STATISTICAL HYDROCHEMISTRY AS TOOL TO FACILITATE THE CONTROL OF MINE WATER*

Ben Keet**

*Study performed at Institute of Earth Sciences, Free University Amsterdam, The Netherlands

**Hydrogeologist, Petroleum Engineer, Shell Internationale Petroleum Maatschappij B.V. P.O. Box 162, 2501 AN The Hague, The Netherlands.

ABSTRACT

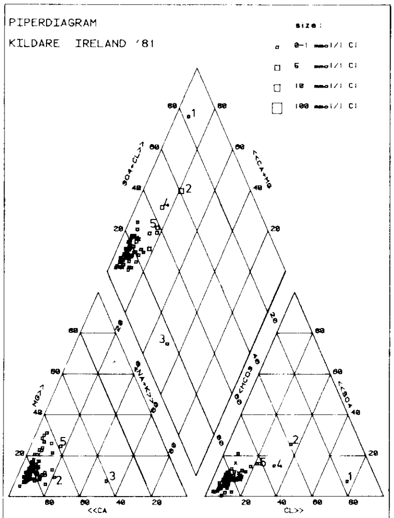
Principal component factor analysis with the varimax orthogonal rotation criterion applied to hydrochemical data is shown to be a promising tool in determination of the origin of water. With advance dewatering becoming more important, this is illustrated in a case study of an area in pre-mining state having an almost homogeneous water chemistry. The R-mode classification reveals the ruling chemical processes, enabling selection of ions to study the areal distribution of a reduced number of processes.

INTRODUCTION

Determination of the origin of water entering mines can add to the development of more effective and economic schemes for mine dewatering [1]. In addition to active dewatering, infiltration control and grouting of source aquifers may help to reduce water inflow and consequently the need for costly water treatment plants [2]. A way to identify the origin of the water is to determine its chemical composition. Due to the strict water pollution controls in many countries, the chemistry of water from mines and surrounding dewatering wells must be analysed frequently. Therefore only a few additional analyses of samples are needed from water inrush areas inside the mine and from surface waters in the mine area.

Common techniques to obtain a regional picture such as a Stiff Symbol Map (effective when large variations occur [1]) or a Piper Plot only take into account the major ions. Furthermore they are time consuming to construct and when this is done by hand, errors can be made. Moreover the information is often limited if the samples show only a minor differentiation in chemistry, i.e. samples from still unpolluted water. In conclusion, a grouping technique which incorporates all available information in the grouping procedure and with a sensitivity to minor changes in composition is needed.

Such techniques are provided by the vast number of multivariate statistical analyses [3,4,5,6]. The application of factor analysis is relatively simple and unambiguous compared with other multivariate techniques such as cluster analysis [4,5,7]. Although the latter and discriminant analysis provide useful grouping facilities [4,7], factor analysis is chosen for the ease of application and interpretation of the results. The objective of factor analysis is to represent a large number of variables (or cases) by a reduced number of hypothetical entities or factors. Consequently, factor analysis tries to reduce the data by detecting the possible underlying pattern of relationships, such that the given data is rearranged (reduced) into a smaller number of factors or components, accounting for the interrelations among data variables (or cases) [5]. An objective grouping of the samples or ions is simply obtained by running the factor analysis program on a computer, however; for a physically significant interpretation, hydrogeological and hydrochemical knowledge of the area is essential [4,5]. The use of factor analysis applied to hydrochemical data from a chemically almost homogeneous area, as indicated by the Stiff Diagram Map [7] and the Piper Plot (Fig.1) is presented in a case study.



