Commun. Fac. Sci. Univ. Ank. Series C
V. 6, pp 313-323 (1988)

PARTITIONING OF GEOCHEMICAL POPULATIONS BY SINCLAIR'S METHOD

(An Application on a Geochemical Stream Sediment Data from the Belgian Ardennes)

HALİL ARAL and CEM SARAÇ

Hacettepe University, Engineering Faculty, Department of Geological Engineering

ABSTRACT

Sinclairs's method of separation of statistical populations is successfully applied on numerical stream sediment data (hot extractable lead data) from the Belgian Ardennes. The application showed that the data consisted of one background (values less than 60 ppm) and two anomalous population (a) between 60 and 390 ppm and (b) values greater than 390 ppm. Samples representing these populations are shown with different symbols on a map (discrete map) so that delineation of the anomalous areas is possible on the map.

INTRODUCTION

Sinclair's method (Sinclair 1976) of partitioning statistical populations on cumulative probability plots is applied on geochemical stream sediment data from the Ardennes, Belgium. This method includes the calculation of limits of class intervals. cumulative frequency percentiles, their plot on arithmetic /logarithmic probability graph papers, partitioning of constituent populations and estimation of threshold values. Interpretation of the field and analytical data was published elsewhere (Aral, 1987 a, b) and is not repeated here.

The survey area, Couvin, is located in the Namur Province of Southern Belgium between Chimay on the west and the Belgium–France border on the east (Fig. 1). The survey of the stream sediment was conducted by the senior author over an area of about 210 sq. km and 754 samples were collected from streams, tributaries and gullies at a distance

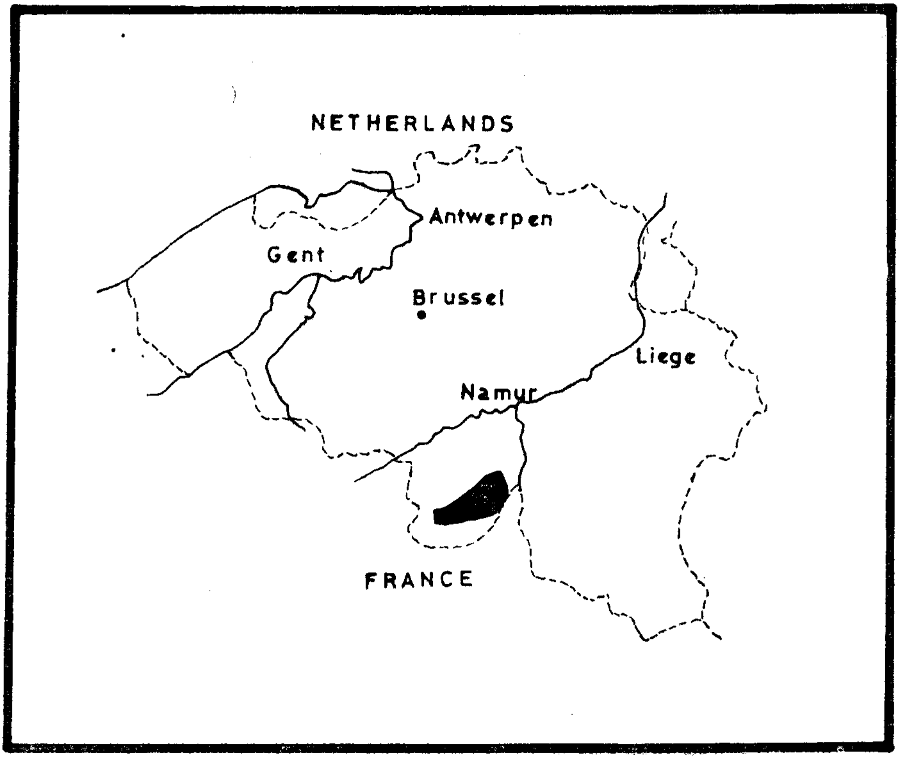

Figure 1. Location map of the study area.

of 200 m. intervals. The stream sediments were analysed for Pb, Zn, Cu, Ni and Mn. However, this article considers only the stream sediment survey of hot extractable lead values. Lead is chosen as the main metal to do this evaluation because the lithostratigraphically controlled mineralization is essentially a lead bearing type. For all analyses atomic absorption spectrophotometry was used.

