

A
THESIS

entitled

APPLICATION OF DISCRIMINANT AND CLUSTER
ANALYSIS TO REGIONAL GEOCHEMICAL SURVEYS

Submitted for the
degree of

DOCTOR OF PHILOSOPHY

in the
FACULTY OF SCIENCE OF THE UNIVERSITY OF LONDON

by

ROLANDO CASTILLO-MUÑOZ

Royal School of Mines,
Imperial College.

January, 1973

TO:

MARIA EUGENIA,

Olga Eugenia,
Rolando Jose and
Ana Lucia

ABSTRACT

Discriminant and cluster analysis have been applied to the interpretation of the geochemistry of 2440 regional reconnaissance stream-sediment samples from Denbighshire, Derbyshire and Devonshire

Problems investigated include: 1) classification of large data sets into "natural" groups related to the main geological features; 2) determination of the usefulness of these methods in routine geochemical survey interpretation; 3) selection of the "best" elements and "training" information for performing these tasks.

Empirical discriminant function (EDF) analysis was applied to the whole data sets based on small "training sets" representing the main geological formations using a priori information. Cluster-seeking methods sought to distinguish groups of samples related to the geology solely on their chemical composition. In both cases quantitative assessment of the results was based on analysis of correct sample classifications.

Both techniques have shown promise as tools for the delineation of broad-scale geological features, mineralization, contamination, and unusual secondary environment conditions.

The problem of large data sets was overcome by the use of "linkage samples" to achieve a general result from those obtained with subsets of data. Cluster separability was subsequently corroborated with EDF analysis. A "majority-vote" smoothing technique was applied to minimize noisy patterns in the original classification maps.

It is suggested that cluster and EDF analysis are used as complementary tools to assist in the interpretation of large multi-element regional geochemical data.

ACKNOWLEDGEMENTS

This study forms part of an Applied Geochemistry Research Group programme investigating the general utility of computer methods in the interpretation of geochemical survey results under the general direction of Professor J.S. Webb and supported by the Natural Environment Research Council.

Grateful thanks are due to the Ministry of Overseas Development of Great Britain, the Ministerio de Economia, Industria y Comercio de Costa Rica and the Universidad de Costa Rica for providing the financial support for the writer.

The author particularly wishes to thank Dr. Richard J. Howarth under whose supervision this study was carried out, and to Professor John S. Webb for his cooperation.

My special gratitude and love go to my wife Maria Eugenia for her encouragement and help; also to my parents, relatives and friends who interceded for the accomplishment of my studies.

Thanks are also given to Dr. E.A. Edmonds (I.G.S., Exeter Office) and Dr. E.B. Selwood (Exeter University) for geological information concerning the Devonshire area.

Computer time on the CDC 6400 and 6600 machines was made available by the Imperial College and the University of London Computer Centres.

LIST OF TABLES

<u>Table No.</u>	<u>Title</u>	<u>Page</u>
CHAPTER 1		
1-1	Analytical precision (%) calculated from duplicate analysis of drainage sediments	8
1-2	Sampling and analytical precision (%) calculated from duplicate analysis of drainage sediments	9
CHAPTER 3		
3-1	Range and mean content of minor elements in stream sediments associated with major geological formations, Denbighshire	45
3-2	Training area boundaries for bedrock classes, Denbighshire	54
3-3	Table of average training area compositions, Denbighshire	55
3-4	Classification of 565 stream-sediment samples in bedrock category testing set, EDF untransformed data, Denbighshire	59
3-5	Training area boundaries for mineralization classes, Denbighshire	61
3-6	Classification of 659 stream-sediment samples based on bedrock and mineralization training sets. EDF untransformed data, Denbighshire	62
3-7	Percentage of testing samples assigned to each class, EDF log-10 transformed data, Denbighshire	63
3-8	Classification of 659 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data, Denbighshire	66

<u>Table No.</u>	<u>Title</u>	<u>Page</u>
3-9	Mean metal values of stream sediment samples collected in areas geochemically similar according to the EDF analysis, Denbighshire (Spectrographic analysis)	67
3-10	Mean metal values of 3 limestone and 3 siltstone samples, Clwydian Range, Denbighshire, (Spectrographic analysis)	68
3-11	Classification of 681 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data, Denbighshire	71
3-12	Classification of 585 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data, Denbighshire	74
3-13	Classification of 594 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data, Denbighshire	76
3-14	Classification of 715 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data, Denbighshire	78
3-15	Arithmetic means and standard deviations, Denbighshire, dendrogram representation, untransformed data	88
3-16	Group classification results by EDF analysis, Denbighshire, dendrogram representation, untransformed data results	90
3-17	Arithmetic means and standard deviations, Denbighshire, dendrogram representation, log-10 transformed data	96
3-18	Group classification results by EDF analysis, Denbighshire, dendrogram representation, log-10 transformed results	97
3-19	Arithmetic means and standard deviations, Denbighshire, NLM algorithm, untransformed data	106

<u>Table No.</u>	<u>Title</u>	<u>Page</u>
3-20	Group classification results by EDF analysis, Denbighshire, NLM algorithm, untransformed results	111
3-21	Comparative successful recognition rates (%) of the "anomalous" samples in relation to different geochemical environments, Denbighshire	121
CHAPTER 4		
4-1	Range and mean metal content of stream sediments derived from the principal geological formations, Devonshire	127
4-2	Average training area compositions, Devonshire	134
4-3	Classification of 910 stream-sediment samples based on bedrock training set only, EDF untransformed data, Devonshire	135
4-4	Classification of 894 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data, Devonshire	136
4-5	Arithmetic means and standard deviations, Devonshire, dendrogram representation, untransformed data	140
4-6	Chemical composition (spectrographic analysis) of two Westphalian and Namurian sandstones collected south and southwest of Bude	142
4-7	Group classification results by EDF analysis, Devonshire, dendrogram representation, untransformed data results	146
4-8	Arithmetic means and standard deviations, Devonshire, dendrogram representation, log-10 transformed data	152

<u>Table No.</u>	<u>Title</u>	<u>Page</u>
4-9	Group classification results by EDF analysis, Devonshire, dendrogram representation, log-10 transformed data results	153
4-10	Arithmetic means and standard deviations, Devonshire, NLM algorithm, untransformed data	163
4-11	Group classification results by EDF analysis, Devonshire, NLM algorithm, untransformed data results	167
4-12	Comparative successful recognition rates (%) of the "anomalous" samples in relation to different geochemical environments, Devonshire	177
CHAPTER 5		
5-1	Range and mean metal content of stream sediments derived from the principal geological formations	182
5-2	Training area boundaries for bedrock classes	188
5-3	~~Average training area compositions, Derbyshire~~	189
5-4	Classification of 173 stream-sediment samples in bedrock category testing set, EDF untransformed data	193
5-5	Percentage of testing samples assigned to each class, log-10 EDF analysis, Derbyshire	196
5-6	Classification of 262 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data	198
5-7	Classification of 276 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data	199
5-8	Classification of 295 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data	201

<u>Table No.</u>	<u>Title</u>	<u>Page</u>
5-9	Classification of 270 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data	203
5-10	Arithmetic and standard deviations, Derbyshire, dendrogram representation, untransformed data	208
5-11	Group classification results by EDF analysis, Derbyshire, dendrogram representation, untransformed data results	209
5-12	Arithmetic and standard deviations, Derbyshire, dendrogram representation, log-10 transformed data	213
5-13	Group classification results by EDF analysis, Derbyshire, dendrogram representation, log-10 transformed data results	214
5-14	Arithmetic means and standard deviations, Derbyshire, NLM algorithm, untransformed data	221
5-15	Group classification results by EDF analysis, Derbyshire, NLM algorithm, untransformed data results	222
5-16	Comparative successful recognition rates (%) of the "anomalous" samples in relation to different geochemical environments	233
CHAPTER 6		
6-1	Overall correct recognition rates (%) of the EDF analysis applied to the cluster-seeking results	238
6-2	Mean metal values of 6 shale and 13 sandstone samples of the Namurian (Crackington) Formation, Devonshire (Spectrographic analysis)	243

LIST OF FIGURES

<u>Figure No.</u>	<u>Title</u>	<u>Page</u>
CHAPTER 1		
1-1	Location of study areas	6
CHAPTER 2		
2-1	Selected points from the Hertzsprung-Russell diagram for those stars observable with modern telescopes.	13
2-2	Behaviour of the overall recognition success rate for the testing set according to different combinations of variables	16
2-3	Non-linearly separable classes	18
2-4	Interpolated one-dimensional probability density function from set of five training samples with increasing values of smoothing parameter	20
2-5	Two-dimensional distance measurement d , for samples i and j in (x,y) space	24
2-6	Mount Shoobridge granite, Northern Territory of Australia, Dendrogram for samples from the Mount Shoobridge granite	27
2-7	Mount Shoobridge granite Q-mode dendrograms derived from data transformed by dividing by the maximum value	29
2-8	Mount Shoobridge granite Q-mode dendrograms derived from data transformed to range 0.0-1.0	30

<u>Figure No.</u>	<u>Title</u>	<u>Page</u>
2-9	Mount Shooobridge granite R-mode dendrograms derived from raw data and transformed by dividing by the maximum value, and to range 0.0-1.0	31
2-10	Mount Shooobridge granite, NLM algorithm, untransformed data	34
2-11	Mount Shooobridge granite, NLM algorithm, normalized data	35
2-12	Mount Shooobridge granite, NLM algorithm, log-10 transformed data	36
2-13	Mount Shooobridge granite, Factor Analysis untransformed data (First two factors only)	37
2-14	Mount Shooobridge granite, NLM algorithm, untransformed data from uncorrelated variables from factor analysis	38
CHAPTER 3		
3-1	Drainage system and sampling localities, Denbighshire	41
3-2	Major soil groups of the western half of Denbighshire	43
3-3	Distribution of the principal geological units, Denbighshire	44
3-4	Distribution of principal mineral veins and known smelting sites, Denbighshire	49
3-5	Variation in overall classification success rate for testing samples with increase in smoothing parameter (σ), Denbighshire	56
3-6	Map of classification results, Denbighshire, discriminant analysis, untransformed data	57
3-7	Map of classification results, Denbighshire, discriminant analysis, log-10 transformed data	64

<u>Figure No.</u>	<u>Title</u>	<u>Page</u>
3-8	Map of classification results, Denbighshire, discriminant analysis, log-10 transformed data	73
3-9	Basic and subsequent dendrograms, Denbighshire, untransformed data	82
3-10	Original map of classification results, Denbighshire, dendrogram representation, untransformed data	86
3-11	Smoothed map of classification results, Denbighshire, dendrogram representation, untransformed data	87
3-12	Map of the classification results of "anomalous" samples, Denbighshire, dendrogram representation, untransformed data	92
3-13	Original map of classification results, Denbighshire, dendrogram representation, log-10 transformed data	94
3-14	Smoothed map of classification results, Denbighshire, dendrogram representation, log-10 transformed data	95
3-15	Diagram of classification results, Denbighshire, NLM algorithm, untransformed data	105
3-16	Original map of classification results, Denbighshire, NLM algorithm, untransformed data	108
3-17	Smoothed map of classification results, Denbighshire, NLM algorithm, untransformed data	109

<u>Figure No.</u>	<u>Title</u>	<u>Page</u>
3-18	Diagram of classification results, Denbighshire, NLM algorithm, log-10 transformed data	114
3-19	Diagram of the classification results of "anomalous" samples, Denbighshire, NLM algorithm, log-10 transformed data	116
3-20	Diagram of classification results, Denbighshire, NLM algorithm, normalized data	118
CHAPTER 4		
4-1	Drainage system and sampling localities, Devonshire	124
4-2	Distribution of the principal geological units, Devonshire	126
4-3	Location of mineral deposits, Devonshire	131
4-4	Original map of classification results, Devonshire, dendrogram representation, untransformed data	137
4-5	Smoothed map of classification results, Devonshire, dendrogram representation, untransformed data	138
4-6	Schematic basic dendrograms, Devonshire, untransformed and log-10 transformed data	139
4-7	Original map of classification results, Devonshire, dendrogram representation, log-10 transformed data	149
4-8	Smoothed map of classification results, Devonshire, dendrogram representation, log-10 transformed data	150

<u>Figure No.</u>	<u>Title</u>	<u>Page</u>
4-9	Diagram of classification results, Devonshire, NLM algorithm, untransformed data	160
4-10	Original map of classification results, Devonshire, NLM algorithm, untransformed data	161
4-11	Smoothed map of classification results, Devonshire, NLM algorithm, untransformed data	164
4-12	Diagram of classification results, Devon- shire, NLM algorithm, log-10 transformed data	169
4-13	Diagram of the classification of "anomalous" samples, Devonshire, NLM algorithm, log-10 transformed data	171
4-14	Diagram of classification results, Devon- shire, NLM algorithm, normalized data	173
CHAPTER 5		
5-1	Drainage system and sampling localities, Derbyshire	179
5-2	Distribution of the principal geological units, Derbyshire	181
5-3	Distribution of principal mineral veins and known smelting sites, Derbyshire	186
5-4	Variation in overall classification success rate for testing samples with increase in smoothing parameter (σ), Derbyshire	191

<u>Figure No.</u>	<u>Title</u>	<u>Page</u>
5-5	Map of classification results, Derbyshire, discriminant analysis, untransformed data	192
5-6	Map of classification results, Derbyshire, discriminant analysis, log-10 transformed data	197
5-7	Original map of classification results, Derbyshire, dendrogram representation, untransformed data	206
5-8	Schematic basic dendrograms, Derbyshire, untransformed and log-10 transformed data	207
5-9	Original map of classification results, Derbyshire, dendrogram representation, log-10 transformed data	212
5-10	Diagram of classification results, Derbyshire, NLM algorithm, untransformed data	219
5-11	Original map of classification results, Derbyshire, NLM algorithm, untransformed data	220
5-12	Diagram of classification results, Derbyshire, NLM algorithm, log-10 transformed data	224
5-13	Diagram of classification results, Derbyshire, NLM algorithm, normalized data	225
5-14	Diagram of the classification of "anomalous" samples, Derbyshire, NLM algorithm, log-10, transformed data (12 variables)	227
5-15	Diagram of the classification of "anomalous" samples (mineralized and contaminated areas) Derbyshire, NLM algorithm, log-10 transformed data (12 variables)	228

<u>Figure No.</u>	<u>Title</u>	<u>Page</u>
5-16	Diagram of the classification of "anomalous" samples (mineralized and contaminated areas) Derbyshire, NLM algorithm, log-10 transformed data (7 variables)	229
APPENDIX 1		
1	Extrapolation effect of two unimodal probability density functions for categories Ω_1 and Ω_2 with increase of σ	251

CONTENTS

	Page
ABSTRACT	i
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	iii
LIST OF FIGURES	viii
CONTENTS	xv
 CHAPTER 1 INTRODUCTION	
1-1 General statement	1
1-2 Aims of this study	4
1-3 Study areas and geochemical data	5
 CHAPTER 2 INTRODUCTION TO DISCRIMINANT AND CLUSTER ANALYSIS TECHNIQUES	
2-1 Introduction	10
2-2 Discriminant analysis	14
i) Introduction	14
ii) Methods	17
2-3 Cluster analysis	19
i) Average linkage clustering	19
ii) Non-linear mapping algorithm	23
iii) Hierarchical cluster analysis of the Mount Shooobridge granite. Difficult- ies in interpretation. Comparative results.	25
 CHAPTER 3 DISCRIMINANT AND CLUSTER ANALYSIS OF DENBIGHSHIRE DATA	
3-1 Location and description of the study area	40
i) Geography	40
ii) Geology and Regional Geochemistry	42
3-2 Empirical discriminant analysis	51
i) Introduction	51
ii) Results	52
a) Classification based on the untransformed data	52
b) Classification based on the log-10 transformed data	60

	Page
iii) Discussion	77
iv) Conclusions	80
3-3 Hierarchical cluster analysis:	81
i) Method	81
ii) Q-Mode results	84
a) Classification based on the untransformed data	85
b) Classification based on the log-10 transformed data	93
iii) R-mode results	100
iv) Discussion	101
3-4 Non linear mapping algorithm (NLM)	103
i) Method	103
ii) Results	104
a) Classification based on the untransformed data	104
b) Classification based on the log-10 transformed data	113
c) Classification based on the normalized data	117
iii) Discussion	117
3-5 Conclusions: Hierarchical cluster analysis and non-linear mapping algorithm	120
CHAPTER 4	DISCRIMINANT AND CLUSTER ANALYSIS OF DEVONSHIRE DATA
4-1 Location and description of the study area	123
i) Geography	123
ii) Geology and Regional Geochemistry	125
4-2 Empirical discriminant analysis	132
4-3 Hierarchical cluster analysis	133
i) Q-mode results	133
a) Classification based on the untransformed data	133
b) Classification based on the log-10 transformed data	148
ii) R-mode results	156
iii) Discussion	157

	Page
4-4 Non-linear mapping algorithm (NLM)	159
i) Results	159
a) Classification based on the Untransformed data	159
b) Classification based on the log-10 transformed data	168
c) Classification based on the normalized data	172
ii) Discussion	172
4-5 Conclusions: Hierarchical cluster analysis and non-linear mapping algorithm	175
 CHAPTER 5 DISCRIMINANT AND CLUSTER ANALYSIS OF DERBYSHIRE DATA	
5-1 Location and description of the study area	178
i) Geography	178
ii) Geology and Regional Geochemistry	180
5-2 Empirical discriminant analysis	185
i) Introduction	185
ii) Results	190
a) Classification based on the untransformed data	190
b) Classification based on the log-10 transformed data	195
iii) Discussion	202
iv) Conclusions	204
5-3 Hierarchical cluster analysis	205
i) Q-mode results	205
ii) R-mode results	215
iii) Discussion	216
5-4 Non-linear mapping algorithm (NLM)	217
i) Results	217
ii) Discussion	230
5-5 Conclusions: Hierarchical cluster analysis and non-linear mapping algorithm	231

	Page
CHAPTER 6 GENERAL CONCLUSIONS AND RECOMMENDATIONS	
6-1 Introduction	234
6-2 General conclusions	234
i) Empirical discriminant analysis	234
ii) Hierarchical cluster analysis (dendrogram representation) and non-linear mapping algorithm	235
6-3 Recommendations	245
APPENDIX 1 THE EMPIRICAL DISCRIMINANT FUNCTION	249
APPENDIX 2 THE NON-LINEAR MAPPING ALGORITHM	253
REFERENCES	255

CHAPTER 1

INTRODUCTION

1-1 General statement.

The most difficult problem faced by an applied geochemist is the interpretation of large quantities of element determinations which could be derived from studies in the fields of mineral exploration, agriculture, medical geography and marine geochemistry. The need for computer aided solutions to assist in the interpretation of a complex multi-element data set is acute. Many methods suitable for use with small data sets are unsuitable for those which might involve upwards of 1000 samples. The regional geochemist could be concerned with data sets in the order of 10,000 samples per year for which perhaps 3 to 30 element determinations per sample are obtained. Clearly, the human mind can not comprehend more than the most obvious features of such data.

In this study multi-element regional geochemical data derived from stream sediments is evaluated. The basic idea in the use of drainage sediments is that they are composed of both the mechanically dispersed products of bedrock weathering (incorporated into the streams by surface runoff waters) and the soluble products (leached from the bedrock weathering products and which exist at some stage as solutes in the soil or drainage waters). The drainage sediments are thus representing all geological (lithological) features occurring in the catchment area of a stream. Stream sediments used in regional geochemistry are usually sampled from the tributary drainage since this restricts the size of the catchment area and thus allows the data to be more easily interpreted.

The advantage of the stream sediment material is that it enables large areas to be geochemically surveyed at a minimum cost because of the relatively low sample density and rapidity of sampling. The disadvantage of this method is that the data can be

rather more difficult to interpret than that derived from rock, soil and vegetation because: 1) the drainage sediment is not only influenced by the bedrock materials but also by the overburden, the secondary environment and by any sort of artificial and natural contamination (e.g. the presence of smelting sites, industrial and domestic activities, glacial drift, etc., and 2) the bedrock generally suffers complex geochemical changes before it is transformed to drainage sediment and, therefore, the relationship between the bedrock and stream sediment geochemistry may be considerably modified.

The aims of a regional geochemical survey are:

- 1) to determine the areal distribution of each of the elements considered, their mutual relationships, and the mechanisms governing their distribution;
- 2) to determine whether there is any relationship between the different geological features (e.g. bedrock, mineralization, secondary environment, etc.) and the distribution of each one of the elements or associations of elements in geologically "well-understood" areas;
- 3) the recognition of geochemical patterns that might have some connection with mappable geological features in geologically unknown (virgin) or poorly known areas; and
- 4) to delineate those areas with good possibilities of mineralization or other important geochemical patterns for subsequent follow-up surveys.

In very rugged and highly promising virgin areas of Central America, particularly Costa Rica, where this author has been involved for several years in geological mapping, major difficulties arise in regional mapping and mineral exploration. The exuberant tropical vegetation and the deeply weathered overburden with a thick cover of organic material (humus) prevents the efficient determination of the "fresh" lithology along crests and flanks of the mountain ridges; hence regional geological mapping and mineral exploration using multi-element stream sediment geochemistry could be more practicable when detailed lithologic observations and stream

sediment sampling at broad regular intervals (e.g. one sample each 1 or 2 Kms or more) are both carried out simultaneously on streams flowing along generally "fresh" exposures at the bottom of steep and deep valleys which are commonly difficult to penetrate. This information used in addition to some photogeologic control (e.g. regional physiographic features) could be a valuable aid in "pattern recognition" methods for mapping the regional geology and possible delineation of mineral target areas, thus avoiding long, dangerous, and sometimes impossible traverses along streams with lithologies that would have little or no effect on the final form of a regional geological map and mineral exploration features. Some of the rock sequences of these regions consist of thin to medium bedded interstratified sandstones, siltstones and shales; acid igneous facies of granite intrusions; volcanic rocks of different periods of vulcanism consisting of interbedded lavas, tuffs, agglomerates, etc., that could be perfectly well broadly mapped by means of the stream sediment geochemistry.

Current research carried out by the Applied Geochemistry Research Group, has enabled the development of computer-based methods suitable for multi-element data interpretation and display. The initial results of this "pattern recognition" approach using methods of discriminant and cluster analysis have proved extremely encouraging.

It is highly desirable that we now attempt to show the feasibility of these methods for use by the geochemist in the interpretation of any routine geochemical survey by evaluation of these techniques using "well understood" data sets. It has been said (Wilkinson, 1970) that "our ability to produce good quality geochemical data has outstripped our facilities for processing and making optimum use of the data". The further development of computer aided methods of data interpretation is regarded as an essential step in the expansion of research in applied geochemistry.

1-2 Aims of this study.

The aims of the research are to investigate methods for:

- 1) obtaining the optimum classification of data into groups of samples¹ as far as possible related to the main geological formations as an attempt to map geological and mineral exploration features on the basis solely of the chemical composition of the derived stream sediments,
- 2) the use of geological knowledge in addition to the geochemical data. Results based on areas of known geology could then be extrapolated into adjoining areas with relatively little field control;
- 3) obtaining the best results for classification of the geochemical data when no geological knowledge is available. The chemical composition of the samples is used alone without taking into consideration the a priori geological knowledge. This approach will have significant application to undeveloped areas.
- 4) the main goal in the pattern recognition task as directed towards geological mapping and mineral exploration, is the discrimination between bedrock and mineralization classes.

The main problems encountered in trying to achieve these aims are:

- 1) the size of the data sets. The large data sets will need to be processed in the minimum amount of computer time and with maximum efficiency (a typical commercial rate is at least £50/hr.);
- 2) the selection of the "best" variables (elements) for showing up the natural multi-element structure of the data in order to obtain the most representative geological and geochemical results;
- 3) the selection of "typical" areas to represent the main geological formations as "training" information which is supplied to those

¹What is meant by a group of samples ("natural groups") is discussed in more detail in Chapter 2.

statistical techniques requiring a priori geological knowledge.

1-3 Study areas and geochemical data.

This evaluation has been based on regional geochemical data from three areas of the United Kingdom (Denbighshire, Devonshire and Derbyshire) covering a total surface area of approximately 3.000 square miles (Fig.1-1). These test areas were chosen because their basic geological and geochemical features are already "well understood" from earlier works (Horsnail, 1968; Khaleelee, 1969; Nichol and others, 1970a, 1970b, 1971) which have been used as the major sources of reference for the background geological and geochemical information in chapters 3, 4 and 5. These areas also show a variety of geological features which allow comparative evaluation of the performance of each statistical method. The Denbighshire and Devonshire areas have well defined broad geological formation and localised contamination from mining activities (and glacial drift in Denbighshire) and drainage systems which in general favour transportation of similar geological materials, and constitute suitable areas for the evaluation of the methods in relatively favourable conditions. On the other hand, the Derbyshire area has a more complicated spatial distribution with more rapid changes of mapped formations, widespread contamination (mining, industrial, glacial drift) over large parts of the area, and a drainage system which mainly dissects different geological formations as well as transporting contaminated detritus. This area offers ideal material to test the behaviour of these statistical methods in relatively unfavourable conditions.

A knowledge of the regional distribution of the elements and specific associations, and their correlation with the main geological features were required to carry out an evaluation of the statistical methods under study in order to know how best to apply them subsequently in virgin areas.

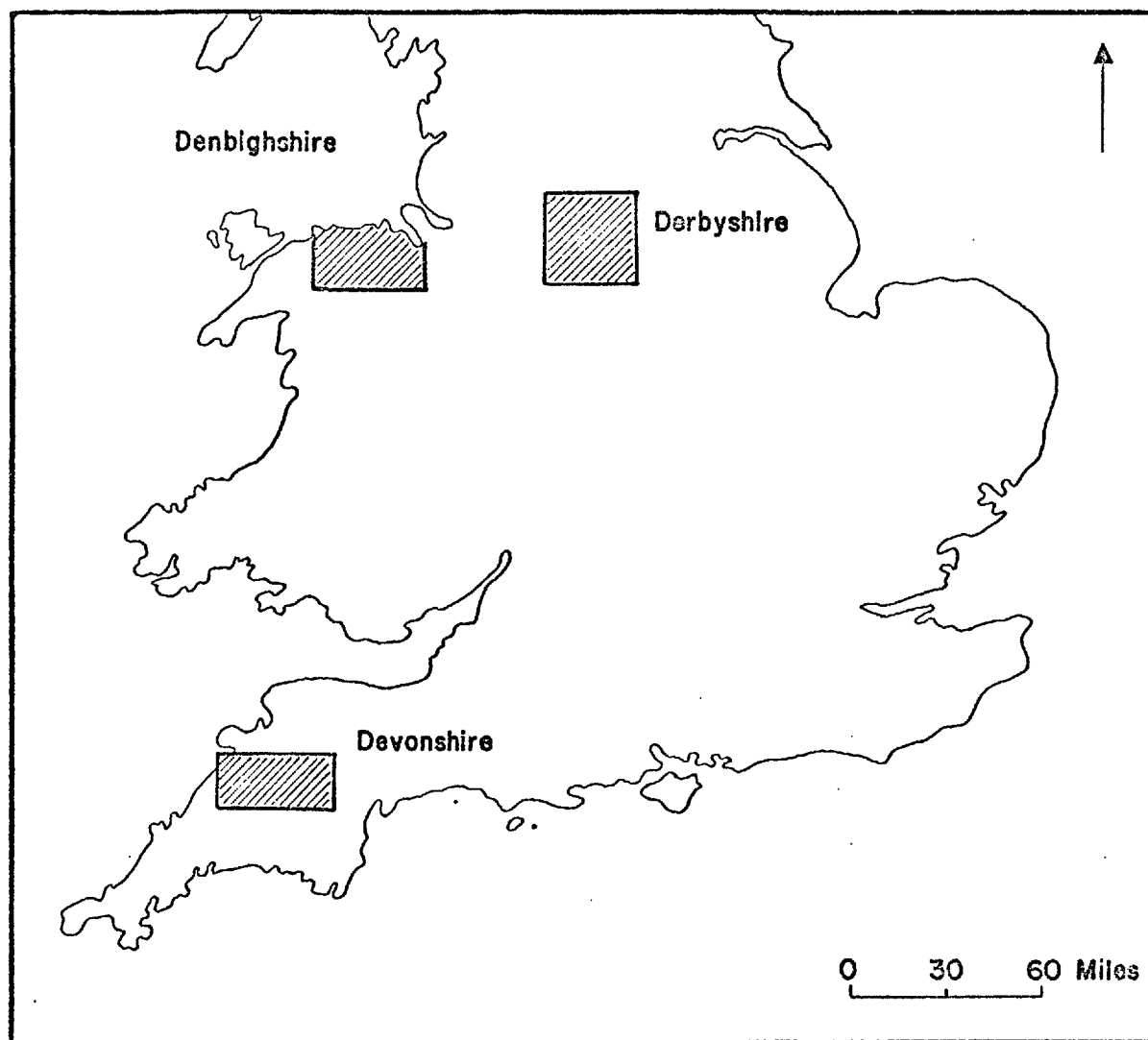


FIG. I-1
LOCATIONS OF STUDY AREAS

The geochemical data are represented by multi-element analysis of 2440 stream sediment samples collected by members of the Applied Geochemistry Research Group at road/track and stream intersections, at a mean density of one sample per square mile. The -80 mesh (c.200 μ) fraction of the oven-dried samples was analysed spectrographically for Pb, Sn, Ga, V, Mo, Cu, Ti, Ni, Co, Mn, Ag, Cr, and Fe_2O_3 , and colorimetrically for As and Zn; mean element precisions at the 95 per cent confidence level are $\pm 50\%$ and $\pm 25\%$ for spectrographic and colorimetric determinations, respectively (Nichol and others, 1970-71). The analytical precision (Table 1-1) and the combined sampling and analytical precision (Table 1-2) in stream sediments were previously determined by Khaleelee (1969) from duplicate analysis. It is evident that the sampling error will in most cases account for a large proportion of the data variability.

The problems usually found by previous workers when interpreting multi-element regional geochemical data (particularly in areas of the United Kingdom) have been principally related to the influence of contamination caused by past mining and smelting activities and to regional variations in the secondary environment. Contamination has originated from the presence in the stream sediments of anomalously high levels of elements associated with mineral extraction. The effect of contamination is to obscure the "natural" bedrock composition of the detritus for the contaminated elements. The secondary environment is commonly controlled by acid and chemically reducing poorly drained soils which facilitate the leaching of the mobile elements (Fe, Mn, Ni, Co, Zn, Cu and As) from these soils by percolating waters. Drainage channels with higher Eh and pH levels than those in the soil waters tend to precipitate manganese and iron and simultaneously to coprecipitate the rest of the mobile elements, leading to anomalously high levels of these elements in the drainage channels of stream derived from water logged soil areas. The effect of this enhancement is again to obscure the trace element concentration of the detritus derived from the bedrock of these elements.

TABLE 1 - 1

Analytical precision (%)* calculated from duplicate analysis of drainage sediments (After Khaleelee, 1969, table 33).

Element	Mean	Standard deviation
Ag	1.0	3.3
As	23.9	26.3
Co	3.2	3.6
Cr	4.5	8.7
Cu	7.1	5.8
Fe	5.6	4.5
Ga	3.4	3.5
Mn	8.0	7.4
Mo	7.9	9.8
Ni	3.6	3.7
Pb	8.3	6.9
Sn	15.0	23.3
Ti	4.0	2.9
V	4.8	4.2
Zn	12.0	15.0
NO OF DUPLICATES	11	

* Defined by Khaleelee (1969, p.342) as the relative difference between two duplicate determinations x_1 and x_2 :

$$\left(\frac{1}{2}(x_1 - x_2)\right) / \left(\frac{1}{2}(x_1 + x_2)\right) \times 100 \text{ per cent}$$

TABLE 1 - 2

Sampling and analytical precision (%) calculated
from duplicate analysis of drainage sediments
(After Khaleelee, 1969, table 34).

Elements	Mean	Standard deviation
Ag	5.0	17.3
As	34.2	21.1
Co	16.5	15.2
Cr	26.4	23.0
Cu	17.6	13.8
Fe	24.7	19.7
Ga	15.7	9.3
Mn	27.4	20.0
Mo	27.3	18.7
Ni	23.5	16.8
Pb	27.5	16.8
Sn	5.0	9.1
Ti	13.6	10.9
V	19.5	23.9
Zn	34.7	25.1
NO OF DUPLICATES	12	

CHAPTER 2

INTRODUCTION TO DISCRIMINANT AND CLUSTER ANALYSIS TECHNIQUES

2-1 Introduction

In the discussion which follows we can distinguish between two different aspects of treatment of the regional geochemical survey data. In the first case, we have a priori knowledge of the classes in which we are interested and are concerned with the classification of the sample of unknown type, assigning it to one of these classes; this task is treated by discriminant analysis. In the second case a priori knowledge of the classes inherent in the data is not available, and we are concerned with data analysis, using cluster-seeking methods in an attempt to identify the groups of samples present, on the basis of the measurements on each sample. Fundamental to both approaches is the nature of a group, or cluster, which is discussed below.

In the biological sciences, Taxonomy is the theoretical study of classification, including its bases, principles, procedures and rules (Simpson, 1961, p.11). In the past most of the problems of classification were concerned with biological data, but the general principles of classification have been extended to treat other types of numerical data and the term Numerical Taxonomy has arisen. Sokal and Sneath (1963, p.48) define numerical taxonomy as the "numerical evaluation of the affinity between taxonomic units and the ordering of these units into taxa on the basis of their affinities".

Numerical Taxonomy is based on the ideas of Adamson. The following axioms (modified from Sneath (1958, p.5)) define his ideas:

- 1 - The ideal taxonomy is that in which the taxa have the greatest content of information and which is based on as many characters as possible.
- 2 - A priori, every character is of equal weight in creating natural taxa.
- 3 - Overall similarity (or affinity) between any two entities is a function of the similarity of the many characters in which they are being compared.

- 4 - Distinct taxa can be constructed because of diverse character correlations in the groups under study.
- 5 - Taxonomy as conceived by us is therefore a strictly empirical science.
- 6 - Affinity is estimated independently of phylogenetic considerations.

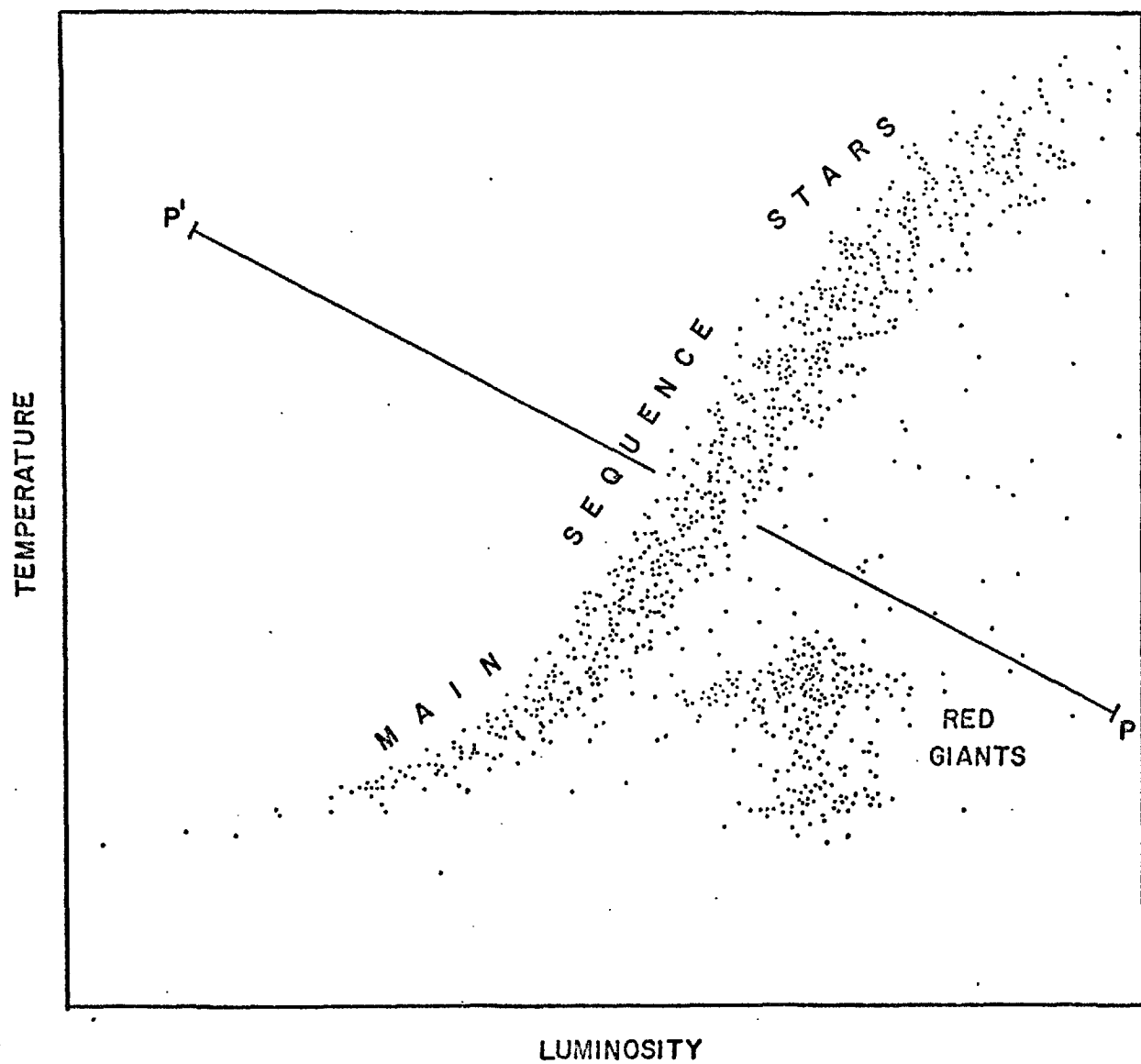
Once the numerical data have been computed the main problem arises whether the defined taxonomic units are really representative of the natural conditions, in other words, if they represent "natural" groups and not "arbitrary" groups as a product of human bias. In the field of biology, Danser (1950) realizes the difficulty of defining natural groups, but is not able to state any exact or scientific definition for them, ending with the hope that "some day systematics will arrive at a more exact stage, but this does not alter the fact that already now we are more intuitive but nevertheless scientific in manner". Simpson (1961, p.57) agrees that "in fact much of the theoretical discussion in the history of taxonomy has, beneath its impersonal language and objective facade, been an attempt to find some theoretical basis for these personal and subjective results".

Gilmour (1937, 1940, 1951, 1961b) has emphasized: "we could arrange living creatures in many ways but we chose one way because we think it is best for some purpose. If the purpose is restricted, then the classification is a special classification often called "arbitrary". Such a classification conveys less information than a general or "natural" one. Gilmour agrees that a "natural" taxonomy is the main goal that can be achieved for all scientists.

The first attempts to separate data into "natural" groups or clusters using numerical methods were carried out in the biological sciences using hierarchical clustering with dendrogram representation of the results.

Forgy (1963) pointed out the problems of defining "natural" groups in multi-dimensional space by saying that it is necessary to relate the ideas of mixture and "natural" groups to some aspect of the observed measurements. He visualized the naturalness of a group as a dense agglomeration of individual sample points distributed in n -dimensional (Euclidean) space (where n is the number of variables that define those individuals) separated by regions of thinner density containing few individuals. Each one of the dense groups corresponds to a population, each having a specific mode; a single natural group or cluster would therefore explain a uni-modal population. In order to detect the presence or absence of a natural group in the imaginary multi-dimensional space of the mathematician, Forgy proposed the use of a matrix of (Euclidean) distances between all pairs of individuals as the best point of departure, because this procedure gives all the information concerning the relative positioning of points in the space.

Subsequently, Forgy (1965) says that an efficient classification system for assigning all multivariate samples into natural groups implies compact, homogeneous classes. The best statistical partition on this basis is achieved by minimizing the within-group sum of squares. Referring to the Herzprung-Russell diagram (Fig.2-1) of the distribution of stars on the two variables of absolute luminosity and temperature, both you and the astronomer can see two obvious "natural" clusters or groups: the so-called "main sequence" stars and the "red giants", both groups separated by thinner mixtures of stars. The minimum within-group sum of squares for the best classification is produced by the "minimum variance partition" cutting artificially across the "natural" grouping. Thus, a clustering method suitable for objective classification purposes may not detect the "natural" grouping and vice versa. We must therefore aim to find cluster-seeking techniques which minimise this type of artificial "chopping-up" of the data.