No	in Fig. 2
1	53
2	28
3	42
4	72
5	37

Figure 1. Piper plot of the samples

CASE STUDY ; AREA DESCRIPTION

The study area is located in County Kildare 50 km southwest of Dublin on the eastern margin of the Central Lowland of Ireland. The location of the area, as well as the location of the sample points and their numerical designations are given in Fig.2. Topographically the area can be divided into (1) a flat central part whose elevation gradually decreases to the southwest from 100 m to 60 m O.D., (Irish Ordnance Datum is 2.554 m below mean sea level) in this area water levels are high and extensive bogs exist, (2) some hillocks in the north rising up to 135 m and (3) a chain of hills in the southeast and east, some of which reach heights of more than 150 m, groundwater is usually found deeper than 10 meters. The lowest part is a valley of the river Barrow in the western part of the area (60 m to 55 m O.D.). In the northeast we find a valley of the river Liffey, its elevation is between 95 m and 85 m O.D. Considering only the area between these two rivers, we find three main aquifers: (1) starting at surface we find unconfined aquifers in the north-eastern and southern part, which are respecticaly ice-pushed Weichselian kame deposits (Curragh aquifer) and moraines, both containing a large amount of limestone fragments.

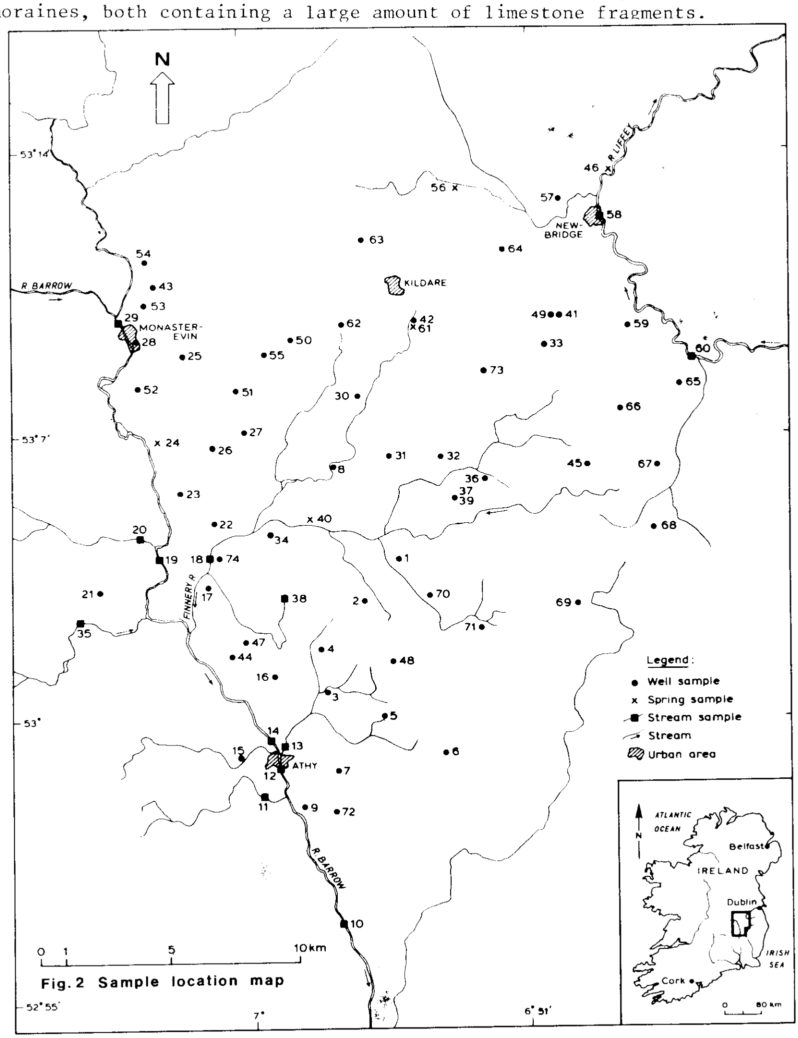


Fig. 2 Sample location map

A large part of this aquifers is unsaturated. A semi-confined strip aquifer [2] is found in a paleo-channel (Fig.3), the former course of the Liffey flowing southwest to the Barrow. This Paleo-channel aquifer is composed of sand and limestone gravels of the Elsterian or Saalian glaciation. The limestone basement is most likely karstified, therefore the hydraulic boundaries of this limestone aquifer are difficult to define. Outside the flat central plain where the Paleo-channel aquifer acts as a drain, groundwater flow is controlled by the topography (Fig.4).