The purposes of this article are to give a brief information about the Sinclair's method of the statistical population partitioning and to apply this method on the data of the author from the Belgian Ardennes.

SINCLAR'S METHOD OF PARTITIONING

Procedure:

The procedure that is concerned with the estimation of constituent populations from a combination of two or more density distribution is known as Sinclair's method of partitioning which is based on the recognition of inflection point (s) along a curve defined by plotting data points (cumulative percentile frequencies against class intervals) on a log–normal probability paper.

When the inflection points are determined the estimation of constituent populations is done on the probability graph according to the following simple relationships:

$$P = (Pm/f)*100 \quad \ldots \quad (1)$$

$$P = (100 - Pm)/f \quad \ldots \quad (2)$$

$$P = 100 - \left(\frac{100 - Pm}{f}\right) \quad \ldots \quad (3)$$

where Pm = probability percentile of any point on the main curve at a given ordinate level,

P = probability percentile of the population to be partitioned at the same ordinate level of Pm, and

f = percentage of the population to be partitioned in comparison to the rest.

To illustrate the use of these relationships let's consider the curved distribution of Figure 2 which is given in Sinclair (1976). An inflection point is apparent at the 30 cumulative percentile indicating 30 percent

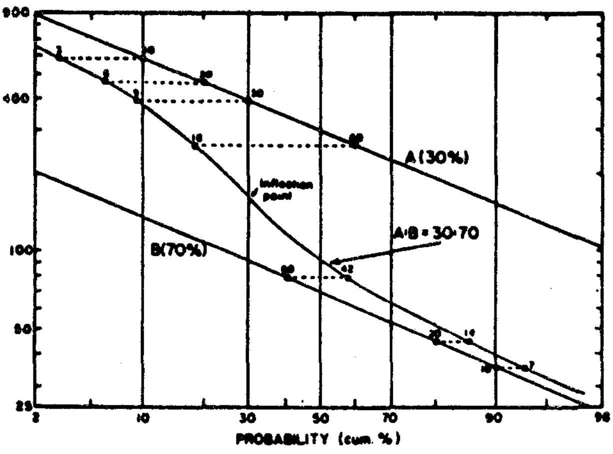

Figure 2. Illustration of partitioning proscedure utilizing the two non-intersecting populations.

of an upper population and 70 percent of a lower population. Now, let's consider the cumulative percentile at any ordinate level near the upper extremity of the curve, say the 3 percentile for ease of calculation. This point represents only 3 percent of the total data but is (3 /30) * 100 = 10 percent of the upper population (see equation 1 above). Hence, at this ordinate level a point has been defined at the 10 percentile that plots on the upper population. This procedure can be repeated for various ordinate levels that correspond to 6, 9, etc., cumulative percent on the curve, to define points on the line describing the upper population at 20, 30 etc., cumulative percent. The procedure can only be repeated until the effect of the lower population becomes significant at which point the calculated values depart from a linear pattern. In practice, generally, sufficient points with a linear trend can be determined to extrapolate the trend throughout the entire probability range and thus define the upper population quite precisely.

The points defining the lower population on the same graph paper can be partitioned by the same procedure providing the probability scale is read in a complementary fashion (e.g., 80 cumulative percent is read as $100_{+}80=20$ cumulative percent; see equation 2 above).

When the points are determined, a smooth line is plotted through the data points. If the points of a partitioned population do not plot on a straight line, a new inflection point must be chosen as a basis for new calculation. In Figure 2, the sloping central segment of the main curve indicates that there is appreciable overlap of the ranges of the two populations. If no overlap existed the central segment would be near vertical.

Normally, if three populations (i. e. two inflection points) are present and do not overlap too extensively, their partitioning procedure is like that of two population curves, but involving an additional step to discern the third population. Considerable difficulty arises, however, if relatively few data values (generally less than 75) are available, or if intersecting populations are present (for example see Figure IV–10 in Sinclair, 1976).