Selected points from the Hertzsprung-Russell diagram for those stars observable with modern telescopes. The line PP' shows the minimum-variance partition into two classes. After Forgy (1965)

FIG. 2-1

The general principles applied to the classification of biological data can also be applied to the classification of geochemical data. Using geochemical data we have the advantage of always having numerical data, however, we may expect similar problems with the definition of natural geochemical groups.

In applied geochemistry the natural groups are directly (objectively) controlled by bedrock geology, mineralization, artificial contamination (e.g., the presence of smelting sites, industrial wastes, etc.), natural contamination (e.g., glacial drift) and the secondary environment. However, the same natural groups can be influenced subjectively by the choice of which variables are to be included in the data set, sampling and analytical errors, and the choice of the statistical methods used (e.g. use of different parameters), and so on. The main problem in the delimitation of natural geological and geochemical groups is therefore directly connected with the presence of human bias.

2-2 Discriminant analysis

i) Introduction

Discriminant analysis is a classification tool for use when we have a priori knowledge of the classes which we are interested in distinguishing between, and we aim to find a decision rule which discriminates well from these classes on the basis of the observed measurements. Application of this decision rule to an unknown data set will assign each sample to one of the categories purely on the basis of its measurements.

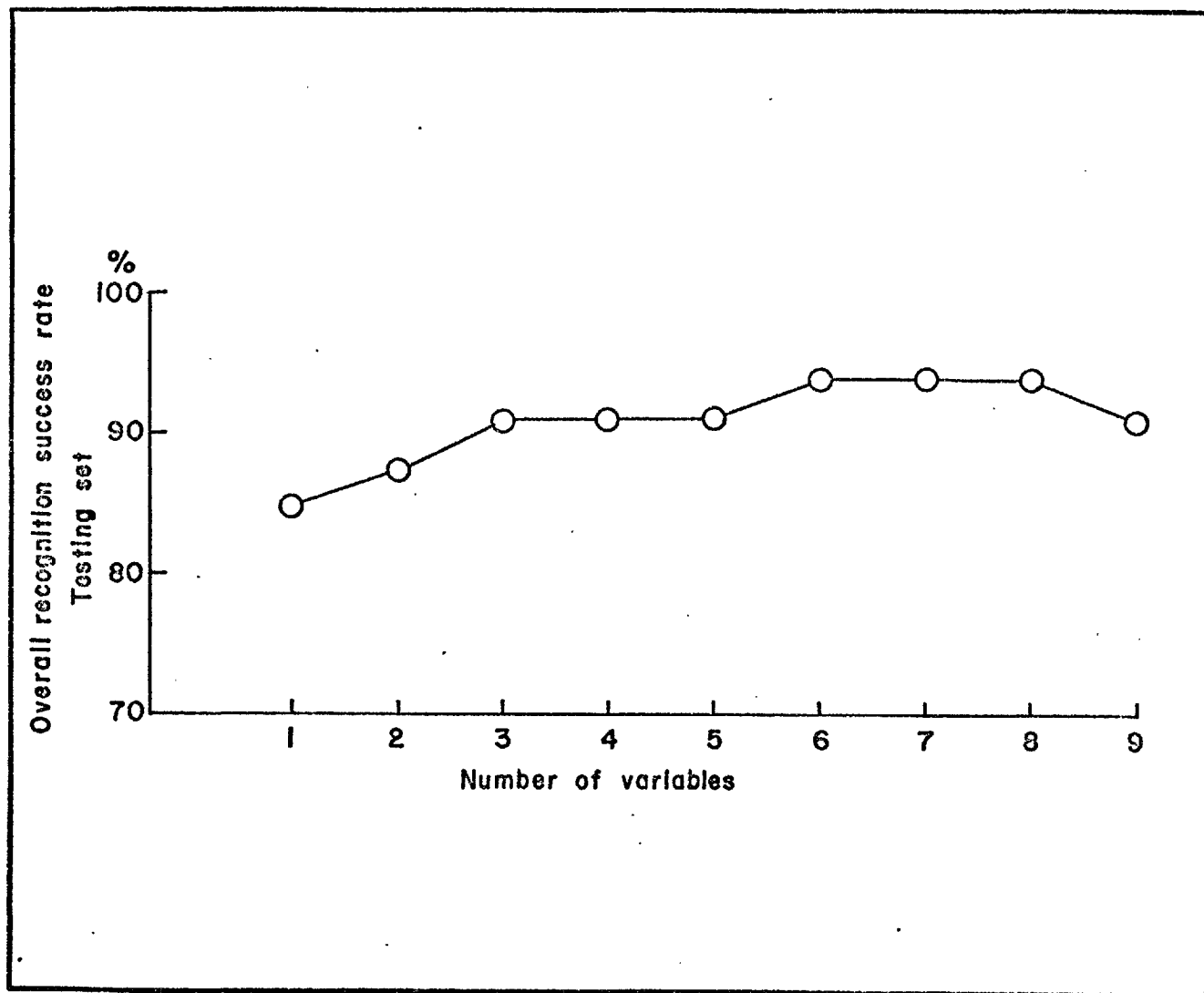
In general, the discriminant function $D_i(x)$ for the measurement set x associated with the i -th category is such that the value $D_i(x)$ is largest for all samples belonging to this class, i.e. $D_i(x) > D_j(x)$ for all $i \neq j$. In measurement space the decision boundary is given by $D_i(x) - D_j(x) = 0$. Forms of this discriminant function include the linear discriminant (LDF) and empirical discriminant (EDF) functions.

To set up an ideal discriminant function the classes should be mutually exclusive and exhaustive and the measurements chosen should permit the assignment of each sample to one or other of the possible classes. In practice, the classes chosen are usually exhaustive but possess various degrees of overlapping so that they are not mutually exclusive.

The selection of the measurements which produce the "best" discrimination between classes is one of the most important steps and one of the major difficulties faced in the application of the discriminant analysis. In applied geochemistry the chemical composition (measurements) of the samples is the most common and valuable information obtained by the geochemist. Such measurements which could be represented by, say, up to 15 or 30 elements, are not necessarily going to produce the "best" classification results of "unknown" samples presented to the classifier following the training phase; a lesser number of variables could give an improved successful recognition rate for each class if some variables in the original data set did not have good discriminatory power.

To illustrate this case, a set of analysis of 69 ortho- and para- leptynites with nine major elements determined by Davoine (1969) have been studied. The results obtained with a linear discriminant function (LDF) are presented in figure 2-2. The graph clearly shows that the maximum recognition success rate for samples of known type presented to the classifier as "unknown" (the testing set) is attained with particular combinations of either 6, 7 or 8 variables. This is a good example of the importance of selecting the "best" measurements and has a direct bearing on the efficiency and economy of every project. This problem of feature selection in a geochemical context has been discussed at length by Howarth(1973).

The investigator faces another problem in the selection of the nature of the classes (training set) based on the a priori information. While it is hoped that the boundaries of each class may be exclusive, in a real-world environment, partial overlap of the classes is relatively common.



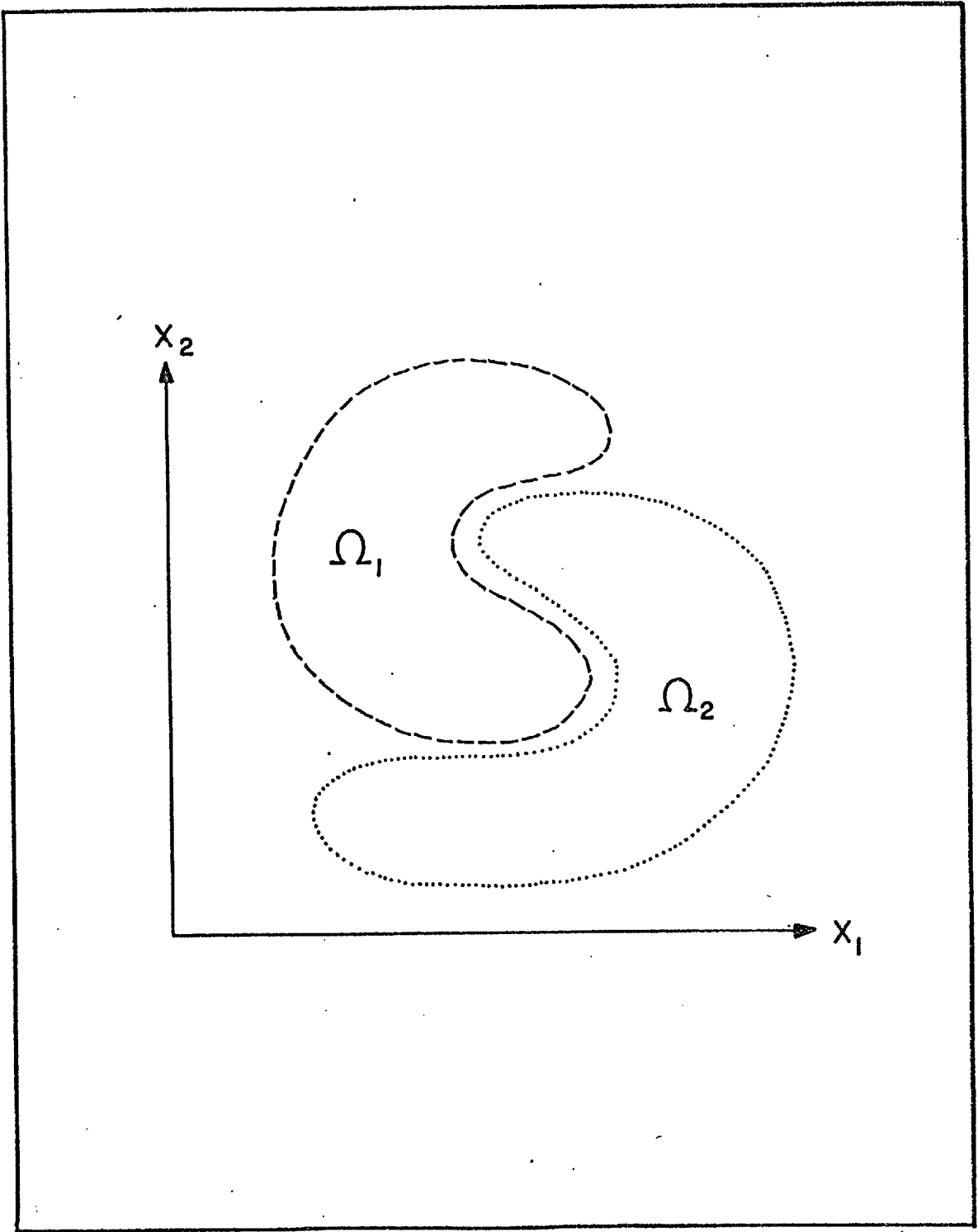
Behavior of the overall recognition success rate for the testing set according to different combinations of variables

FIG. 2-2

The choice of the training and testing sets is another problem. This selection is particularly difficult in regional geochemical data because the geochemist chooses the "training sets" on the basis of the mapped broad-scale geology which is in itself commonly incomplete, or of insufficient detail for accurate interpretation. That will undoubtedly produce appreciable variations in the evaluation of the "goodness" of the classification results. From a comparison of the true classes of the testing set with those assigned by the classifier one arrives at a measure of the "goodness" of the results as shown by the overall correct recognition rate of the classification system (one may construct a classification success table for all classes). An unsuitable choice of the training set would undoubtedly produce results which were not representative of conventional geological mapping establishing a training area over, for example, sediments intruded by "not reported" basic igneous rocks; although the results will be correct in terms of a real-world natural group on the basis of the geochemical measurements, they would not be in agreement with conventional geological mapping rules. In the definition of the "training set" the first problem to be solved is to determine the type of geologically "natural" environments which it is intended to discriminate between (e.g. sedimentary formations with one or several lithologies from well differentiated igneous masses; bed-rock units from mineralization; mineralization from secondary environment, and so on). Subsequently, to establish the training areas which best represent the natural (geological) features previously established by the geochemist.

ii) Methods

The linear discriminant function (LDF) has been used in few published applied geochemistry studies (Wignall, 1969; Cameron, Siddeley and Durham, 1971) although it has been widely used in pure geochemical studies. However, as figure 2-3 shows, it may not always be possible to completely separate two (or more) classes with the LDF method. A good theoretical explanation of the LDF is given by Harbaugh and Merriam (1968). Howarth (1971a, 1971b) used the empirical discriminant function (EDF) as an alternative in a preliminary investigation. In this study the EDF has been applied to the classification of regional geochemical data of the United Kingdom in order to evaluate its applicability in different



Non-linearly separable classes

FIG. 2-3

geological environments.

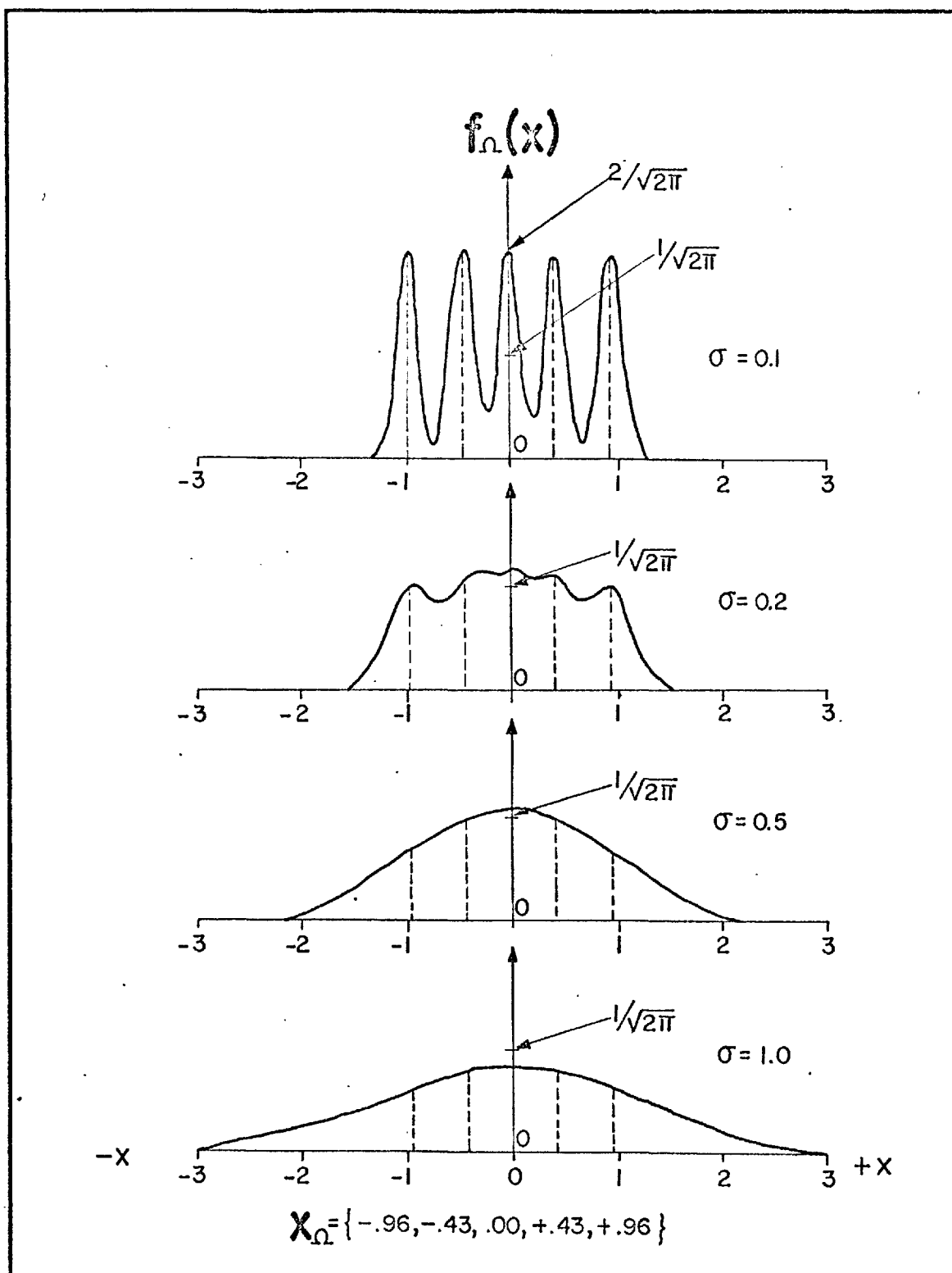
This method is the exponential form of the polynomial method of Specht (1967). It is based on the non-parametric estimation of a probability density function for each category to be classified so that the Bayes decision rule may be applied. A smoothing parameter (σ) is applied so that a density function is estimated from the training sample points in measurement space. The figure 2-4 shows how the five distinct modes (corresponding to the position of training samples) are progressively smoothed by increasing σ for a one-dimensional case. Detailed mathematical considerations are given in Appendix 1.

The first preliminary geochemical investigation was successfully carried out in sedimentary-rock major-element geochemistry (Howarth, 1971a) and, subsequently, in stream-sediment regional data of Devonshire (Howarth, 1971b).

2-3 Cluster analysis.

i) Average linkage clustering.

Hierarchical cluster analysis has been widely used in Numerical Taxonomy, and is attaining major importance in the geological sciences, particularly in applied geochemistry although the published literature in this field is still relatively small (Rhodes, 1969; Obial, 1970; Obial and James, 1973; Nichol and Larson, 1973). Little or no a priori geological knowledge is required and the samples are grouped on the basis of their measurements (element compositions) alone into clusters which (hopefully) represent an inherent "natural" category of the existing environment. Similarity coefficients are then computed between each pair of samples or variables depending on whether samples (Q-mode) or variables (R-mode) are being clustered. Cluster analysis groups together those variables (or samples) with a high mutual similarity into groups or clusters in a hierarchical fashion in order of decreasing similarity, so that their interrelationships are completely contained in two dimensions, graphically represented as a dendrogram. In practice, however, it may be difficult or impossible to provide effective representation of multi-dimensional data in a two-dimensional diagram of this type, and a



Interpolated one-dimensional probability density function for set of five training samples with increasing values for smoothing parameter σ (redrawn from Specht, 1967)

FIG. 2-4

little distortion is not therefore surprising. Methods for the evaluation of the amount of distortion are available (Sokal and Sneath, 1963, p.312-313).

There are a wide diversity of clustering methods and in most geochemical applications agglomerative (clumping) methods have been more widely used. Amongst the most widely used are the "group" methods proposed by Sokal and Michener (1958). The simplest of these methods are the "pair-group" methods in which the new groups produced at any stage in the clustering process contain only two members.

In this study clustering by average linkage (Sokal and Sneath, 1963, p.182) has been used with the weighted pair-group method (WPGM) of Sokal and Michener (1958).

In the dendrogram the variables (samples) are represented by the tips of two or more branches issuing from any bifurcation point, and the level at which two branches join is the similarity coefficient at which the groups of variables (samples) at a higher level on each branch join. Any variables (samples) occupying positions intermediate between two "natural" clusters will be affected, as overlapping of clusters is not permitted, and they are forced into one or other group. The WPGM divides the variables (samples) objectively into group structure and represents it as a dendrogram but there is then a subjective interpretation of where to put the group boundaries. Gower (1967) remarks that cluster analysis of this type builds clusters by searching for hyperspheres; forces entries into hierarchical classification regardless of "natural" groupings; and may have considerable distortion in the dendrogram representation. Also, the results vary depending on the data transforms used; weighting of variables; and type of similarity coefficient used.

Sammon (1969) points out the following deficiencies which apply to cluster-seeking algorithms.

- 1) When using a particular algorithm, the resulting cluster configuration is highly dependent upon a set of control parameters which must be fixed by the user. - Some examples are:
 - a) the similarity measures;
 - b) various similarity thresholds;
 - c) number of iterations required;
 - d) thresholds which control the increase or reduction of the number of clusters;
 - e) the minimum number of vectors (samples) required to define a cluster.
- 2) Most of the existing clustering algorithms are particularly sensitive to hyperspherical structure and are inefficient in detecting more complex relationships in the data.
- 3) Perhaps the most serious deficiency involving present-day clustering algorithms is that there do not exist really good ways for evaluating a resultant cluster configuration.
- 4) When two clusters are close, samples lying between them tend to form a bridge and cause spurious mergers (Nagy, 1968). This is one of the main difficulties found in this investigation because the WPGM method forces transitional members into a nearby cluster.

In this study the data for the Q-mode and R-mode analysis is usually transformed in order to give each variable equal weighting by dividing each set by the highest value of the variable for the set, so that the values of each transformed variable range between 0 and 1.

Various methods are available for determining the similarity coefficients. They can be grouped into two main categories: those using numerical data such as Pearson product-moment (linear) correlation coefficient, distance coefficient or $\cos \theta$ coefficient, and those using non-numerical data coded into two-or-multi-state form such as matching and association coefficients. Details of

these coefficients are given by Sokal and Sneath (1963) and Harbaugh and Merriam (1968).

This study has been carried out using the distance similarity coefficient for inter-sample comparisons. This is based on the concept that a quantitative measure of the degree of similarity between two samples is provided by the distance that separates them in a rectangular cartesian coordinate system; the shorter the distance the greater the degree of similarity and vice-versa.

Unless otherwise stated (e.g. Rhodes (1969) paper) the distance coefficient is defined as follows: the data is transformed to lie in range 0.0 - 1.0 by dividing through for each variable by the maximum value for that variable. Let d_{ij} = distance between samples i and j , then (Fig.2-5):

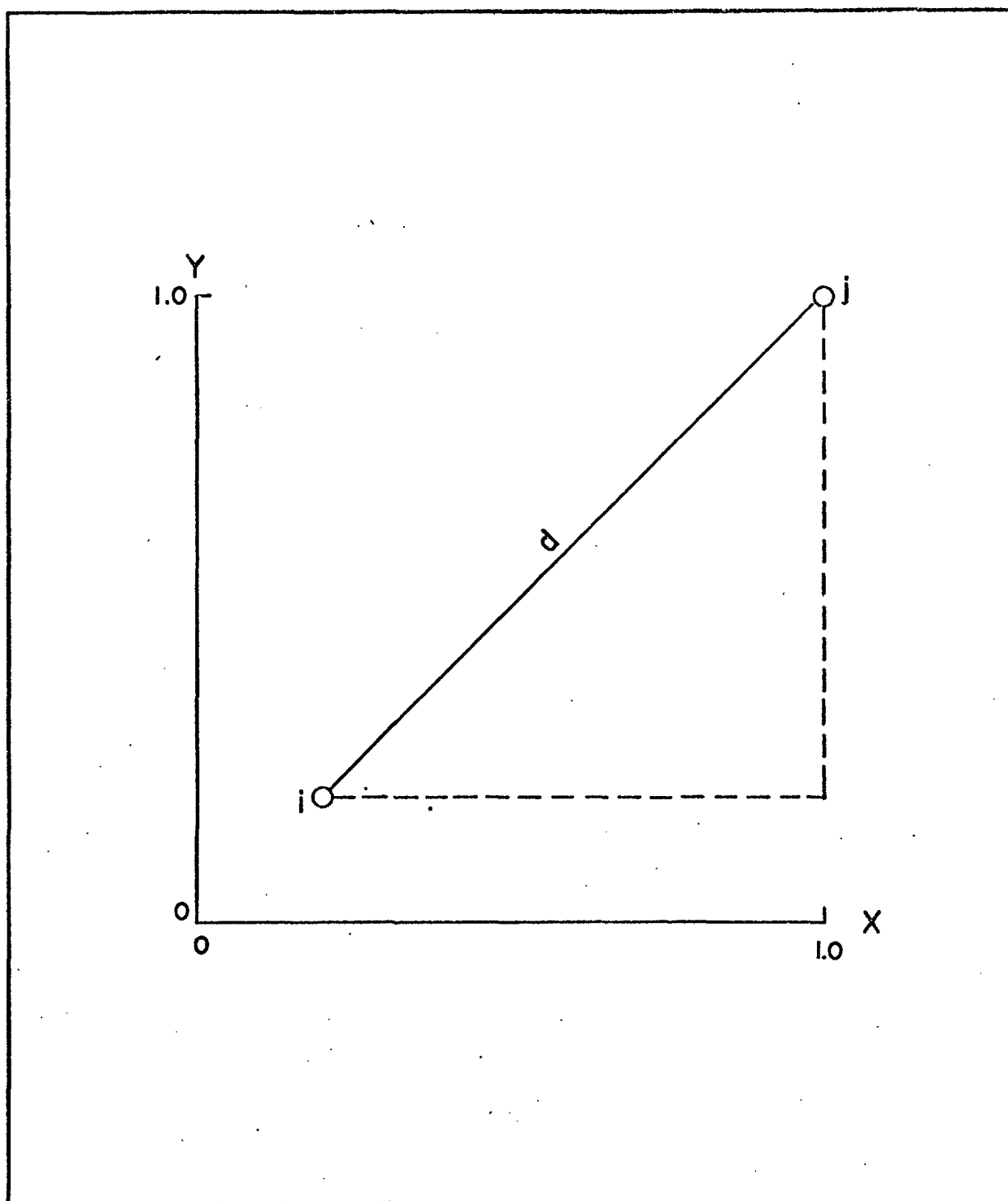
$$d = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2},$$

where $X_{ij} \leq 1$ and $Y_{ij} \leq 1$. In general, for p measurements the maximum similarity is $d_{ij} = 0$ and the minimum similarity is $d_{ij} = \sqrt{p}$. In order to have a similarity coefficient with 1 as the maximum similarity, inverse distance is used, $d^* = 1-d$. In this thesis "distance coefficient" refers to d^* rather than d unless otherwise stated.

Two "large sample" clustering programmes, one developed by Parks (1970) and the other by Valentine and Peddicord (1967) were not used because in preliminary tests the first one was extremely time consuming and produced poor results and the second contained an insoluble programme error (Howarth, 1972, pers. comm.).

ii) Non-linear mapping algorithm.

The non-linear mapping algorithm (NLM) has been developed by Sammon (1969) in order to minimize the distortions of representing the structure of multi-element data in two dimensions. The algorithm carries out a mapping (projection) of N -samples, for which p elements are determined, distributed in p -dimensional space into a two-dimensional space such that the original structure in the multi-



Two-dimensional measurement d for samples i and j
in (X, Y) space

FIG. 2-5

dimensional space is preserved with a minimum mapping error E (Sammon, 1969). This is achieved by fitting the N -points (samples) in the 2-dimensional space such that their interpoint distances approximate the corresponding inter-point distance in the p -space. The mathematical considerations are given in appendix 2.

Sammon (1969) feels that the NLM algorithm is highly promising because it is not affected by the general deficiencies of clustering algorithms discussed earlier, rather: 1) the routine does not depend upon any control parameters that would require a priori knowledge about the data; 2) it is highly efficient in identifying hyperspherical, hyperellipsoidal, and other complex data structures; 3) the resulting two-dimensional diagram (scatter diagram) is easily evaluated by the researcher, thereby taking advantage of the excellent ability of the human eye to detect and identify data structure in two-dimensional scatter plots (c.f. Hertzsprung-Russell diagram, Fig. 2-1, referred to earlier); 4) the problem concerning extraneous data and spurious mergers is not present since humans easily eliminate troublesome data points by making global evaluations (machines have difficulty performing this function); 5) the algorithm is simple and efficient.

- iii) Cluster analysis of the Mount Shooobridge granite.
Difficulties in interpretation. Comparative results.

When the WPGM is applied to classification problems, as was previously pointed out, difficulties regarding distortion of structure generally arise. The difficulties take place when different type of transformations, similarity coefficient and data input (ordering of samples) are considered. An example is provided here to illustrate this and to compare the results with those obtained using the NLM algorithm. For this purpose a set of data for the Mount Shooobridge granite (Rhodes, 1969) has been re-examined.

Rhodes mapped a small diapiric stock of granite in the Northern Territory of Australia about 100 miles (160 kms) south of Darwin. The granite occupying an area of only about one square mile was divided on the basis of field mapping into three distinct types: 1) a marginal granodiorite, 2) a porphyritic granodiorite and 3) a small central core of leucocratic adamellite. (Fig.2-6). In order to apply discriminant and cluster analysis to the granite sixteen samples were collected from the three main rock types over the best outcrops and analysed for major elements and for some trace elements. The cluster analysis results obtained by Rhodes using all the variables, except H_2O and CO_2 , are illustrated in the dendrogram of figure 2-6. As can be seen in the Q-mode dendrogram the groups show a direct relationship with the mapped geological types. At a fixed distance coefficient of about 0.4 samples 11, 33 and 35 are clustered together representing the leucocratic adamellite. The remaining samples constitute a single group that includes simultaneously the porphyritic granodiorite and the marginal granodiorite. Lowering the distance coefficient to about 0.3 the marginal granodiorite is well separated from the porphyritic granodiorite and some intermediate samples (25, 28, 21) between both petrologic types, but the leucocratic adamellite group becomes split. On decreasing the distance coefficient to about 0.2, the intermediate samples constitute a group (which was not recognised previously by the field mapping) representing a petrologic type intermediate between the porphyritic granodiorite and the marginal granodiorite. At this stage the clustering of the remaining groups was destroyed causing misleading groupings. Poor selection of the different correlation levels may cause very important details to be missed, producing groups which are not really representative of the mappable geology.

Q- and R-mode analysis of Rhode's data, taking into consideration only ten major element oxides (Si, Al, Fe, Ca, Na, K, Ti, Mn, Mg, P) and using different transformations was carried out. The data matrix was computed from the raw data either: 1) by dividing the data values by the maximum value or 2) by compressing the data values

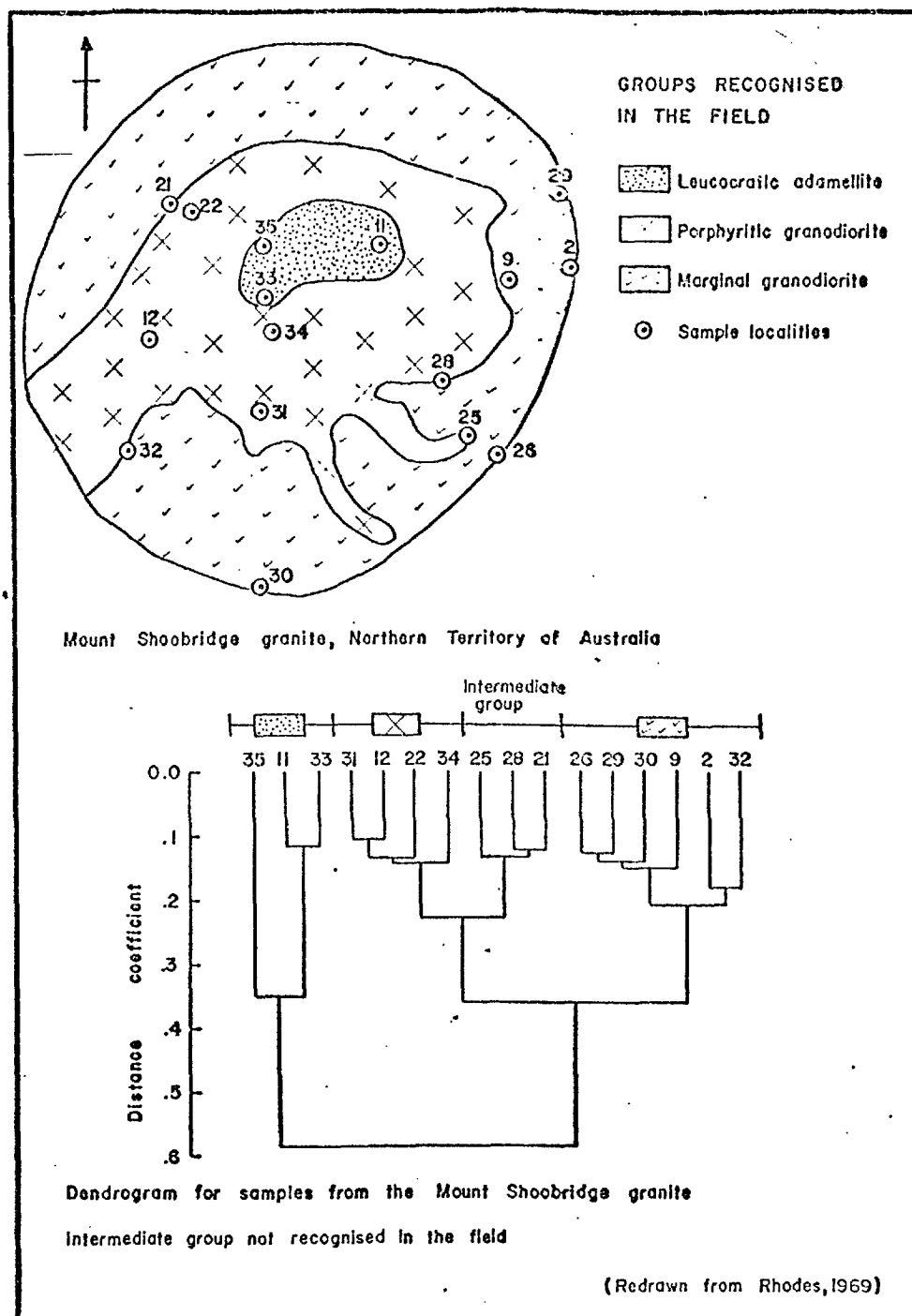


FIG. 2-6

into the range 0.0 - 1.0. Correlation and distance coefficients were computed in all cases. The results are presented in figures 2-7, 2-8 and 2-9, and the Q-mode results are summarised below:

1) All petrologic groups were perfectly separated at a distance coefficient of about 0.85 when the data were transformed by dividing by the maximum value (Fig. 2-7B). However, when the samples were sorted in different order (data input) and processed in the same way, sample 9 was misclassified (Fig. 2-7B').

2) All petrologic groups were perfectly separated at a constant distance coefficient of 0.73 when the data were compressed into the range 0.0-1.0 (Fig. 2-8D'). However, when the samples were sorted into a different order the porphyritic granodiorite group and the intermediate group were not well separated (Fig. 2-8D), and instead they were grouped as a single cluster probably because of their geochemical similarity, an affinity already established in Rhode's dendrogram (Fig. 2-6).

3) The petrologic groups were not well separated when the data were transformed either by dividing by the maximum value or compressing into the range 0.0-1.0, and the correlation coefficient computed (Figs. 2-7A, A'; 2-8C, C'). Correlation coefficients established at constant levels generally displayed groupings which were not in complete agreement with the mapped petrologic types; instead each petrologic group is defined by specific coefficients lying within a "correlation coefficient range" that contains all groups.

4) The petrologic groups were not delineated when the set of 10 uncorrelated variables from a factor analysis were transformed by dividing by the maximum value and the distance coefficient computed.

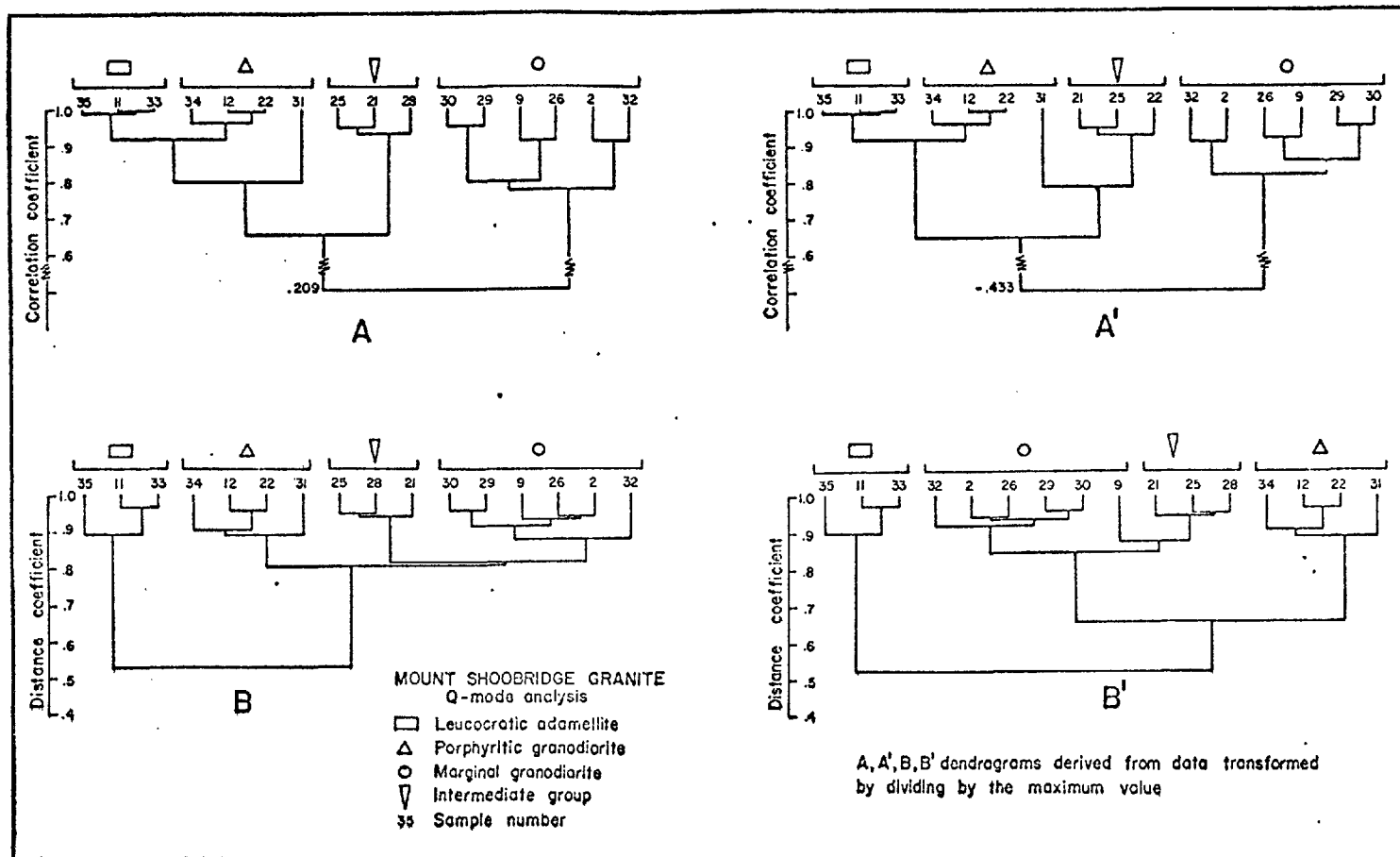


FIG. 2-7

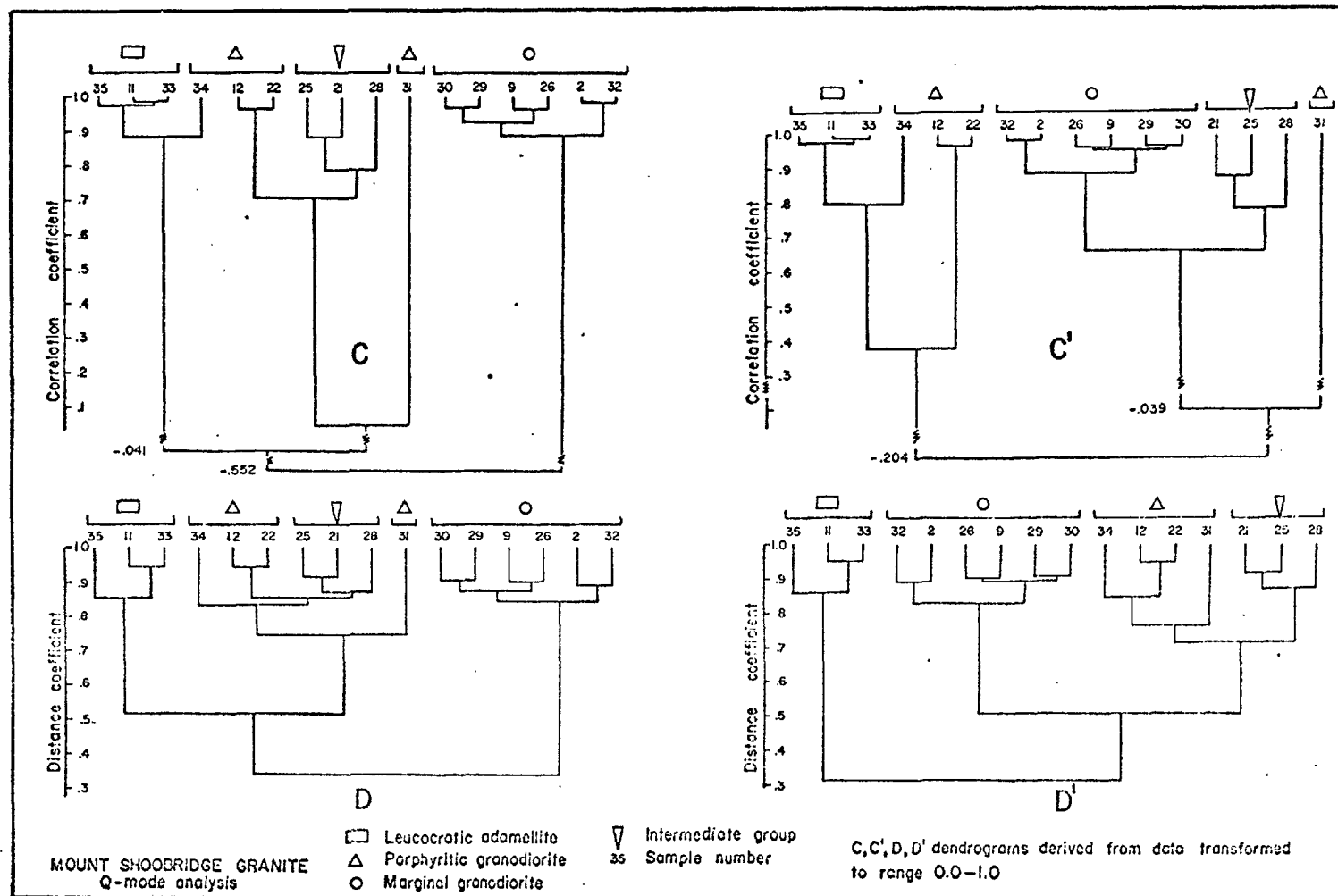
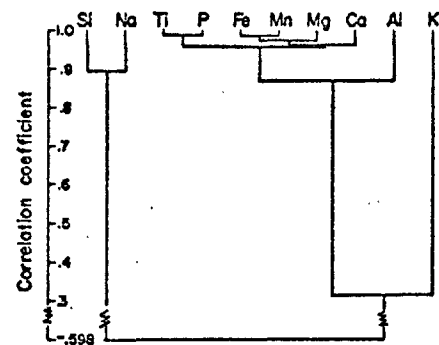


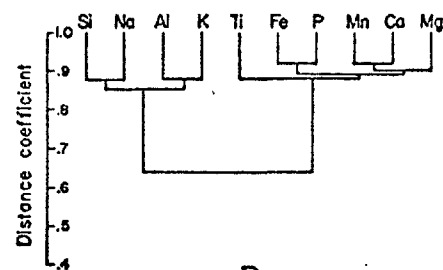
FIG. 2-8



A

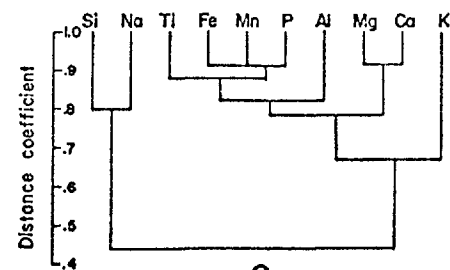
Dendrogram A derived from:

- 1) Data transformed by dividing by the maximum value
- 2) Data transformed into range 0.0-1.0
- 3) Untransformed data (Raw data)



B

Data transformed by dividing by the maximum value



C

Data transformed into range 0.0-1.0

MOUNT SHOOBRIDGE GRANITE
R-mode analysis

FIG. 2-9

More stable dendrogram structures (Fig. 2-7B, B') yielding the best solutions were obtained by dividing the data values by the maximum value and computing the distance coefficient.