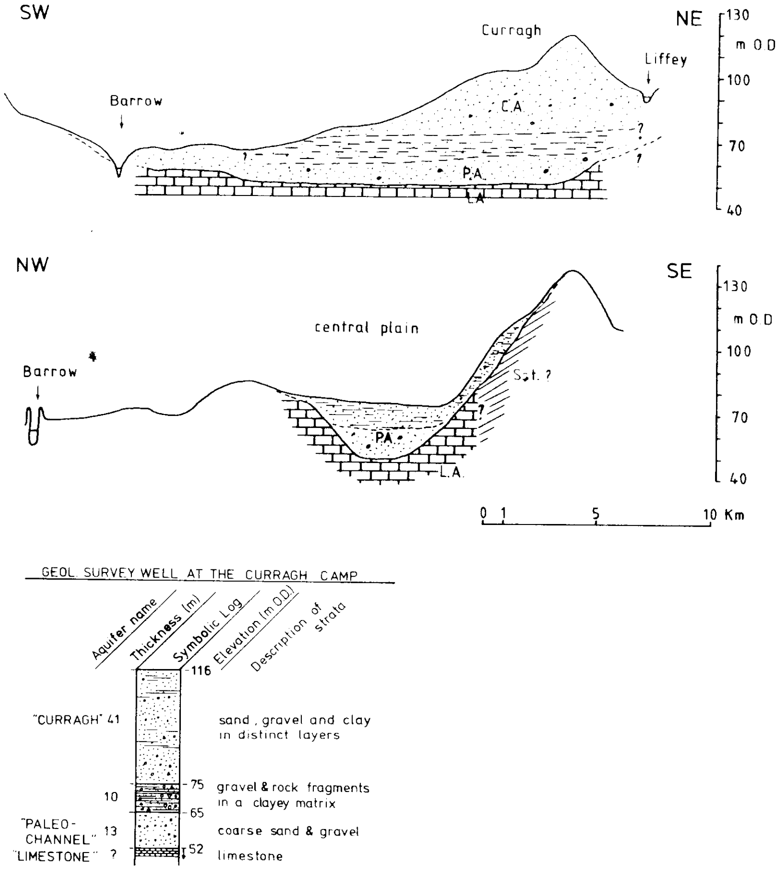


Figure 3: Two sketch profiles running through the middle of the area based on resistivity soundings [7] and one borehole log at the Curragh

INPUT DATA MATRICES

In the study area, 74 samples were taken (Fig.2) from wells (56), springs (5), and streams (13); 14 variables were quantified, among which the temperature, pH and Electrical conductivity were measured in the field and the ion concentrations determined in the laboratory (Institute of Earth Sciences, Free University, Amsterdam). These 14 times 74 numbers constitute the raw data matrix (Table 1).

Pre-processing of this raw data matrix is necessary due to (a) the requirement that the values are in the same order of magnitude and (b) the variables should have a normal distribution.

A transformed data matrix is used for the Q-mode analysis, which involves the grouping of cases. Transformation involves the values of all variables in the raw data matrix being divided by their respective maximum value.