The quality of partitioning in the overlapped areas is controlled by applying the following relationships:

For two populations: $P_{(A \S B)} = P_A f_A + P_B f_B$ ....(4)

For three populations: $P_{(A \S B \S C)} = P_B f_B + P_{((A \S C)} . f_{(A \S C)}$ ..... (5)

For detailed elaboration of this method, one is advised to refer to Sinclair (1976).

Estimation of threshold values:

When there is only one normally distributed population the traditional procedure recommended to estimate threshold values by Hawkes and Webb (1962) is to make the assumption that the upper 2.5 % of values (i.e. mean + 2* standard deviation) are anomalous. The standdard deviation of a partitioned population is read from the x–axis of the probability curve at 16 and 84 cumulative percentiles and the mean is read from 50 percentile. The 2*standard deviation value is calculated by substracting the value of 16 percentile from that of 84 percentile. When two populations are involved in a set of data the threshold values can be estimated from the cumulative frequency percentile curve by plotting a horizontal line to separate best the populations (sloped straight lines) which are plotted by applying Sinclair's partitioning method.

The thresholds are used to code values by symbols and plotted on a map in an attempt to define the fundamental geologic /geochemical significance of each population.

AN APPLICATION; PARTITIONING OF GEOCHEMICAL POPULATIONS OF THE DATA FROM THE ARDENNES

The analytical hot extractable Pb data, listed against a corresponding number of samples (frequencies) in Table 1, indicate that the lowest value is 14 ppm and the highest value 2410 ppm. The mean value of all samples (754 samples) is 81.86 ppm, the standard deviation of which is 150.24 ppm. The mean value falling to the area between the 5 the and 95 th percentiles is 60 ppm. The data is classified into 27 class intervals, the square root of the number of samples, i. e., $\sqrt{754} = 27$, and for each arithmetic class interval frequency and cumulative percentile frequencies are calculated (Table 2).

Plotting the cumulative percentile frequency values against arithmetic class limits (see Table 2) Figure 2 shows a concave upward curved line rather than a straight line therefore suggesting that the distribution of the data is log normal. For that reason, the logarithmic data (Table 3) is used as the information to partition the populations. The cumulative percentile frequencies (last column in Table 3) are plotted against logarith-

Table 1. List of hot extractable Pb data.

ppm	grequency	ppm	frequency	ppm	frequency	ppm	frequency	ppm	frequency
14	1	56	7	97	2	167	2	360	1
16	1	57	8	98	4	168	1	363	1
17	5	58	7	100	4	178	2	370	1
18	10	59	5	102	1	185	1	372	1
19	5	60	12	103	2	188	1	1733	1
20	4			106	1				
21	6	61	2	107	3	189	1	374	1
22	3	62	2	108	2	190	1	378	1
23	7	63	7	109	1	191	1	414	1
24	9	64	3	111	1	200	2	419	1
25	12	65	6	113	4	201	2	465	1
26	17	66	2	114	1	203	1	505	1
27	15	67	2	115	1	205	2	510	1
28	25	68	2	116	2	210	1	570	1
29	10	69	3	117	3	211	1	690	1
30	11	70	1	118	1	213	1	700	1
31	7	71	23	119	2	214	2	785	1
32	14	72	1	120	1	215	1	980	1
33	16	73	3	122	2	216	1	1010	1
34	10	74	4	125	1	220	1	2240	1
35	24	75	2	128	3	223	1	2410	1
36	21	76	3	129	2	236	1		
37	13	77	3	130	2	243	1		
38	17	78	2	131	1	244	2		
39	12	79	3	136	1	247	2		
40	24	80	2	137	2	1248	2		
41	18	81	4	138	1	254	1		
42	16	82	3	135	2	255	1		
43	22	83	1	140	2	262	1		
44	15	84	5	143	2	263	1		
45	15	85	1	144	2	269	1		
46	12	86	3	145	2	274	1		
47	10	87	5	146	1	285	1		
48	8	88	1	148	1	286	1		
49	9	89	4	151	1	293	1		
50	17	90	1	152	1	308	1		
51	8	91	5	155	1	323	1		
52	6	92	2	156	2	337	1		
53	9	93	1	161	1	339	1		
54	9	94	1	162	1	342	1		
55	8	96	1	165	1	358	1		

mic class intervals (first column in Table 3) and a slightly bent curve (Figure 4) is obtained.