The results of the R-mode analysis are shown in figure 2-9. The different ordering of the samples did not affect the structure of the dendrograms. Considering either the raw data, the data values divided by the maximum value, or compressed into the range 0.0-1.0, using a correlation coefficient matrix, identical dendrograms are obtained. Figure 2-9A shows the association of Si-Na (basic elements of felsic minerals), Ti - P - Fe - Mn - Mg - Ca (basic elements of mafic minerals) defined at a correlation level of 0.88, and the Al and K (basic elements of felsic minerals) behaving independently at lower levels (0.3) and unexpectedly showing more connection with typical elements of mafic minerals.

Figure 2-9B shows the R-mode results based on computing the distance coefficient from a data matrix obtained by dividing by the maximum value.

The felsic element association (representing quartz, feldspars, feldspathoids, muscovite) is perfectly well differentiated from the mafic elements association (pyroxenes, amphiboles, olivine) at a constant distance coefficient of about 0.85 or under. If the data set is compressed into the range 0.0 - 1.0, the structure becomes more expanded (Fig. 2-9C) and the element associations do not strictly reflect the mineralogical characteristics of the acid igneous rocks. The associations resemble those obtained with the correlation coefficient (Fig. 2-9A); in both cases a constant correlation level (correlation and distance coefficients) does not properly achieve separation of the groups according to the mineralogy of the acid igneous rocks.

The best picture of the element associations related to the principal mineralogical components of the petrologic types of the Mount Shooobridge granite was then obtained (as in the Q-mode analysis) by dividing the data values by the maximum value and computing the

distance coefficient. In this case too difficulties of choosing natural groups representing the main geological features of the granite were minimized; also, samples constituting the main taxonomic groups showed a higher compression in a "distance coefficient range" in such a manner that in all cases a generally constant and higher distance coefficient value outlined the main taxonomic and petrologic groups.

No consistent element association was obtained when the uncorrelated variables from factor analysis were transformed by dividing by the maximum value and the distance coefficient computed.

The Mount Shooobridge granite data were subsequently studied with the NLM algorithm and factor analysis. The results are shown in Figures 2-10, 2-11, 2-12 and 2-13, and the most important features were:

- 1) With the NLM algorithm all petrologic types were perfectly grouped considering the raw data (Fig. 2-10). Different ordering of the samples gave similar results.

- 2) With the NLM algorithm the petrologic types were poorly grouped with normalization and log-10 transformation of the data (Figs. 2-11 and 2-12).

- 3) Similar results can be seen between factor analysis (Fig. 2-13) and the NLM algorithm applied to uncorrelated variables derived from factor analysis (Fig. 2-14). Both results not as good as when the NLM algorithm was applied to the raw data.

In this particular example the NLM algorithm has given the most consistent results with a minimum of distortion and has proved to be more effective than the WPGM dendrogram when the raw data have been considered.

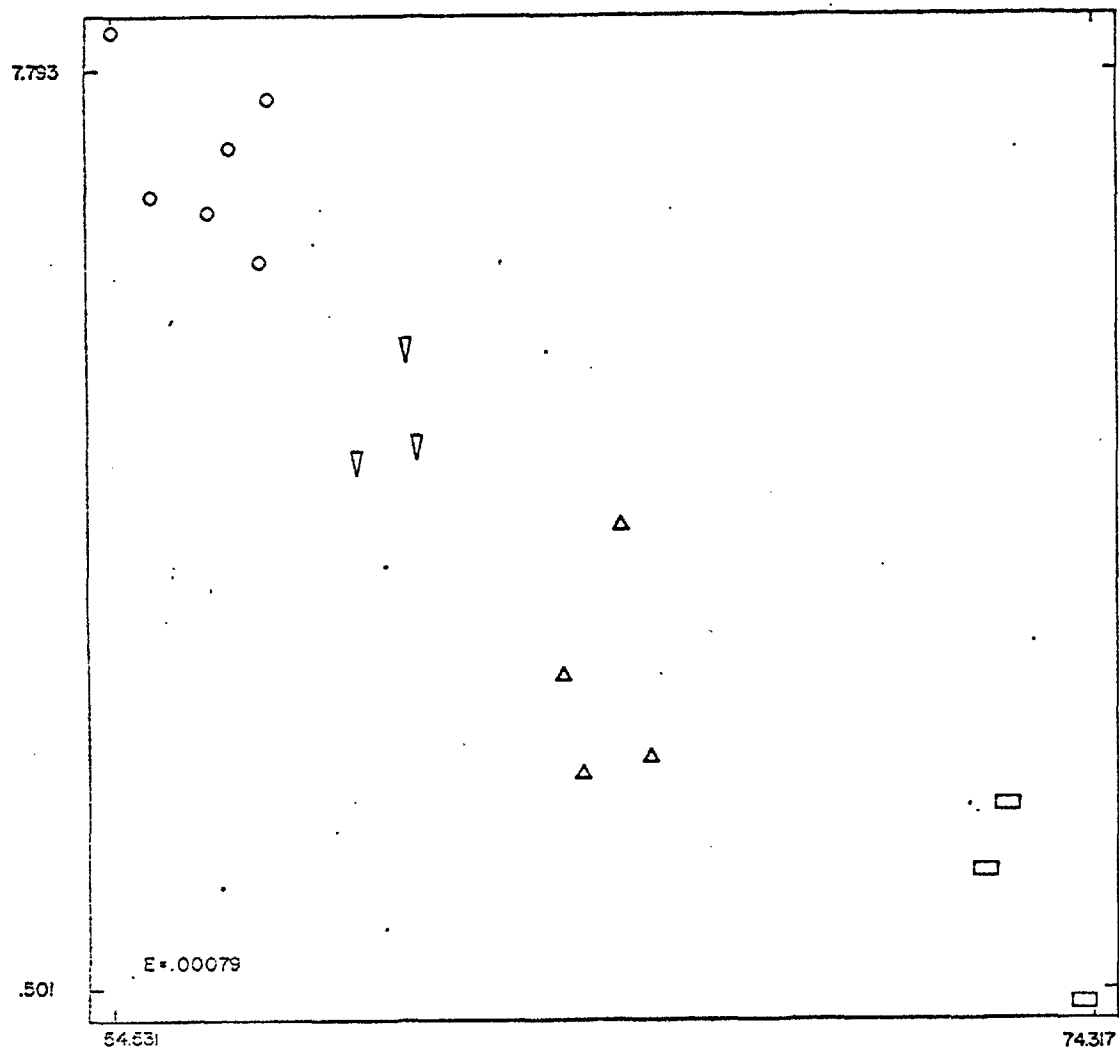


FIG. 2-10
MOUNT SHOOBRIDGE GRANITE
NLM algorithm
Untransformed data

- Leucocratic adamellite
- △ Porphyritic granodiorite
- Marginal granodiorite
- ▽ Intermediate group

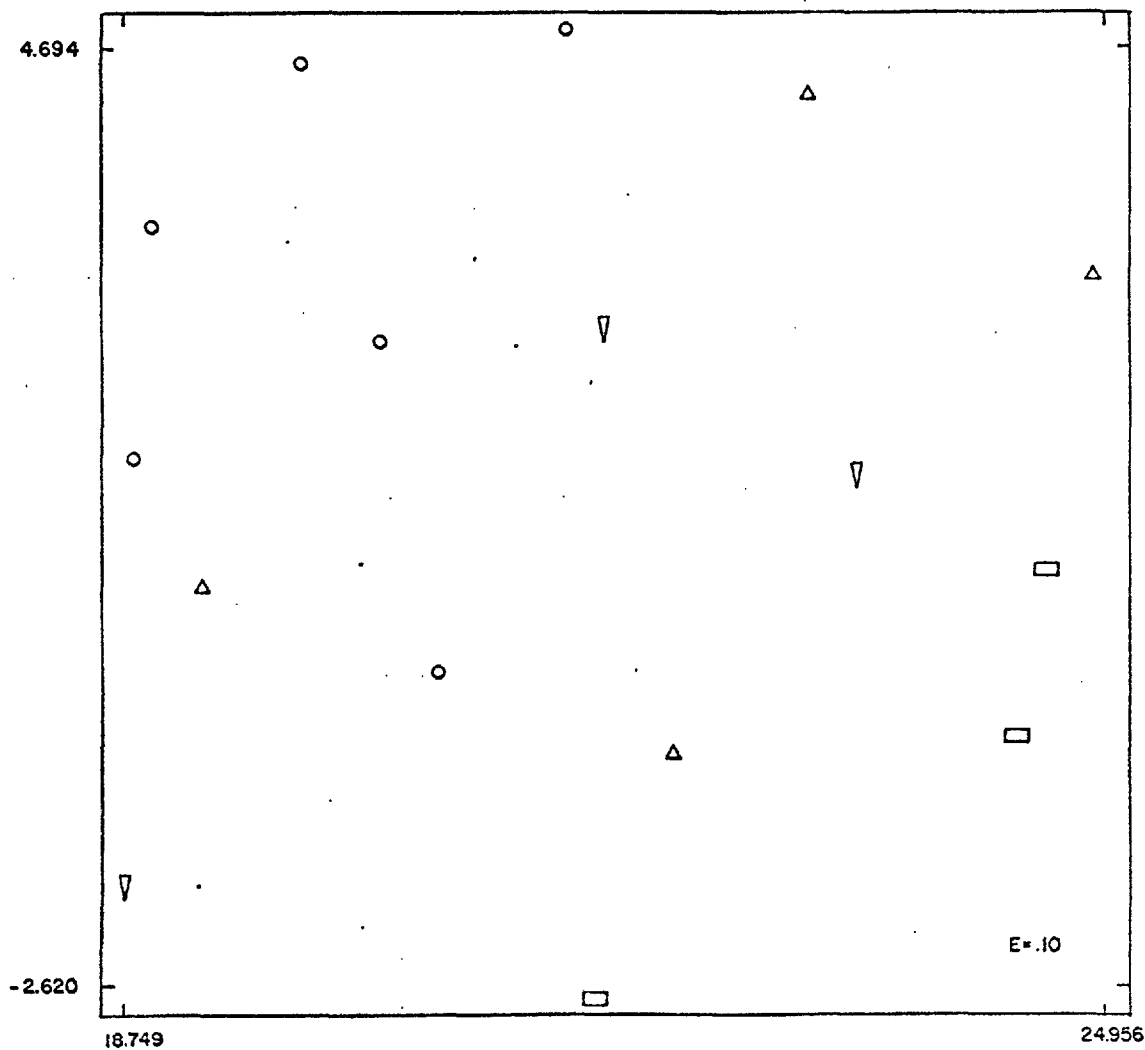


FIG. 2-II
MOUNT SHOOBRIDGE GRANITE
NLM algorithm
Normalized data

- Leucocratic adamellite
- △ Porphyritic granodiorite
- Marginal granodiorite
- ▽ Intermediate group

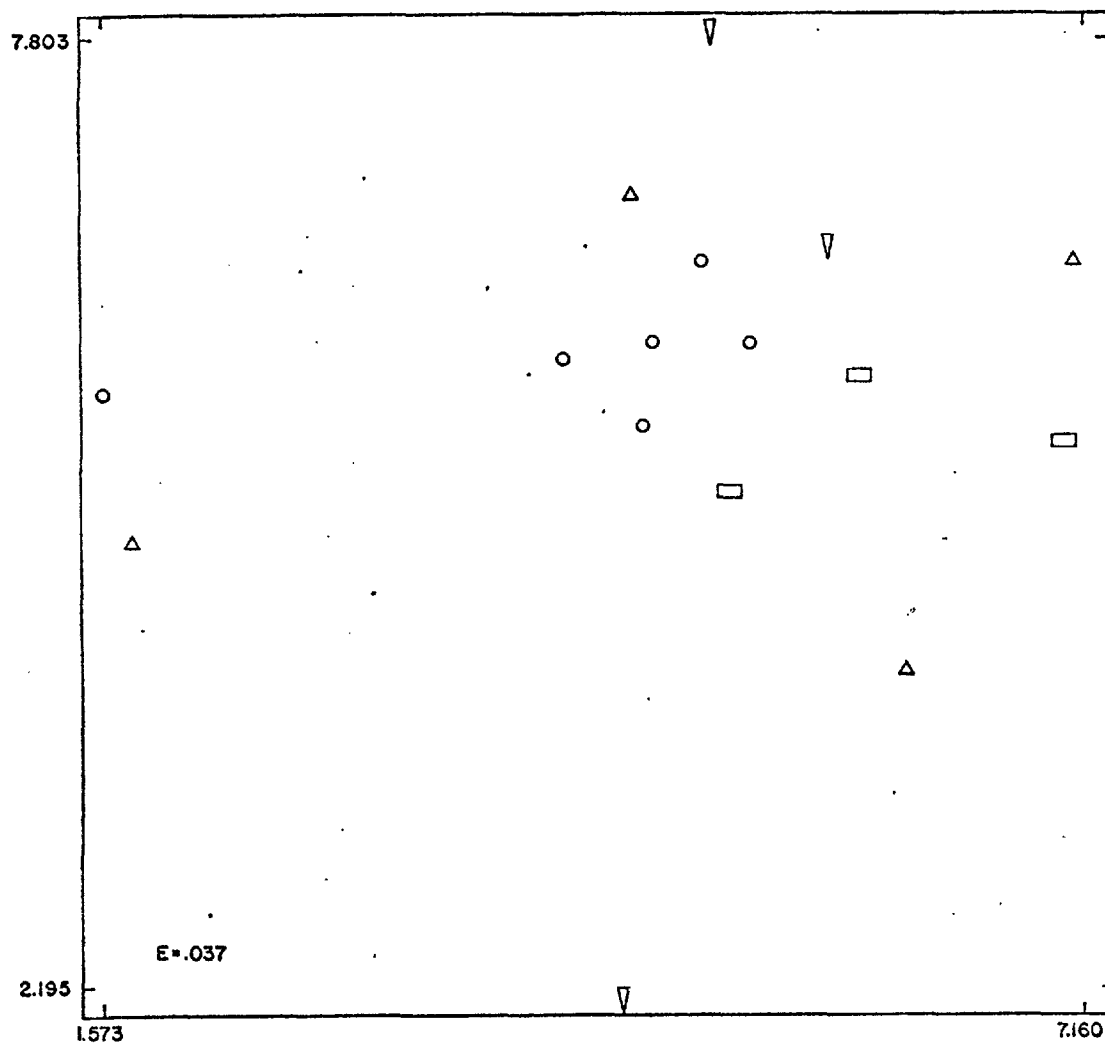
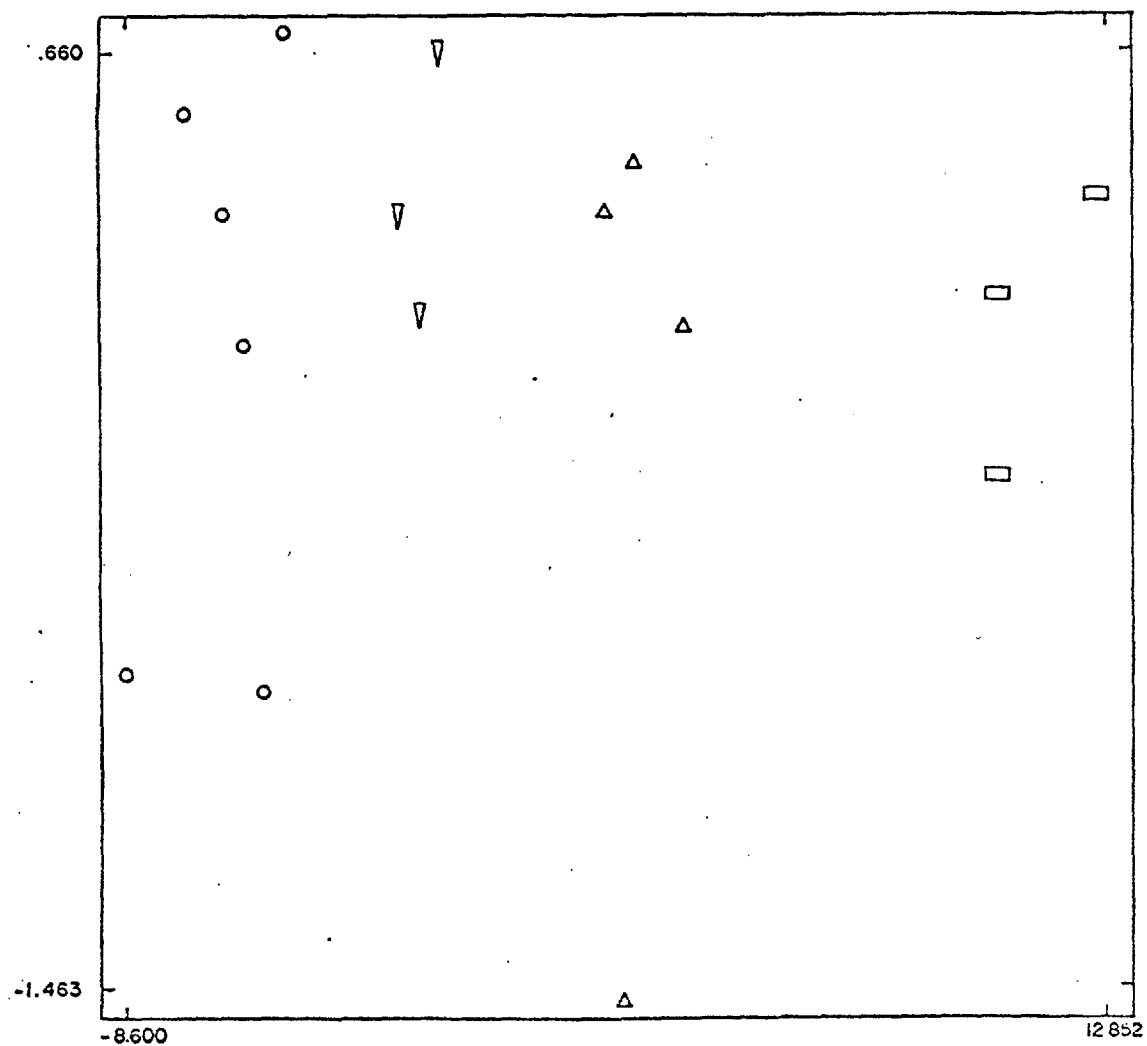


FIG. 2-12
MOUNT SHOOBRIDGE GRANITE
NLM algorithm
Log-10 transformed data

- Leucocratic adamellite
- △ Porphyritic granodiorite
- Marginal granodiorite
- ▽ Intermediate group



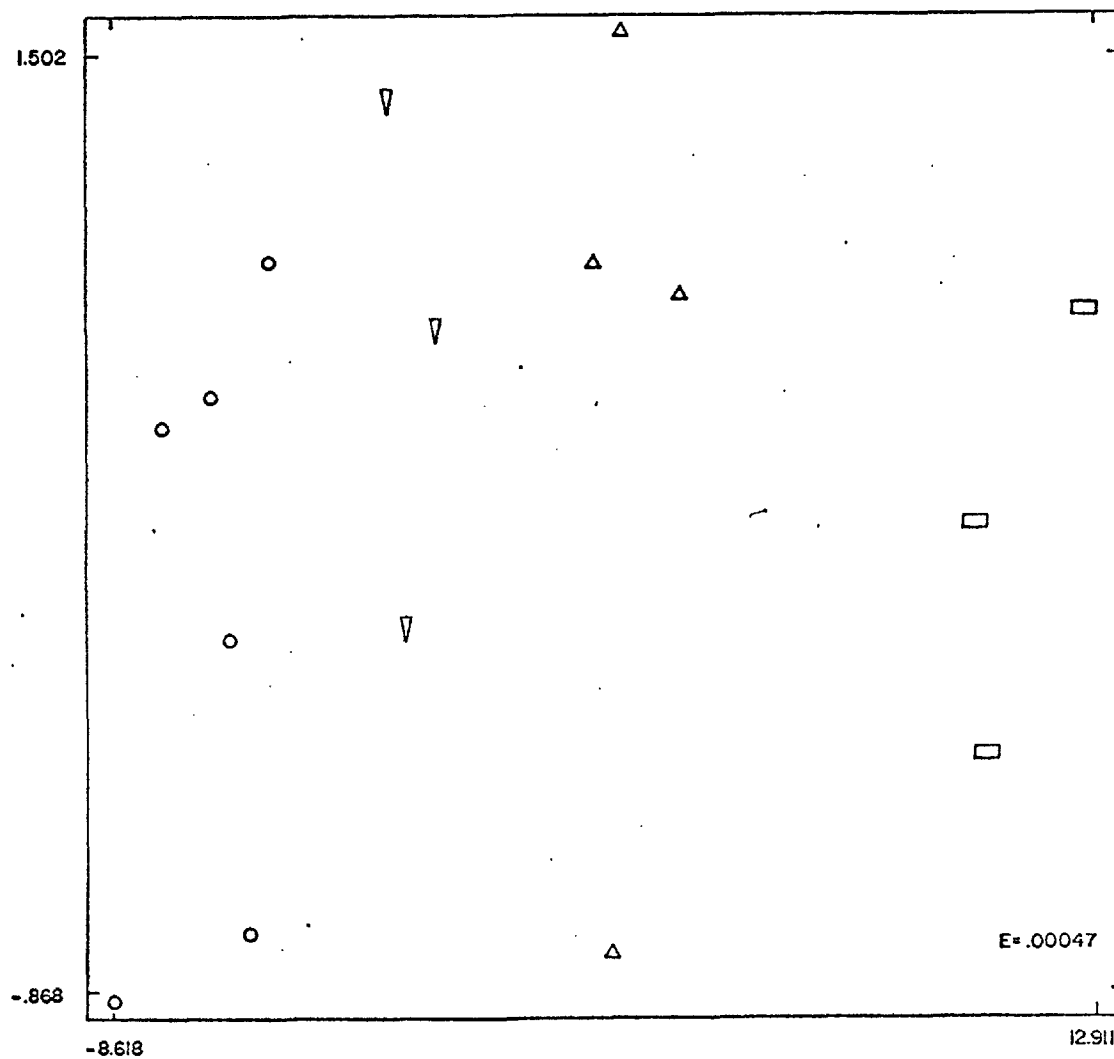


FIG. 2-14
MOUNT SHOOBRIDGE GRANITE
NLM algorithm, untransformed data
from uncorrelated variables of
factor analysis

- Leucocratic adamellite
- △ Porphyritic granodiorite
- Marginal granodiorite
- ▽ Intermediate group

These remarks must not be considered as the general rule when the NIM algorithm is applied because (as it will be subsequently shown) in the classification of the regional geochemical data excellent results were obtained with the log-10 transformation of the data; this suggests that previous orientation tests will be required to establish the most suitable transformations according to the sort of geochemical data under study (e.g. major elements, trace elements, pH and Eh measurements, etc.).

CHAPTER 3

DISCRIMINANT AND CLUSTER ANALYSIS OF DENBIGHSHIRE

3-1 Location and description of the study area.

i) Geography

The study area (Fig. 3-1, 3-3) includes most of Flintshire, northern Denbighshire and a small portion of eastern Caernavonshire and is bounded by the National Grid coordinates 3400 (south), 3858 (north), 2680(west) and 3400(west) covering a total surface area of 850 sq. miles. Some of the principal towns are Denbigh, Llangollen, Wrexham, Chester, Mold, Prestatyn, Conway and Llanrwst.

There are six well defined physiographic regions controlled by the geology and climate. These are, from west to east: North Snowdonia, the Conway Valley, the Denbighshire Uplands and Moorland, the Vale of Clwyd, the Clwydian Hills and the Flintshire plain.

The rugged topography of North Snowdonia attains a maximum altitude of 2860 feet over the western side of the area and is bounded at its eastern side by the Conway Valley. The Denbighshire Uplands lie between 500 and 1000 feet over the north-central part of the area and show a gradual increase southward until they constitute the Denbighshire Moorland (Denbigh Moorland) with elevations of between 1000 and 1400 feet. The Denbigh Moorland forms a monotonous area of heathy plateau and deeply dissected rivers and show a decrease in elevation toward the rivers Conway, Clwyd and those valleys developed in the south forming in the adjoining areas an approximately radial drainage system. The Vale of Clwyd occupies a stretch of country around the central part which had sunk as a result of faulting; it is a V-shaped valley with numerous truncated spurs and ending in an extensive coastal plain at the mouth of the vale. East of the Clwydian Hills which attain an elevation of between 500 and 670 feet there is a gradual decrease in elevation until the Flintshire is reached drained by the river Dee and its tributaries. The greater part of Denbighshire is drained by the Clwydian system, although in the west it is drained by the Conway. The overall

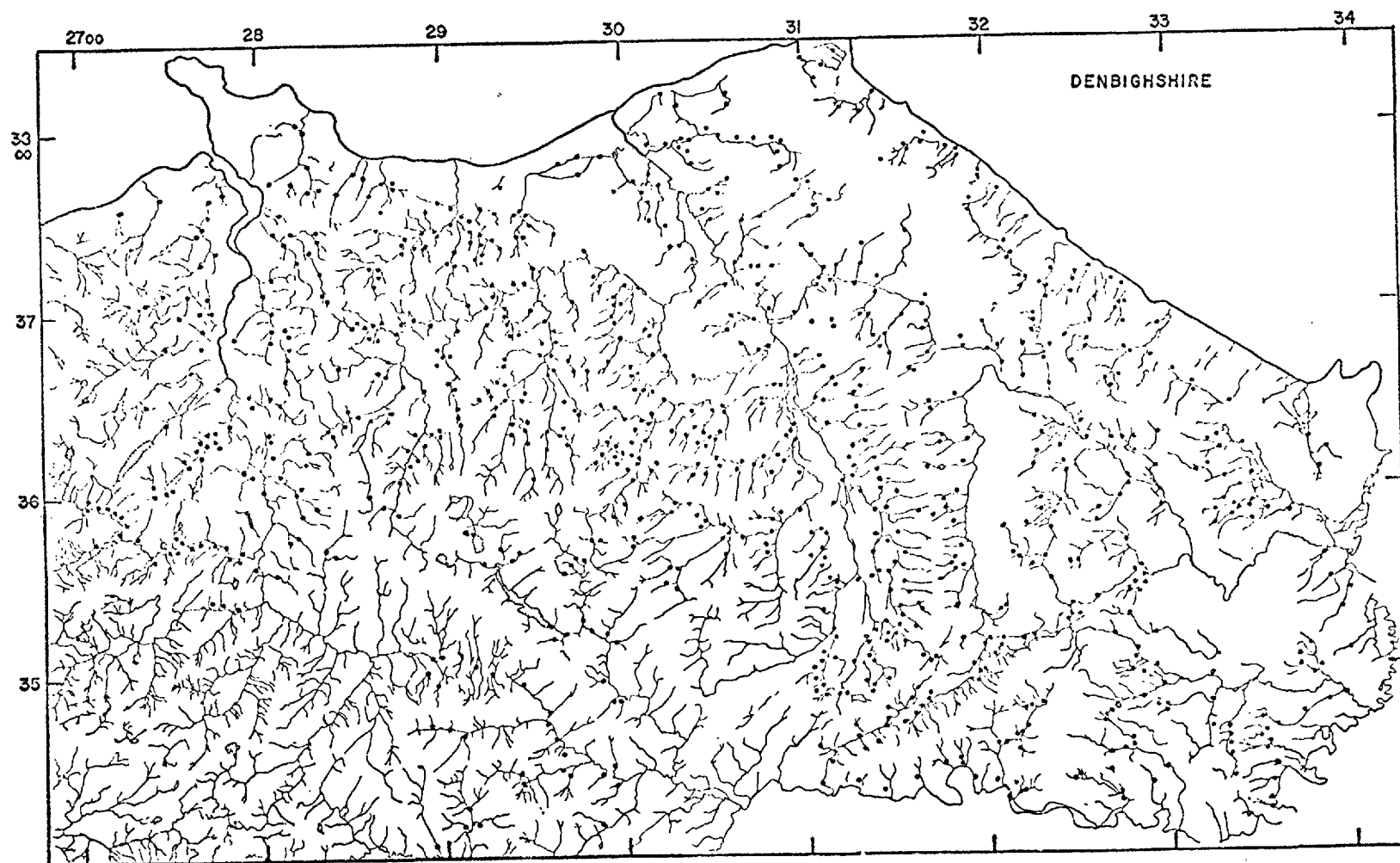


FIG. 3-1

DRAINAGE SYSTEM AND SAMPLING LOCALITIES
(After Khaloos, 1969)

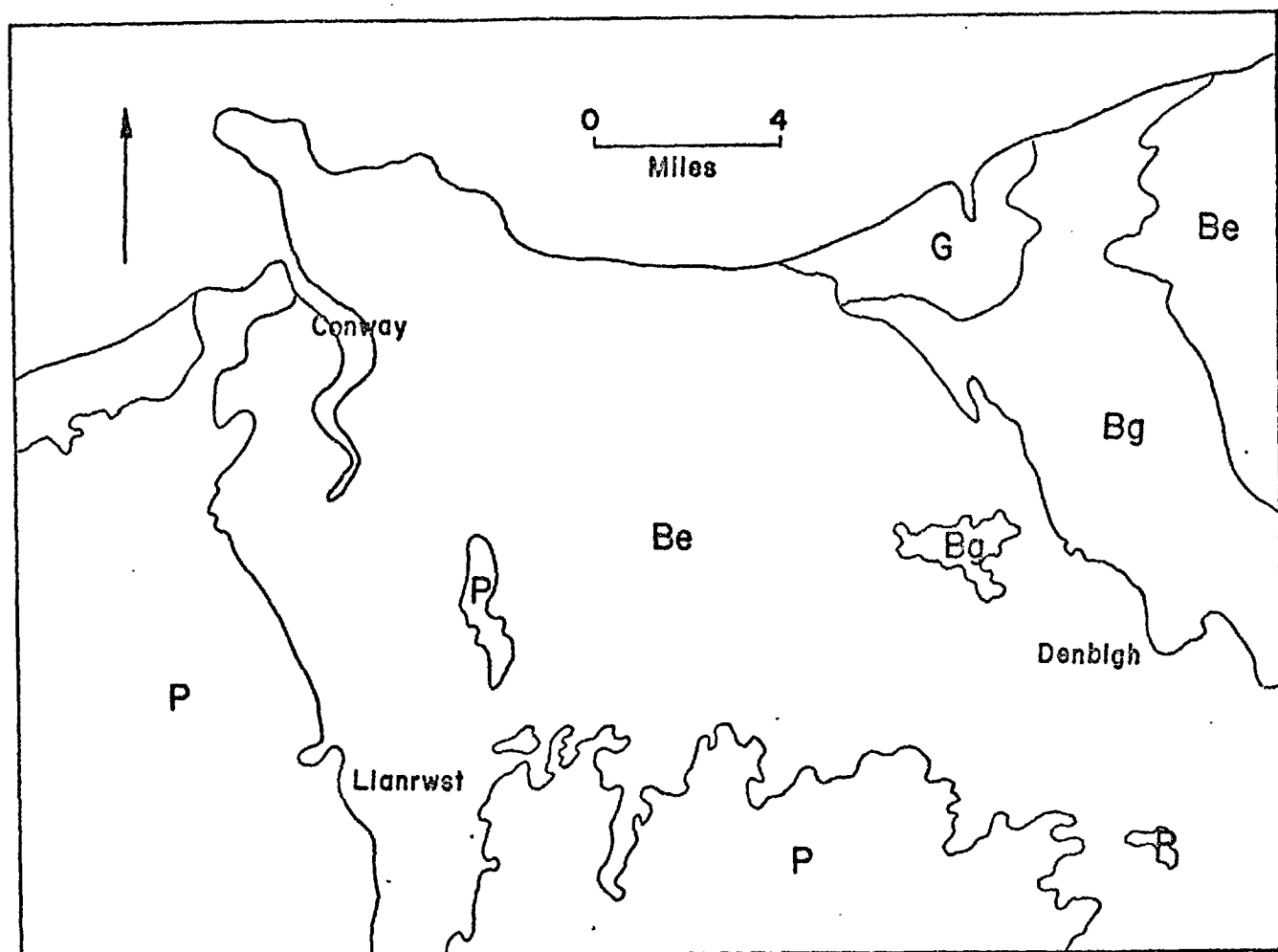
linear density of tributary drainage is approximately 3 miles per square mile (Fig. 3-1).

The climate is maritime, cold to temperate, with high humidity. The rainfall is high in the west and gradually decreases towards the east. The highest rainfall occurs in Snowdonia (60-100 inches), decreasing to 40-55 inches over the Denbigh Moorland, 30-40 inches over the Denbigh Uplands and is lowest at the coast with 25 inches at the mouth of the Vale of Clwyd. The temperature is lower over the higher ground.

Ball (1960, 1963) has mapped the soils that cover two thirds of the study area (Fig. 3-2). Podzolic and peaty gley soils have developed over the Ordovician rocks west of the Conway Valley. Acid peaty gleys and podzolic soils are distributed over the Denbigh Moorland with hill peat deposits in the topographic depressions. Free-drained brown earth of low base status has developed over Silurian rocks of the Denbigh Uplands. Gley soils and imperfectly drained brown earth types form the main soils covering the Denbigh-shire coastal plain and the Vale of Clwyd; the brown earth soils are typically developed over the Northern Drift and alluvial deposits. The Dinantian rocks are covered by freely-drained, high base status brown earths.

ii) Geology and Regional Geochemistry

Rocks of Ordovician, Silurian, Carboniferous, Permo-Triassic and Quaternary age have been mapped in the study area (Smith and George, 1961). A geologic map is given in Figure 3-3. Cambrian rocks are exposed in the south western corner of the study area but they were not included in the original regional survey (Horsnail, 1968). The regional geochemistry of the study area is a brief discussion of Nichol and others (1970b). The geochemical characteristics of stream sediments derived from the principal geological formations are given in Table 3-1 in order to show the relationships between the geochemical



P Peats, peaty gleys
and podzols

Be Brown earths

G Gleys

Bg Brown earths
with gleying

MAJOR SOIL GROUPS OF THE WESTERN HALF OF DENBIGHSHIRE

(Modified from Ball 1960,63)

FIG. 3-2

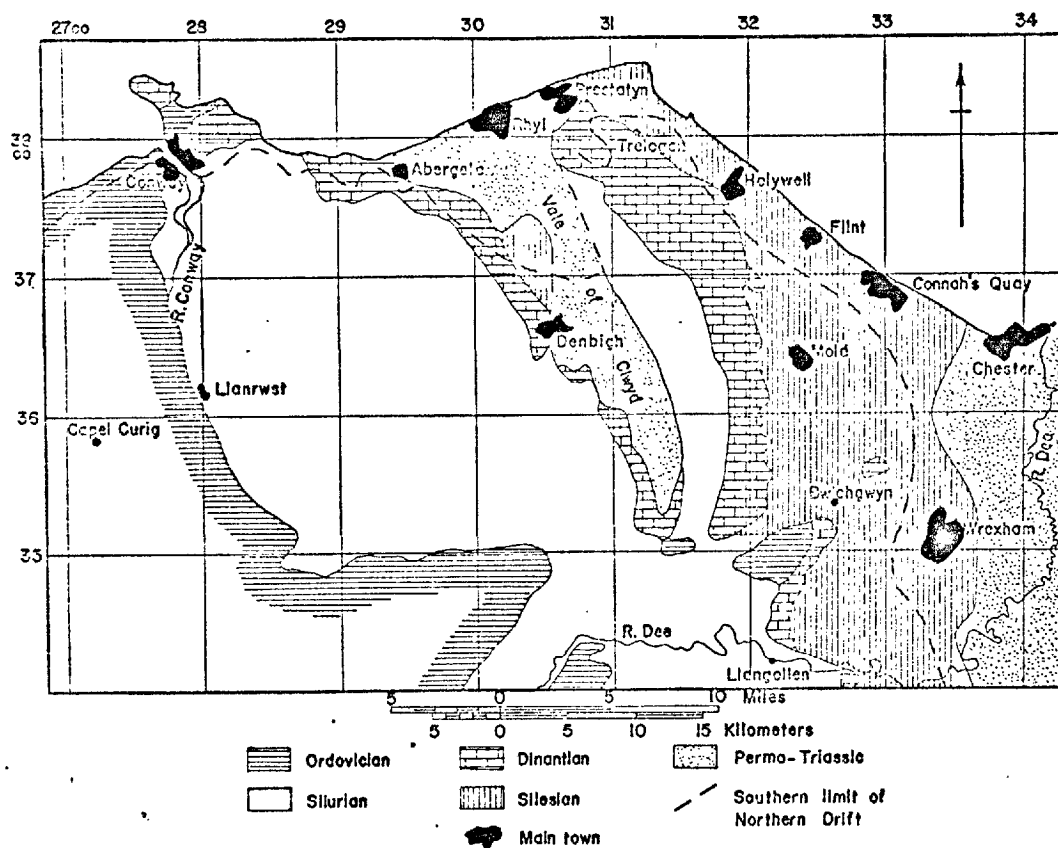


FIG. 3-3

DISTRIBUTION OF THE PRINCIPAL GEOLOGICAL UNITS, DENBIGHSHIRE
(After I.G.S., Nichol and others, 1970b)

TABLE 3 - 1

Range and mean* content of minor elements in stream sediments associated with major geological formations. (After Nichol & others, 1970b, p.11).