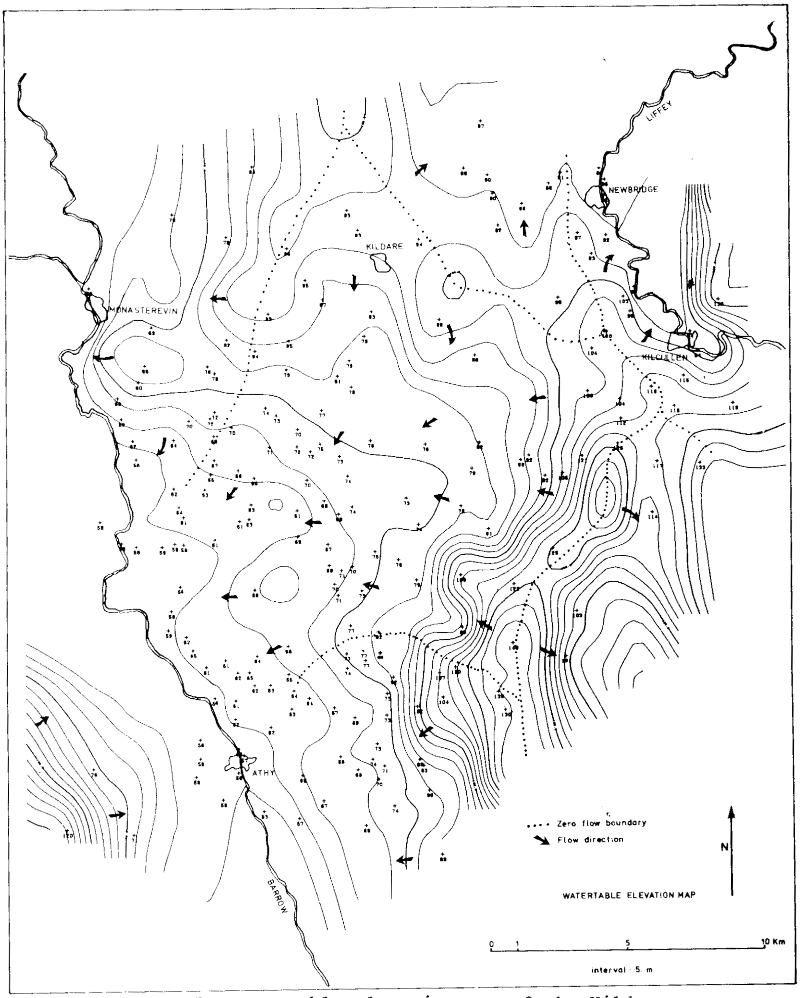


Figure 4. The watertable elevation map of the Kildare area

SAMP. No.	variables 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	11.1	7.18	770.	404.	388.	147.	605.	339.	598.	495.	217.	348.	111.	10.
2	10.2	7.06	930.	511.	574.	368.	586.	391.	1210.	570.	365.	471.	123.	12.
3	10.0	7.10	785.	643.	899.	294.	460.	332.	592.	646.	433.	447.	159.	5.
4	10.8	7.48	505.	369.	558.	33.	685.	166.	530.	456.	50.	55.	152.	48.
5	11.3	7.28	650.	445.	334.	22.	1060.	230.	474.	551.	201.	234.	148.	7.
6	10.5	7.29	605.	488.	357.	119.	410.	253.	355.	490.	226.	577.	142.	.5
7	10.3	7.09	715.	534.	572.	32.	510.	273.	547.	733.	395.	495.	216.	.5
8	9.7	7.29	800.	444.	439.	76.	373.	377.	942.	523.	310.	648.	129.	7.
9	10.3	7.39	820.	714.	874.	320.	369.	349.	1070.	628.	325.	1050.	155.	6.
10	13.5	7.92	510.	474.	341.	77.	319.	250.	429.	511.	363.	113.	75.	.5
11	14.2	8.43	695.	435.	379.	43.	408.	319.	654.	562.	348.	234.	92.	.5
12	13.4	7.92	530.	471.	319.	63.	298.	246.	417.	495.	349.	150.	69.	.5
13	12.9	7.40	820.	506.	435.	74.	661.	332.	728.	688.	333.	234.	91.	8.
14	13.4	7.76	540.	458.	342.	62.	294.	247.	412.	490.	331.	116.	70.	10.
15	13.0	7.31	790.	550.	515.	277.	601.	310.	733.	550.	674.	650.	87.	.5
16	9.8	7.14	720.	401.	344.	16.	663.	307.	626.	561.	218.	176.	96.	.5
17	10.2	6.84	1080.	593.	562.	18.	656.	495.	1270.	952.	744.	126	183.	230.
18	14.5	8.02	765.	658.	444.	83.	618.	326.	615.	721.	367.	208.	105.	.5
19	13.5	7.66	520.	444.	288.	58.	281.	228.	384.	472.	345.	100.	58.	8.
20	13.8	8.29	690.	512.	380.	55.	397.	312.	626.	624.	413.	281.	85.	7.
21	11.6	7.35	700.	626.	391.	61.	262.	318.	1070.	547.	295.	556.	126.	9.
22	10.3	7.25	625.	500.	414.	27.	663.	236.	621.	600.	218.	.5	126.	61.
23	11.2	6.92	805.	379.	427.	26.	598.	367.	536.	511.	440.	234.	140.	.5