The visual examination of the curve given in Fig. 4 indicates the presence of two inflection points having the ordinate values at 2.60 and 1.72. This means that the present data includes three populations, A.,

Table 2. Arithmetic class limits, frequency and cumulative percentile frequency values.

Arithmetic Class Limits (ppm Pb)	Freq.	Sum of Freq.	Cum. % Freq.
> 270.5	30	30	3.98
260.5—270.5	3	33	4.38
250.5—260.5	2	35	4.64
240.5—250.5	7	42	5.57
230.5—240.5	1	43	5.70
220.5—230.5	1	44	5.84
210.5—220.5	7	51	6.76
200.5—210.5	6	57	7.56
190.5—200.5	3	60	7.96
180.5—190.5	4	64	8.49
170.5—180.5	2	66	8.75
160.5—170.5	6	72	9.55
150.5—160.5	5	77	10.21
140.5—150.5	8	85	11.27
130.5—140.5	9	94	12.47
120.5—130.5	10	104	13.79
110.5—120.5	16	120	15.92
100.5—110.5	10	130	17.24
90.5—100.5	20	150	19.89
80.5—90.5	28	178	23.61
70.5—80.5	26	204	27.06
60.5—70.5	30	234	31.03
50.5—60.5	79	313	41.51
40.5—50.5	142	455	60.34
30.5—40.5	158	613	81.30
20.5—30.5	115	728	96.55
≤ 20.5	26	754	100.00

B, C, representing 70 %, 28.5 % and 1.5 % of the overall data, respectively.

The presence of more than one population in the data is also supported from the application of chi–square test. The application of chi–square test on the logarithmic data indicated that at 99 % confidence level the distribution was consisted of more than one population as calculated value 214.68 was greater than the value 18.48 given in standard chi–square test tables.

The partitioning of each population is done by applying equation 1 for population A, equation 2 for population B and equation 3 for population C . In this way, data points for each partitioned population are calculated and straight lines (see dashed lines in Fig. 4) are fitted to each of them. The mixture of the partitioned populations in the middle part (Fig. 4) is checked against the main curve drawn by using the real data by applying equation 4 given above. Such control points are shown with triangular symbols in Fig. 4.

Table 3. Logarithmic class limits, freguencies and cumulative percentile frequency values. f1 = 1.72 and f2 = 2.60 indicate the breaking points between population A, AB. and C.

Logarithmic Class Limits (log ppm Pb)	Arithmetic Class Limits (ppm Pb)	Freq.	Sum of Freq.	Cum. % Freq.
3.0626	1155.1	2	2	0.27
2.9828—3.0626	961.1—1155.1	2	4	0.53
2.9029—2.9828	799.6— 961.1	0	4	0.53
2.8230—2.9029	665.3—799.6	3	7	0.93
2.7432—2.8230	553.6—665.3	1	8	1.06
2.6633—2.7432	460.6—553.6	3	11	1.46
2.5835—2.6633	383.3— 460.6	2	13	1.72
2.5036—2.5835	318.9— 383.3	12	25	3.32
2.4238—2.5036	265.3— 318.9	6	31	4.11
2.3439—2.4238	220.8—265.3	13	44	5.84
2.2641—2.3439	183.7— 220.8	20	64	8.49
2.1842—2.2641	152.8—183.7	11	75	9.95
2.1044—2.1842	127.2—152.8	26	101	13.40
2.0245—2.1044	105.8—127.2	26	127	16.84
1.9447—2.0245	88.0—105.8	28	155	20.56
1.8648—1.9447	73.2—88.0	42	197	26.13
1.7850—1.8648	60.9—73.2	35	232	30.77
1.7051—1.7850	50.7—60.9	81	313	41.51
1.6252—1.7051	42.2—50.7	108	421	55.84
1.5454—1.6252	35.2—42.2	121	542	71.88
1.4655—1.5454	29.2—35.1	82	624	82.76
1.3857—1.4655	24.3—29.2	79	703	93.24
1.3058—1.3857	20.2—24.3	25	728	96.55
1.2260—1.3058	16.8—20.2	24	752	99.73
1.1461—1.2260	14.0—16.8	2	754	100.00

Two threshold values are estimated by plotting imaginary horizontal lines separating best the populations obtained from the application of the Sinclair's method. According to this, the threshold value between population A and B is 1.78 (or 60 ppm) from the horizontal line separating population A and B and 2.58 (or 390 ppm) from the horizontal line separating population B and C . In the identification of these threshold values the overlap (A+B and B+C; see Fig. 4) in the tails of the populations is not taken into consideration. The threshold values arbitrarily represent the upper limits of the overlapping areas.