ELEMENT**	ORDOVICIAN ppm	SILURIAN ppm	DINANTIAN ppm	SILESIA ppm
As	12.5 5-40	<5 -	<5 -	<5 -
Cr	75 50-107	80 50-130	45 30-75	45 28-75
Co	40 20-70	20 12-40	12 5-20	15 10-25
Cu	40 25-70	35 20-50	30 20-45	30 20-50
Fe ₂ O ₃ %	5.2 3.4 - 7.8	3.1 1.9 - 5.0	1.6 0.9 - 2.8	1.3 0.9 - 1.8
Pb	85 25-290	40 20-80	75 25-240	210 45 -900
Mn	1900 650-5500	450 160-1100	275 120-635	300 150-600
Mo	2 2-20	<2 -	<2 -	<2 -
Ni	45 15-120	45 30-60	30 20-50	35 25-45
Sn	10 5-30	5 5-10	<5 -	7 5-25
Ti	4800 3200-7000	3700 2500-6000	2000 100-4000	1400 850-2500
V	120 80-180	110 70-175	55 25-105	35 20-55
Zn	375 120-1100	170 95-280	160 95-685	260 60-1200
No of Samples	103	358	25	42

* Geometric mean calculated from data obtained on minus 80-mesh fraction.

** Range = geometric mean $\pm$ log standard deviation.

** Element concentrations expressed as p.p.m. unless otherwise stated.

patterns and the geological formations taking into account the mean values of the elements.

The Ordovician rocks consist chiefly of slates, sandstones, and acidic volcanics and few small intrusive bodies of augite dolerite and rhyolitic rocks. The stream sediments derived exclusively from these rocks show a marked enrichment of Mn, As, Co, Fe, Zn, and to a lesser extent Cu, Ti, and Sn in contrast with those from Silurian strata. As, Cu, Mn, Mo, Pb, Sn, Ti, and Zn attain higher mean levels in the west near Llanrwst than in the south-east section of the outcrop. As, Pb, and Zn contents are two to three times higher than in the south-east. A decrease in Cr, Ni, and V over the southwest part of Llanrwst has been reported and high As and Mo values occur in metal-rich pyritic and black shale horizons within the Crafnant Volcanic and Llanrhychwyn slate series. In the same area the high sediment values of Pb and Zn and of Mn, Co, Ni, and Fe are related to minor mineralization and features of the secondary environment, respectively. These lithological types do not exist in the south-east portion of the Ordovician. The Ordovician west of the Conway river has widespread horizons of both basic and acidic igneous rocks which are essentially absent from the Ordovician in the south-east.

The Silurian rocks consist of interstratified shales, mudstones, siltstones, sandstones, and grits with some local and restricted calcareous facies. Silurian grits and shales occur over and around the Denbigh Moorland and mudstones and shales on the south-east portion of the major outcrop, south of the Clwydian Range. In the stream sediments the mean levels of Cr, Fe, Mn, Ti, and V, and to a lesser extent all other elements present in detectable concentrations, other than Pb, have a higher concentration over the major Silurian outcrop area than over that on the eastern side of the Vale of Clwyd and no major lithological changes are known between the two areas which might originate such features. Mn, Fe, As, Co, and to

a lesser extent, Ni and Zn show concentrations of higher value over the south and south-west sectors of the Silurian, as well as Ordovician rocks, coincident with areas of poorly-drained soils.

The Carboniferous rocks include three lithological units: the Carboniferous Limestone Series (Dinantian), the Millstone Grit Series (Namurian) and the Coal Measures (Westphalian). Both the latter units are of Silesian age. The Dinantian consists of basement beds of shales, sandstones and conglomerates followed by the limestone. The stream sediments exclusively derived from these rocks are characterized by generally low content of Co, Ni, Cr, Ti, V and Fe_2O_3 ; the same element depletion occurs in younger rocks relative to the Silurian and Ordovician. The Namurian consists mainly of shale (Holywell shales) and sandstone, and changes laterally in lithology from shale in the north to sandstone in the south. These rocks appear as small local and restricted outcrops around the Dinantian on its eastern side. The Westphalian consists of shales and sandstones with coal seams. For the purpose of this study the Namurian and the Westphalian are considered as a whole lithologically similar unit referred to as the Silesian. Pb and Zn are enhanced over the Dinantian and Silesian due in part at least to the effects of mineralization and contamination.

The Permo-Triassic rocks consist predominantly of reddish sandstones and pebble beds of the Bunter with local conglomerates and breccias. The stream sediment geochemistry of these rocks is not reported because the outcrops lie in areas generally covered by Quaternary sediments derived from adjoining older formations and the glacial drift. Enhanced contents of Cu, Pb, Zn, Fe, Ti and V in stream sediments over the sub-outcrops of Permo-Triassic occur at the mouth of the Vale of Clwyd relative to comparable rocks in the main area of the Vale.

The Quaternary glacial deposits are derived from two sources. The Irish Sea (Northern) Drift is a reddish silty clay largely derived from Permo-Triassic material. The Welsh till is a dark blue-grey stony loam consisting of material locally derived from the Silurian and Dinantian rocks constituting the adjoining uplands. Considerable thicknesses of fluvio-

glacial materials accumulated in the Vale of Clwyd and the Dee estuary contemporaneous with the retreat of the ice sheets, and the Wrexham Delta terrace being formed at the close of this period.

In general the outstanding regional geochemical feature is the general decrease from Lower Palaeozoic to Upper Palaeozoic along the line of the Silurian/Dinantian contact west of the Vale of Clwyd. This feature is most strongly displayed by V, Cr, Ti and, to a lesser extent, by Fe, Ni, Co, Mn, Cu and Zn.

According to Smith and George (1961, p.40) the Caledonian mineralization of the Lower Palaeozoic rocks (Ordovician-Silurian) of Denbighshire (including the Dolgelley gold belt) is for the most part to be attributed to deep-seated solutions associated with the Ordovician intrusions (Fig. 3-4). Fluids introduced into Ordovician and Silurian strata contained sulphur compounds that precipitated pyrites, sphalerite (blende) and galena. The mineralization of the Silurian rocks in the Trecastell and Llanrwst districts occurs as syngenetic stratiform pyrite deposits (Sherlock, 1918), and as N-S and E-W fracture infillings bearing galena and sphalerite usually within Craftnant Volcanic Series (Archer, 1959). The mineralization of Silurian rocks in the Llanfair-Talhaiarn district contains chalcopyrite with subsidiary sphalerite and galena occurring as fracture infilled lodes crossing sandstone-siltstone strata (Archer, 1959). Gold bearing quartz veins occur in the Clwydian Range Hills (Khaleelee, 1969).

The mineralization in the Carboniferous rocks occurs mainly along faults and fractures that cut the Dinantian and are of much greater importance than the mineralization in the Ordovician-Silurian rocks. The metallic ores occur abundantly along the eastern Dinantian outcrop, particularly at Halkyn Mountain and in the neighbourhood of Minera. The chief minerals are sphalerite and galena, with little silver, and associated copper pyrites, barytes and fluorspars.

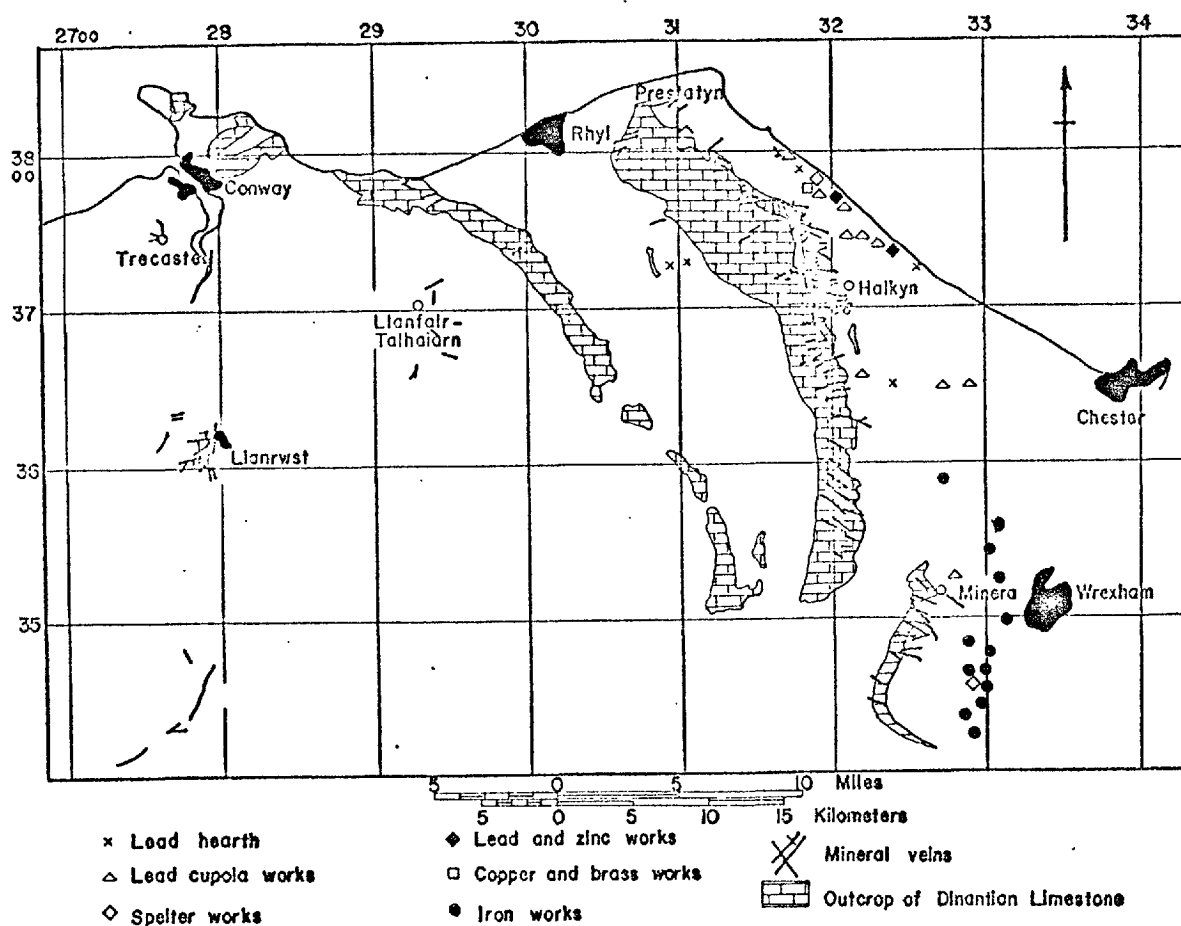


FIG. 3-4

DISTRIBUTION OF PRINCIPAL MINERAL VEINS AND KNOWN SMELTING SITES, DENBIGHSHIRE
(After I.G.S., Nichol and others, 1970b)

Khaleelee (1969, p.51-52) summarized the information concerning the mineralization as follows: 1) the steeply dipping tabular lodes represent infilled open fractures, 2) the principal lode trend is E-W with N-S cross-cut veins of only local importance, and 3) veins exhibit regional zoning; sphalerite is the principal mineral in the north and south, whilst galena is more important in the central region.

As a result of the long continued mining activity since Roman times, most streams in the vicinity of the ore fields are contaminated by mine waste and mill tailings. The history of the smelting is not known in detail but it is probable that contaminated smelter sites are also widespread. Nichol and others (1970b, p.5) gave a good account of the most conspicuous contaminated sites in the study area (Fig.3-4). Geochemically high base-metal content, variously associated with other elements, occur over adjacent areas to the zones of known mineralization. Anomalous Cu, Pb, and Zn together with Mo are recorded in the vicinity of Llanrwst mining district. Anomalous Cu, Pb and Zn characterize the Llanfair-Talhaiarn mineralized area, while extensive Pb-Zn and minor Cu anomalies are associated with the Halkyn-Minera field in the east.

3-2 Empirical Discriminant Analysis.

i) Introduction

It was necessary to use the available geological information regarding the bedrock geology and mineralization for evaluating the application of discriminant analysis to the regional geochemical data of Denbighshire as well as Derbyshire. Limitations of the results are expected because the bedrock geology is mainly represented on the available geological maps by "stratigraphical units" with relatively little detailed lithological control. Therefore, the majority of the training areas chosen over the different mapped geological formations are more representative of mixed lithologies (e.g. the interbedded Silurian sandstones, grits, shales and mudstones). The interpretation of the classification patterns is then carried out on the basis of the "stratigraphical units", but special attention is also directed towards the local lithological types (facies) shown on the recent Institute of Geological Sciences one inch to a mile geological maps. The mineralized areas as a whole do not present problems in their classification because they are sufficiently geochemically distinct from the unmineralized bedrock. In the subsequent analysis (including the cluster-seeking techniques) the term "mineralization" or "mineralized area" must be understood to mean those areas mainly influenced by the mineralized veins, and "artificial contamination" produced by the mining activity (e.g., smelting sites) and relatively farther away from, and less influenced by the mineralized veins.

Thus, the large set of geochemically "well understood" data (895 stream sediment samples) collected over an area of relatively good geological control provided ideal source material for testing the empirical discriminant function (EDF) algorithm as applied to regional geochemical surveys.

In order to carry out the application the following information must be provided by the geochemist as data for the computer programme:

1) the number of variables (elements) to be considered; 2) the number of classes into which it is desired to classify the data set; 3) whether the data are to be untransformed or log-10 transformed before the application of the discriminant analysis (this will depend on the nature of the problem, and is discussed later); 4) the a priori probability of the occurrence of each of the classes specified; 5) since there is a finite probability of a wrong classification being made, a loss value (penalty) incurred for each incorrect decision made is given for each class for the misclassification of a member of that class as a member of another class; 6) a smoothing parameter (σ) for the estimation of the empirical probability density function for each variable in each class, and 7) the boundaries of one or more "training areas" are specified for each class: the program will then automatically find all samples included within these limits and assign them to the training set for the appropriate class.

The algorithm will then classify all the samples on the basis of the relative probability that an unassigned sample resembles each of the defined classes. The sample may be assigned to the "unknown" class, if the absolute probability that the unassigned sample belongs to one of the given categories is below a threshold value. The resulting classification of all the samples, together with the relative probabilities of assignment to each class, are listed in the output and the training samples are distinguished from the rest. A map of the classification results is produced by an off-line Calcomp plotter. In the Denbighshire area the complete operation, including generation of the magnetic tape for off-line plotting, took approximately 41 to 66 sec. of central-processor time for classifying 895 samples on the basis of 12 elements.

ii) Results.

a) Classification based on untransformed data.

Seven training areas were originally selected to represent the major geological formations (classes) and were initially processed

without log-10 transformation of the data. The training areas for the classes of stream sediments derived from the bedrock geology (country rock) were selected by use of 1/4 inch to one mile geological maps (Institute of Geological Sciences; Khaleelee, 1969) such that lithologically (and geochemically) homogeneous geological formations were represented as far as possible. The coordinates for each training area are given in table 3-2. Twelve variables were considered in the analysis: Pb, Ga, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, and Fe_2O_3 . Sn, Ag and As were eliminated because they were not considered "good" discriminators due to their very similar concentrations over all classes, or their virtual absence as occurred with Ag (Table 3-3). Comparison of Khaleelee's (1969) maps for As with the results of the current A.G.R.G. Wolfson Geochemical Atlas for As (1972) over similar areas suggests that the earlier maps were based on seriously erroneous data and for this reason showed little discrimination between rock types.

Since it was not possible to assign realistic a priori probabilities for the occurrence of these classes, nor to weight the penalties for committing a misclassification error unequally, these attributes were made equal for each class throughout the investigation.

The optimum smoothing parameter (σ) was chosen on the basis of the behaviour of the testing set classification results as the smoothing parameter was iteratively increased (Fig.3-5). The number of samples assigned to each class refers to the testing set only, but the derived percentage is relative to the size of the total data set. An optimum result appears to be achieved with σ in the region of 20.0, prior to an abrupt change in slope of the Silurian and Dinantian curves, which suggests that at high σ some overlap between these two classes occurs. The smoothing parameter of 20.0 was therefore chosen in accordance with the stable behaviour of the classes and a suitable small number of unknown samples.

The results of the first classification can be seen in figure 3-6 and the percentage of correct classifications and misclassifications

TABLE 3 - 2

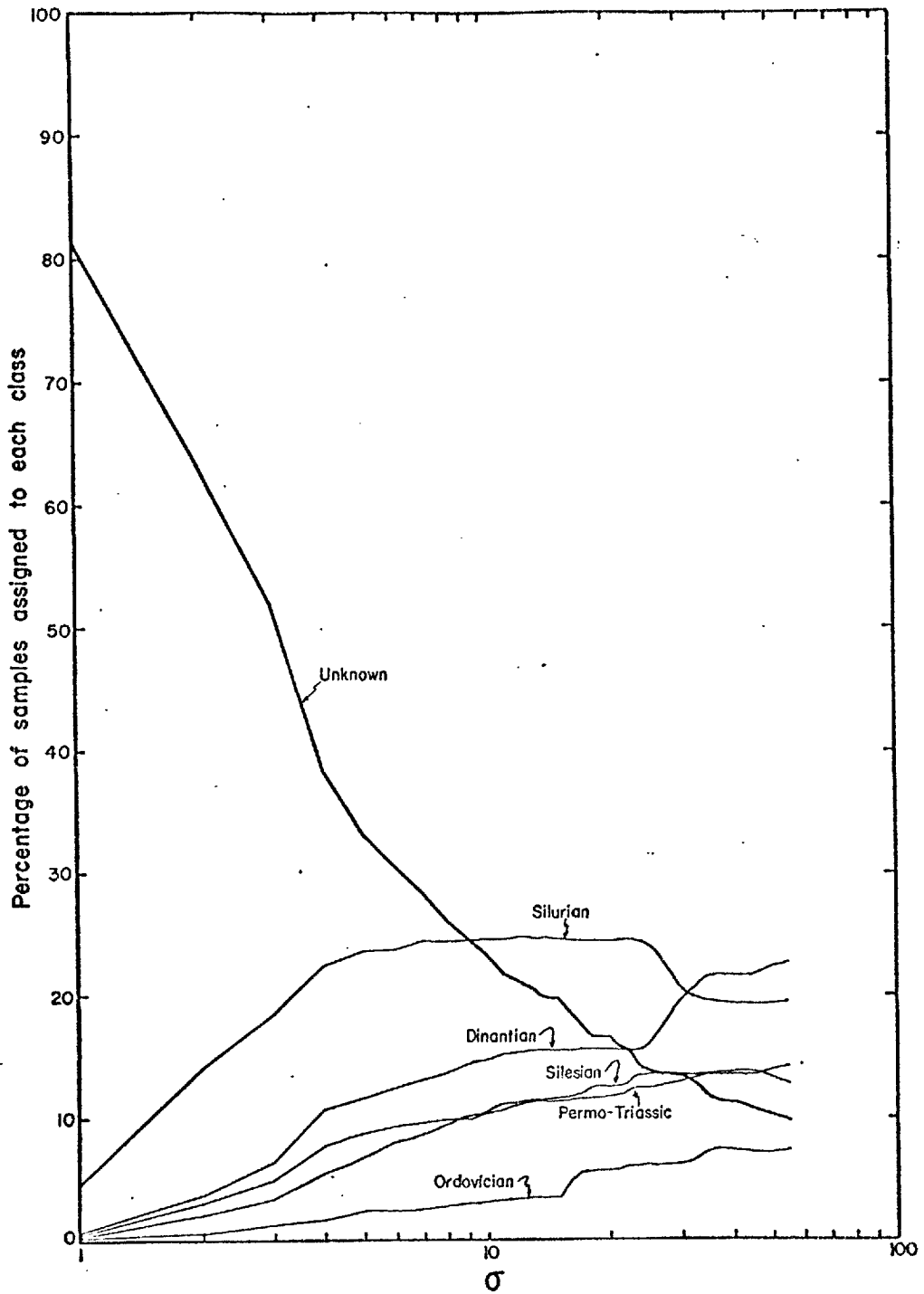
Training area boundaries for bedrock classes

CLASS	WEST	EAST	SOUTH	NORTH	SIZE OF TRAINING SET
ORDOVICIAN	2900	3000	3420	3490	16
SILURIAN	2800	2900	3660	3750	54
DINANTIAN	3090	3130	3750	3780	11
DINANTIAN	3060	3100	3780	3800	
SILESIAN	3220	3270	3550	3600	24
SILESIAN	3240	3270	3600	3700	
PERMO-TRIASS- IC	3060	3100	3635	3700	12

TABLE 3 - 3
Average Training Area Compositions, Denbighshire

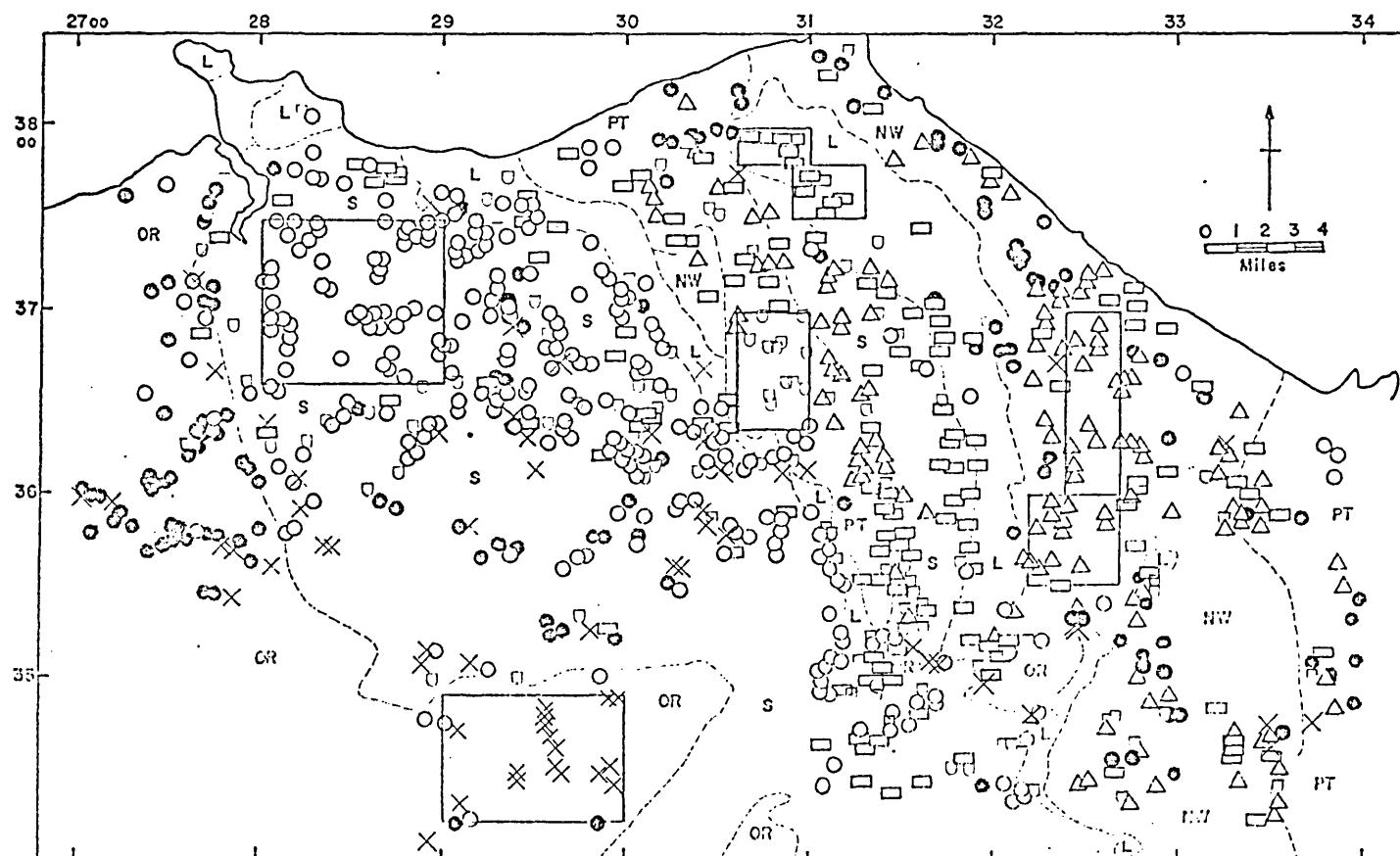
TRAINING AREA	ppm												SAMPLE SIZE
	Pb	Ga	V	Mo*	Cu	Zn	Ti	Ni	Co	Mn	Cr	Fe ₂ O ₃ %	
ORDOVICIAN	36	35	138	1	34	175	4094	71	40	1288	84	6.4	16
SILURIAN	38	28	138	1	39	152	4778	47	24	610	91	3.9	54
DINANTIAN	50	12	63	1	29	155	2273	36	15	332	47	1.9	11
SILESIAN	99	5	32	1	26	121	1225	28	17	311	44	1.1	24
PERMO-TRIASSIC	55	18	88	1	42	162	3108	39	15	413	65	2.9	12

*Mo is included because of its significantly higher concentration (> 1 p.p.m.) in the Llanrwst area (Nichol & others, 1970b, p.8, 12-13).



Variation in overall classification success rate for testing samples with increase in smoothing parameter (σ), Denbighshire

FIG. 3-5



DENDIGHSHIRE
 DISCRIMINANT ANALYSIS
 Variables: Pb, Ga, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, Fe₂O₃
 Untransformed data

GEOLOGY
 PT Permian-Triassic
 NW Silurian
 L Dinantian
 S Silurian
 OR Ordovician

CLASSES
 X Ordovician
 O Silurian
 □ Dinantian
 △ Silurian
 ◻ Permian-Triassic
 ● Unknown
 Training area

FIG. 3-6

for the testing set samples assigned to each class in table 3-4. The following observation can be made:

- 1) a close correlation exists between the samples classified as Silurian and the Silurian bedrock to the west and south of the Vale of Clwyd where 49% of the testing samples were correctly classified. The relatively "good" classification of the major Silurian outcrop is attributable to the homogeneous lithology of the geological formation and to the drainage over the Silurian rocks which favours the transportation of mainly Silurian stream sediments. Those samples assigned to other classes are probably due to a variety of causes. For example, over the Denbigh Uplands those samples classified as Permo-Triassic (and Dinantian) are probably due to the influence of the Permo-Triassic training area which is composed of samples derived from Silurian and Dinantian rocks. Those samples classified as Ordovician, mainly over the Denbighshire Moorland, are probably a response to the Ordovician training area located over Ordovician moorland terrains.

- 2) A general lack of coincidence exists between the rest of the classification patterns and the boundaries of the geological formations which is probably mainly due to the strong masking influence of the areas of mineralization and the widespread presence of glacial drift. However, in the Clwydian Range, over the central part of the Dinantian and on the southern end of the integrated Dinantian-minor Silurian outcrop the samples are mainly classified as Dinantian, whilst over the northern part of the minor Silurian outcrop (Clwydian Range) a predominantly Silesian pattern exists. Samples classified as Dinantian or Silesian appear intermixed (see table 3-4) over the eastern Silesian outcrop and to a lesser extent over the Permo-Triassic of the Dee valley, suggesting a close chemical resemblance between the stream sediments of both geological units. The Permo-Triassic patterns are irregularly distributed throughout the area without being specifically differentiated, probably because the stream sediments in the training area for this category were at least in part, derived from the Dinantian, Silesian and Silurian sources.

Since no "mineralization" classes had been included in the training set, a large quantity of stream sediments were correctly recognized

TABLE 3 - 4

Classification of 565 stream-sediment samples in bedrock category testing set, EDF untransformed data.

TRUE CLASS	ASSIGNED CLASS						Sample Size	% Correct
	Ordovician	Silurian	Dinantian	Silesian	Permo-Triassic	Unknown		
Ordovician	10	12	2	-	1	44	69	14
Major Silurian	26	140	39	-	48	30	283	49
Minor Silurian*	-	4	25	11	10	1	51	49** 8***
Dinantian	-	1	3	3	2	4	13	23
Silesian	3	1	18	20	3	6	51	39
Permo-Triassic	-	9	25	36	11	17	98	11

Minimum overall correct recognition rate = 37% (Including minor Silurian within the Dinantian (Minor Silurian + Dinantian)).

Minimum overall correct recognition rate = 33% (Including minor Silurian within major Silurian (Minor Silurian + major Silurian)).

* Silurian rocks of the Clwydian Range.

**Minor Silurian stream sediments evaluated within the Dinantian because of their geochemical resemblance.

***Minor Silurian stream sediments evaluated within the Silurian.

as not resembling any of the bedrock-derived categories, and were consequently classified as "unknown". The 86 of 138 unknown samples (62%) show a close correlation with the mineralized areas. Other "unknown" samples are related to those areas strongly influenced by the secondary environment. For example, over the Denbigh Moorland, which is covered by poorly-drained peat and peaty gley soils, the secondary environment has produced higher concentration of Mn, Fe, As, Co and, to a lesser extent Ni and Zn.

It was then decided to attempt a classification of the stream sediments in which both the bedrock geology (country rock) and mineralization-derived categories were included. The new mineralization training areas (Table 3-5) were chosen assuming the knowledge of the Pb - Zn - Cu - Mo association that characterizes the western mineralization (Llanrwst, Trecastell, Llanfair - Talhaiarn) and the Pb-Zn association in the east (Halkyn-Minera). The percentage of correct classifications assigned to each class for the testing set is shown in Table 3-6. The classification of the bedrock geology is practically the same as in the previous treatment (Table 3-4), however, those samples classified as "unknown" (some of which were considered to be related to the mineralization) were now assigned to the Pb-Zn class established and some still remained classified as "unknown" mainly regarding the Pb-Zn-Cu-Mo class, probably because of the extrapolation effect of the empirical probability density function (See Appendix 1).

b) Classification based on log-10 transformed data.

In order to try to improve the success of the classification a log-10 transformation of the data was implemented. A smoothing parameter (σ) of .04 was iteratively computed using both bedrock and mineralization classes and the same criteria as before, and the corresponding proportion of samples assigned to each class is shown in Table 3-7. The results of the classification are seen in Figure 3-7. The success of the testing set classification for each class is

TABLE 3 - 5

Training area boundaries for mineralization classes

CLASS	WEST	EAST	SOUTH	NORTH	SIZE OF TRAINING SET
Pb-Zn-Cu-Mo	2730	2800	3595	3650	21
Pb-Zn	3140	3240	3705	3800	25

TABLE 3 - 6

Classification of 659 stream-sediment samples based on bedrock and mineralization training sets, EDF untransformed data.

ASSIGNED CLASS											
TRUE CLASS		Ordovician	Silurian	Dinantian	Silesian	Perm - Triassic	Mineralization Pb-Zn Pb-Zn Cu-Mo		Unknown	Sample Size	% Correct
Ordovician		9	10	1	-	1	4	-	24	49	18
Major Silurian		23	152	34	-	44	3	2	25	283	54
Minor Silurian*		-	3	25	11	9	-	2	1	51	49** 6***
Dinantian		-	1	2	3	2	-	-	4	12	17
Silesian		3	1	18	19	3	-	3	4	51	37
Permo-Triassic		3	7	28	29	11	-	7	13	98	11
Mineral- ization	Pb-Zn	10	9	1	-	1	-	3	21	45	0
	Cu-Mo										
	Pb-Zn	2	1	5	17	6	2	14	23	70	20

Minimum overall correct recognition rate (bedrock) = 40%
 (Minor Silurian + Dinantian) (bedrock and mineralization) = 35%

Minimum overall correct recognition rate (bedrock) = 36%
 (Minor Silurian + Major Silurian) (bedrock and mineralization) = 32%

* Silurian rocks of the Clwydian Range.

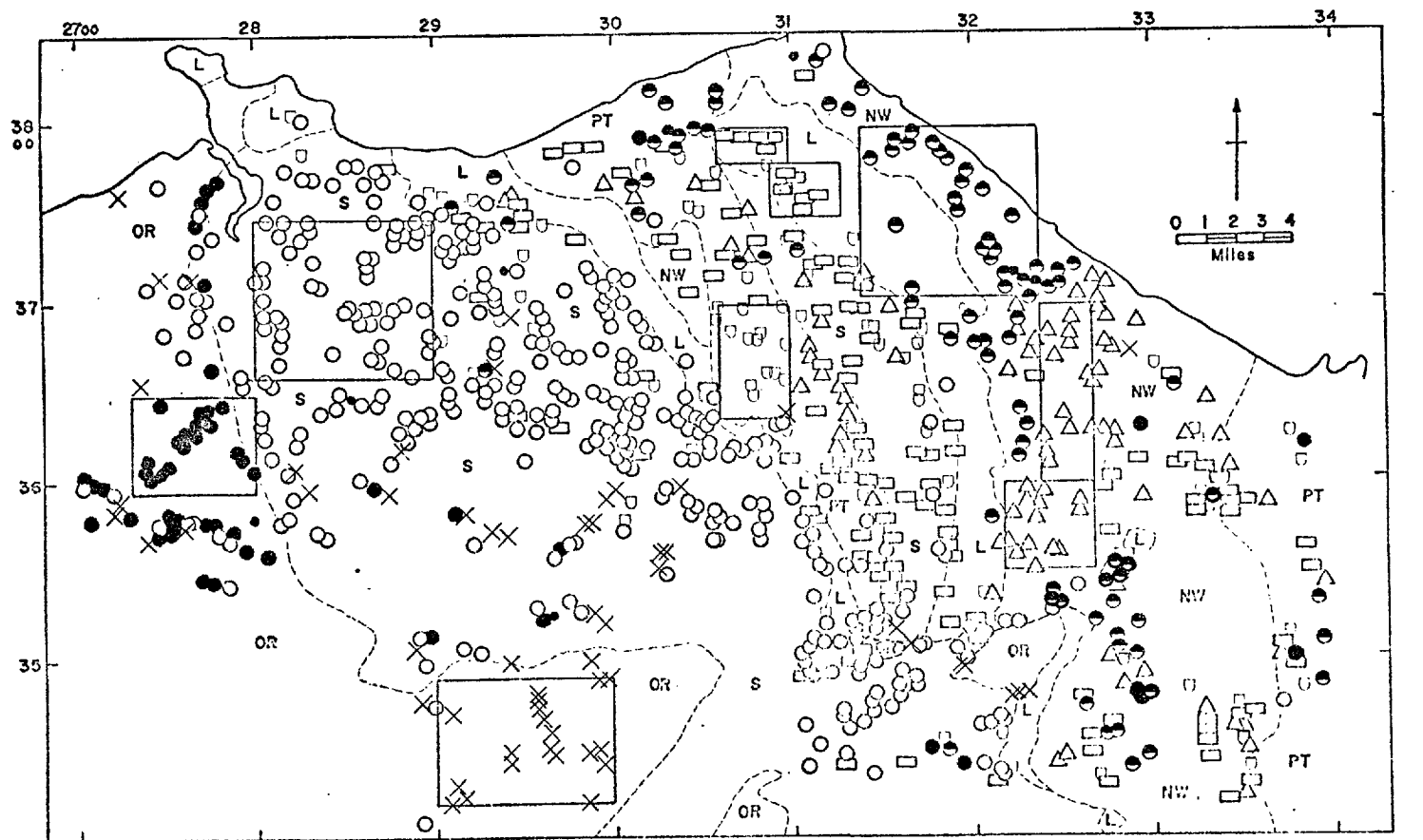
** Minor Silurian stream sediments evaluated within the Dinantian because of their geochemical resemblance.

*** Minor Silurian stream sediments evaluated within the Silurian.

TABLE 3 - 7

Percentage of testing samples assigned to each class

σ	ORDOVICIAN	SILURIAN	DINANTIAN	SILESIAN	PERMO- TRIASSIC	Pb-Zn- Cu-Mo	Pb-Zn	UNKNOWN
.04	5.1	34.0	13.4	8.3	8.5	4.4	7.5	0.7



DENBIGHSHIRE
 DISCRIMINANT ANALYSIS
 Variables: Pb, Ga, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, Fe₂O₃
 Log-10 transformed data

GEOLOGY
 PT Permo-Triassic
 NW Silurian
 L Dinorrian
 S Silurian
 OR Ordovician

CLASSES
 X Ordovician
 O Silurian
 □ Dinorrian
 △ Silurian

□ Permo-Triassic
 ● Mineralization (Pb-Zn-Cu-Mo)
 ● Mineralization (Pb-Zn)
 • Unknown

□ Training area

FIG. 3-7

given in Table 3-8. Compared with the untransformed results (Table 3-6) the classification of the bedrock geology and mineralization showed a significant increase in the successful recognition rate for the testing set for each geological formation, with the exception of the Ordovician and Silesian, with an overall increase of about 20 percent. The phenomenon of apparent misclassification over the minor Silurian outcrop is clearly explained by the fact that the mean chemical composition of these rocks is closely similar to the mean values of the Dinantian (Table 3-1). According to M.J.C. Nutt (1971, I.G.S., Llanrwst Office, personal communication) there are no major lithological changes between the major and minor Silurian outcrops.

Individual patterns corresponding to the Dinantian rocks are this time more clearly shown on and around the contact between the Silesian and the Permo-Triassic in the eastern part of the area around Mold, Chester and Wrexham (N.G.R., 3320-3370/3580-3600) and over the Silesian south of Wrexham (N.G.R., 3325-3370/3425-3475). These patterns might be a response to some local limestone or calcareous outcrops in the vicinity, but field investigation failed to confirm the presence of such material because exposures are extremely poor, even along streams. The similar Dinantian classification patterns around Mold, Chester and Wrexham and south of Wrexham were confirmed by a detailed sampling of stream sediments (Table 3-9). These values and those of the Clwydian Range show a general chemical resemblance confirming the EDF analysis. Comparing these mean stream sediment values with those rocks (3 siltstones and 3 limestones, Table 3-10) collected along the Clwydian Range, it is observed that the mean values of the stream sediments are in general closer to the mean values of the limestone and siltstone samples taken together, or to the siltstone samples alone, than to the mean values characterizing the limestone samples alone. Stream sediments over the Clwydian Range show an

TABLE 3 - 8

Classification of 659 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data.

ASSIGNED CLASS										
TRUE CLASS	Ordovician	Silurian	Dinantian	Silesian	Permo-Triassic	Mineralization		Unknown	Sample Size	% Correct
						Pb-Zn	Pb-Zn			
						Cu-Mo				
Ordovician	8	17	-	-	-	23	-	1	49	16
Major Silurian	-	226	18	3	24	4	5	3	283	80
Minor Silurian*	-	7	24	4	14	-	2	-	51	47**
										14***
Dinantian	-	1	3	2	3	-	2	-	11	27
Silesian	1	-	18	14	10	2	6	-	51	27
Permo-Triassic	-	9	32	20	16	2	19	-	98	16
Mineralization	Pb-Zn	7	14	-	-	23	-	1	45	51
	Cu-Mo	-	-	-	-	-	-	-	-	-
	Pb-Zn	-	2	8	11	2	47	-	70	67

Minimum overall correct recognition rate (bedrock categories) = 53%
 (Minor Silurian + Dinantian) (bedrock mineralization categories) = 55%

Minimum overall correct recognition rate (bedrock categories) = 50%
 (Minor Silurian + major Silurian) (bedrock mineralization categories) = 52%

* Silurian rocks of the Clwydian Range.

** Minor Silurian stream sediments evaluated within the Dinantian because of their geochemical resemblance.

*** Minor Silurian stream sediments evaluated within the Silurian.

TABLE 3 - 9

Mean metal values of stream-sediment samples collected in areas geochemically similar according to the EDF analysis, Denbighshire (Spectrographic analysis).

ELEMENT	AROUND MOLD, CHESTER & WREXHAM*	SOUTH OF WREXHAM**	CLWYDIAN RANGE***
	$\bar{x}$	$\bar{x}$	$\bar{x}$
Pb	386	57	202
Ga	10	10	13
V	38	47	49
Cu	21	20	18
Zn	558	190	379
Ti	1196	1400	1619
Ni	43	35	36
Co	27	14	17
Mn	803	775	374
Cr	39	35	38
Mo	2	2	2

Stream sediment sample locations:

* No. 1, N.G.R., 3237/3526 - 2, 3260/3537 - 3, 3225/3555 - 4, 3250/3558 - 5, 3276/3570 - 6, 3283/3557 - 7, 3297/3585 - 8, 3262/3620 - 9, 3240/3590 - 10, 3310/3625 - 11, 3331/3605 - 12, 3335/3587 - Sample Size = 12

** No. 28, 3355/3457 - 29, 3360/3445 - 30, 3351/3337 - 31, 3348/3423 - 32, 3342/3417 - Sample Size = 5

*** No. 15, 3162/3589 - 16, 3188/3612 - 17, 3189/3627 - 18, 3175/3648 - 19, 3172/3675 - 20, 3140/3709 - 21, 3112/3721 - 22, 3088/3715 - 23, 3079/3746 - 24, 3109/3678 - 25, 3100/3649 - 26, 3125/3615 - 27, 3142/3610 - Sample Size = 13

TABLE 3 - 10

Mean metal values of 3 limestone and 3 siltstone samples, Clwydian Range, Denbighshire (Spectrographic analysis).