24	10.8	7.15	685.	417.	475.	35.	482.	301.	733.	615.	287.	329.	118.	.5
25	10.6	7.06	900.	770.	743.	32.	824.	356.	1260.	700.	548.	550.	127.	10.
26	11.6	7.15	840.	507.	1000.	129.	435.	350.	587.	770.	359.	526.	123.	8.
27	10.5	6.93	915.	527.	817.	54.	634.	378.	1400.	690.	372.	395.	110.	.5
28	11.2	6.97	1430.	1890.	5200.	172.	1180.	878.	480.	533.	1740.	9890.	204.	20.
29	13.5	7.51	538.	451.	348.	62.	261.	251.	389.	492.	353.	113.	70.	8.
30	10.6	7.26	640.	376.	389.	26.	1080.	218.	412.	538.	148.	113.	126.	65.
31	10.8	7.70	1060.	614.	684.	492.	523.	503.	745.	1010.	655.	55.	262.	22.
32	10.1	7.43	610.	381.	367.	80.	732.	213.	564.	567.	14.	24.	87.	75.
33	10.2	7.29	630.	531.	209.	57.	341.	308.	226.	669.	89.	161.	104.	12.
34	11.7	7.36	785.	413.	345.	24.	857.	311.	654.	587.	363.	23.	116.	114.
35	13.3	8.11	570.	450.	365.	64.	247.	280.	542.	538.	161.	239.	61.	9.
36	12.3	7.85	465.	364.	484.	126.	206.	177.	728.	347.	191.	165.	33.	.5
37	10.7	7.83	660.	486.	1040.	182.	879.	210.	1480.	405.	532.	5.	49.	20.
38	11.2	7.34	240.	203.	384.	16.	135.	96.2	361.	187.	129.	113.	33.	16.
39	10.7	7.18	1000.	647.	994.	179.	887.	268.	1350.	498.	545.	1280.	151.	.5
40	12.6	7.33	790.	406.	442.	33.	909.	284.	609.	455.	231.	344.	132.	9.
41	10.0	7.90	730.	643.	494.	33.	553.	372.	513.	749.	362.	374.	122.	.5
42	11.0	6.50	280.	294.	2470.	25.	183.	101.	282.	172.	22.	11.	159.	16.
43	10.6	7.70	720.	441.	377.	33.	1270.	297.	542.	482.	13.	123.	149.	7.
44	9.7	8.00	700.	424.	453.	51.	546.	319.	818.	413.	363.	432.	90.	.5
45	10.6	8.00	750.	471.	393.	40.	595.	338.	739.	470.	229.	632.	140.	.5
46	13.3	7.80	735.	443.	458.	26.	975.	242.	417.	564.	256.	82.	156.	.5
47	10.0	7.50	780.	402.	452.	52.	535.	320.	874.	436.	249.	140.	92.	.5
48	13.0	7.60	810.	582.	1050.	535.	573.	304.	1380	413.	526.	990.	103.	.5