The 60 ppm threshold value coincides with the actual mean value falling to the area between 5 th and 95 th percentiles and represents 90 % of the data . Therefore, any value less than 60 ppm is classified as background and those larger than 60 ppm as anomalous. The existence of another threshold value at 390 ppm indicates that the anomalous values are not only from one source.

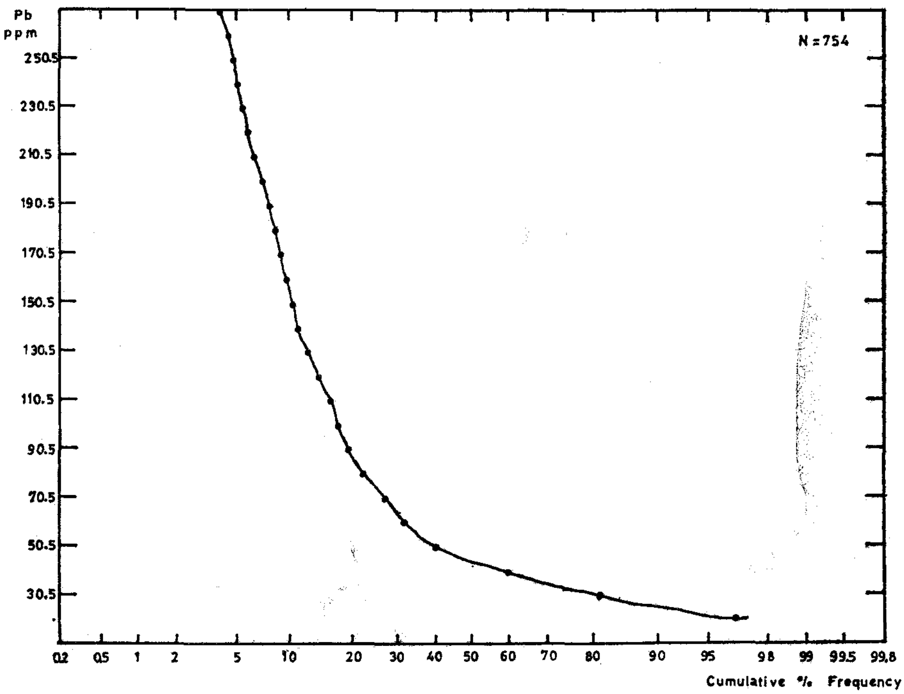

Figure 3. Plot of arithmetic class intervals against cumulative percentile frequency values.

Finally the background and anomalous values which are represented with different symbolcs are plotted to their proper locations on the sample collection map. This map is named as discrete map and shown in Fig. 5. The clustering of the anomalous symbols on this map allows to delineate target areas for detailed follow - up work.