ELEMENT	LIMESTONE *	SILTSTONE**	LIMESTONE + SILTSTONE
	$\bar{x}$	$\bar{x}$	$\bar{x}$
Pb	8	43	26
Ga	1	23	12
V	38	95	67
Cu	3	40	21
Zn	100	233	142
Ti	287	3333	2478
Ni	5	73	40
Co	5	19	11
Mn	183	400	292
Cr	17	77	47
Mo	2	2	2

Rock sample locations and descriptions:

* No.7, N.G.R., 3203/3630. Limestone, light grey, unweathered, hard, microfractures filled with calcium carbonate.

No.8, 3172/3675. Limestone, idem as above.

No.10, 3082/3732. Limestone, idem as above.

** No.6, 3160/3585. Siltstone, medium grey, poorly weathered.

No.9, 3135/3715. Siltstone, light grey, medium weathered.

No.11, 3090/3759. Siltstone, idem as above.

abundance of clay fraction, even in those sediments derived from limestone terrains. The abundance of siltstone strata (argillaceous fine-grained sandstone) in the Silesian sequence probably has produced an appreciable proportion of samples assigned to the Dinantian class. In those stream sediments derived from Silesian rocks on the eastern side of the area, a significant amount of clay fraction was observed, probably derived from the interestratified siltstone beds.

Several samples over the Denbigh Moorland, formerly classified as unknown were now mainly classified as Ordovician or, to a lesser extent, as the Pb-Zn-Cu-Mo class. The classification of the latter samples is probably a response to the enhancement of Mn, Fe, As, Co, and to a lesser extent Ni, and Zn, coinciding with areas of poorly drained soils.

Two types of mineralizations are well established in the area and are separated by a contact line between the Silurian and the Carboniferous west of the Vale of Clwyd. These types are: 1) the Pb-Zn association in rocks formed in a Middle Palaeozoic geosyncline in the east disturbed by the Hercynian Orogeny, and 2) the Pb-Zn-Cu-Mo association in rocks formed in a Lower Palaeozoic geosyncline in the west, and disturbed by the Caledonian Orogeny. Nichol and others (1970b, p.4) point out that the mineralization of the area originated during the Hercynian Orogeny, however, Smith and George (1961, p.40 and 73) established a separation in time between both types of mineralization.

Only one sample from the Llanfair-Talhaiarn district was classified as belonging to the Pb-Zn class which characterized the eastern mineralization, and another sample from this district in a lower drainage network was classified as unknown.

Five of the samples along the western flank of the Clwydian Range, formerly assigned to the Pb-Zn class were this time more correctly classified with the adjacent bedrock geology: four to the Dinantian and one to the Permo-Triassic class. The samples classified as belonging to the Pb-Zn class at the mouth of the Clwyd river are

probably due to the mineralization of the Clwydian Range and to the occurrence of variously enhanced contents of Cu, Pb, Zn, Fe, Ti and V in stream sediments over the Permo-Triassic strata (Nichol and others, 1970b).

The map of the distribution of principal base metal lodes and known smelting sites in North Wales (Fig. 3-4) shows that most of the known contaminated areas are distributed at the eastern side of the study area. Taking this feature into consideration, the Pb-Zn training area was eliminated in order to minimize as far as possible the influence of contaminated samples. The resulting percentage of correct recognition for the testing set classes is shown in Table 3-11. The classification of the bedrock geology is very similar to that of Table 3-8, except for the Silesian which suffers an appreciable increase in the percentage of correct classifications owing to a large number of samples, which had been previously assigned to the Pb-Zn class, located east and northeast of Halkyn and east of Minera.

The correct classification of the samples in the mineralized areas was significant in the western mining district of Llanrwst and Trecastell and in the eastern district of Minera. This is probably due to the smaller occurrence of contaminated samples over the areas previously mentioned in contrast with the eastern and northeastern parts of the study area.

The classification of the samples previously assigned to the Pb-Zn class, in the Halkyn district and adjoining areas and along the coast between Flint and Rhyl was not successful. The samples were now either classified as "unknown" or Silesian. However, the samples classified as Silesian are probably unrepresentative due to the northernmost Silesian training area which appears to be strongly influenced by the mineralization and contamination. Only one sample in the area was classified as belonging to the Pb-Zn class. Only two samples at the mouth of the river Clwyd were now classified as mineralized whereas all the samples had been previously assigned to this category. The remaining samples were classified as "unknown".

TABLE 3 - 11

Classification of 681 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data.

TRUE CLASS	ASSIGNED CLASS							Sample Size	% Correct
	Ordovician	Silurian	Dinantian	Silesian	Permo-Triassic	Mineralization	Unknown		
Ordovician	9	17	-	-	-	22	1	49	18
Major Silurian	25	206	20	1	19	8	4	283	73
Minor Silurian*	1	6	25	5	13	-	1	51	49** 12***
Dinantian	-	1	3	4	4	-	1	13	23
Silesian	1	-	17	20	10	2	1	51	39
Permo-Triassic	1	7	35	29	18	3	5	98	18
Mineralization	7	20	8	43	10	28	20	136	21

Minimum overall correct recognition rate (bedrock categories) = 52%
 (Minor Silurian + Dinantian) (bedrock mineralization categories) = 45%

Minimum overall correct recognition rate (bedrock categories) = 48%
 (Minor Silurian + major Silurian) (bedrock mineralization categories) = 43%

* Silurian rocks of the Clwydian Range.

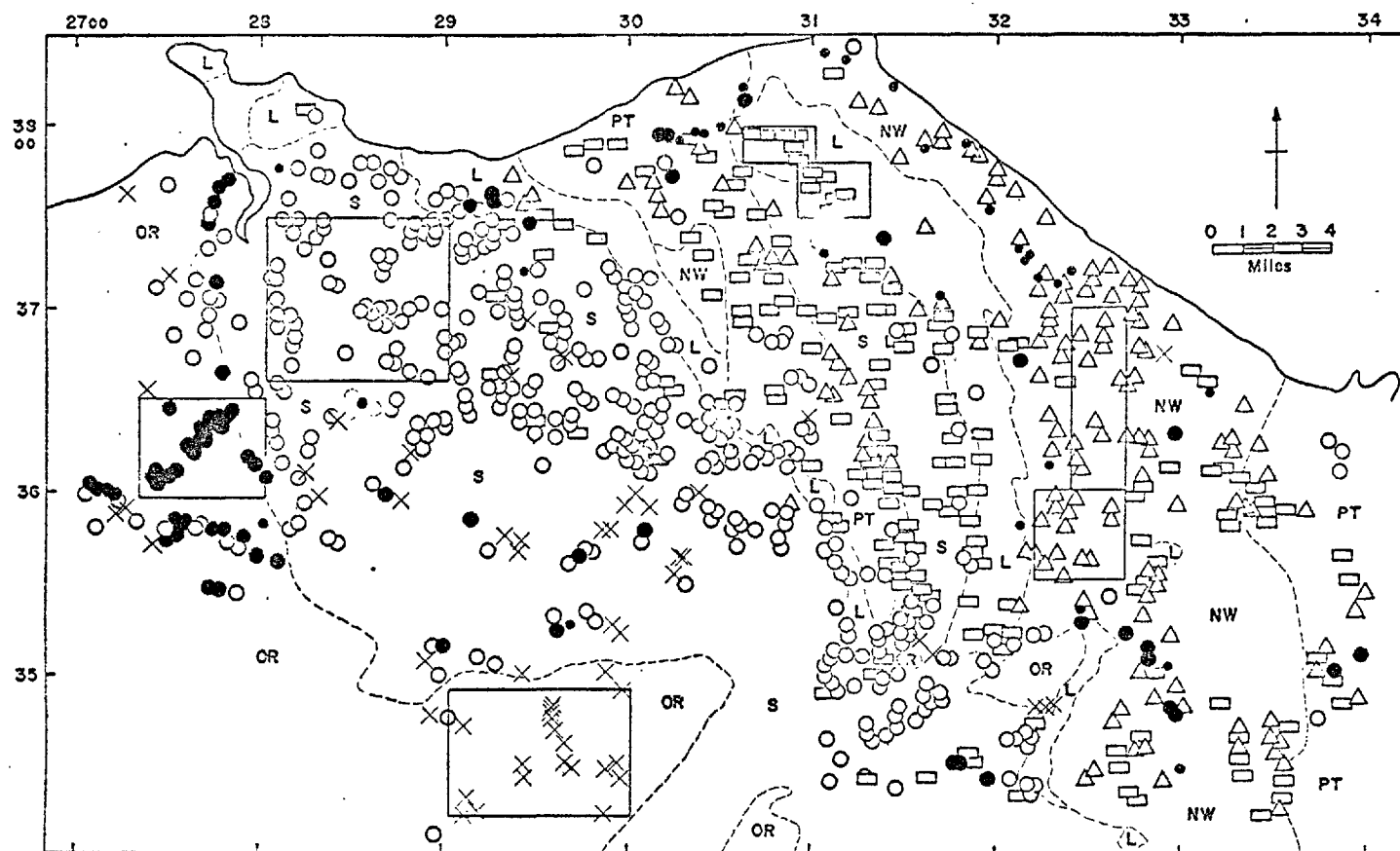
** Minor Silurian stream sediments evaluated within the Dinantian because of their geochemical resemblance.

*** Minor Silurian stream sediments evaluated within the Silurian.

Six samples in the Denbigh Moorland were classified as belonging to the "mineralization" class. This is probably a consequence of secondary environment of the poorly-drained peat and peaty gley soils (Nichol and others, 1970b p.9) discussed earlier.

As the streams sampled over the Permo-Triassic rocks have their headwater catchments in other adjacent formations it was now decided to eliminate the Permo-Triassic training area in the Vale of Clwyd. The results are presented in figure 3-8 and classification results for the testing set are shown in Table 3-12.

The classification of the bedrock geology shows a marked increase in correct recognition for the Ordovician, minor Silurian outcrop, Dinantian and Silesian classes, although the classification patterns (Fig. 3-8) show a poorer correlation with the bedrock geology over areas near mineralization or in the vicinity of other geological formations, older bedrock covered by recent alluvium, or influence of the secondary environment, glacial drift, etc. This low correlation can also be seen in the earlier figures. The bedrock classification patterns in the Vale of Clwyd and the Dee Valley also reflected heterogeneity because the composition of the Quaternary stream sediments overlying the Permo-Triassic rocks is influenced by the upstream geological formations and mineralization. As seen in Figure 3-8 the samples that constituted the Permo-Triassic training area were this time classified as 50% Silurian and 50% Dinantian. Apparently a near-by restricted outcrop of Dinantian rocks (~~3290/5560~~) has not significantly influenced the chemical composition of these samples. The remaining Permo-Triassic testing samples were mainly assigned to the Silurian and Dinantian classes. The results demonstrate that the samples classified as Silurian were mostly derived from the western flank where the drainage mainly transports Silurian sediments, and those classified as Dinantian were mainly derived from the eastern side where rocks of Dinantian composition are exposed.



DENBIGHSHIRE

DISCRIMINANT ANALYSIS

Variables: Pb, Ga, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, Fe₂O₃

Log-10 transformed data

GEOLOGY

PT Permo-Triassic

NW Silurian

L Dinantian

S Silurian

OR Ordovician

CLASSES

× Ordovician

○ Silurian

□ Dinantian

△ Silurian

● Mineralization

• Unknown

□ Training area

FIG. 3-8

TABLE 3 - 12

Classification of 585 stream-sediment samples based on bedrock and mineralization training sets. EDF log-10 transformed data.

TRUE CLASS	ASSIGNED CLASS						Sample Size	% Correct
	Ordovician	Silurian	Dinantian	Silesian	Mineralization	Unknown		
Ordovician	10	18	-	-	20	1	49	20
Major Silurian	26	224	20	3	6	4	283	79
Minor Silurian*	1	11	33	5	-	1	51	65** 22***
Dinantian	-	1	5	5	1	1	13	38
Silesian	1	-	20	27	2	1	51	53
Mineralization	7	20	13	50	25	23	138	18

Minimum overall correct recognition rate (bedrock categories) = 67%
 (Minor Silurian + Dinantian) (bedrock mineralization categories) = 55%

Minimum overall correct recognition rate (bedrock categories) = 62%
 (Minor Silurian + major Silurian) (bedrock mineralization categories) = 52%

* Silurian rocks of the Clwydian Range.

** Minor Silurian stream sediments evaluated within the Dinantian because of their geochemical resemblance.

*** Minor Silurian stream sediments evaluated within the Silurian.

East of the study area, streams draining Silesian rocks have probably been influenced by glacial drift, upstream mineralization, and by Dinantian sediments derived from the west. Samples overlying the Permo-Triassic of the Dee Valley were mainly assigned to the Dinantian and Silesian categories and to a lesser extent to the Silurian or mineralization categories, demonstrating the strong influence of the near-by Carboniferous sources. Samples overlying the Ordovician rocks at the western side of the study area, were mainly classified as mineralized because of the proximity of mineralization and were to a lesser extent classified as Silurian or Ordovician.

The classification patterns over the Dinantian of the Clwydian Range also reflects a certain degree of heterogeneity, probably due to the proximity of the mineralized veins.

The classification patterns related to the mineralized areas show correlation with the western Llanrwst and Trecastell districts and with the Minera district in the east. All the main mineralized areas were in general well detected by the samples classified as "unknown".

Assuming the contamination in the eastern part of the area is due to the past mining activity and the influence of the glacial drift, it was decided to eliminate one of the two Silesian training areas which was believed to be the most influenced by these effects. The training area limited by the coordinates 3240 west; 3270 east; 3700 north; 3600 south was eliminated.

The resulting testing set classifications are shown in Table 3-13. The results for the bedrock geology (except the Silesian) and mineralization categories are similar to those of Table 3-12. A slight but interesting variation of the patterns was seen over the samples that constituted the training area now eliminated. Some of these samples were always classified as Silesian but a significant proportion were now classified as Dinantian; only one sample was assigned to the Silurian category. This suggests that in spite of the extensive influence of mineralization, glacial drift and contamination, the stream sediment samples over the Silesian rocks have a chemical

TABLE 3 - 13

Classification of 594 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data.

TRUE CLASS	ASSIGNED CLASS						Sample Size	% Correct
	Ordovician	Silurian	Dinantian	Silesian	Mineralization	Unknown		
Ordovician	9	18	-	-	21	1	49	18
Major Silurian	26	224	22	-	7	4	283	79
Minor Silurian *	-	9	38	3	-	1	51	75** 18***
Dinantian	-	2	5	4	1	1	13	38
Silesian	1	1	43	14	3	1	63	22
Mineralization	7	21	17	37	29	24	135	21

Minimum overall correct recognition rate (bedrock categories) = 63%
 (Minor Silurian + Dinantian) (bedrock mineralization categories) = 54%

Minimum overall correct recognition rate (bedrock categories) = 57%
 (Minor Silurian + major Silurian) (bedrock mineralization categories) = 49%

* Silurian rocks of the Clwydian Range.

** Minor Silurian stream sediments evaluated within the Dinantian because of their geochemical resemblance.

*** Minor Silurian stream sediments evaluated within the Silurian.

composition very similar to the Dinantian stream sediments. The close resemblance in mean chemical composition between the stream sediments collected over both geological formations is confirmed by Table 3-1.

This large increase in the number of samples classified as Dinantian occurred over the whole of the eastern Silesian outcrop with a consequent decrease in the number of Silesian samples. (see Tables 3-11 and 3-12). This behaviour is perhaps due to an effect of the Dinantian exposures on the nearby Silesian training samples, in spite of the fact that the drainage system would appear to transport only Silesian-derived stream sediments.

When all the training areas over the Silesian rocks were eliminated, the majority of the stream sediment samples over the Silesian were reassigned to the Dinantian class owing to their geochemical resemblance (Table 3-14). With this new classification the whole of the Dinantian as well as the minor Silurian outcrop along the Clwydian Range in the eastern half of the area became perfectly separated from the Silurian of the western half of the area.

The mining districts, mainly Llanrwst, Trecastell and Minera were very well detected, and also to some extent the Halkyn-Prestatyn and adjoining mineralized areas, by "mineralization" patterns with "unknown"-type samples, specially in the last area. The Denbigh Moorland was also covered with some "mineralization"-type samples probably for the reasons discussed before.

iii) Discussion.

If the discriminant analysis results are considered from a quantitative point of view it is possible to assess important information for the practicality of the method. The lithological control was based on the 1/4 inch to a mile Geological Survey map, and since accurate lithological control was not therefore possible, the classification success rate was based on the geological formation type. In the quantitative evaluation of the classification

TABLE 3 - 14

Classification of 715 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data.

ASSIGNED CLASS							
TRUE CLASS	Ordovician	Silurian	Dinantian- Silesian- Minor Silurian	Mineralization	Unknown	Sample Size	% Correct
Ordovician	10	18	-	20	1	49	20
Major Silurian	25	222	21	11	4	283	78
Dinantian- Silesian- Minor Silurian	1	23	178	18	21	241	74
Mineralization	7	24	53	38	20	142	27

Minimum overall correct recognition rate (bedrock categories) = 72%
 (bedrock mineralization categories) = 63%

results, the minor Silurian outcrop has also been considered to be more similar geochemically to the Dinantian than the main area of Silurian rocks.

In the case of the untransformed data (Table 3-4) in which no mineralization classes were included, the minimum overall successful classification rate is 38 per cent if the Dinantian and minor Silurian outcrop are taken together, rising to 42 per cent if the Silesian is added to cover the eastern half of the study area underlain by the Carboniferous rocks. If the minor Silurian outcrop is now considered together with the major Silurian outcrop then the minimum overall successful classification is 35 per cent. No major changes of the minimum overall success rate occur if the minor Silurian outcrop is taken together with either the Silurian or Dinantian classes. The significant change occurs with the very low percentage of samples correctly assigned to the minor Silurian category (8 per cent, Table 3-4) as a response to the non-Silurian chemical composition. Slightly higher minimum overall recognition rates were obtained when "mineralization" training areas were included (Table 3-6). 15 per cent of the whole data set was classified as "unknown" and 62 per cent of these "unknown" samples was found to be correlated with the main mineralized areas; the remaining "unknown" samples were mainly related to contamination on the eastern part of the area and to the secondary environment over the Denbigh Moorland. Although only 62 per cent of the sediments in streams draining known mineralization were classified as "unknown" there is still a fairly good chance of detecting mineralization (if present) in an area of similar geology (with good control of the secondary environment and sources of contamination), for which only unmineralized bedrock training areas were available, provided all samples assigned to the "unknown" class were to be investigated.

In the case of log-10 transformation of the data (Table 3-12) the minimum overall successful classification rate for the bedrock classes is 67 per cent if the Dinantian and the minor Silurian outcrop are taken together, rising to 74 per cent if the Silesian is

added to cover the eastern half of the study area. If the minor Silurian outcrop is now considered together with the major Silurian outcrop then the minimum overall successful classification rate is 62 per cent. The percentage of correct classification (excluding the Permo-Triassic training area) (Table 3-14) rose to 83 per cent for the Lower Palaeozoic bedrock and to 74 per cent for the Upper Palaeozoic, and then the minimum overall correct classification rate for the bedrock also rose to 79 per cent. The success rate for the mineralization over the eastern part of the area was very low (10 per cent) in contrast with the higher rate over the western mineralized areas (49 per cent). The success rate for the mineralization in the eastern mining districts was significantly increased to 47 per cent when a Pb-Zn training area representing that type of mineralization was established.

It is highly probable that some of the apparent misclassifications are correct, but without extremely detailed lithological control the task of checking these would not be possible.

iv) Conclusions

1) The bedrock geology was better classified with the log-10 transformed data than with the untransformed data.

2) The classification assigned samples distributed over different geological formations which are chemically similar to a single category. This illustrates that in the pattern recognition task the resultant classification patterns are not always going to bear a close relationship to the mappable geological units because the patterns are a direct response to the chemical composition of the samples, and therefore constitute real geochemical facies that may or may not correlate with the mapped or mappable geology.

3) With the untransformed data, the mineralized areas were generally detected by either the "unknown" and/or "mineralization" class samples depending on the use of mineralization training areas or not. The mineralization types were better detected with the log-10

transformed data than with the untransformed data when using the appropriate training areas.

3-3 Hierarchical cluster analysis.

i) Method.

The procedure adopted for the Denbighshire area, described below, is also applicable to the Devonshire and Derbyshire data.

Two hundred samples randomly chosen from the entire data set (without replacement) were processed at a time and for each set both untransformed and log-10 transformed data were used. Only one association of variables was considered for each study area. The computer operation using a program to process up to 200 samples took 33 sec. of central-processor time for classifying the complete set of 200 samples on the basis of 12 elements (variables).

The data were further transformed in order to equally weight the variables prior to clustering by dividing the values observed for each variable by the highest value of that variable, thus making the range 0.0-1.0. The distance coefficient matrix was computed for each pair of the 200 samples. In the case of either the untransformed or log-10 transformed data, in the first dendrogram (derived from processing the first group of two hundred samples selected) the more conspicuous groups of samples or "basic groups", were chosen by visual inspection of the dendrogram. This point is returned to below.

This first dendrogram is referred to as the "basic dendrogram" and one can be seen in figure 3-9. Subsequently, forty "linkage samples" (20%) were selected in order to represent the "basic groups" and to provide the basis for the subsequent delimitation of similar groups in successive sets of two hundred samples. These sets therefore consisted of successive groups of 160 new samples and the same

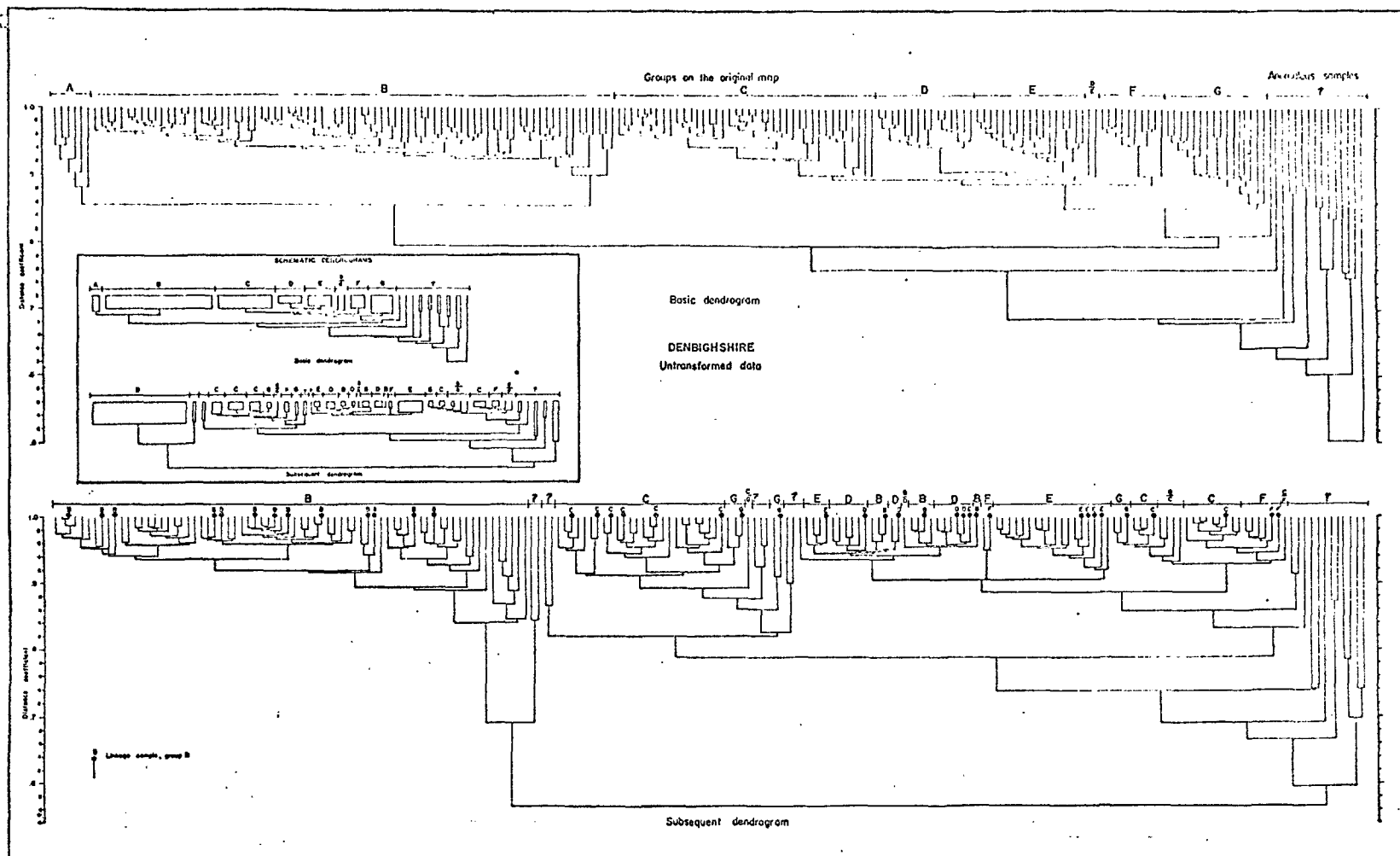


FIG. 3-9

40 linkage samples, and form the "subsequent dendrograms" (Figure 3-9). A generalized or schematic display for each basic and subsequent dendrograms is also seen in Figure 3-9. The distance coefficient scale (vertical dimension) is not shown in the schematic dendrogram due to the necessity of exaggerating the distance scale in places in order to show the linking of the groups correctly in the "basic" and "subsequent" dendrograms. The size of the groups (horizontal dimension) was taken proportional to the number of samples in each group. Specific linkage sample sets were chosen for each one of the untransformed and log-10 transformed data applications.

The procedure adopted in the delimitation of the groups is illustrated in Figure 3-9 and it reflects part of the results for the Denbighshire area. Samples distinguished with a question mark (?) are considered as "anomalous samples" because they were not assigned to any specific group on the basis of the linkage samples. Instead they were joined to more than two "basic groups" and generally at lower similarity levels. The anomalous samples were generally found associated with "anomalous" geochemical environments, as for example, mineralization, contamination and secondary environment. The majority of the remaining samples were found to be commonly associated with the bedrock geology, these groups occupied most of the dendrogram structure and were generally linked at higher levels of similarity. Samples joined at a lower similarity level to composite groups composed of two different groups (on the basis of the linkage samples) were considered as "intermediate samples" (e.g. the groups of samples denoted C/G and C/F in Figure 3-9).

The samples belonging to basic groups, the intermediate samples and the anomalous samples were all plotted on a one inch to a mile geological map and subsequently transferred to a 1/4 inch to a mile map. The few intermediate samples were not plotted on the latter map as a consequence of their small number and because of the difficulties

in their graphical representation, but in spite of this they were considered in the final interpretation of the results.

Because of the noisy nature of the original data, the spatial patterns of the samples in their classification groups were also rather irregular (see for example Figure 3-10) and spatial overlapping of some of the groups occurred. A technique of smoothing by "majority-vote" (moving mode) was carried out, in a similar manner to that used for moving average smoothing, in order to clarify the spatial classifications patterns. The sample classification plot at a scale of 1 in: 1 mile was overlaid by a 1 mile x 1 mile side grid of cells. A moving window of 3 miles x 3 miles was scanned over the base grid with a 1 mile overlap between successive grid positions. The most frequent class present in the window position was taken to be that representing the current location, and this class was plotted at the "centre of gravity" of the sample spatial coordinates. Thus, samples in minority groups were eliminated from the spatial patterns. The anomalous samples were not taken into consideration during the smoothing because they constitute a group of samples significantly different from the rest.

In order to provide statistical estimation of probable misclassification error over all groups identified in each area, EDF analysis of the classification results was carried out in which all training samples (in this case, all samples in their classification groups) are used as testing samples with successive elimination of each sample in order to optimise the estimation of the probable misclassification error.

ii) Q-mode results.

The results from both untransformed and log-10 transformed data are based on 800 samples and on the 12 variables: Pb, Ga, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr and Fe_2O_3 . The distance coefficient in the "basic" dendrograms for the bedrock units ("basic" groups) generally ranged between 0.87 and 0.92 with the exception of the "anomalous" samples. Coefficients ranging between 0.84 and 0.95 were estimated

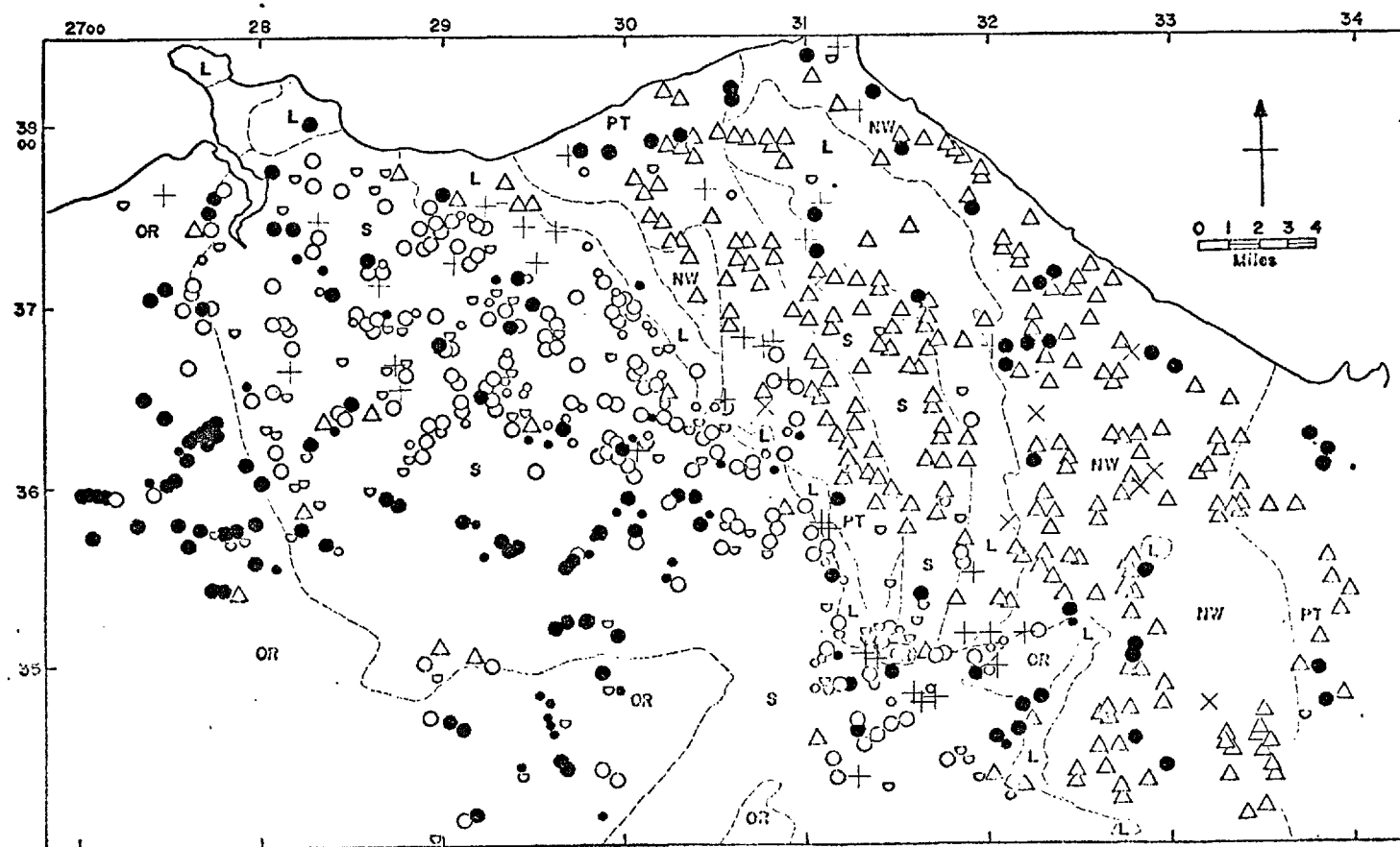
in the "subsequent" dendrograms.

a) Classification based on the untransformed data.

The original classification map for Denbighshire can be seen in Figure 3-10 and the smoothed map in Figure 3-11. The statistical properties of each group are included in Table 3-15.

The whole group B (Fig. 3-11) is distributed over the Carboniferous (Dinantian, Silesian), the minor Silurian outcrop and the Permo-Triassic. Group A in the original map (Fig. 3-10), is similar statistically to group B (Table 3-15), and was eliminated during smoothing. The difference between B and A is due to a higher mean content of Zn, lower mean values of Pb, Mn, Ti, and to a lesser extent Ni, Cu, Co and V. The remaining elements maintain closer or identical mean concentrations (Ga, Mo, Cr, and Fe_2O_3). The EDF results (Table 3-16) clearly show the similarity between both groups. The percentage of correct classifications for both groups taken together rises to 81 per cent. The remaining groups are mainly distributed over the Silurian and Ordovician rocks. Samples collected over the Permo-Triassic, in the Vale of Clwyd and the Dee Valley are stream sediments derived mainly from the adjoining Carboniferous rocks. However, on the western margin of the Vale of Clwyd several stream sediments were assigned to the Silurian groups as a response to the major Silurian outcrop.

Over the southern tip (Fig. 3-10) of the Clwydian Range stream sediments over the Dinantian which are strongly influenced by the minor Silurian outcrop have outlined a Silurian classification pattern (E and C) higher in Ga, V, Cu, Ti, Ni, Co, Mn, Cr and Fe (Table 3-15) relative to the surrounding B pattern, that could suggest a rapid leaching rate over the lower minor Silurian strata, perhaps as a response to the presence of more permeable strata and soils, if the original chemical composition of the minor Silurian outcrop was the same as the major Silurian exposures west of the Vale of Clwyd. More detailed geochemical investigation is needed to explain this restricted geochemical phenomenon which frequently appears in the cluster analysis and NLM (Fig. 3-16) results for this area.



DENDIGHSHIRE
 DENDROGRAM - REPRESENTATION
 ORIGINAL MAP
 Variables: Pb, Ga, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, Fe₂O₃
 Untransformed data

GEOLOGY
 PT Permian-Triassic
 NW Silurian
 L Devonian
 S Silurian
 OR Ordovician

GROUPS
 X A
 △ B
 ○ C
 + D
 □ E
 ○ F
 • G
 ● Anomalous samples

FIG. 3-10

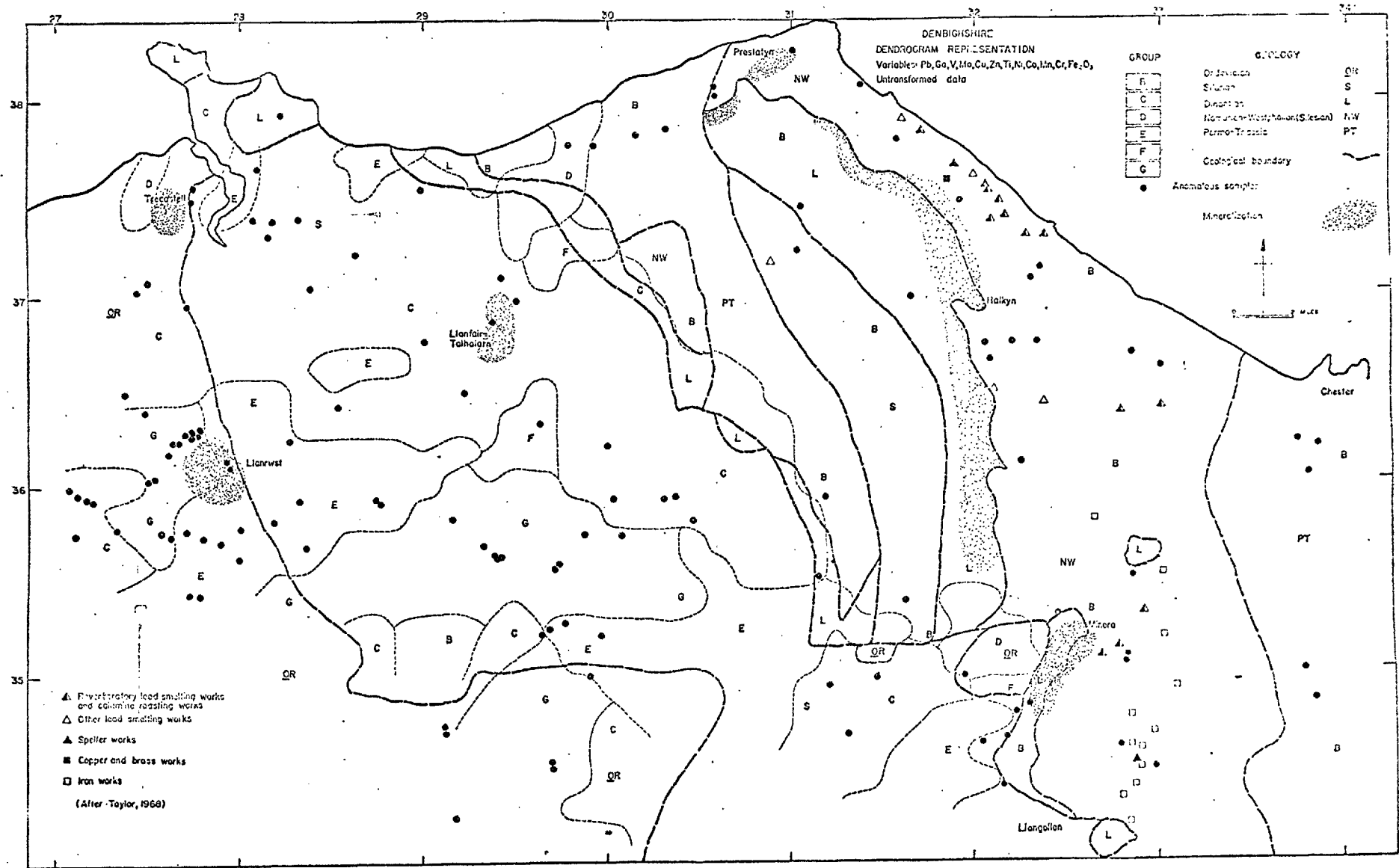


FIG. 3-11

TABLE 3 - 15

Arithmetic means ($\bar{x}$) and standard deviations (s), Denbighshire, dendrogram representation, untransformed data; total sample size = 748.

GROUPS	A	B	C	D	E	F	G	Anomalous Samples
Sample Size	8	278	152	33	84	37	36	120
Pb $\bar{x}$	321	227	63	80	64	97	51	937
s	466	586	140	172	114	320	27	2243
Ga	11 4	10 6	31 4	18 3	29 4	20 3	31 5	25 10
V	50 20	42 20	200 610	97 8	96 16	150 0	146 23	116 50
Mo	1 0.3	1 0.3	2 11	1 0	1 0.1	1 1	1 2	4 9
Cu	60 15	26 14	44 62	32 9	37 12	31 8	36 13	61 59
Zn	246 125	312 997	295 1431	171 59	285 638	177 62	253 192	1881 5785
Ti	1975 897	1772 1011	4641 1135	3818 833	3821 762	3986 642	4611 1208	4283 2059
Ni	45 10	30 9	56 120	40 7	44 11	39 5	70 28	58 33
Co	18 7	14 10	23 9	19 7	24 10	19 9	38 18	58 112
Mn	481 278	411 474	878 4495	383 242	773 941	447 394	1469 1430	2973 4437
Cr	49 17	49 23	197 1276	66 18	82 28	83 13	102 55	105 88
Fe ₂ O ₃	2 1	2 1	4 1	3 1	3 1	3 1	5 2	5 3

Mean values in p.p.m., except Fe₂O₃(%).

Over the north and southeast part of the major Silurian outcrop, where well-drained brown earth soils have developed, group C predominates. In the smoothed map (Fig. 3-11) group C covers a larger surface area whilst groups E and, to a lesser extent, F constitute marginal groups, both generally with lower metal values (Table 3-15) by comparison with group C. Groups C and E could be considered as a single group representative of the Silurian bedrock covered by well-drained brown earth soils due to their mean chemical similarity. These groups are apparently not influenced by secondary environment and mineralization. However, this association is not confirmed by the EDF analysis (Table 3-16) probably because of the higher variability of the data in group C; instead group C appears to be more associated with group F but no obvious explanation can be given.

South and southwest of the major Silurian outcrop the distribution of some groups (particularly G) is clearer. This group is distributed over the zone influenced by the Llanrwst mineralization and by the western Ordovician exposures (Fig. 3-11). The group shows significant concentrations in the elements characteristic of the secondary environment: Fe, Mn, Co and Ni (Table 3-15). The EDF analysis results (Table 3-16) show group G to be strongly associated with the group of "anomalous" samples.

Group D, scattered over different parts of the area, shows a general depletion of Ga, V, Cu, Zn, Ti, Ni, Co, Mn and Cr, compared with groups C and E of the Silurian, perhaps due to some differential leaching on the periphery of the higher ground. According to the EDF analysis results (Table 3-16) this group is closely associated with group E, probably because of the closer mean values mainly of V, Mo, Cu, Ti, Ni, Co and Fe_2O_3 and similar variability in both groups of data (Table 3-15).

