, 87––100.

BRULAND K. W., FRANKS, KNAUER G. A., MARTIN J. H., 1979: Sampling and analytical methods for the determination of copper, cadmium, zinc and nickel at nanogram per liter level in seawater. Analytica Chimica Acta, 105, 233––245.

CHARKHABI A.H., SAKIZADEH M., 2006: Assessment of spatial variation of water quality parameters in the most polluted branch of the Anzali wetland, Northern Iran. Polish J. Environmental Studies, 15, 395–403.

EINAX J. W., TRUCKENBRODT D., KAMPE O., 1998: River pollution data interpreted by means of chemometric methods. Microchemical J., *58*, 315–324.

HELENA B., PARDO R., VEGA M., BARRADO E., FERNANDEZ J.M., FERNA´NDEZ L., 2000: Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. Water Research., *34*, 807–816.

KIM J.H., KIM R.H., LEE J., CHEONG T.J., YUM B.W., CHANG H.W., 2005: Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea. Hydrological Processes, 19, 1261–1276.

KIM J.O., MUELLER C.W., 1987: Introduction to factor analysis: what it is and how to do it. Quantitative Applications in the Social Sciences Series. Sage University Press, Newbury Park.

KOWALKOWSKI T., ZBYTNIEWSKI R., SZPEJNA J., BUSZEWSKI B., 2006: Application of chemometrics in river water classification. Water Research., *40*, 744–752.

LEE J.Y., CHEON J.Y., LEE K.K., LEE S.Y., LEE M.H., 2001: Statistical evaluation of geochemical parameter distribution in a ground water system contaminated with petroleum hydrocarbons. Journal of Environmental Quality, *30*, 1548–1563.

LIU C. W., LIN K. H., KUO Y. M., 2003: Application of factor analysis in the assessment of groundwater quality in black foot disease in Taiwan. The Science of the Total Environment., 313, 77–89.

MARQUES DA SILVA A. M., SACOMANI L. B., 2001: Using chemical and physical parameters to define the quality of Pardo river water (Botucatu – SP-Brazil). Water Research., *35*, 1609–1616.

MENDIGUCHIA C., MORENO C., GALINDO RIANO M. D., GARCIA VARGAS M., 2004: Using chemometric tools to assess anthropogenic effect in river water, a case study: Guadalquivir River (Spain). Analytical Chimica Acta, 515, 143–149.

NEERI, 1991: Manual of Water and Pollution control. (Vol. *1.*)

OLDHAM R. D., 1917: The structure of the Himalayas and of the Gangetic Plain, as elucidated by geodetic observations in India. Mem. Geol. Survey of India, *XLII*, 2, Calcutta.

OUYANG Y., 2005: Evaluation of river water quality monitoring stations by principal component analysis. Water Research., *39*, 2621–2635.

OUYANG Y., NKEDI KIZZA P., WU Q. T., SHINDE D., HUANG C. H., 2006: Assessment of seasonal variations in surface water quality. Water Research., *40*, 3800–3810.

PASCOE E.S.,1968: A manual of the geology of India and Burma, Government of India, Pub.

REGHUNATH R., MURTHY T.R.S., RAGHAVAN B.R., 2002: The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. Water Research., *36*, 2437–2442.

SHRESTHA S., KAZAMA F., 2007: Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environmental Modeling and Assessment, *22*, 464–475.

SIMEONOV V., EINAX J. W., STANIMIROVA I., KRAFT J., 2002: Environmetric modeling and interpretation of river water monitoring data. Analytical and Bioanalytical Chemistry, 374, 898–905.

SIMEONOV V., SARBU C., MASSART D. L., TASAKOVSKI S., 2001: Danube river water data modeling by multivariate data analysis. Microchimica Acta, 137, 243–248.

SIMEONOV V., SIMEONOVA P., TSITOURIDOU R., 2004: Chemometric quality assessment of surface waters: two case studies. Chemical and Engineering Ecology, *11*, 6, 449––469.

SIMEONOV V., STEFANOV S., TASAKOVSKI S., 2000: Environmetrical treatment of water quality survey data from Yantra River, Bulgaria. Microchimica Acta, 134, 15–21.