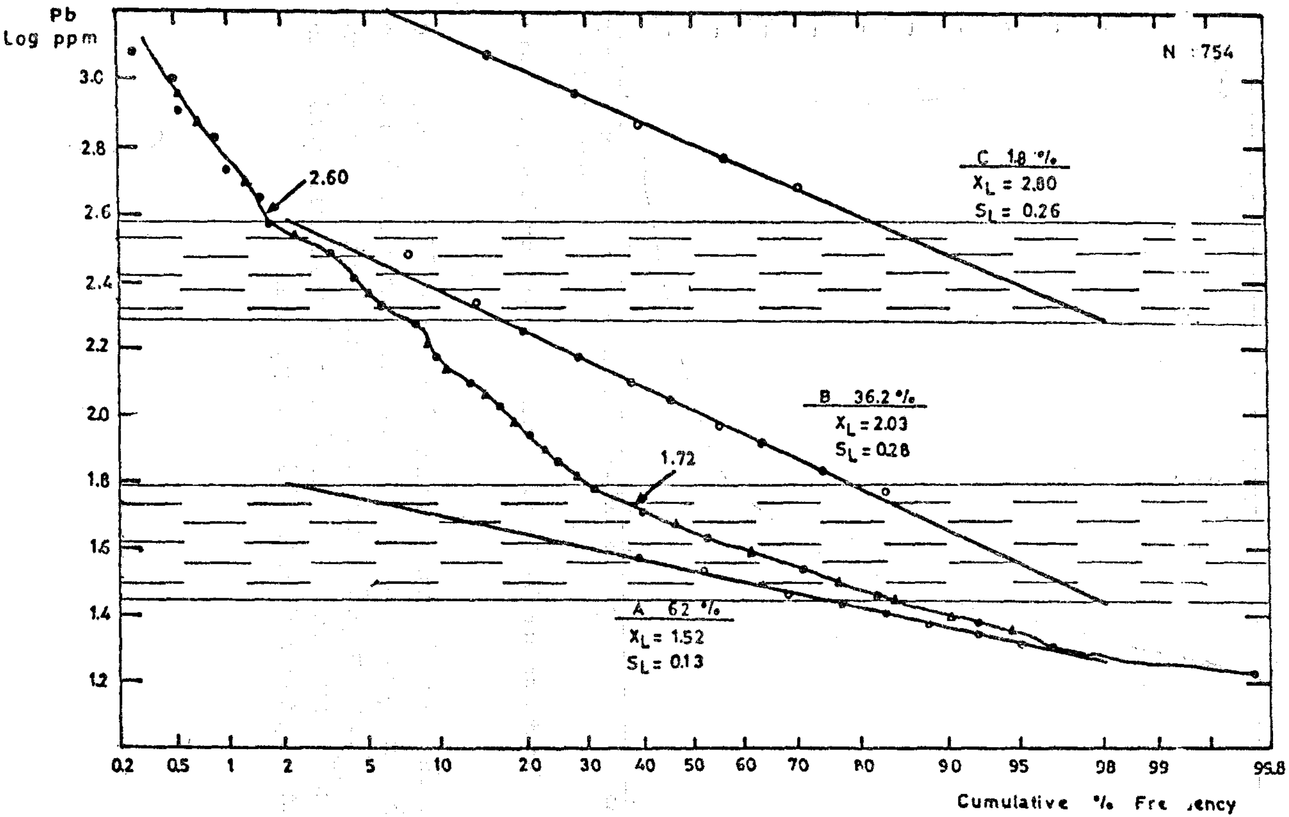

Figure 4. Plot of logarithmic class intervals aganist cumulative percentile frequency values of all data.