49	12.5	7.65	650.	374.	369.	18.	630.	301.	338.	420.	242.	323.	152.	.5
50	11.2	7.85	560.	468.	602.	50.	228.	244.	474.	524.	159.	13.	101.	23.
51	11.3	7.42	770.	602.	551.	33.	509.	331.	920.	652.	339.	297.	101.	6.
52	10.6	7.25	750.	626.	509.	97.	342.	353.	451.	751.	229.	65.	163.	8.
53	10.7	7.56	640.	768.	484.	23.	789.	245.	649.	14.4	31.	6320.	135.	.5
54	10.4	7.21	580.	533.	297.	83.	231.	290.	316.	608.	230.	27.	76.	.5
55	10.2	7.35	1160.	595.	2210.	159.	394.	510.	1830.	770.	655.	527.	118.	.5
56	10.5	7.12	790	359.	396.	57.	409.	341.	558.	462.	156.	284.	126.	.5
57	10.7	7.42	920.	666.	830.	117.	554.	376.	829.	731.	375.	814.	173.	.5
58	14.0	7.98	310.	246.	273.	24.	168.	115.	305.	226.	125.	69.	56.	.5
59	11.2	7.31	735.	482.	471.	68.	584.	283.	621.	574.	224.	503.	123.	.5
60	13.5	7.80	270.	207.	253.	20.	137.	98.9	282.	192.	106.	61.	61.	.5
61	10.7	7.16	790.	442.	384.	25.	735.	313.	485.	538.	180.	258.	128.	.5
62	11.7	7.47	635.	288.	254.	44.	435	264.	338.	538.	106.	213.	128.	7.
63	11.3	7.27	740	595.	488.	190.	499.	272.	733.	600.	147.	300.	361.	.5
64	11.9	7.37	695.	452.	302.	11.	482.	347.	305.	669.	85.	21.	131.	24.
65	11.4	7.19	710.	353.	309.	19.	216.	391.	327.	502.	137.	185.	158.	.5
66	11.5	7.22	760.	404.	402.	84.	429.	323.	429.	690.	101.	132.	214.	.5
67	9.3	7.05	730.	582.	373.	15.	280.	352.	519.	646.	137.	739.	124.	.5
68	9.9	7.22	615.	508.	315.	13.	335.	307.	395.	613.	132.	264.	143.	8.
69	11.4	7.29	745.	477.	690.	25.	653.	345.	677.	626.	244.	184.	115.	.5
70	10.7	7.25	810.	670.	647.	117.	766.	329.	931.	672.	425.	542.	142.	16.
71	11.0	7.28	720.	420.	424.	18.	526.	320.	542.	542.	191.	306.	121.	.5
72	13.0	7.60	810.	697.	1410.	13.	570.	359.	1830.	315.	441.	2940.	88.	.5
73	10.7	7.40	740.	610.	396.	71.	589.	347.	480.	542.	192.	321.	127.	.5
74	12.0	7.75	785.	661.	452.	29.	743.	277.	666.	644.	278.	113.	141.	.5