SIMEONOV V., STRATIS J.A. SAMARA C., ZACHARIADIS G., VOUTSA D., ANTHEMIDIS A. et al., 2003: Assessment of the surface water quality in Northern Greece. Water Research., *37*, 4119–4124.

SIMEONOVA P., SIMEONOV V., ANDREEV G., 2003: Environmetric analysis of the Struma River water quality. Central European J. Chemistry, 2, 121–126.

SINGH K.P., MALIK A., SINGH V.K., 2005a: Chemometric analysis of hydro-chemical data of an alluvial river – A case study. Water, Air and Soil Pollution, 170, 383–404.

SINGH K.P., MALIK A., SINHA S., 2005b: Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques– A case study. Analytical Chimica Acta, 538, 355–374.

SINGH K.P., MALIK A., MOHAN D., SINHA S., 2004: Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – A case study. Water Research., *38*, 3980–3992.

SOTELO F.M., ANDRADE J.M., CARLOSENA A., TAULER R., 2007: Temporal characterization of river waters in urban and semi-urban areas using physico-chemical parameters and chemometrics methods. Analytical Chimica Acta, 583, 128–137.

SUNDARAY S.K., PANDA U.C., NAYAK B.B., BHATTA D., 2006: Multivariate statistical techniques for the evaluation of the Mahanadi river-estuarine system (India) – A case study. Environmental Geochemistry and Health, 28, 317––330.

SUESS E., 1904–1924: The Face of the Earth, 5 Vols., Clarendon press, Oxford.

TARRADO M., BARCELO D., TAULER R., 2006: Identification and distribution of contamination sources in the Erbo river basin by chemometrics modeling coupled to geographical information systems. Talanta, 70, 691–704.

VEGA M., PARDO R., BARRADO E., DEBAN L., 1998: Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research., 32, 3581–3592.

WUNDERLIN D.A., DIAZ M.P., AME M.V., PESCE S.F., HUED A.C., BISTONI M.A., 2001: Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba, Argentina). Water Research., 35, 2881–2894.

ZEILHOFER P., LIMA E.B.N.R., LIMA G.A. R., 2006: Spatial patterns of water quality in the Cuiaba river basin, Central Brazil. Environmental Monitoring and Assessment, 123, 41–62.

MONITOROVÁNÍ PODZEMNÍ VODY V ČASE A V PROSTORU POMOCÍ MULTIVARIAČNÍCH STATISTICKÝCH METOD V OBLASTI BAREILLY, UTTAR PRADESH, INDIE

Sudhir Kumar Singh, Chander Kumar Singh, Kewat Sanjay Kumar, Ramvtar Gupta, Saumitra Mukherjee

Multivariační statistické metody jako shluková analýza a analýza základního prvku byly aplikovány na změřené hodnoty kvality podzemní vody v oblasti Bareilly, Uttar Pradesh, Indie. Monitorování bylo aplikováno na deseti odběrech s využitím 15 parametrů v letech 2005–2006. Kvalita vzorků podzemní vody obvykle podává informaci o přírodních a antropogenních zdrojích znečištění. Vzorky vody po fyzikálně chemickém vyhodnocení vykázaly zhoršení v některých parametrech jako tvrdost, rozpuštěné látky (TDS), obsah tuhých částic (TS), obsah kovů jako železo, zinek a jiné. Tyto parametry vyznačují úroveň kvality podzemní vody pro možnost použití jako pitné nebo závlahové vody, případně pro jiné účely.

Po fyzikálně chemickém rozboru byly výsledky zpracovány metodou analýzy hlavního prvku (principal component analysis PCA) a metodou shlukové analýzy (cluster analysis CA). Aplikací statistických metod na naměřené výsledky byly získány tři složky, které určily strukturalizaci těchto výsledků. Jako principiální prvky byly pojaty stopy skupiny kovů (prosakování z půdy, místa ukládání odpadů), skupina organických znečištění (komunální a průmyslové odpady) a skupina živin (organických i anorganických). Shoda odběrných míst byla získána pomocí shlukové analýzy na základě podobnosti mezi nimi. Studie ukazuje užitečnost multivariačních metod pro vyhodnocení a interpretaci kvality podzemních vod. Proporcionální rozdělení zdrojů znečištění lépe informuje o kvalitě podzemních vod a umožňuje účinný způsob jejich využití.