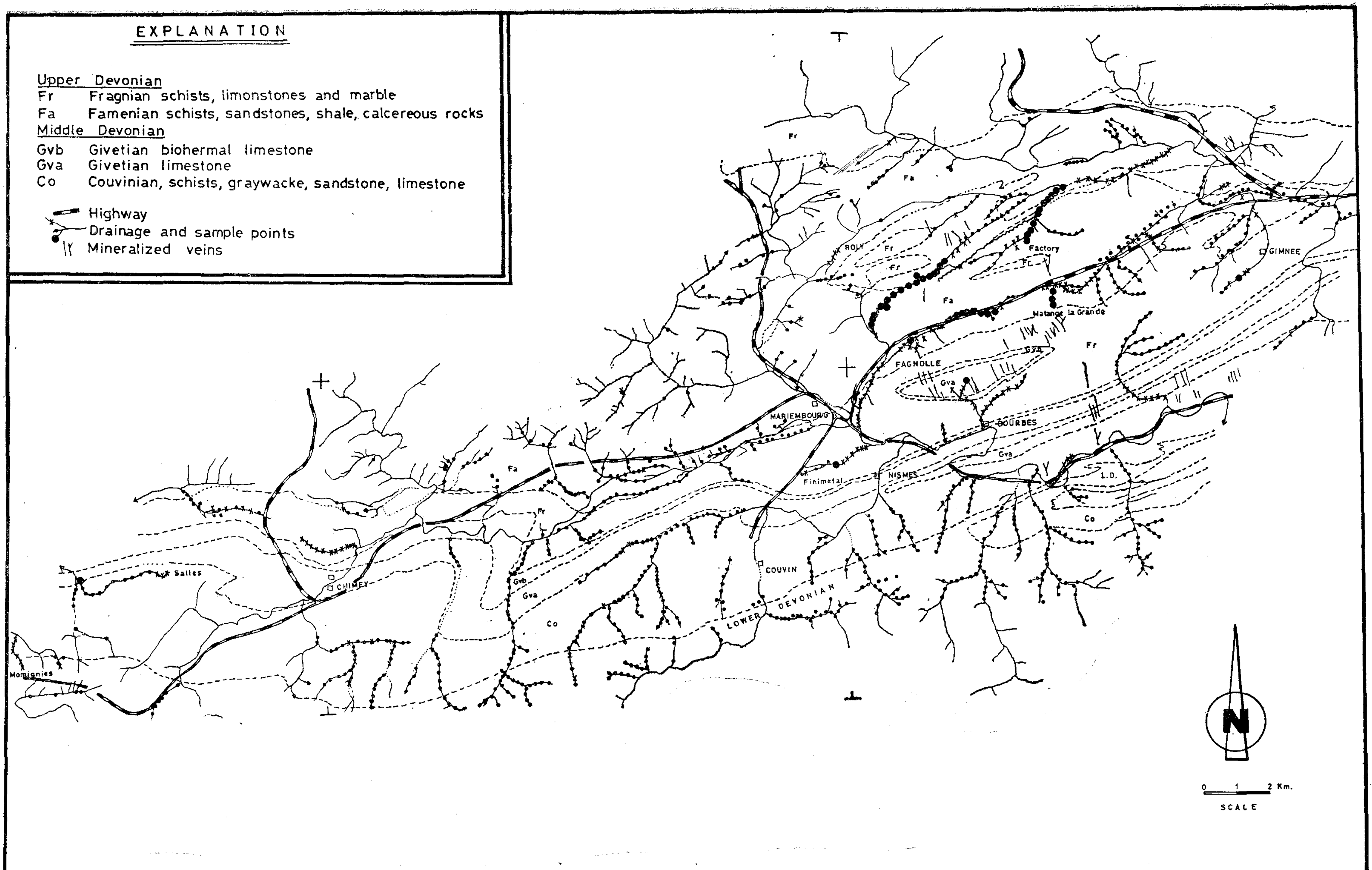

Figure 5. Discrete map. Note the clustering of the anomalous samples in the northeastern part of the study area.

SUMMARY

Sinclair's method of separation of statistical populations is successfully applied on the stream sediment data from the Belgian Ardennes.

The partitioning procedure included the following stages in sequence:

(a) drawing a smooth curve through plotted data points,

(b) picking the inflection points,

(c) partitioning all the populations by applying the relationships given in equations 1 to 3,

(d) checking ideal mixtures of partitioned populations with the original curve (main curve describing the real data)

(e) estimateing threshold values from partitioned straight lines, and

(f) based on the threshold values preparation of a discrete map.

This application showed that the data included one background and more than one anomalous populations.

ACKNOWLEDGEMENTS

We are grateful to Prof. Dr. Mümin Köksoy from Hacettepe University who introduced the Method of Sinclair first time to us.

REFERENCES

ARAL, H., 1987 (a) Evaluation of geochemical stream desediment data from the Belgian Ardennes: COMMUNICATIONS, c. 5, n. 1–2, 47–67, University of Ankara Faculty of Sciences, Ankara.

ARAL, H., 1987 (b). Application of two–population discriminant analysis method to geochemical stream sediment data from the Belgian Ardennes: COMMUNICATIONS, v. 5, n. 1–2, 67–95, University of Ankara Faculty of Sciences, Ankara.

SİNCLAİR, A.J., 1976. Application of probability graphs in mineral exploration. The , Association of Exploration Geochemists. Special Volume No. 4, 95 pages.