Samples classified as "anomalous" showed correlation with the main mining districts and with the terrains influenced by the nature of the widespread secondary environment over the Silurian and Ordovician rocks.

TABLE 3 - 16

Group classification results by EDF analysis, Denbighshire, dendrogram representation, untransformed data results, $\sigma = 20.0$.

ASSIGNED GROUP											
	A	B	C	D	E	F	G	Anom.	?	% Correct	
TRUE GROUP	A	2	5	0	0	0	0	0	0	1	25
	B	8	218	0	14	16	1	5	7	9	78
	C	0	2	84	3	2	48	8	4	1	55
	D	0	3	2	16	10	1	0	1	0	48
	E	0	10	12	15	36	2	4	2	3	43
	F	0	1	12	3	2	18	1	0	0	49
	G	0	2	5	1	2	2	11	10	3	31
	Anom.	1	12	12	2	4	7	15	27	40	23

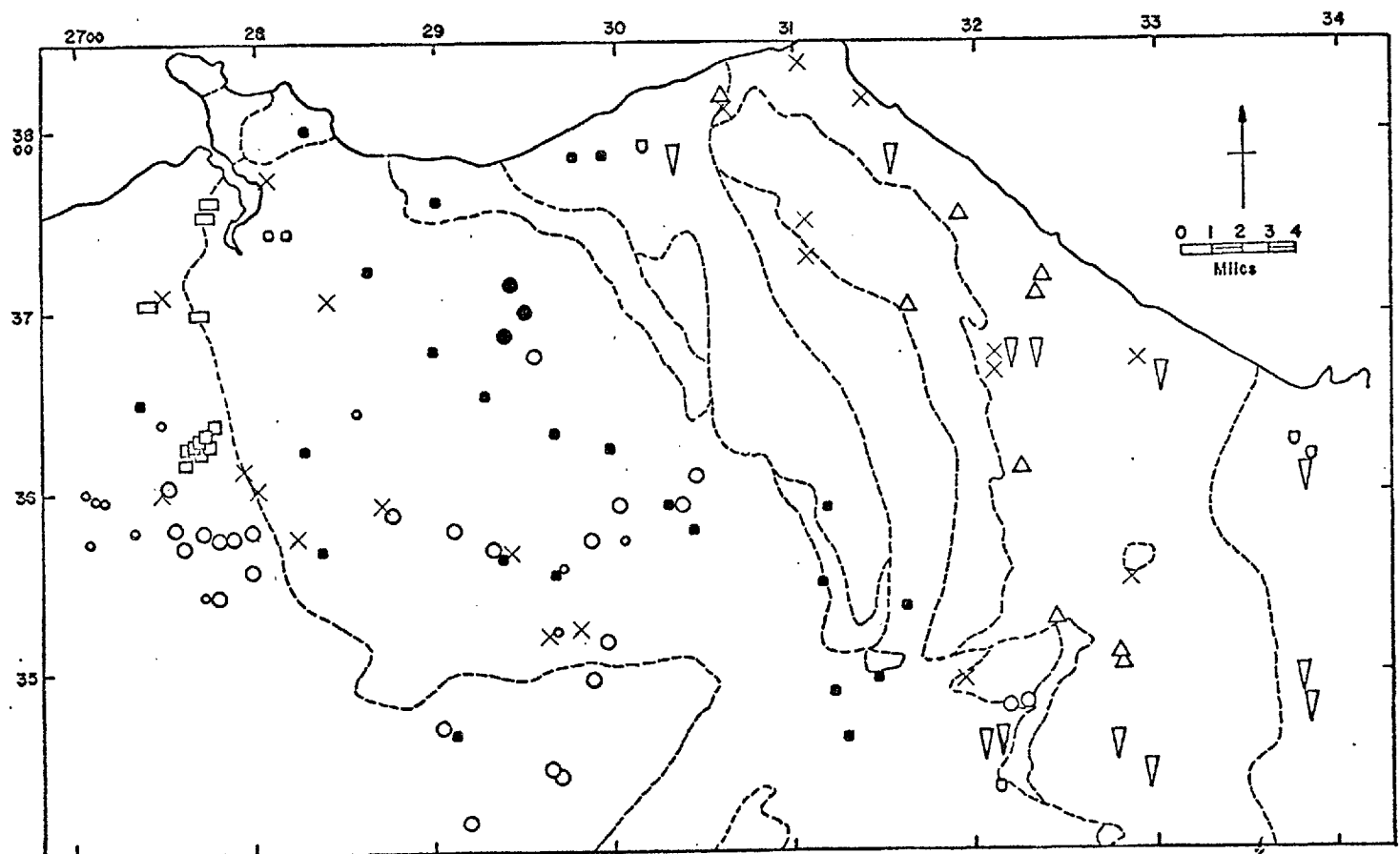
Overall correct recognition rate
(Excluding anomalous group) = 61%

Overall correct recognition rate
(For all groups) = 55%

The group of "anomalous" samples shows the highest mean metal values for Pb, Mo, Cu, Zn, Co, and Mn, and to a lesser extent Fe, all representing either mineralization or secondary environment conditions, and with the highest variability for Pb, Ga, Mo, Zn, Ti and Co (Table 3-15). Again the EDF analysis results (Table 3-16) confirm the close association with group G, both found associated with geochemically anomalous areas. The erratic "anomalous" samples mainly distributed over well-drained brown earth soils overlying the Silurian are of uncertain origin; individually most of the samples showed chemical compositions generally closer to the composition of the stream sediments derived from the Silurian bedrock. Those "anomalous" samples over the Dee Valley showed a generally higher content of V, Cu, Zn, Ti, Ni, Mn, Cr and Fe compared with the average chemical composition of the Carboniferous bedrock. Within the group of "anomalous" samples, 40 (33%) were classified by the EDF analysis as "unknown" and they are probably more representative of geochemically anomalous areas (e.g. mineralization and secondary environment).

Subsequently, classification of the group of "anomalous" samples alone was carried out. Sixteen groups were found, and of these the ten most important are shown in Figure 3-12. Groups "b", "h" and "i" are mainly distributed over the Ordovician rocks, surrounding the Llanrwst mining district although "b" and "i" also occur over the Silurian of the Denbigh Moorland. Groups "b" and "i" may therefore be more influenced by the nature of the secondary environment compared with group "h". Group "h" is more representative of the Llanrwst mineralization. Group "c" shows a close connection with the Trecastell district and group "e" with the Llanfair-Talhaearn district.

The restricted but irregular distribution of groups "f", "g" and "d" in the eastern part of the area is explained by the occurrence of a different type of mineralization from that of the west and contamination by mining and smelting activities.



DENDROGRAM REPRESENTATION
CLASSIFICATION OF ANOMALOUS SAMPLES
Variables: Pb, Ga, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, Fe_2O_3
Untransformed data

- GROUPS
- a
 - b
 - c
 - ◻ d
 - e
 - △ f
 - ▽ g
 - ◻ h
 - i
 - × Samples constituting groups of very small size

FIG. 3-12

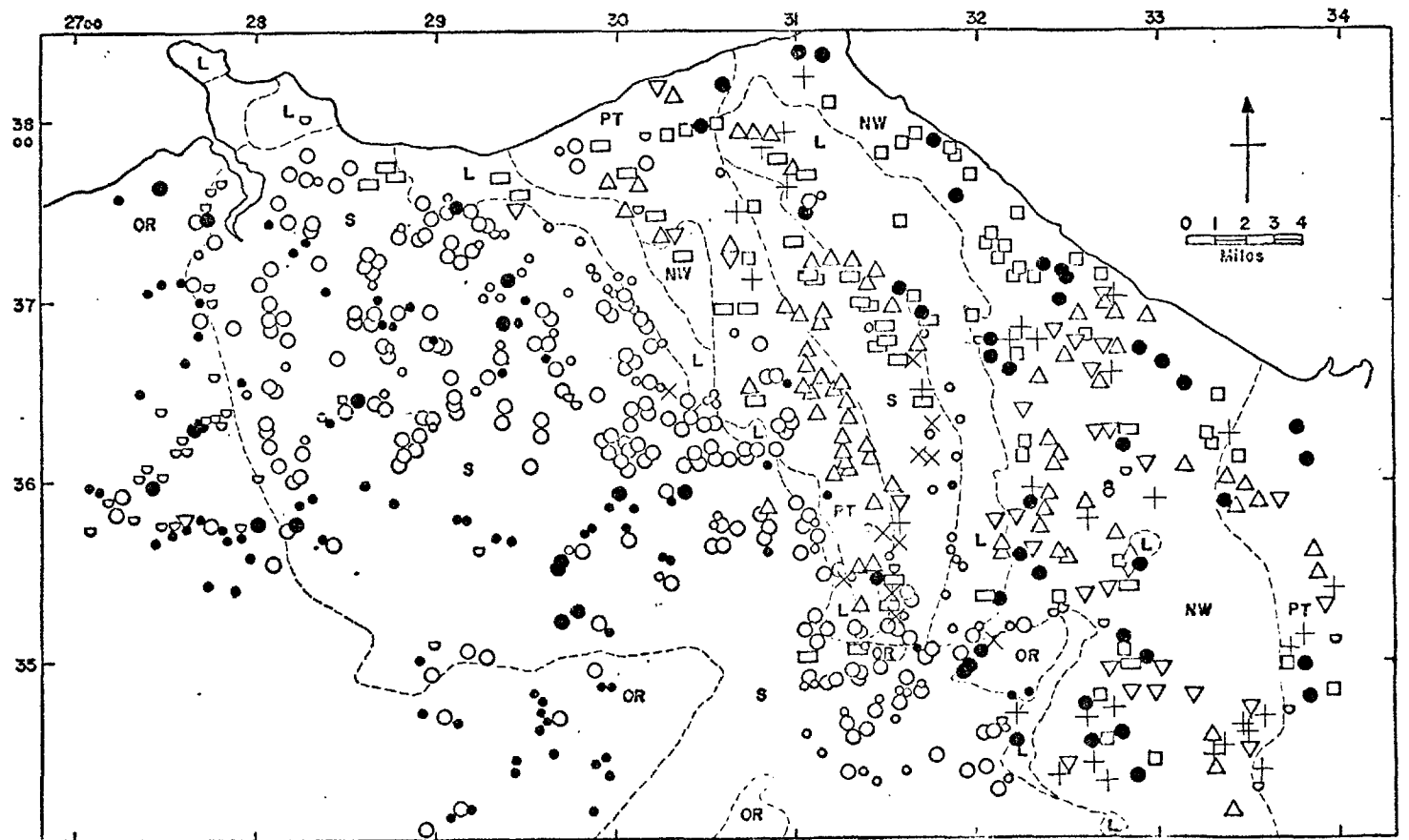
b) Classification based on the log-10 transformed data.

The original classification map can be seen in Figure 3-13 and the smoothed map in Figure 3-14. Group D was eliminated during "smoothing" because of its relatively small size. The statistical properties of the groups are seen in Table 3-17.

The Silurian-Ordovician sequence is mainly covered by groups, B, C, F and G, whilst the Carboniferous (Dinantian, Silesian) and Permo-Triassic sequences are represented by groups A, D, I, J and K. Group E appears to be a transitional group. These associations are in general also shown by the EDF analysis (Table 3-18). The "anomalous" samples are irregularly distributed throughout the area, but show a better correlation with the eastern mining districts (Halkyn, Minera). The mining districts of the western part (Llanrwst, Trecastell, and Llanfair-Talhaiaim) were not well shown by the presence of "anomalous" samples.

The northern half of the major Silurian outcrop is predominantly covered by group C and to a lesser extent by group B (Fig. 3-14); the proportion of groups C and B is equal in the southeastern part of the same major outcrop. The close association between both groups is well shown by the EDF analysis (Table 3-18). Groups F and G with higher contents of Mn, Fe, Co, and to a lesser degree Ni, Cu and Ti (Table 3-17) occur on the southern half of the major Silurian outcrop and on the southern and western Ordovician exposures, together with a few "intermediate" samples A/I which were eliminated during the smoothing procedure. Groups C and B, are lower in most of the elements (Table 3-16) compared to groups F and G, and coincide with the occurrence of well-drained brown earth soils in the northern and southeastern parts of the major Silurian outcrop, whilst in the southern and southwestern parts of the area groups F, G and a high proportion of "anomalous" samples correspond to poorly-drained peat and peaty gley soils.

The distribution of groups C and B on the major Silurian outcrop are well shown by the smoothed map (Fig. 3-14); C occupies most of the outcrop, whilst B is distributed on the margins: one classific-



DENDIGHSHIRE
 DENDROGRAM REPRESENTATION
 ORIGINAL MAP
 Variables: Pb, Ga, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, Fe₂O₃
 Log-10 transformed data

GEOLOGY
 PT Permian-Triassic
 NW Silurian
 L Devonian
 S Silurian
 OR Ordovician

GROUPS
 ▽ A
 ○ B
 ○ C
 X D
 □ E
 • F
 ○ G
 □ I
 + J
 △ K
 ● Anomalous samples

Note: Group H is not plotted

FIG. 3-13

TABLE 3 - 17

Arithmetic means ($\bar{x}$) and standard deviations (s), Denbighshire, dendrogram representation, log-10 transformed data; total sample size = 768.

GROUP	A	B	C	D	E	F	G	I	J	K	Anomalous Samples
Sample Size	44	94	236	11	40	85	43	38	37	82	58
Pb $\bar{x}$ (log)	2.1	1.5	1.6	1.6	1.6	1.7	2.3	2.7	1.8	1.6	2.5
s (log)	0.4	0.2	0.3	0.1	0.3	0.2	0.6	0.6	0.4	0.3	0.8
Ga	1.0	1.4	1.4	1.3	1.2	1.5	1.4	0.9	1.0	0.7	1.1
	0.2	0.1	0.2	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.3
V	1.7	2.0	2.1	1.8	1.8	2.1	2.0	1.5	1.5	1.5	1.8
	0.2	0.1	0.2	0.1	0.2	0.2	0.2	0.1	0.1	0.2	0.3
Mo	0.02	0.0	0.01	0.0	0.0	0.0	0.6	0.02	0.0	0.0	0.1
	0.1	0.03	0.05	0.0	0.0	0.03	0.5	0.1	0.0	0.0	0.3
Cu	1.5	1.5	1.6	1.4	1.5	1.6	1.7	1.5	1.4	1.2	1.7
	0.2	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.3
Zn	2.2	2.2	2.3	2.3	2.1	2.3	2.7	2.7	2.1	1.9	2.6
	0.3	0.2	0.2	0.2	0.2	0.2	0.5	0.9	0.6	0.2	0.8
Ti	3.2	3.6	3.6	3.4	3.4	3.7	3.7	3.1	3.2	3.0	3.4
	0.3	0.1	0.2	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.3
Ni	1.5	1.6	1.7	1.6	1.6	1.7	1.7	1.5	1.5	1.4	1.7
	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2
Co	1.0	1.2	1.4	1.2	1.1	1.6	1.5	1.1	1.0	0.9	1.4
	0.2	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.4
Mn	2.6	2.5	2.7	2.5	2.3	3.2	3.1	2.5	2.5	2.2	2.8
	0.3	0.2	0.3	0.3	0.4	0.4	0.5	0.3	0.3	0.3	0.5
Cr	1.7	1.9	1.9	1.7	1.7	2.0	1.9	1.7	1.7	1.5	1.8
	0.1	0.1	0.2	0.1	0.3	0.2	0.2	0.2	0.2	0.2	0.3
Fe ₂ O ₃	0.2	0.4	0.5	0.4	0.2	1.0	0.7	0.2	0.1	-0.01	0.4
	0.2	0.06	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.3

Mean values in p.p.m. (log), except Fe₂O₃ (log%)

TABLE 3 - 18

Group classification results by EDF analysis, Denbighshire, dendrogram representation, log-10 transformed data results, $\sigma = 0.04$.

ASSIGNED GROUP														
	A	B	C	D	E	F	G	I	J	K	ANOM.	?	% CORRECT	
TRUE GROUP	A	16	0	0	1	1	1	2	4	5	10	4	0	36
	B	0	66	22	0	3	1	2	0	0	0	0	0	70
	C	0	28	162	0	2	24	6	0	2	1	10	2	69
	D	0	3	1	2	5	0	0	0	0	0	0	0	18
	E	2	5	1	5	20	0	0	1	4	2	0	0	50
	F	2	0	25	0	0	53	1	0	0	0	4	0	62
	G	2	1	6	0	0	2	29	0	0	0	3	0	67
	I	4	0	0	1	0	0	0	19	6	0	8	0	50
	J	5	0	0	0	4	0	0	6	11	11	0	0	30
	K	3	0	0	0	2	0	0	0	10	67	0	0	82
Anom.	6	2	12	0	3	6	5	5	2	2	13	2	22	

Overall correct recognition rate
(Excluding anomalous group) = 63%

Overall correct recognition rate
(For all groups) = 60%

ation pattern immediately northeast of Llanfair-Talhaiarn district and the other at the end of the southeastern corner crossed by the Dee river. The available geological information (Smith and George, 1961) does not mention the existence of similar lithologies in both areas, suggesting that similarity is due to the chemical composition of the stream sediments alone. Over the area of poorly-drained peat and peaty gley soils, groups F and G show a very well defined distribution. Group F is mainly associated with the Denbigh Moorland on the Silurian and Ordovician rocks, extending over the western Ordovician exposures in stream sediment samples collected along Afon Llugwy river (2730-2790/3530-3575) southwest of Llanrwst and in streams draining terrains west of the Conway river (2730-2770/3650-3710). Group G is composed of stream sediments collected along a tributary of the Conway river, draining the lakes Llyn Craftnant and Llyn Geirionydd and is apparently strongly influenced by secondary environment, molybdeniferous shales and adjoining mineralization. Group G has a significantly higher concentration of Pb, Zn and Mo and to a lesser extent Cu compared with group F. The EDF analysis (Table 3-18) has shown a substantial dissimilarity between groups F and G probably caused by the higher Mo content in Group G over the area of molybdeniferous black pyritic shales within the Craftnant Volcanic and Llanrhychwyn slates series southwest of Llanrwst. Group G samples near the mouth of Conway river are stream sediments collected along a tributary of the Conway river draining Ordovician bedrock influenced by the Trecastell mineralization.

The clear delimitation of groups F and G west of Llanrwst shows approximately the same results obtained with the classification of the "anomalous" samples mentioned earlier and shown in Figure 3-12. Comparing both results, a close similarity can be seen between group F and groups "b" and "i" of Figure 3-12 which probably represent the influence of the secondary environment conditions; and between group G and "h" reflecting a geochemical environment strongly influenced by the mineralized Ordovician bedrock of Llanrwst district. The "anomalous" samples are mainly associated with group G west of Llanrwst and Trecastell whereas on the Denbigh Moorland they are associated with the prevailing group F.

In the eastern half of the study area group K shows a close correlation with the northern part of the Dinantian, with the minor

Silurian outcrop along the Clwydian Range, and with the alluvial sediments derived from both geological units and subsequently deposited over the Permo-Triassic rocks of the Vale of Clwyd. This distribution could be expected because of the chemical similarity between both formations. Similar results were obtained with discriminant analysis.

As with the untransformed data, in the southern crest of the Clwydian Range (3170-3190/3530-3660) classification pattern, group B is associated with the Silurian. Stream sediments comprising this group are mainly influenced by the minor Silurian outcrop and to a lesser extent by the Dinantian. The higher metal content for Ga, Cu, Zn, Ti, Ni, Co, Mn, Cr and Fe (Table 3-16) of group B at the top of the crest compared to the relatively lower content of the stream sediments of group K derived from the lower strata of the minor Silurian outcrop, seems to confirm the mechanism of differential leaching suggested earlier, perhaps as a response to the presence of more permeable strata and soils in the lower levels of the minor Silurian outcrop.

With the EDF analysis (Table 3-18) groups K and J have shown a very close similarity which is not explicable by any obvious geological cause, except that they are closely associated in chemical composition (Table 3-17).

Groups A, I and J are mainly associated with mineralization and contamination over the bedrock of the eastern part of the area (Fig. 3-14). Group I is also associated with several "anomalous" samples and constitutes a well defined classification pattern on the north-east side of the Clwydian Range, closely correlated with the Halkyn-Prestatyn districts and their smelting sites. The same correlation can be made in Minera, southwest and west of Prestatyn and to a smaller extent along the middle of the eastern Dinantian margin on the Clwydian Range. A higher Pb and Zn content (Table 3-17) compared to the chemically similar groups A and J suggests a major influence of mineralization. The "anomalous" samples also show a significantly higher content of Zn, Pb, Cu and Mo, elements associated with mineralization in groups I and of the "anomalous" samples; they also

show the higher degree of variability of the metal content. The EDF analysis (Table 3-18) shows the closer association between groups I and the group of "anomalous" samples. Groups A and J also show a close connection with the mineralized areas mainly near the Halkyn and Minera districts. However, the same groups are also distributed farther east than the mineralization and are probably more influenced by industrial contamination. For example, where groups I and A occur in an area strongly contaminated by smelter sites west of Chester. The taxonomic proximity between groups A and J is shown by the EDF analysis (Table 3-18) as well as a lesser similarity with the group of "anomalous" samples which in this case seem to be more connected with mineralization features; this would support the previously proposed stronger influence of contamination than mineralization.

Subsequently the classification of the "anomalous" samples derived from the original untransformed results was carried out. The classification results were mainly erratic over the Silurian and Ordovician exposures; only one specific group delimited the area northwest of Llanrwst (2720-2780/3600-3640) and another the Denbigh Moorland over the Silurian. On the eastern side a specific group was associated with the mineralization, but the rest of the groups were also observed scattered on the western side of the study area.

iii) R-mode results.

The R-mode results for the variable associations have been interpreted by studying each 200 sample R-mode dendrogram independently. The most common associations were considered together, thus eliminating to some extent sampling variability, and are described below.

With the untransformed data, the most conspicuous associations were: Pb-Zn, Ga-V-Ti, and the Co-Mn association which is closely associated with the remaining elements: Mo, Cu, Cr, Ni and Fe.

With the log-10 transformed data the associations varied substantially, with the exception of the Ga-V-Ti association. A new Ni-Cr association frequently occurred. The Pb-Zn association was not clearly defined, and was replaced by a less common Pb-Zn-Co-Cu-Fe association. The remaining elements showed different associations in each one of the dendrograms and were impossible to integrate in a single association. Mo generally behaved as an independent element but was slightly associated with Fe in some dendrograms.

If the results are compared with the R-mode factor analysis results obtained by Khaleelee (1969, p.295-299) the following can be pointed out: 1) The Pb-Zn association, due to mineralization and contamination, was mainly defined with the untransformed data and was associated with Co, Cu and Fe in the log-10 transformed data; 2) the new Ga-V-Ti association found by Khaleelee over the Lower Palaeozoic was also detected both with the untransformed data and log-10 transformed data, and 3) the Co-Mn group which is also associated with Mo, Cu, Cr, Ni and Fe in the untransformed data dendrograms, apparently coincides with the As-Co-Fe-Mn and Ni and Zn associations reported by Khaleelee and appears to be related to the secondary environment. The Ni-Cr association of the log-10 transformed data dendrograms are difficult to correlate with the factor analysis results, however, such association could represent the known association of high Cr and Ni over the Silurian-basal Dinantian contact near Llangollen (Nichol and others, 1970b p.9). The observed instability of the other associations suggests that they could be artificial products of the sampling, and may not therefore be significant in the factor analysis.

iv) Discussion.

The dendrogram representation has proved to be effective for the location of geochemically anomalous environments (e.g. mineralizations, contaminations and secondary environment), whereas it has not been so successful in the delimitation of the main bedrock units as it has been restricted to broad geological units (e.g. the Lower Palaeozoic sequence from the Upper Palaeozoic sequence) if no a priori geological knowledge is taken into consideration.

In the untransformed data results 120 samples (15%) of the whole data set were considered as "anomalous" samples; 65 out of 120 "anomalous" samples (54%) were found to be connected with mineralization or areas of unusual secondary environment. However, only 18% of the "anomalous" samples were distributed over the moorland.

In the case of the log-10 transformed data 58 samples (7.0%) out of the whole data set were considered as "anomalous" samples, but were mainly restricted to the eastern mineralization and the Denbigh Moorland. 34 out of 58 "anomalous" samples (59%) were found to be related to mineralization or secondary environment influence, however, only 10% of them were distributed over the moorland. Good control of pH and Eh would probably assist in the further discrimination of mineralization from secondary environment conditions.

In comparison to the untransformed data results, the log-10 transformed data yielded a larger number of groups, some of which were more closely related to bedrock units, but mainly over those areas conspicuously influenced by secondary environment and mineralization (contamination). However, more detailed lithological information than was available would be needed in order to evaluate quantitatively each one of the bedrock groups present. A quantitative evaluation, such as that done with the discriminant analysis results, is a difficult task given the presence of several taxonomic groups over a single geological unit. Therefore, in future investigations better geological information (e.g. physiographic features) and more detailed field control during sampling (e.g. detailed lithological description of the stream bed and banks at the sampling site; soil type and drainage conditions) would be necessary if such an evaluation is going to be exhaustively done. Such information would undoubtedly assist the delimitation of the main bedrock features (groups) in areas with little or no geological knowledge. For comparison and testing purposes, discriminant analysis could be applied using the main groups outlined by the cluster analysis as "training areas".

Similar, or closely overlapping classification patterns obtained with the combination of both techniques would probably increase the possibility of detecting patterns representing true ("natural") groups or classes correlated with the main bedrock features of the area under investigation.

3-4 Non-linear mapping algorithm

i) Method

The procedure followed with this technique in order to process and classify the Denbighshire regional geochemical data is also applicable to the Devonshire and Derbyshire data.

Five hundred samples chosen at random (without replacement) from the total data set were processed at a time for the untransformed, log-10 and normalized transformed data. Two associations of variables were considered: one the same as that used in the discriminant analysis, and the other formed by elements representative of the secondary environment: Mn, Ni, Co and Zn (except when the data was normalized). The computer program took up to 1150 seconds of central-processor time to classify the complete set of 500 samples on the basis of 12 elements.

According to the untransformed results a variable number of suitable "linkage" samples for each area were selected on the basis of the groups present in the first "basic" two dimensional mapping. Linkage samples were chosen from the untransformed results because these showed structures consisting of several separate clusters in order to identify the same groups in the second or "subsequent" mapping. In general the linkage samples remained associated in the same groups in the subsequent mapping. The log-10 transformed and normalized results did not provide a basis for the selection of linkage samples due to their characteristic appearance of single, rather diffuse, structures.

With the untransformed results, samples widely scattered outside the main groups (clusters) and not assigned to any specific group were considered as "anomalous" samples, however several samples estimated to be transitional were not assigned to any of the main groups.

The main groups derived from the untransformed data were plotted on one inch to a mile geological maps and subsequently transferred to 1/4 inch to a mile maps. The untransformed data mapping was also used to distinguish with a specific symbol those samples draining a single geological formation or any other valuable geological feature according to the a priori geological knowledge in order to study the separation among natural groups.

The "single-cluster" structures of both the log-10 transformed and normalized results in the absence of a priori information prevented the plotting of a classification map. However, once the known geological information was used, applying the same symbol identification as before, the main geological features were distinguished on the mapping.

ii) Results

Results are based on 895 samples and on the combination of the 12 variables: Pb, Ga, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr and Fe_2O_3 , and on the 4 variables: Mn, Ni, Co and Zn.

a) Classification based on the untransformed data.

Figure 3-15 shows the groups defined when the first 500 samples were processed. The remaining samples of the whole data set were tested in the same manner but using "linkage samples" to locate the same "basic" groups on subsequent mappings. The statistical properties of the main groups ("basic" and "anomalous" samples groups) are included in table 3-19.

In figure 3-15 groups Fa and Ga are not discussed in detail because of their small size and they were not plotted on the map. Group D represents the Silesian bedrock, group H and to a lesser extent group G correspond to the Silurian category and group J to the secondary environment of the Denbigh Moorland over the Ordovician rocks. The remaining groups display a clear overlapping among the main geological units. It is quite obvious that group E and F are transitional clusters where the main geological units are

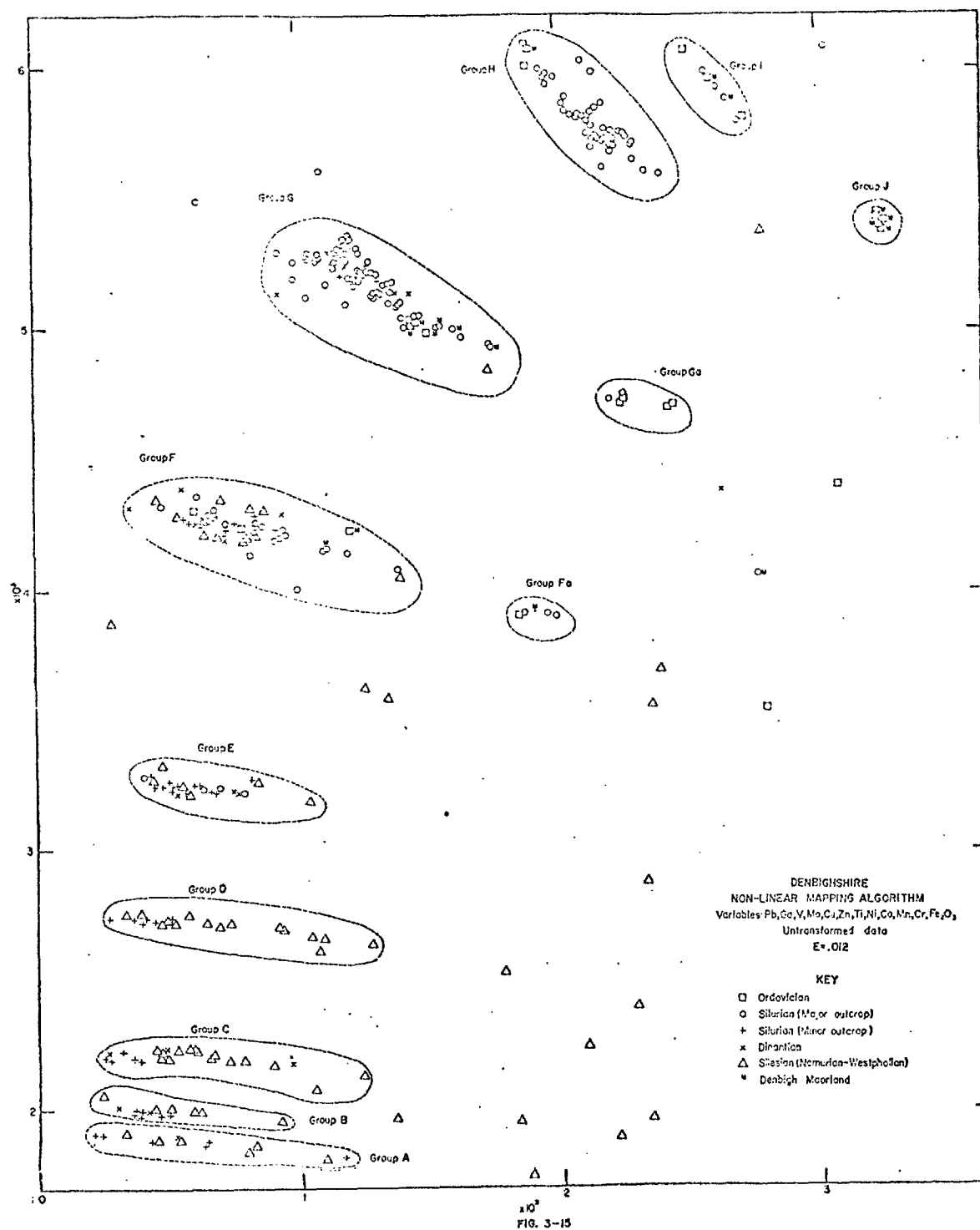


TABLE 3 - 19

Arithmetic means ($\bar{x}$) and standard deviation (s), Denbighshire, NLM algorithm, untransformed data; total sample size = 772

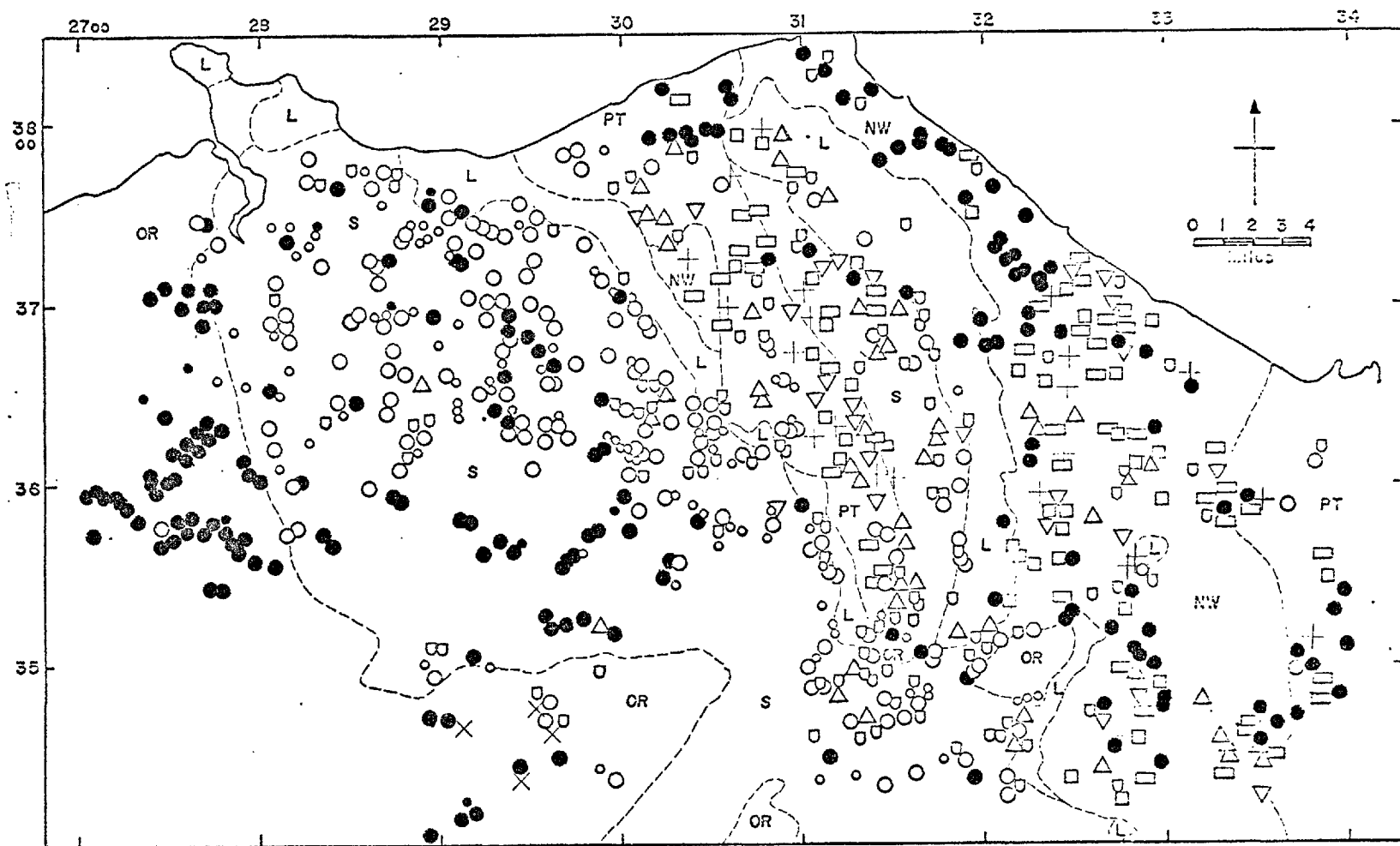
GROUPS	A	B	C	D	E	F	G	H	I	J	ANOM. SAMPLES.
SAMPLE SIZE	23	25	39	54	55	115	175	94	8	5	179
Pb $\bar{x}$	95	73	100	136	83	65	59	76	49	42	906
s	172	64	106	145	138	74	91	171	12	15	1838
Ca	5	5	6	9	15	22	26	29	31	29	21
	2	2	3	4	6	8	7	7	8	8	11
V	28	32	32	38	58	92	125	138	136	130	99
	11	9	8	12	29	41	32	30	44	25	54
Mo	1	1	1	1	1	1	1	1	1	1	3
	0.4	0	0.2	0.1	0.1	1	1	1	0	0	7
Cu	18	21	29	24	32	35	39	40	40	30	49
	9	10	18	16	15	19	35	12	13	6	47
Zn	140	107	124	147	180	180	192	199	202	183	1002
	179	80	77	97	91	70	85	157	58	46	5052
Ti	722	784	1013	1506	2018	3022	4023	4913	5000	5000	3863
	66	37	79	73	134	351	213	660	0	0	2472
Ni	26	26	29	30	36	42	46	49	59	66	43
	7	11	10	8	7	11	15	14	21	28	32
Co	8	11	14	12	15	19	22	23	34	48	49
	4	7	10	5	6	8	10	9	12	12	96
Mn	241	270	295	344	322	376	425	529	1500	2000	2792
	145	215	214	278	197	216	240	335	0	0	4135
Cr	36	37	48	47	60	74	88	98	109	90	84
	23	17	33	23	42	51	42	23	26	20	151
Fe ₂ O ₃	1	1	1	1	2	3	3	4	6	6	4
	1	1	0.4	1	1	1	1	1	1	1	3

Mean values in p.p.m., except for Fe₂O₃ (%)

intermingled. Groups A, B, C and D are also not strictly representative of any specific geological unit nevertheless they show a general separation between the Silurian and the geochemically related minor Silurian outcrop-Dinantian composite group. In groups C and D, the Silesian samples predominate and, they are therefore considered more representative of this geological unit. In general, the groups mentioned above, with the exception of D, H and J, have restricted correlation with the bedrock geology categories because of the composite geological nature of each group. By normalizing the data an attempt was made to eliminate or minimize the overlap between geological units in each group but results showed instead a higher degree of overlap than before.

With the exception of sub-group Fa and Ga, the "basic" groups and the "anomalous" samples were plotted on 1 inch to a mile geological map and then transferred to 1/4 inch to a mile map. The original results can be seen in figure 3-16 and the smoothed results in figure 3-17.

In the original map (Fig. 3-16) and more clearly in the smoothed map (Fig. 3-17) groups G, H, I and J cover the western half of the study area, with the exception of group J which is mainly distributed over the Ordovician rocks of the moorland, and group I over both Ordovician and Silurian rocks influenced mainly by the nature of the secondary environment. Group J and I are significantly higher in Mn, Fe, Co, Ni, Ti and Ga (Table 3-19) and generally show similar variability of the metal values. The remaining groups (mainly G and H) predominate over the Silurian with well-drained brown earth soils. Group G apparently represents the geochemical characteristics of those stream sediments derived from the Silurian bedrock, whereas group H relatively higher in all the elements considered, except Mo, is apparently highly responding to the nature of the secondary environment on the Denbigh Moorland over the Silurian bedrock. Both groups (G and H) are higher in Ga, V, Cu, Zn, Ti, Ni, Co, Mn, Cr and Fe_2O_3 with respect to groups A, B, C, D and E.