Column		
1	Temp.	
2	pH	
3	Ec field	(mmho/cm)
4	Ec lab	(mmho/cm)
5	Na	(μmol/lt)
6	K	(μmol/lt)
7	Mg	(μmol/lt)
8	Ca	(10 μmol/lt)
9	Cl	(μmol/lt)
10	HCO_3	(10 μmol/lt)
11	SO_4	(μmol/lt)
12	NO_3	(μmol/lt)
13	Si	(μmol/lt)
14	Fe tot	(μmol/lt)

Table 1 Raw data matrix showing all variables of the samples used in this case study

A logarithmic input data matrix is used for the R-mode analysis, the grouping of variables, since these show a better defined normal distribution after logarithmic transformation [7].

R-MODE RESULTS

Using the logarithmic input data matrix, the principle components of factor analysis are applied to the input data correlation matrix (Pearson product moment) with the varimax orthogonal rotation criterion [5,6]. The resulting simplified rotated factor score coefficient matrix is presented in Table 2, in which factor scores below 0.600 have been depleted as being not significant [4]. The factors, representing independent new variables, will be discussed in order of decreasing importance (determined by the percentage of explained variance).

Factors	I	II	III	IV	V
Variables					
1 Temp		.777			
2 pH		.826			
3 Ec field	-.843				
4 Ec lab					
5 Na				.795	
6 K				.695	
7 Mg					
8 Ca	-.847				
9 Cl					.723
10 HCO_3	-.738				
11 SO_4	-.600				
12 NO_3			-.788		
13 Si		-.622			
14 Fe tot			.788		
Cum. % expl.var	37	51	61	70	78

Table 2. R-mode simplified rotated factor score coefficient matrix

Factor I : Ec field, Ca, HCO_3, and SO_4

The *limestone solution factor* is the most important factor which is hardly surprising in a limestone area where the reaction

$$CaCO_3 + CO_2 + H_2O \rightarrow Ca^{2+} + 2HCO_3^-$$

is the ruling process. This is in the same way (same sign) responsible for variations in electrical conductivity (which, having a linear relation with all ions, is a factor in itself). The presence of SO_4 is less obvious and dissolved gypsum could account for this.

Factor II : Temp, pH, and Si

The *diatom factor* is of secondary importance, since only 13 surface water samples are present due to the homogeniety of the samples. The ruling reaction is growth of diatoms due to the higher temperature of surface waters in springs, thereby reducing the Si concentration (opposite sign). Furthermore, surface waters generally have a higher pH.

Factors samples	I	II	III	IV	V	VI
1				-.533		
2		-.628				
3	.734					
4				-.713		
5				-.801		
6	.611		-.564			
7	.706					
8	.534					-.540
9	.634	-.503				
10			-.743			
11			-.674			
12			-.755			
13			-.512	-.526		
14			-.757			
15			-.537			
16				-.612		
17					-.593	
18			-.581			
19			-.776			
20			-.678			
21			-.563			-.540
22				-.683		
23	.527			-.571		
24	.523		-.502			
25				-.571		-.651
26	.649		-.509			
27						-.721
28					.917	
29			-.761			
30				-.865		
31	.859					
32			-.502	-.696		
33	.629		-.577			
34				-.756		
35			-.753			
36			-.771			
37					-.580	
38			-.861			
39				-.556		-.554
40				-.732		
41	.592					
42			-.668			
43				-.826		
44			-.526	-.507		-.511
45				-.569		
46				-.736		
47				-.504		-.541
48		-.829				
49			-.544	-.622		
50	.500		-.705			
51			-.504			-.553
52	.738					
53				-.544		
54	.575		-.657			
55						-.817
56	.557		-.522			
57	.670					
58			-.848			
59	.513		-.502	-.550		
60			-.849			
61				-.660		
62	.546		-.594	-.520		
63	.715					
64	.582			-.533		
65	.647		-.570			
66	.734					
67	.645		-.507			
68	.642		-.529			
69				-.577		
70	.508			-.589		
71	.513		-.522	-.546		
72						-.738
73	.549			-.547		
74				-.624		
cum % of expl var	83	87	90	93	95.4	97.3

Table 3 Q-mode simplified rotated factor score coefficient matrix

Factor III : Fe tot, and NO_3

The *anaerobic factor* in this boggy area only takes third place. The ruling process is denitrification according to

$$1/5NO_3^- + 6/5H^+ + e \rightarrow 1/10N_2\ (g) + 3/5H_2O$$

which is due to the breakdown of organic matter. In this reducing environment (where low, up to negative, pe-values prevail) iron may dissolve :

$$\tfrac{1}{2}Fe_2O_3 + 3H^+ + e \rightarrow Fe^{2+} + 3/2H_2O$$

While the signs of the factor scores are opposite, this reaction is possible.

Factor IV : K, and Na

The *fertilizer factor* represents an addition of unpurified K-fertilizer. Still explaining 9% of the variance, it is a significant factor.

Factor V : Cl

This factor, being the last and having only one variable, is not significant, although it still explains 8% of the variance. Pollution, caused by seepage of manure through the lining of brick, into the dugwells might be one of the processes leading to variance in the data which is explained by this factor.