DENBIGHSHIRE
 NON-LINEAR MAPPING ALGORITHM
 ORIGINAL MAP
 Variables: Pb, Ga, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, Fe₂O₃
 Untransformed data

GEOLOGY
 PT Permian-Triassic
 NW Silurian
 L Dinantian
 S Silurian
 OR Ordovician

GROUPS
 + A
 ▽ B
 □ C
 ▢ D
 △ E
 □ F
 ○ G
 ● H
 • I
 × J
 ● Anomalous samples

Note: Groups Fa and Ga are not plotted

FIG. 3-16

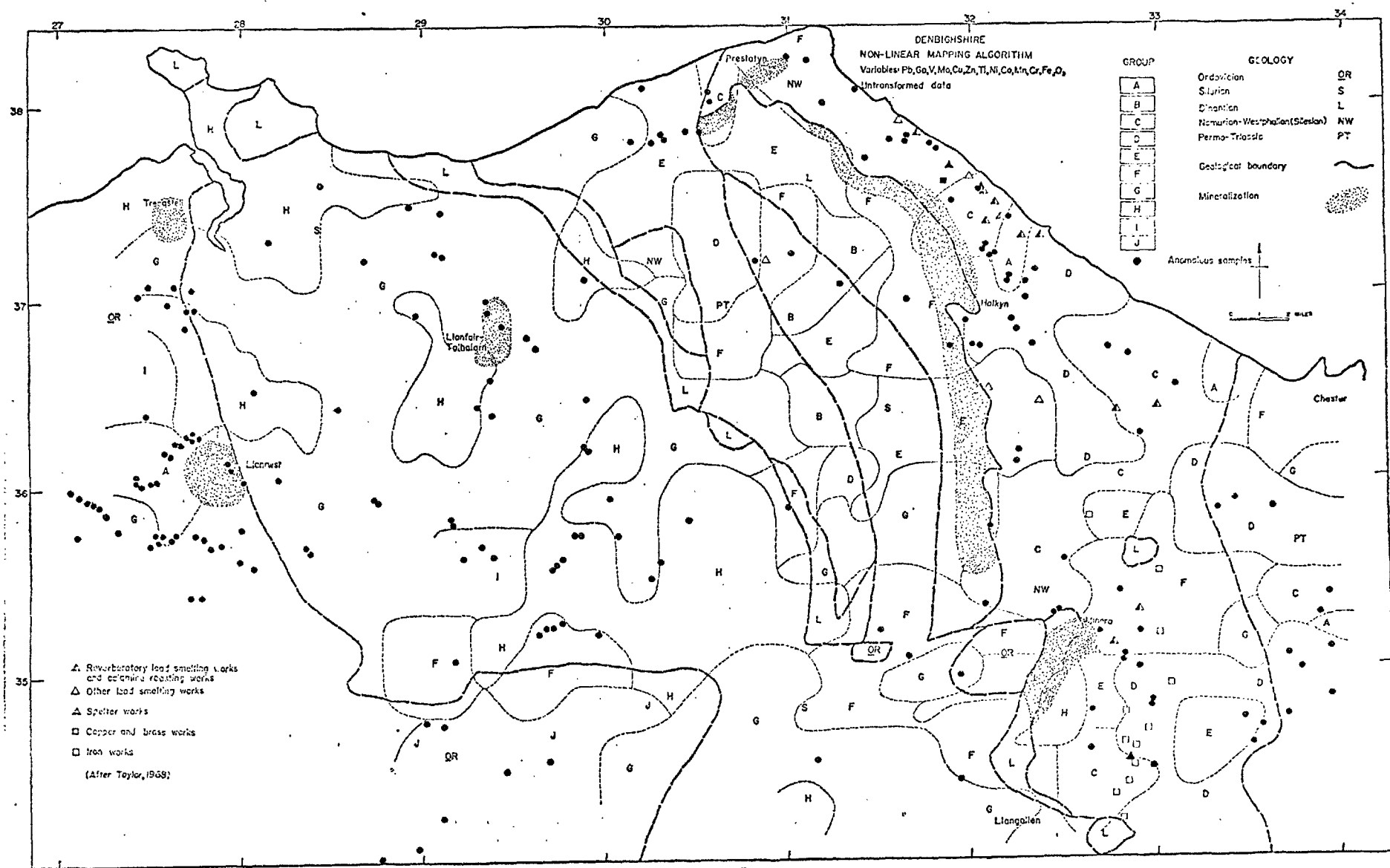


FIG. 3-17

In the eastern half of the area, groups A, B, C, D and E are irregularly distributed on the Carboniferous, minor Silurian outcrop, and Permo-Triassic rocks. These groups reflect stream sediments derived from the minor Silurian outcrop and Carboniferous exposures (Dinantian, Silesian) and show the chemical similarity already shown by hierarchical cluster analysis and discriminant analysis. They are generally lower in metal content (Table 3-19) but relatively higher in Pb probably because of the influence of mineralization and contamination.

No well defined associations between the groups were shown by the EDF results (Table 3-20) (although visual inspection of the maps suggested the presence of some group associations related to the main geological formations), probably because of the chemical composition in terms of the untransformed data. On the basis of the sensitivity of the method all clusters are geochemically well separated "natural" groups not necessarily reflecting the main geological features of the area, but instead well defined "geochemical facies". The higher correct recognition rates for each group confirm the existence of these "natural" groups (geochemical facies) defined strictly on the basis of the chemical composition. However, only a slight A and B group association appears to be confirmed by the EDF analysis.

The "anomalous" samples on the Ordovician rocks predominate on and around the Llanrwst and Tre Castell mining districts and to a lesser extent southeast of the outcrop in areas influenced by the secondary environment of poorly-drained peat and peaty gley soils. Over the Silurian rocks the "anomalous" samples predominate over the Denbigh Moorland where poorly-drained peat and peaty gley soils are also developed. However, several "anomalous" samples are scattered on the northern part of the major Silurian outcrop, some of them evidently connected with the influence of the Llanfair-

TABLE 3 - 20

Group classification results by EDF analysis, Denbighshire, NLM algorithm, untransformed results, $\sigma = 20.0$

ASSIGNED GROUP														
	A	B	C	D	E	F	G	H	I	J	ANOM.	?	% CORRECT	
TRUE GROUP	A	17	3	1	0	0	0	0	0	0	2	0	74	
	B	4	21	0	0	0	0	0	0	0	0	0	84	
	C	1	0	36	2	0	0	0	0	0	0	0	92	
	D	0	0	1	51	1	0	0	0	0	1	0	94	
	E	0	0	0	1	52	2	0	0	0	0	0	95	
	F	0	0	0	1	3	107	2	2	0	0	0	93	
	G	0	0	0	0	0	5	165	2	0	1	2	94	
	H	0	0	0	0	0	1	0	89	0	0	4	95	
	I	0	0	0	0	0	0	0	8	0	0	1	95	
	J	0	0	0	0	0	0	0	0	4	1	0	80	
Anom.	7	0	3	2	0	5	4	10	0	1	87	60	49	

Overall correct recognition rate
(Excluding anomalous group) - 93 %

Overall correct recognition rate
(For all groups) - 83 %

Talhaia district. The rest are of uncertain origin; individually most of the samples showed a chemical composition generally closer to the composition of the stream sediments derived from the Silurian bedrock, and something similar occurred with those "anomalous" samples distributed over the Dee Valley mainly derived from the Carboniferous rocks.

The "anomalous" samples also define very well the Halkyn and Minera mining districts and areas contaminated by mining activity. They also occur in the area southwest of Prestatyn close to the mouth of the Clwyd river where marked concentration of Cu, Pb, Zn, Fe, Ti and V has been reported (Nichol and others, 1970b, p.9) as a response not only to local conditions of the secondary environment but also to the influence of the eastern mineralization on the Dinantian. The group of "anomalous" samples shows the most variability for all the metals and the highest concentrations of Pb, Zn, Cu, Mo, Co and Mn (Table 3-19); the majority of these elements are representative of mineralization and contamination. The EDF analysis results (Table 3-20) show a high proportion of misclassified samples assigned to the other groups which may be expected because of the diversity of different geological sources (geological formations) influencing the "anomalous"-classified samples throughout the area.

In order to investigate the role that a smaller number of variables can play in the classification, only Mn, Ni, Co and Zn were considered. The same number of "basic" groups as those obtained with 12 variables resulted. In general the bedrock units occupied similar positions in the mapping but showed a larger amount of overlapping and a larger number of erratic samples, belonging to different bedrock categories, in all "basic" groups. The 17 groups were subsequently plotted on 1 inch to a mile geological map, but only a poor response to the bedrock geology was obtained. Most of the groups were irregularly distributed throughout the study area making interpretation very difficult. The

"anomalous" samples were found to be related with mineralized areas and to a lesser extent with the Denbigh Moorland. A number of "anomalous" samples were erratically distributed throughout the area probably as a response to the nature of the variables considered, which were not really representative of the general geochemical environment.

Finally the "anomalous" samples derived from the untransformed results with 12 variables were classified on their own using the same untransformed option and the same variables. The purpose of this classification was to determine whether there would be a response to each of the main mineralized areas (mining districts) and the secondary environment. Only two major groups were differentiated: one comprising the Lower Palaeozoic mineralization (Llanrwst, Trecastell and Llanfair-Talhaiarn and secondary environment (Denbigh Moorland), and the other the Upper Palaeozoic mineralization (Halkyn, Minera and southwest of Prestatyn). Separation among individual mining districts was not, however, obtained.

b) Classification based on the log-10 transformed data.

Part of the results obtained are shown in figure 3-18. This mapping illustrates the structure obtained with the same "basic" 500 samples. A similar structure was obtained with the remaining 395 samples using sample identification based on a priori knowledge. The overlapping between classes was minimized. Only the minor Silurian outcrop and the Dinantian show marked overlapping which can be explained by the chemical similarity between both units.

The stream sediments over the Permo-Triassic rocks derived from the uplands show a direct correlation with the main geological features of the area. Stream sediment samples from the Vale of Clwyd are scattered among samples assigned to the major and minor Silurian outcrops and the Dinantian, whilst stream sediments collected along the western margin of the Dee Valley are distributed within the Silesian category.

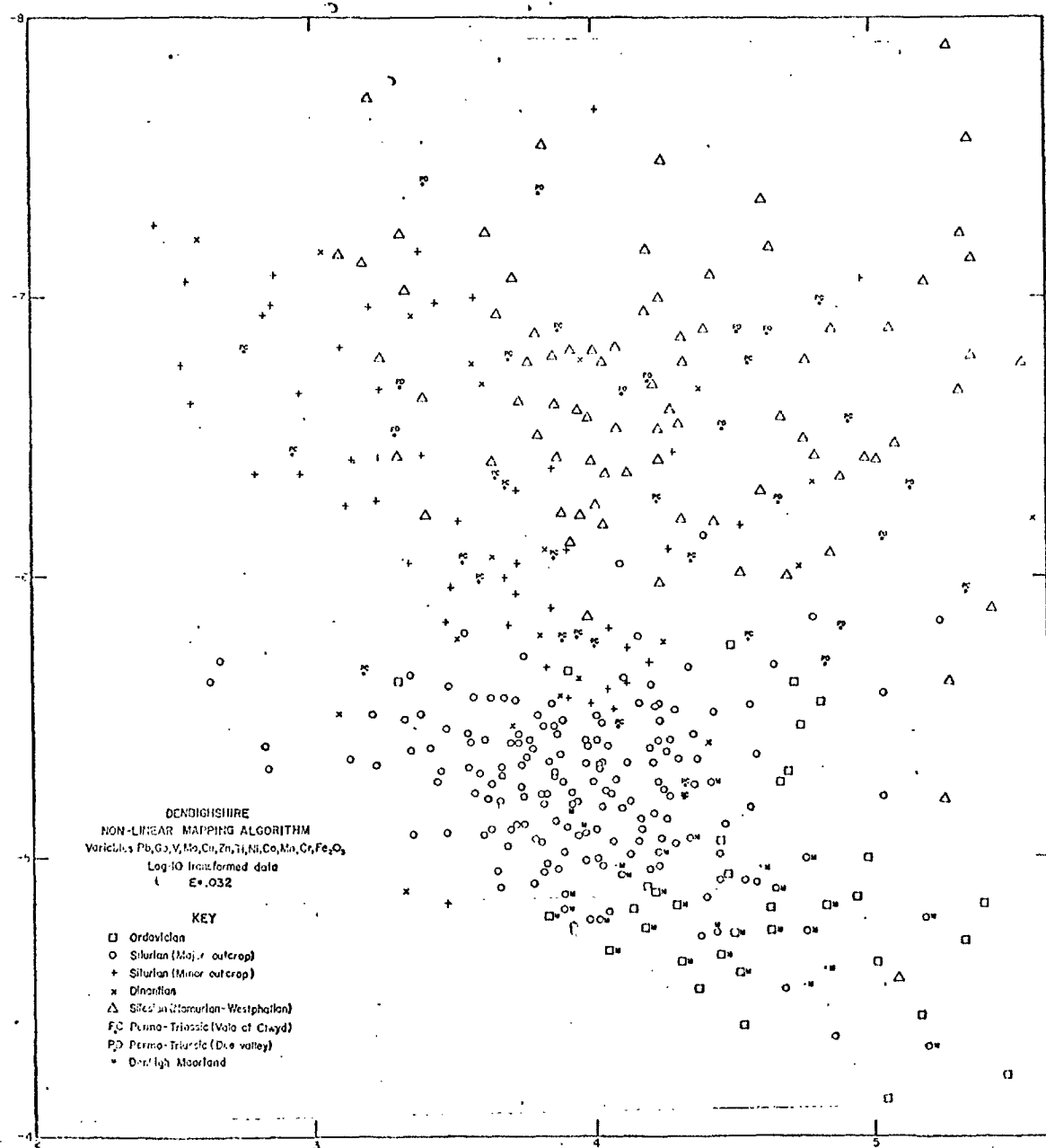


FIG. 3-18

Samples related to anomalous geochemical environments (e.g. mineralized and contaminated areas, secondary environment) were generally distributed around the periphery of the mapping or were isolated from the main clusters.

Elimination of overlapping between geological units was attempted by normalizing the data, but results were similar to those described earlier.

The same data set was subsequently processed using only the 4 variables: Mn, Ni, Co and Zn. A similar diagram to that of figure 3-18 was obtained by with a greater amount of overlapping among bedrock units. The Dinantian category was irregularly distributed among all bedrock classes.

Finally the "anomalous" samples derived from the untransformed results were classified alone using log-10 transformation of the data. The two groups of "anomalous" samples related to the Lower and Upper Palaeozoic mineralizations were very well separated (Fig. 3-19). The classification within the Lower Palaeozoic was improved in comparison with the untransformed results because the samples of the Llanrwst district form an independent group with a gradual overlap into the Denbigh Moorland group. The samples southwest of Llanrwst (N.G.R., 2730-2790/3530-3575), along the Afon Llengwy river, form an independent group closely related to the Denbigh Moorland. This would again explain the geochemical similarity between both groups, a similarity formerly detected by the hierarchical clustering (Fig. 3-12). Samples northwest of Llanrwst and along the outlets of the lakes Llyn Craftnant and Llyn Geirionydd (N.G.R., 2720-2780/3600-3640) (both on the Llanrwst district) are grouped mainly over the periphery of the mapping, more distant from the Denbigh Moorland group and probably more influenced by the mineralization of the district. The samples near Tre Castell and Llanfair-Talhaiarn mining districts were not well classified. The Halkyn and Minera districts and the mineralized area southwest of Prestatyn (Upper Palaeozoic mineralization) were not well differentiated probably due to the strong influence of the highly disseminated contamination.

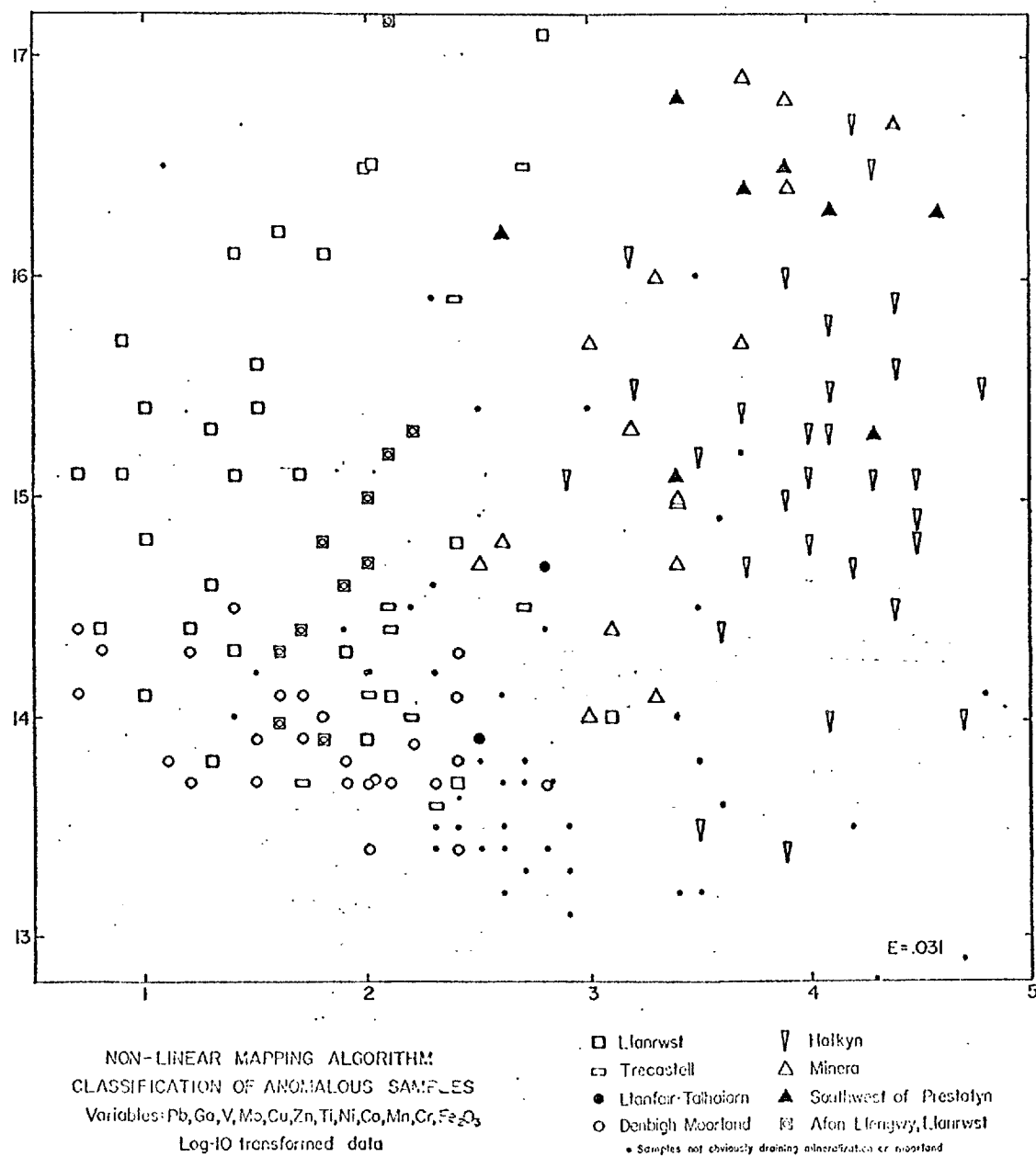


FIG. 3-19

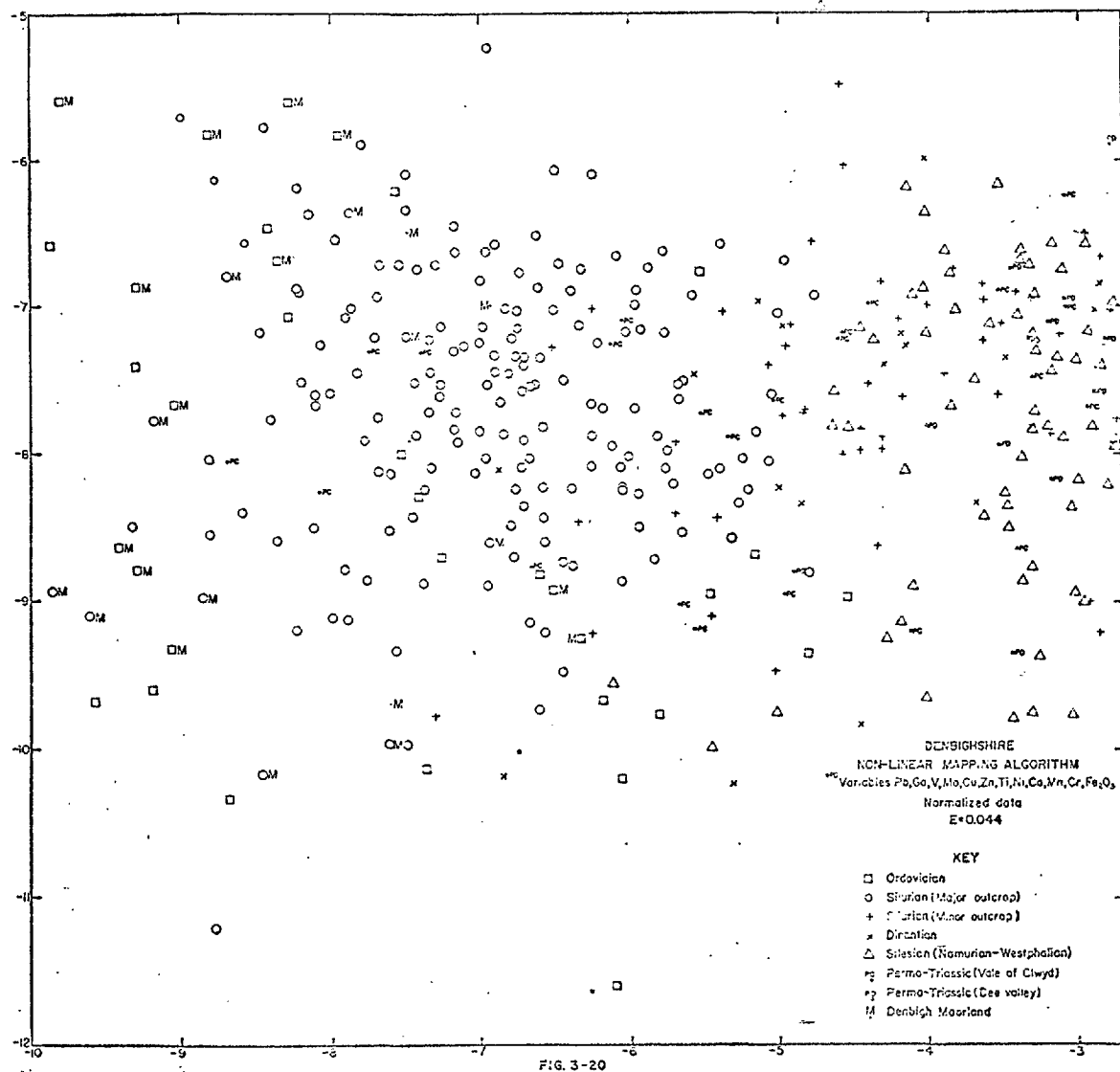
c) Classification based on the normalized data

Part of the normalized raw data results are shown in figure 3-20. This mapping illustrates the structure obtained with the same basic 500 samples classified in previous transformations. A similar structure was obtained with the remaining 395 samples. A high degree of overlap is seen between the minor Silurian outcrop and Silesian category, and between the Silurian and Ordovician classes. The Dinantian category is irregularly distributed throughout the Lower and Upper Palaeozoic which are well separated. Stream sediments from the Denbigh Moorland are widely scattered over a well defined region of the plot, but not showing the clear close agglomeration observed on the log-10 two-dimensional diagram (Fig. 3-18). The stream sediments over the Permo-Triassic rocks derived from the uplands show, as with the log-10 transformed data, correlation with the main geological features surrounding each area of erosion and deposition (Vale of Clwyd and Dee Valley). Those samples distributed along the periphery of the mappings and evidently related with geochemically anomalous environments were considered as "anomalous" samples for further evaluation of mineralization, contamination and secondary environment conditions.

iii) Discussion

For practical purposes the non-linear mapping algorithm has proved to be highly sensitive and effective to locate geochemically anomalous environments (e.g. mineralization, contamination and secondary environment) whether or not log-10 transformation or normalization of the data is used.

In the untransformed data results with 12 variables, 179 samples (20%) of the whole data set were taken as "anomalous" samples. 121 of these 179 (68%) were found to be related to the main mineralized areas or to the unusual secondary environment over the Denbigh Moorland. However, only 15% of the "anomalous" samples were distributed over the moorland; the rest of the "anomalous" samples (32%) were not attributable to any obvious cause.



In the log-10 transformed data results, 172 samples (19%) of the whole data set were taken as "anomalous" samples. 90 of these 172 (52%) were found to be related to the main mineralized (contaminated) areas and the secondary environment; 14 of these 172 samples (8%) were found related with the nature of the secondary environment alone over the Denbigh Moorland. The rest of the "anomalous" samples (48%) were not attributable to any obvious cause.

In the normalized raw data results, 172 samples (19%) of the whole data set were taken as "anomalous" samples. 138 of these 172 (80%) were found to be related to main mineralized (contaminated) areas and the secondary environment; of these, 32 (19%) samples were found related with the nature of the secondary environment alone over the Denbigh Moorland. The rest of the "anomalous" samples (20%) were not connected with any obvious cause.

In general, "anomalous"-classified stream sediments not attributable to any obvious cause, according to each transformation, have a chemical composition resembling the composition of the stream sediments derived from relatively unmineralized (uncontaminated) bedrock sources. In this case, good control of pH and Eh would probably assist in the further discrimination of mineralization from secondary environment conditions.

In the delimitation of the bedrock units the untransformed data option outlined an appreciable number of groups, each one with a high degree of overlap among the main geological units making it difficult to carry out bedrock correlations and hence qualitative evaluation of the classification patterns. However, the log-10 transformed data results produced an almost perfect separation of the main bedrock units with a minimum of overlap. Normalization of the raw data produced a mapping similar to the log-10 diagram but with higher overlap. Better geological

information (e.g. physiographic features) and more detailed field control during sampling (e.g. detailed lithological description of the stream bed and banks at the sampling sites; soil type and drainage conditions) would probably assist in the separation of the currently mapped main bedrock units.

3-5 Conclusions: Hierarchical cluster analysis and NLM algorithm.

1) Broad geological units (e.g. Lower Palaeozoic from Upper Palaeozoic) were perfectly well defined with either the dendrogram representation or the NLM mapping. Features of the bedrock influenced by mineralization and secondary environment, mainly in the Lower Palaeozoic area of peat and peaty gley soils, were better shown by the dendrogram representation, log-10 transformed data, than with the untransformed data, and the NLM algorithm. However, the best results for the bedrock classification were obtained with the NLM log-10 mapping.

2) The mineralization (contamination) and the nature of the secondary environment on the Denbigh Moorland were both (Table 3-21) better detected by the "anomalous" samples classified with the log-10 transformed data dendrogram than with the untransformed dendrogram.

3) Mineralization (contamination) and the nature of the secondary environment on the Denbigh Moorland were both (Table 3-21) better detected by the "anomalous" samples classified with the normalized and untransformed data NLM mappings than with the NLM log-10 transformed data. However, the normalized NLM mapping achieved better quantitative figures than both the NLM untransformed data, and the untransformed and log-10 transformed dendrogram results. The untransformed and log-10 transformed dendrogram results achieved higher quantitative values than the log-10 NLM mapping.

4) The exclusive classification of the "anomalous" samples, derived from the original untransformed data, with either the

TABLE 3 - 21

Comparative successful recognition rates (%) of the "anomalous" samples in relation to different geochemical environments, Denbighshire.

METHOD	DATA OPTION	MINERALIZATION (CONTAMINATION) AND THE DENBIGH MOORLAND %	DENBIGH MOORLAND %	AREAS BARREN OF MINERALIZATION %	NO. OF "ANOMALOUS" SAMPLES
NLM	UNTRANSFORMED (RAW DATA)	68	15	32	179
ALGORITHM	LOG-10	52	8	48	172
	NORMALIZATION	80	19	20	172
	UNTRANSFORMED	54	18	46	120
DENDROGRAM	LOG-10	59	10	41	58

untransformed dendrogram representation or the log-10 NLM mapping was very well performed. The classification with both the log-10 dendrogram representation and the untransformed NLM algorithm was poorly carried out.

CHAPTER 4

DISCRIMINANT AND CLUSTER ANALYSIS OF DEVONSHIRE DATA

4-1 Location and description of the study area.

i) Geography

The study area (Fig. 4-1, 4-2) is situated mainly in north Devon, but also extends into northeast Cornwall, and is bounded by the National Grid coordinates 2100 (west), 2900 (east), 0764 (south) and 1100 (north) covering a total surface area of 820 sq. miles. The principal towns are Bude, Crediton, Holsworthy, Launceston and Okehampton.

The topography of the study area is conspicuously represented by two centres of higher ground: Dartmoor in the east and Bodmin Moor in the west, both lying above 800 feet and attaining a maximum elevation of 2038 and 1375 feet, respectively. The rest of the area, mainly underlain by Devonian and Carboniferous sediments, lies at about 400-800 feet, descending more in elevation towards the river valleys and the coasts.

The major river systems (Fig. 4-1) drain either north (rivers Taw and Torridge) and south (rivers Exe, Teiger, Dart, Plym, Tamar and Lynher); only minor rivers flow to the west coast and some, on the Carboniferous sandstones in the northern part of the area, tend to flow locally subparallel to the regional strike. The drainage over the granite is radial.

The climate is mild and humid, and the highest rainfall rate occurs over the moors (60-100 inches). Elsewhere, the rainfall decreases from the west (40-50 inches) to the east (30-35 inches).

The soils are residual except when developed over "head" deposits. Freely-drained brown earths cover the Devonian shales and slates. Soils originated over the Carboniferous have only been mapped in the Teign Valley by Clayden (1964) who believed that broadly similar soils would

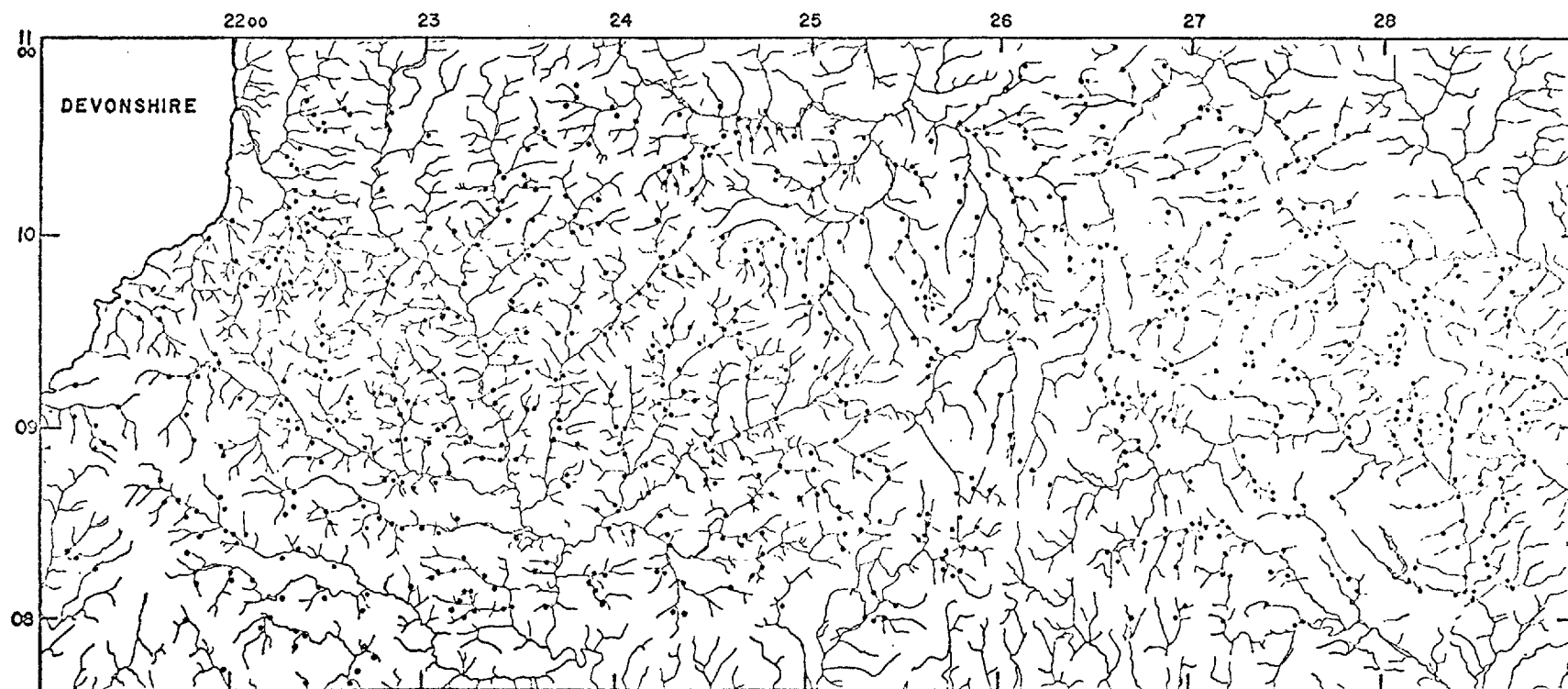


FIG. 4-1
DRAINAGE SYSTEM AND SAMPLING LOCALITIES
(After Khaleelee, 1969)

occur over much of the Carboniferous with less argillaceous brown soils developed on the well-drained slopes and gley soils over flatter, less well-drained areas. According to Clayden and Manley (1964), brown earths and peaty gleyed podzols are developed on the lower ground and around the central core of the granites, respectively; blanket bog and peaty soils cover the central high moorland.

ii) Geology and Regional Geochemistry.

The general description of the geology follows that of Edmonds, McKeown and Williams (1969). Rocks of Devonian, Carboniferous, Permian, and Quaternary age, and granite intrusions have been mapped in the study area. Undifferentiated Permian, Cretaceous and Tertiary rocks outcropping on the eastern side of the study area were not considered in this study. A geological map is given in figure 4-2. The regional geochemistry is a brief discussion of Nichol and others (1971). The geochemical characteristics of stream sediments derived from the principal geological formations are given in table 4-1 in order to show the relationships between the geochemical patterns and the geological formations taking into account the mean values of the elements. The data limits interpretation with respect to lithologies because the stream sediments representing each formation are composed of variable mixtures of material derived from different lithologies within the formations.

The Devonian sediments consist predominantly of slates with shales and intercalated sandstones with some limestones and volcanic rocks in the southern part. Streams draining Devonian strata within the metamorphic aureole of the granite show higher mean values of As, Co, Cu, Pb, Sn, and Zn than those outside the aureole.

The Dinantian rocks consist of dark grey shales with thin sandstones and intercalations of slates, lavas, tuffs, agglomerates, limestones, and cherts. Sediments draining Dinantian rocks within the

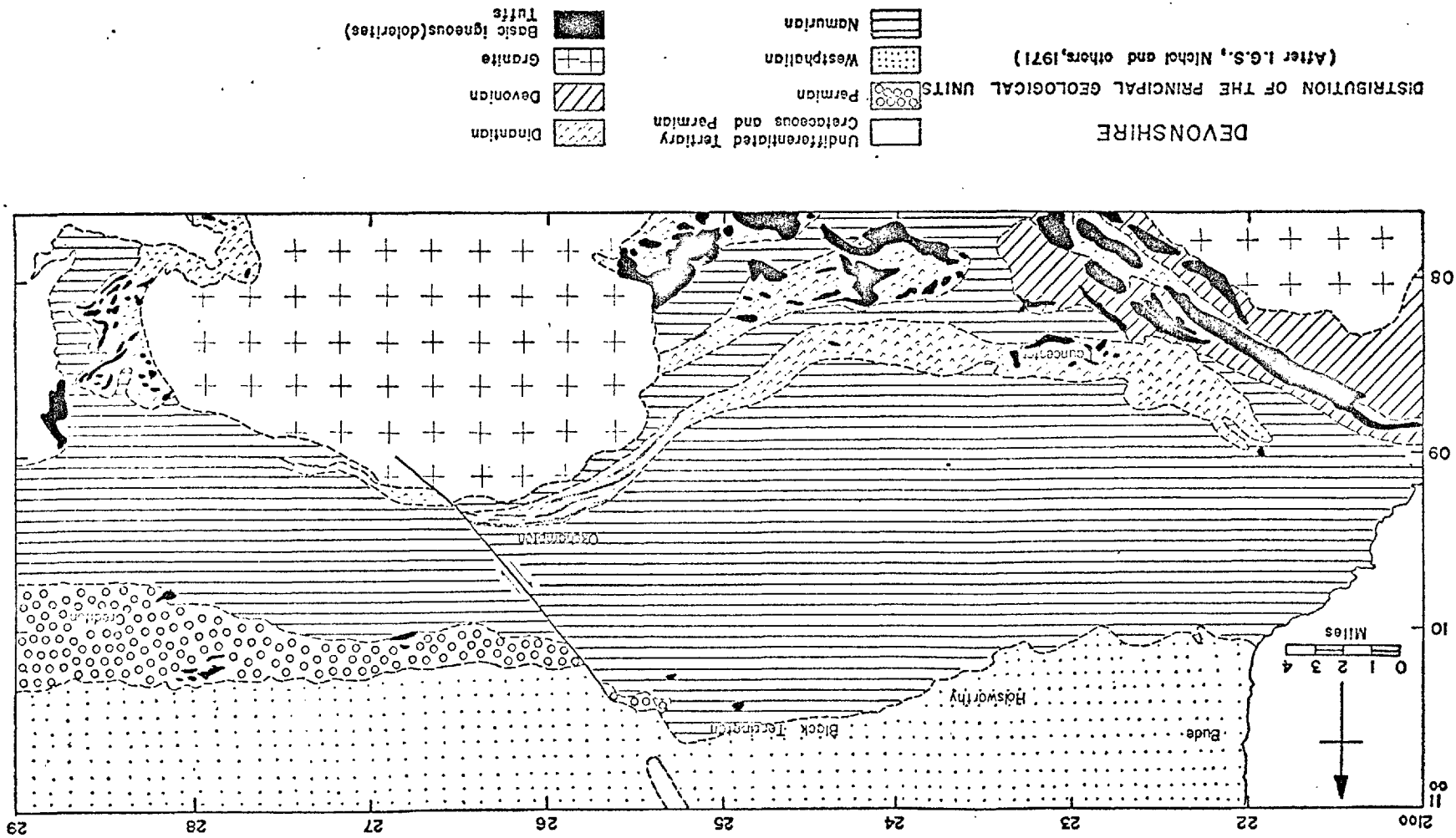


FIG. 4-2

TABLE 4 - 1

Range and Mean* metal content of stream sediments derived from the principal geological formations (After Nichol and others, 1971, p.10).

Values in p.p.m. unless otherwise stated.

ELEMENT	GRANITE	DINANTIAN	NAMURIAN	WESTPHALIAN	PERMIAN
As	18 5-65	40 10-95	12 3-25	<5 <5	7 2-20
Cr	17 9-35	90 60-120	80 65-100	65 45-90	45 30-65
Co	9 4-25	35 15-60	22 15-35	15 10-20	15 8-25
Cu	15 7-30	45 35-70	35 19-45	15 12-20	25 16-40
Fe ₂ O ₃ %	3.0 1.5-5.5	6.6 4.3-9.8	5.5 3.7-8.1	3.0 2.2-4.4	2.5 1.6-3.2
Pb	50 30 - 80	50 22 - 140	35 25 - 45	30 18 - 55	35 30 - 40
Mn	600 220 - 1700	2200 1200-4400	1000 550-1800	1000 650-1700	600 300-1200
Mo	<2 <2	<2 <2-3	<2 <2	<2 <2	<2 <2
Ni	7 4.5 - 12.5	60 40-110	45 30 - 55	35 30 - 40	40 25 - 60
Sn	200 65 - 800	9 4 - 30	5 2 - 10	<5 <5	10 3 - 30
Ti	1650 1100 - 2700	4500 2900 - 6000	3500 2400 - 4500	2500 1800-3500	2000 1300 - 3000
V	30 17 - 45	120 90 - 150	100 65 - 130	55 45 - 75	50 30 - 80
Zn	85 45 - 150	170 105 - 1000	150 80 - 200	90 70 - 120	60 15 - 200
No. OF SAMPLES	101	73	263	82	14

* Geometric mean; range = geometric mean $\pm$ log standard deviation.

metamorphic aureole of the granite show higher mean values of As, Cr, Ni, Sn and Zn over Dinantian rocks. The contrast between high Mo values over Dinantian sediments and lower values over adjacent sediments is most conspicuous between Launceston and the western margin of the Dartmoor granite; these high Mo values have been considered by Fletcher (1968) to be derived from black pyritic shales within the Dinantian rocks. These shales also contain high Cu, Pb, Se and V relative to adjacent strata. The greatest differences in the metal content of Dinantian rocks occurs between streams draining diabase (dolerite) in their catchments and those which do not. Streams draining Dinantian sediments and diabase generally show significant higher concentrations of Mn, Ni, Co, Ti and V in contrast with streams only draining Dinantian sediments.

The Namurian (Crackington Formation) consists of shales with thin turbidite sandstones, and is more argillaceous than the overlying Westphalian (Bude Formation) which is composed of thick-bedded massive sandstones with slates and shales. The Namurian bedrock samples, which are mainly composed of shales, have higher Cr, Cu, Ni, Ti and V contents than those of the Westphalian which are principally sandstones. The thin interbedded turbidite Namurian sandstones are predominantly exposed on the western half of the main outcrop (west of the Sticklepath fault) whereas on the eastern side they are almost absent. The Westphalian exposures show a higher proportion of shale interbedded with sandstones on the eastern side. In broad terms, in the stream sediment geochemistry there is a general decrease in most of the elements over the Carboniferous sequence which is probably unrelated to mineralization. As, Cr, Co, Cu, Fe, V, and, to a lesser extent, Mo, Ni, Ti and Zn tend to be lower over the Westphalian in contrast with the Namurian, particularly in the east; the pattern becomes diffuse in the west. Fe (variously accompanied by Ti, Mn, Co, Ni, and Cu) is seen near Black Torrington and south of Crediton. The lower As, Co and Fe_2O_3 content of Westphalian stream sediments is not reflected in the bedrock.