Now the ruling processes are known, we can reduce the number of variables to be measured in future sampling, e.g. Ec lab, and Mg can be deleted and only one variable of each factor has to be measured; determination of Ca, pH, NO_3, Na, and Cl will suffice. One can also delete the ions of a particular process, for example all variables of the *limestone solution factor* will increase the importance of the other processes on the grouping procedure. In polluted areas or if samples are mixed with a drilling mud, this deletion of variables as a way of digital filtering adds an extra dimension to the multivariate techniques.

Q-MODE RESULTS

Using the transformed data matrix as input for the same analysis as is used to acquire the R-mode results, the simplified rotated factor score coefficient matrix as presented in Table 3 is obtained. If criteria for the factor scores to be significant are set at 0.600, not all samples appear in a factor, however; a significant grouping is obtained. The physical meaning of factors based on the knowledge of the remaining sample points [7] is as follows :

Factor I : samples from wells in the unconfined aquifers, mainly the Curragh aquifer.

Factor II : there is no explanation for the appearance of sample 2 and 48 in a separate factor, when the raw data are used these are grouped in the "unconfined" factor (I).

Factor III : the majority of the surface water samples are clearly present in this group.

Factor IV : samples from wells in the confined aquifers as well as all springs fed by these aquifers; the Paleo-channel and limestone aquifers.

Factor V : one exceptionally polluted well (28) forms this factor.

Factor VI : samples in this factor show high chloride concentrations and are all polluted wells.

The actual importance lies in the *wrongly classified samples* : Note samples 36, 42, 50 and 54 are from wells appearing in the surface water factor (III). Field evidence indicates that surface water has infiltrated in wells 36, 50 and 54. Well 42 contains almost rain water (see also the deviation in the Piper plot, Fig.1). More information can be gained if the significance level is lowered to 0.5 (italic figures in Table 3 are now also considered). Considering these figures, the mixing of water becomes evident. For example, the river water of sample 13 is largely fed by groundwater from the confined aquifer, which is realistic in that artesian area [7]. Deep wells have a filter over their entire depth and then close to rivers, tap confined and unconfined aquifers as well as surface water, as in the case of wells 59, 62 and 71. If these were dewatering wells, river bed sealing might improve mine drainage efficiency. At the 0.500 significance level, factors I and III show an overlap. If we reduce the significance level more, an overlap with other factors will occur, thereby reducing the grouping effect. It is obvious that one should compromise between a well defined grouping and all the cases classified.

CONCLUSION

From this study it is understood that :

(a) Multivariate techniques as factor analysis does appear to be useful and are an extension of the Piper and Duror diagram idea. This is especially true in chemically almost homogeneous areas.

(b) Using factor analysis for an R-mode classification, the ruling chemical processes can be determined in an objective way. All the problems appear to be in the conclusions drawn from the classification results.

(c) The three dimensional distribution of water types, including mixing features is determined by Q-mode classification.

REFERENCES

1. Fernández-Rubio, R., Investigations about the origin of water inflow at Reocin Mine (Santander, Spain), Proceedings of the First International Mine Water Congress, Budapest (1982).

2. Norton, P. J., Advance dewatering and control of groundwater in surface coal mining, First International Mine Water Congress, Budapest (1982).

3. SPSS, Statistical Package for the Social Sciences, edited by : N. H. Nie, C. H. Hull, J. G. Jenkins, K. Strienbenner and D. H. Bent, McGraw-Hill, New York (1975).

4. Seyhan, E. and Keet, B., Multivariate Statistical Analysis (Part I) : Application to hydromorphometrical data (case study : Ahr River Basin, Bolzano, Italy), Aqua-VU, No.8, Rodopi, Amsterdam (1981).

5. Seyhan, E. and Keet, B., Comparative case studies with multivariate techniques, Geol. Mijnbouw 62 (1) (in press).

6. Seyhan, E., Introduction to multivariate statistical analysis in hydrology, Inst. of Earth Sciences, Free University, Amsterdam (1981).

7. Keet, B., Regional groundwater survey in County Kildare, Ireland, Doctoral thesis, Inst. of Earth Science, Free University, Amsterdam (1982).