The Permian rocks that unconformably overlies the Carboniferous rocks consist mainly of breccias, sandstones, and marls, containing Carboniferous pebbles in the lower part (Cadbury Beds) and untourmalinized lava pebbles in the overlying Bow Beds. The presence of detrital

cassiterite, galena and sphalerite in the Permian formation suggest erosion of mineralized ground adjacent to the Dartmoor granite and subsequent deposition in the Crediton trough. Small patches of igneous rocks associated with the Permian are collectively called the Exeter Volcanic Series. The occurrence of high Sn values reflects the presence of cassiterite which has been previously recognized in the Permian.

The Dartmoor and Bodmin Moor granites comprise several petrologic types: 1) porphyritic feldspathic (orthoclase) granite; 2) uniform fine-grained granite, and 3) aplitic and pegmatitic dykes probably of later age and associated with the mineralization. Cr, Co, Cu, Mn, Ni, V and to a lesser extent, Fe, Ti, Zn, and As tend to be low on the granite in relation to the adjoining formations; Sn is the only element consistently higher on the granite than elsewhere. Within the Dartmoor granite there are significant differences: Cr, Ni and V are uniformly low; the lower values of Mn, Co and Fe association over better drained ground of the eastern half of the Dartmoor characterizes a facies within the granite composed of abundant feldspars megacrysts, whereas a marked concentration of high Mn, Fe, Co, Zn and to a lesser extent Mo and lower Ti values over the western half coincides with facies essentially free of feldspars megacrysts and with poorly-drained soils.

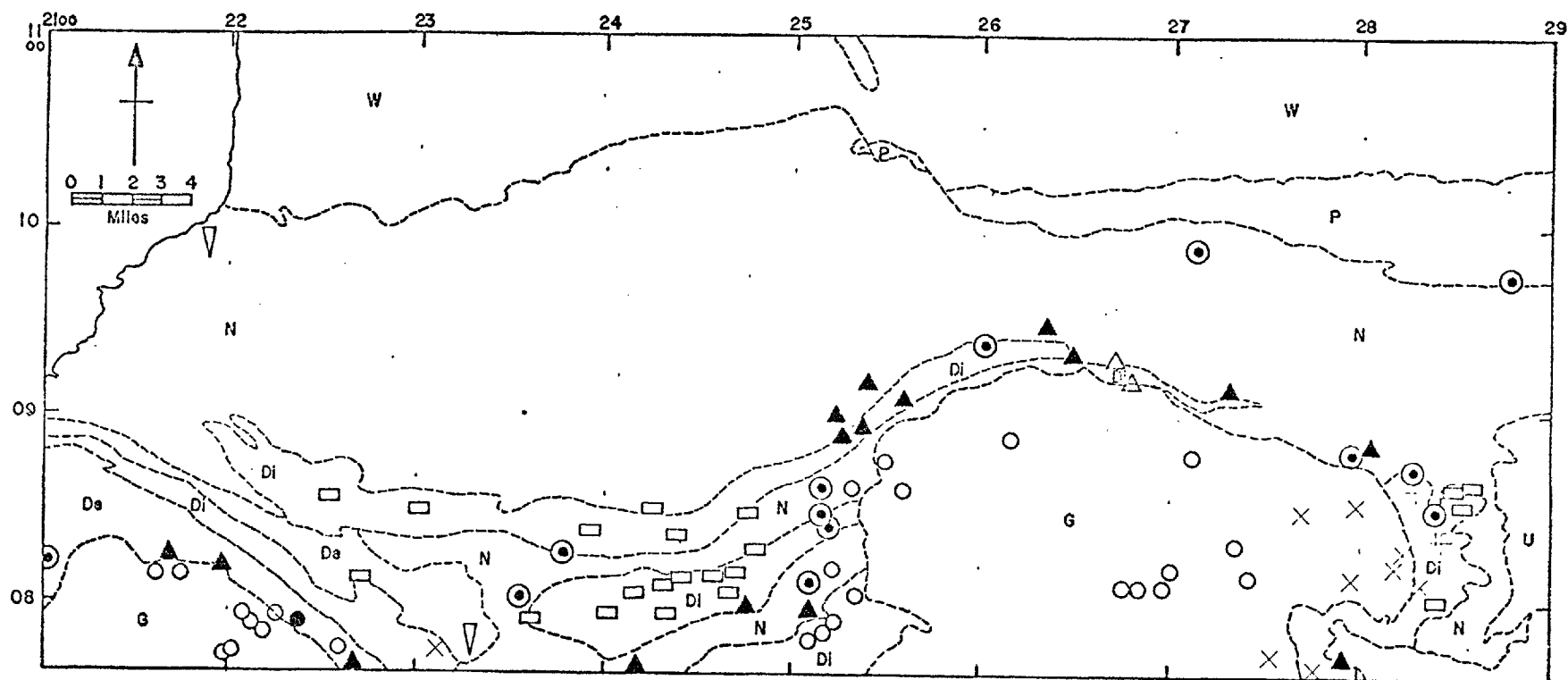
A small outcrop of Undifferentiated Permian, Cretaceous and Tertiary rocks appear to the east but their derived stream sediments were not sampled. Tertiary and Recent overburden is largely residual, with occasional periglacial deposits of "head" (rock debris which has moved down hill by solifluxion).

Differences in metal content, particularly Mn, Fe, Co, Ni and V produced by the secondary environment in other geological formations, beside the granite, are also present. These differences appear mainly between stream sediment, rock and soil metal values and they are mainly caused by those mobile elements in the secondary environ-

ment. This phenomenon usually takes place in Namurian rocks (Nichol and others, 1971, p.13).

As, Zn, accompanied by rather diffuse patterns for Co, Cu, Mn, Mo, Ni and Ti, attain maximum concentration in a broad east-west belt extended along the northern margins of the Bodmin Moor and Dartmoor granites. However, there are certain differences in distribution and metal associations: peak values for Cr, V, Ni and Ti occupy a narrow zone in the south-west and an association of Cu and Mo maxima occur to the north. There are three clusters of high Mn and Ni values east and west of Launceston and east of Dartmoor; the latter cluster is accompanied by high Fe, Ti and erratic Co; Ti is low west of Launceston and none of these three element patterns correlate with Mn and Ni east of the town.

According to earlier works (Dines, 1956; Dearman and El Sharkawi, 1962) the mineralization can be classified into three groups: 1) tin-bearing lodes (containing cassiterite and specularite), filling fractures trending NE-SW, usually within the granites and their metamorphic aureoles. The principal concentrations take place in central Dartmoor (around Birch Tor), north-east of Dartmoor and north Dartmoor (south of Okehampton); a series of ENE veins occur within the Bodmin Moor granite adjacent to its north-east boundary; 2) metasomatic replacement lodes bearing tin, lead, zinc, copper, arsenic, tungsten and barite occur as infilled fractures within the metamorphic aureole around the granite. The main mineralization of this type occurs in the Teign Valley north-east of the Dartmoor granite, and the fractures trend N-S; subsidiary mineralizations are found in the Dinantian calcareous rocks south of Okehampton, and 3) stratiform manganese oxide deposits occurring in the Dinantian cherts are usually associated with a basic period of vulcanism. A sketch map showing the distribution of the principal mineral deposits is given in figure 4-3.



DEVONSHIRE
MINERAL DEPOSITS
(After Khaleel, 1969)

- | | |
|---------|------|
| ○ Sn | ● W |
| ⊙ Pb-Zn | ▽ Sb |
| ▲ Cu | △ As |
| □ Mn | × Fe |
| + Ba | |

GEOLOGY

- | | | | |
|---|---|----|-----------------------------|
| U | Undifferentiated
Tertiary, Cretaceous
and Permian | Di | Dinorthern and
dolerites |
| P | Permian | Da | Devonian and
dolerites |
| W | Westphalian | G | Granite |
| N | Namurian | | |

FIG. 4-3

Undoubtedly the sources of contamination are directly related to the past mining activity that has taken place since the 12th century in the central Dartmoor, metamorphic aureole, the central part of the Tamar Valley (presently excluded from the study area) and the eastern flank of the Bodmin Moor granite. Placer tin deposits were thoroughly worked out in the central Dartmoor area and subsequently during the refining processes, tin-fumes and slag will have contributed significantly toward soil and stream sediment contamination. Lead, zinc and silver mining took place in the Teign Valley. The refining methods were similar to those used for tin and will also have produced contamination for these elements. Copper and arsenic exploitation was carried out mainly on the northern margins of the Dartmoor granite over the metamorphic aureole. Tin, copper and lead-zinc deposits were also worked around the margins of the Bodmin Moor granite but mainly on the northeastern flank.

The marked concentration of the highest Sn contents over the granite, the association of relatively high Pb with the granite and adjacent Dinantian, and the occurrence of the highest mean As, Cu, and Zn contents in the Dinantian adjacent to the granite broadly agree with the mineral zoning with the exception of Pb. The variation of metal content related to the principal geological units as distinct from features related to mineralization are difficult to discriminate due to the superimposition of the two in parts of the Dinantian sequence.

4-2 Empirical discriminant analysis.

Only cluster analysis has been carried out in this area because previously Howarth (1971) successfully applied discriminant analysis to a total of 1238 stream sediment samples. An EDF classification, based on Sn, V, Mo, Cu, Zn, Ti, Ag, Ni and As, and on a training set

of 153 stream sediments derived from Westphalian, Permian, Namurian, Dinantian, granite and basic igneous rocks, achieved a classification success rate of at least 90 per cent correct over these categories. The average training area compositions are given in table 4-2. For comparative purposes the percentages of correct classifications for the testing set samples assigned to each class for both untransformed and log-10 transformed data are given in tables 4-3 and 4-4.

4-3 Hierarchical cluster analysis.

i) Q -mode results.

The results from both the untransformed and log-10 transformed data are based on 840 stream sediment samples on the 10 variables: Pb, Sn, V, Mo, Cu, Zn, Ti, Ag, Ni and As. Mn was not included as variable because of its poor discriminatory power between the training areas found by Howarth (1971b). The ~~correlation~~ distance coefficient in the "basic" dendrograms for the "basic" groups representing bedrock units, with the exception of the "anomalous" samples generally ranged between 0.86 and 0.89. Coefficients ranging between 0.84 and 0.95 were seen in the "subsequent" dendrograms.

a) Classification based on the untransformed data.

The original results can be seen in figure 4-4 and the smoothed results in figure 4-5. The basic dendrogram is shown in figure 4-6. The statistical properties of the groups are included in table 4-5.

The Westphalian (Bude Formation) is represented by group B and to a lesser extent by group D. Some isolated samples of group F and A were eliminated in the smoothing. On the smoothed map (Fig. 4-5) group B shows a discontinuous distribution over the Westphalian although group D overlaps a mile or two north of the contact between the Namurian (Crackington Formation) and the Westphalian in the western half of the region. This is a reflection of streams draining north across this contact. North of Black Torrington (N.G.R.,

TABLE 4 - 2

Average training area compositions, Devonshire
(after Howarth, 1971b)

TRAINING AREA	p.p.m.										SAMPLE SIZE
	Pb	Sn	V	Mo	Cu	Zn	Ti	Ag	Ni	As	
WESTPHALIAN	26	4	52	1.0	15	96	2591	0.9	35	5	33
PERMIAN	38	8	56	1.0	29	80	2429	0.3	40	16	7
NAMURIAN	29	8	107	1.9	30	136	3286	0	44	16	49
DEVONIAN	45	143	99	1.5	53	198	3425	1.5	69	50	8
GRANITE	50	428	31	1.1	22	88	1817	0.1	7	22	30
BASIC IGNEOUS	44	102	125	1.0	41	218	7750	0.2	96	76	16

TABLE 4 - 3

Classification of 910 stream-sediment samples based on bedrock training set only, EDF untransformed data (After Howarth, 1971b, Table 5).

TRUE CLASS	PER CENT CORRECT
PERMIAN	42
WESTPHALIAN (BUDE FORMATION)	82
NAMURIAN (CRACKINGTON FM)	64
GRANITE	68
BASIC IGNEOUS	54

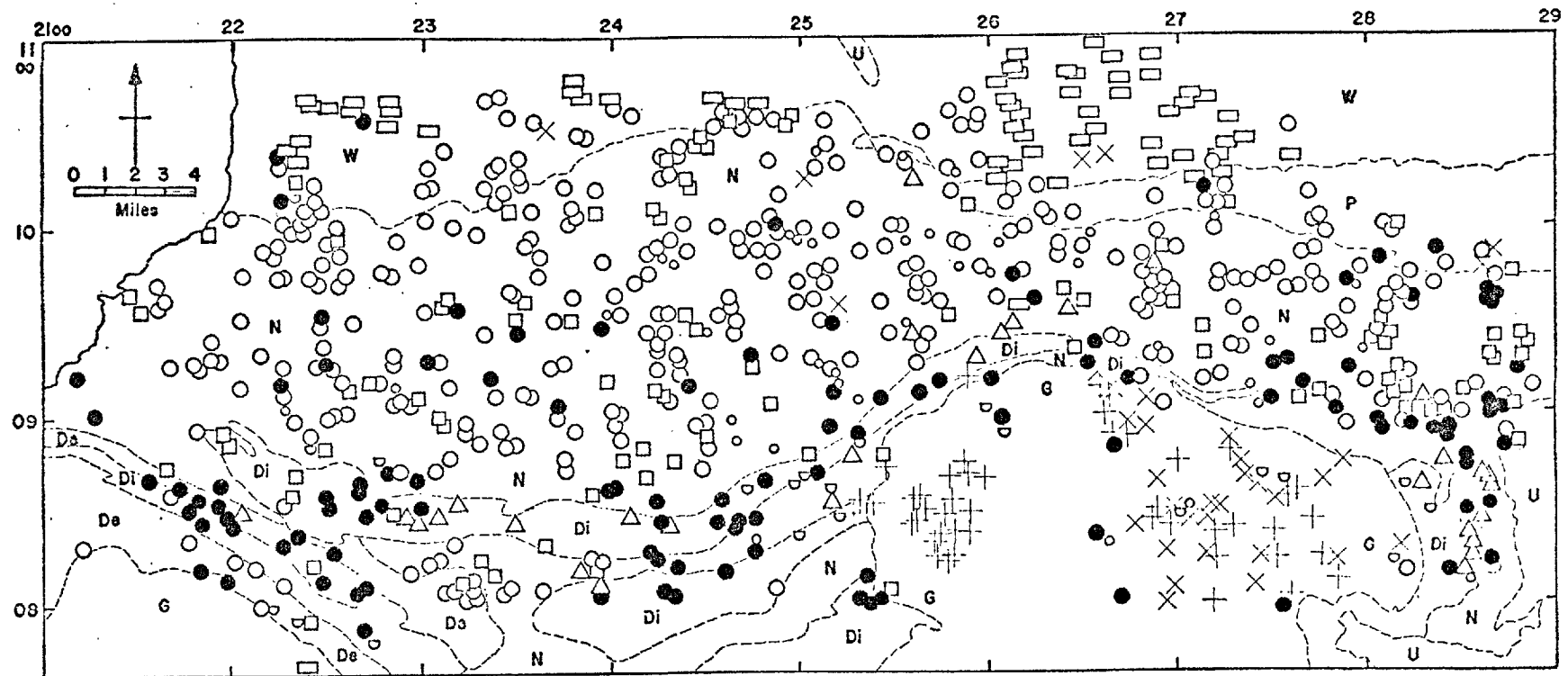
Minimum overall correct recognition rate (bedrock) = 64%

TABLE 4 - 4

Classification of 894 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data. (After Howarth, 1971b, Table 6).

TRUE CLASS		PER CENT CORRECT
PERMIAN		89
WESTPHALIAN (BUDE FORMATION)		62
NAMURIAN (CRACKINGTON FM)		79
GRANITE		79
BASIC IGNEOUS		97
Mineral- ization.	Cu	77
	Pb-Zn	77

Minimum overall correct recognition rate = 81%



DEVONSHIRE
DENDROGRAM REPRESENTATION
ORIGINAL MAP.
Variables: Pb, Sn, V, Mo, Cu, Zn, Ti, Ni, Ag, As
Untransformed data

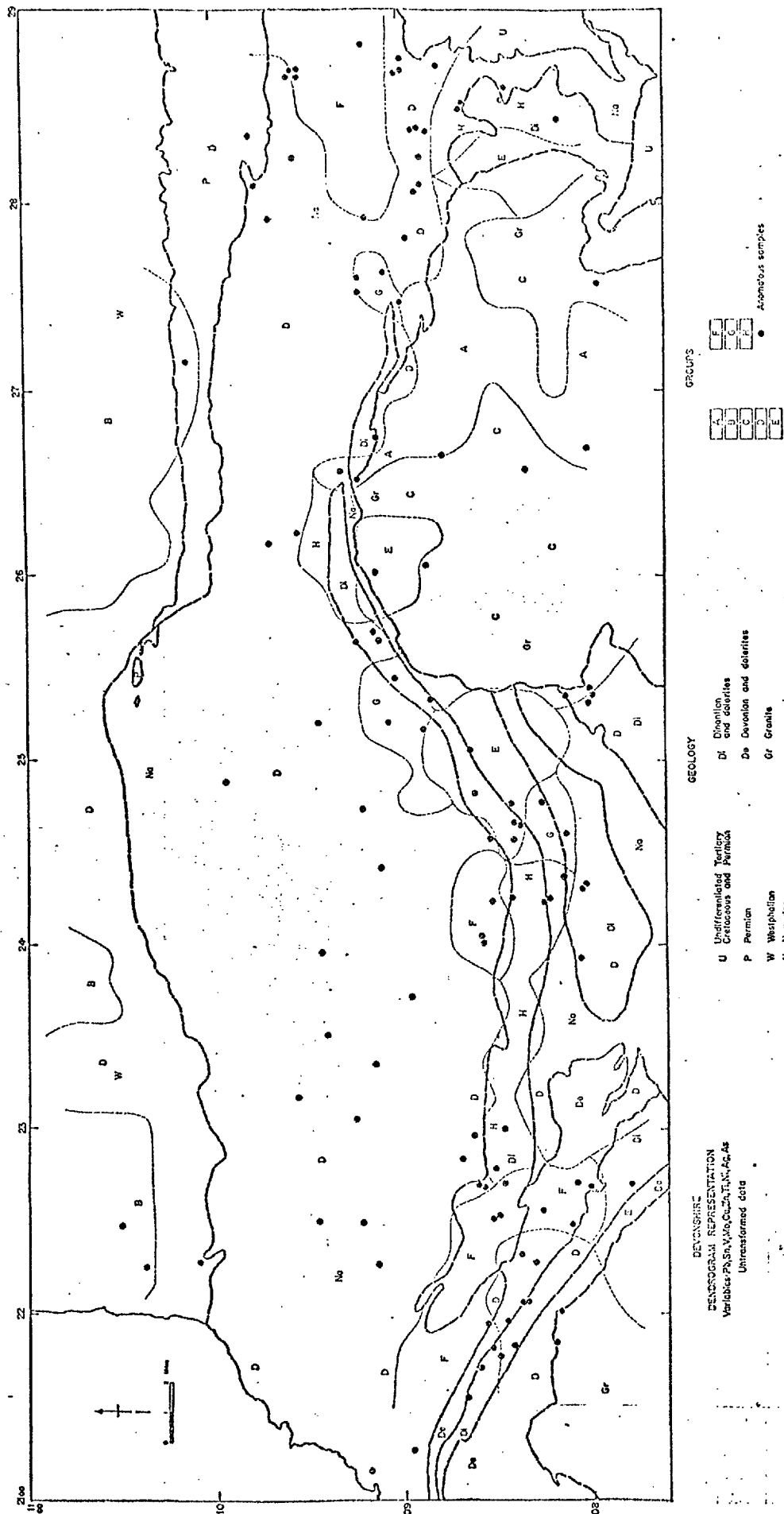
U Undifferentiated
Tertiary, Cretaceous
and Permian
P Permian
W Westphalian
N Namurian

Di Dinantian and
dolerites
Da Devonian and
dolerites
G Granite

GROUPS

X A
□ B
+ C
○ D
◐ E
◑ F
◒ G
△ H
● Anomalous samples

FIG. 4-4



DEVONSHIRE
SCHEMATIC BASIC DENDROGRAMS

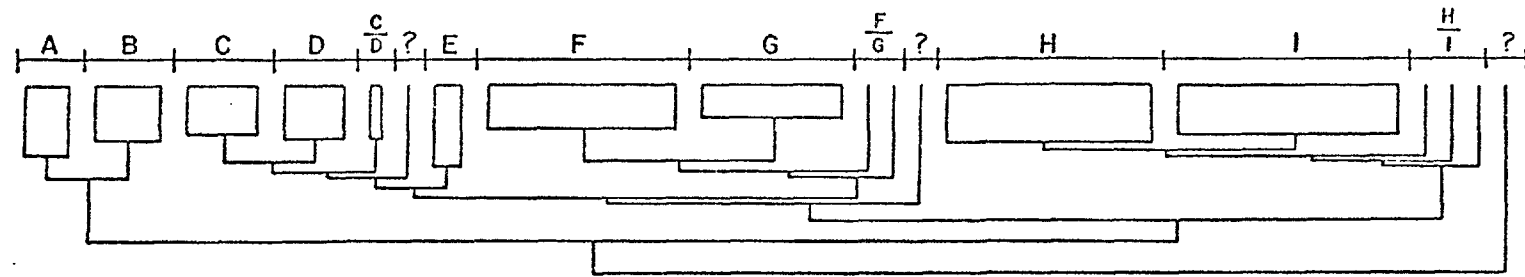
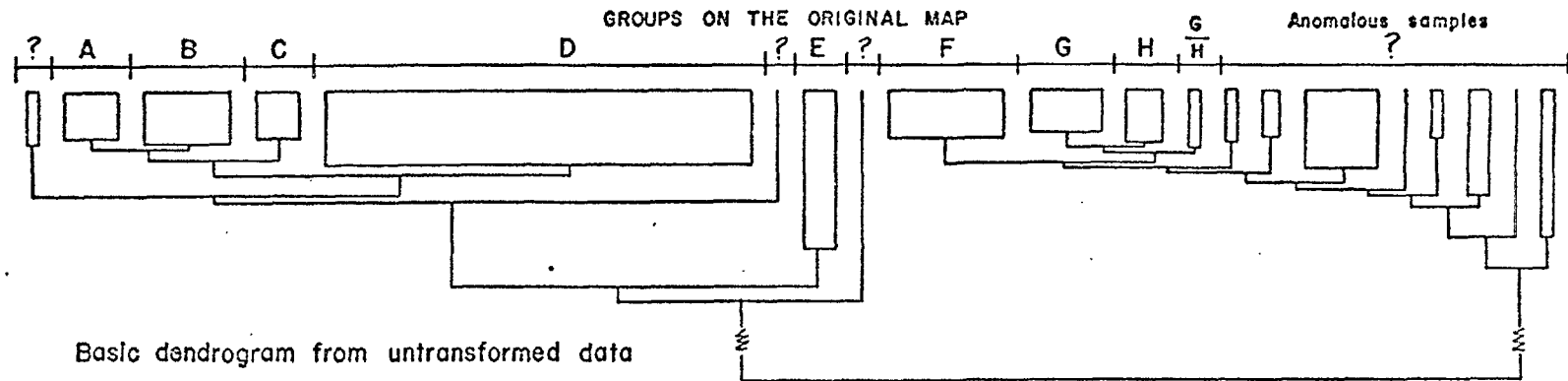


FIG. 4-6

TABLE 4 - 5

Arithmetic means ($\bar{x}$) and standard deviation (s), Devonshire, dendrogram representation, untransformed data; total sample size = 825.

	A	B	C	D	E	F	G	H	ANOMALOUS SAMPLES
SAMPLE SIZE	39	69	47	364	24	101	29	34	118
Pb $\bar{x}$	36	32	56	34	116	44	35	40	59
s	9	12	29	15	121	29	9	13	55
Sn	148 143	15 53	428 376	12 27	422 803	25 106	12 17	25 39	161 633
V	32 14	58 18	27 13	85 23	86 42	137 22	103 14	109 27	111 43
Mo	1 0.4	1 0	2 2	2 1	3 4	2 1	3 1	2 1	2 2
Cu	16 14	17 6	17 14	29 11	89 182	41 13	35 5	43 13	57 95
Zn	81 25	95 27	54 59	136 53	196 199	171 76	150 28	152 38	217 131
Ti	1685 573	2645 826	1703 731	3103 910	3458 1742	4020 1078	3724 690	4058 1397	4850 2634
Ag	0.2 0.3	1 0.2	0.2 0.2	0.1 0.1	1 1	0.1 0.1	0.1 0	1 0.2	1 4
Ni	13 13	33 7	9 8	42 14	51 36	57 19	47 9	59 28	90 64
As	27 29	7 8	15 21	18 27	111 209	23 25	24 21	37 29	76 161

Mean values in p.p.m.

2472/1055, Fig. 4-2) the sparse sampling points over the Westphalian enable group D to extend further north as a consequence of the overlap of successive window positions in the smoothing technique. The same phenomenon occurs on the eastern side of the Permian and Westphalian contact. Group D again appears over the Westphalian terrain between Holsworthy and south of Bude; its occurrence may be a response to the generally similar chemical composition between the Westphalian and Namurian sandstones predominating in the area (Table 4-6). Group B is better defined on the Westphalian rocks in contact with the Permian, suggesting a difference in chemical composition between the eastern and western sediments probably related to the higher proportion of shale interbedded with the sandstones on the eastern side.

Groups D, F and (to a lesser extent) G are mainly distributed over the Namurian except on the eastern side where F predominates. Groups B and D are in general significantly lower in Pb, Zn, Cu and Sn and to a lesser extent in Mo with respect to the other groups closely associated with the rest of the sedimentary formations (Dinantian and Devonian). These groups are chemically similar to the A and C granitic groups with generally lower metal values for some elements, but still higher in Zn, Ti, Ni, V and to a lesser extent Cu. This higher concentration is more conspicuous in group D, probably because of its spatial proximity to the basic igneous rocks, lavas and tuffs, and mineralization on the metamorphic aureole.

The classification of the stream sediments over the Namurian at the head of the Teign Valley near Duns^{Ford} (N.G.R., 2813/0892) as group F (which is different to the Namurian elsewhere) is in agreement with results previously obtained with EDF analysis (Howarth, 1971b). In the EDF results, this pattern was (on the basis of a training area containing sediments with intercalated basic igneous tuffs and lavas) classified as "basic igneous" although no such rocks are currently known in the area and field investigations failed to confirm their presence. Nevertheless, Lower Carboniferous

TABLE 4 - 6

Chemical composition (spectrographic analysis) of two Westphalian and Namurian sandstones collected south and southeast of Bude.
Metal values in p.p.m.

ELEMENT	WESTPHALIAN SANDSTONES		NAMURIAN SANDSTONES	
	No.16	No.13	No.14	No.15
Pb	30	16	20	30
Ga	20	16	13	13
V	130	100	85	60
Cu	20	20	13	16
Zn	200	300	100	200
Ti	2000	1600	1300	1600
Ni	40	40	30	40
Co	20	16	20	20
Mn	85	500	160	200
Cr	60	40	40	40

Rock sample descriptions and locations:

- No.16 Grey reddish medium-grained sandstone, somewhat weathered (N.G.R. 2198/1020).
 No.13 Brown reddish medium-grained sandstone. Railway cutting near (NW) Titson (2240/1018).
 No.14 Grey, medium-grained sandstone. Near (South) Trefouse(2222/0980).
 No.15 Dark grey, medium-grained fresh sandstone with limonitic coatings along fractures. Quarry near Newmill (2219/0988).

(Dinantian) sediments are recently believed to be present in the area (I.G.S., Exeter Office, pers. commun.) This pattern is repeated on a moving average map of the percentage of Devonian EDF classified samples relative to those assigned to the Namurian Crackington Formation (Howarth, 1972). The map reflects to some extent the known distribution of Devonian and Dinantian sedimentary rocks, although there is a transition zone which extends further to the north, and to the east at the head of the Teign Valley, than the northernmost boundary of the Dinantian sediments as mapped by the I.G.S. Howarth concluded that "this transition is not apparent but it may be of interest that the I.G.S. now believes Lower Carboniferous rocks to be present in the area at the head of the Teign Valley where the "majority vote" favours the Devonian rather than the Crackington Formation category".

The correlation of group F in this study with Dinantian and Devonian rocks which are both intruded by basic igneous bodies on the southwestern side of the area again suggests a close geochemical similarity between this group and the Dinantian and Devonian intruded by basic igneous rocks. If Dinantian, Devonian and basic igneous rocks exist at the head of the Teign Valley one may expect that such rocks would have been assigned to the "basic igneous" category by the EDF method (Howarth, 1971 b) because the "basic igneous" training area included the present F and D patterns, which both overlie mixed sedimentary and basic igneous rocks, lavas and tuffs in the southwest corner of the study area.

The Permian rocks are largely assigned to the Namurian category (D) because about 50% of the stream sediment samples over the Permian exposures are influenced by drainage transporting material northward from the Namurian outcrop.

Groups A and C are mainly distributed over the Dartmoor granite but not over the Bodmin Moor granite. The northeast and eastern parts of the granite are covered by freely-drained brown earth soils, and groups A and C form a composite pattern (Fig. 4-4). In the smoothed

map (Fig. 4-5), group C, which has the highest Sn content (Table 4-5), predominates on the eastern side of the Dartmoor granite where most of the tin mineralization occurs delineating almost perfectly the area of mineralization, and also over the western side where poorly-drained peaty gley soils are exposed in which the secondary environment has caused erratically higher concentrations of Mn, Co, Fe and Zn (Nichol and others, 1971); the role played by these elements in the delineation of the C pattern over this area is not known because most of these elements (with the exception of Zn) were not considered as variables. Group A appears to be more representative of the granite bedrock which is less influenced by mineralization because of its relatively low granite-type metal content (Table 4-5).

Groups H, E, G, F and D are restricted in area but occur irregularly distributed over the Dinantian and Devonian in the vicinity of the granites. Group H is distributed as an elongate belt between both granite bodies in the vicinity of Launceston in an area influenced by Dinantian rocks with spilitic and keratophyritic tuffs and lavas. It is again associated with Dinantian sedimentary rocks, lavas and tuffs south of Okehampton and east of the Dartmoor granite in the Teign Valley. Group E, the highest in Pb, Sn, Cu and As, is apparently a response to the dilution of stream sediments derived from the Sn mineralized granitic rocks and by material from the surroundings influenced by Pb-Zn, Cu and As mineralizations or basic igneous rocks. Group (E) behaves as an "anomalous" group because of its connection with the mineralized areas and the high degree of variability for most of the elements (Pb, Sn, Mo, Cu, Zn, As) as a response to the anomalous geochemical environments. Groups H, E, F and G are significantly enriched in V, Ti and Ni (Table 4-5) which seems to confirm the influence of basic igneous rocks (higher in Mn, Ni, Cr, Ti and V) intruding the Dinantian and Devonian sedimentary rocks. The occurrence of group D, typically associated with the Namurian, overlying Devonian and Dinantian sediments in the southwestern side of the area and basic igneous rocks, in contact with the Bodmin Moor granite is difficult to explain. It could be interpreted as a Namurian-type geochemical facies produced by meta-

morphism contemporaneous with the emplacement of the Bodmin Moor granite and the subsequent dilution of stream sediments derived from the surrounding lithologic units. The correlation of group D with the Namurian bedrock in the south-central part of the area is difficult because the stream sediments forming this classification pattern are derived from a mixture of Devonian, Dinantian and Namurian rocks; it is also interpreted as a Namurian-type geochemical facies produced by the natural dilution of different geological units.

Taking into consideration the group associations that can be correlated with each bedrock unit (without a priori geological knowledge) a fairly close visual delineation of the main contacts could be carried out: 1) between the Dartmoor granite (A and C association) and the Namurian-Dinantian category (D, E, F, G and H association), and 2) between the Westphalian (Permian?) (group B) and the Namurian-Dinantian. The almost perfect separation between group D and the association of E, G and H approximately delineates the contact between the Namurian and the Dinantian-Devonian.

The EDF analysis results (Table 4-7) partially confirm the A and C association over the Dartmoor granite because group B characterizing the Westphalian is this time assigned as another member of the association, probably because of the closer chemical similarity between the Granite and the Westphalian (Permian?) sandstones mainly with the elements: V, Mo, Cu, Zn, Ti and As (Table 4-2) in contrast with the Namurian-Dinantian. Group D, associated with most of the Namurian exposures, behaves as an intermediate group (between the A, B and C association and the rest of the groups) with a high variability of misclassification results it is well differentiated from the rest of the groups.

The majority of the "anomalous" samples are distributed along the E-W trending mineralized belt, between both granite intrusions and surrounding most of their margins. The group of "anomalous" samples attains the highest mean metal values for Zn, Ti, and Ni, elements generally associated with the bedrock and mineralization of the metamorphic aureole (Table 4-5). Beside group E (which

TABLE 4 - 7

Group classification results by EDF analysis, Devonshire, dendrogram representation, untransformed data results, $\sigma = 10.0$.

	ASSIGNED GROUP										% CORRECT
	A	B	C	D	E	F	G	H	ANOM.	?	
A	23	7	6	1	0	0	0	0	2	0	59
B	10	42	1	10	0	0	3	1	2	0	61
C	11	2	26	0	3	0	0	0	2	3	55
D	4	78	2	167	4	10	65	17	17	0	46
E	1	0	1	7	2	4	0	0	2	7	8
F	1	3	0	6	4	68	2	6	10	1	67
G	0	1	0	9	0	1	14	3	1	0	48
H	0	1	0	10	0	6	4	7	6	0	21
ANOM.	1	4	4	25	3	16	2	15	33	48	28

Overall correct recognition rate.
(Excluding anomalous group) = 49%

Overall correct recognition rate
(For all groups) = 46%

attains the highest mean values for Pb, Sn, Cu and As), Sn, As and Cu in the group of "anomalous" samples also attain significantly higher mean values in contrast with the relatively ^{un}mineralized areas outside the metamorphic aureole, with the exception of the granite. Pb does not show significantly higher values, probably because only a few samples are correlated with Pb-Zn deposits on the eastern side of the Dartmoor. In general, this group accounts for the highest V, Ti, Ag and Ni variability. The previous observations suggest the influence of either basic igneous rock detritus or mineralized sediments derived mainly from the metamorphic aureole, and to a lesser extent from the granite on the "anomalous" - classified stream sediments.

Individually only two or three "anomalous" samples on the eastern side of the Dartmoor granite (N.G.R., 2840-2860/0810-0850) could be responding to the influence of the known Pb-Zn mineralization upstream and probably to the adjoining Mn mineralization in the Dinantian rocks, which are relatively higher in Cr, Ni, Ti, and V (despite the absence of Mn in the variables considered); one of them on the southwestern tip is apparently more influenced by stream sediments derived from the granite enriched in Sn. One "anomalous" sample close to the northeastern margin of the Dartmoor granite at N.G.R., 2750/0900 is probably a response to the influence of a Cu deposit to the west at N.G.R. 2730/0910. The "anomalous" samples within the metamorphic aureole on the northern margin of the Dartmoor granite are probably a result of the nearby Cu deposits (N.G.R., 2500-2650/0885-0940). Those "anomalous" samples located strictly along the contact as a response to the radial drainage system of the granite, and the majority of those distributed within the granite itself can be correlated with the Sn deposits. Two "anomalous" samples close to the northern contact of the Bodmin Moor granite (2175/0815) fall within the area of influence of the adjoining Cu and Sn mineralization.

The Dinantian rocks immediately westward of the Dartmoor granite (2400-2500/0800-0900) and on the surroundings of Launceston (2200-2300/0850) are relatively higher in Cr, Ni, Ti and V and the conspicuous agglomeration of "anomalous" samples on both places is probably a

result of the nearby Mn mineralization, despite the absence of Mn in the variables considered.

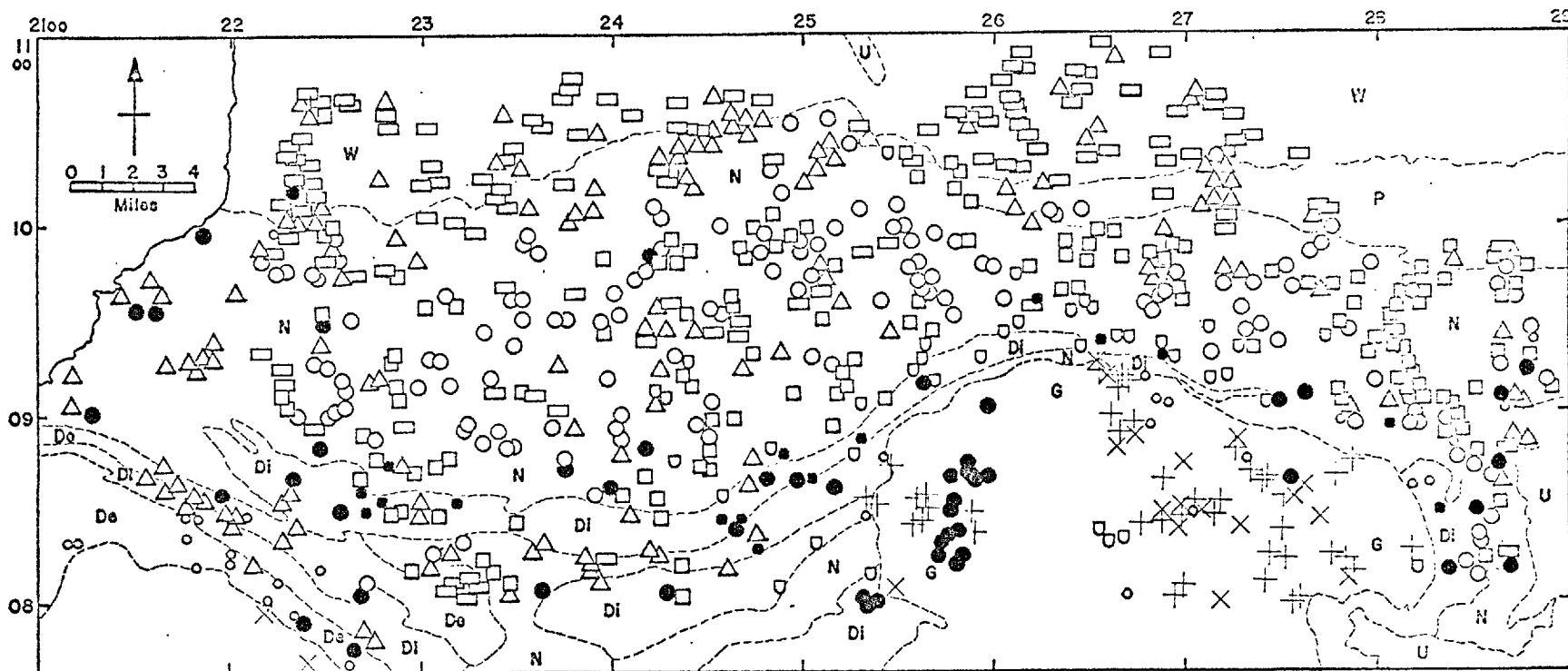
The "anomalous" samples overlying the Dinantian between Launceston and the western side of the Dartmoor granite could well be related to the high Mo values in the black pyritic shales within the Dinantian (Fletcher, 1968).

A significant proportion of "anomalous" samples over Dinantian and Devonian rocks immediately north of the Bodmin Moor granite (2150-2275)/0800-0900) and over the most southern Dinantian outcrop in contact with the western margin of the Dartmoor granite 2530/0800) show stronger correlation with basic igneous rocks, probably due to the relatively higher concentrations of Cr, Ni, Ti and V found in the stream sediments. The smaller number of "anomalous" samples distributed over the central part of the Namurian outcrop are of uncertain origin; those "anomalous" samples close to the head of the Teign Valley show significantly higher values of Zn, Ti, Ni and Mn. The remaining samples individually showed a chemical composition generally closer to the composition of the stream sediments derived from the Namurian bedrock.

A further classification of the "anomalous" samples alone was subsequently carried out. Twenty groups resulted. Of these, one was strongly correlated with the sedimentary and basic igneous rocks, lavas and tuffs of the Lower Carboniferous north of the Bodmin Moor granite. The remaining groups were not specifically related to any mineralization type. Their irregular and somewhat erratic spatial distribution was probably a response to contamination from the mining activity mainly in the southern part of the area.

b) Classification based on the log-10 transformed data.

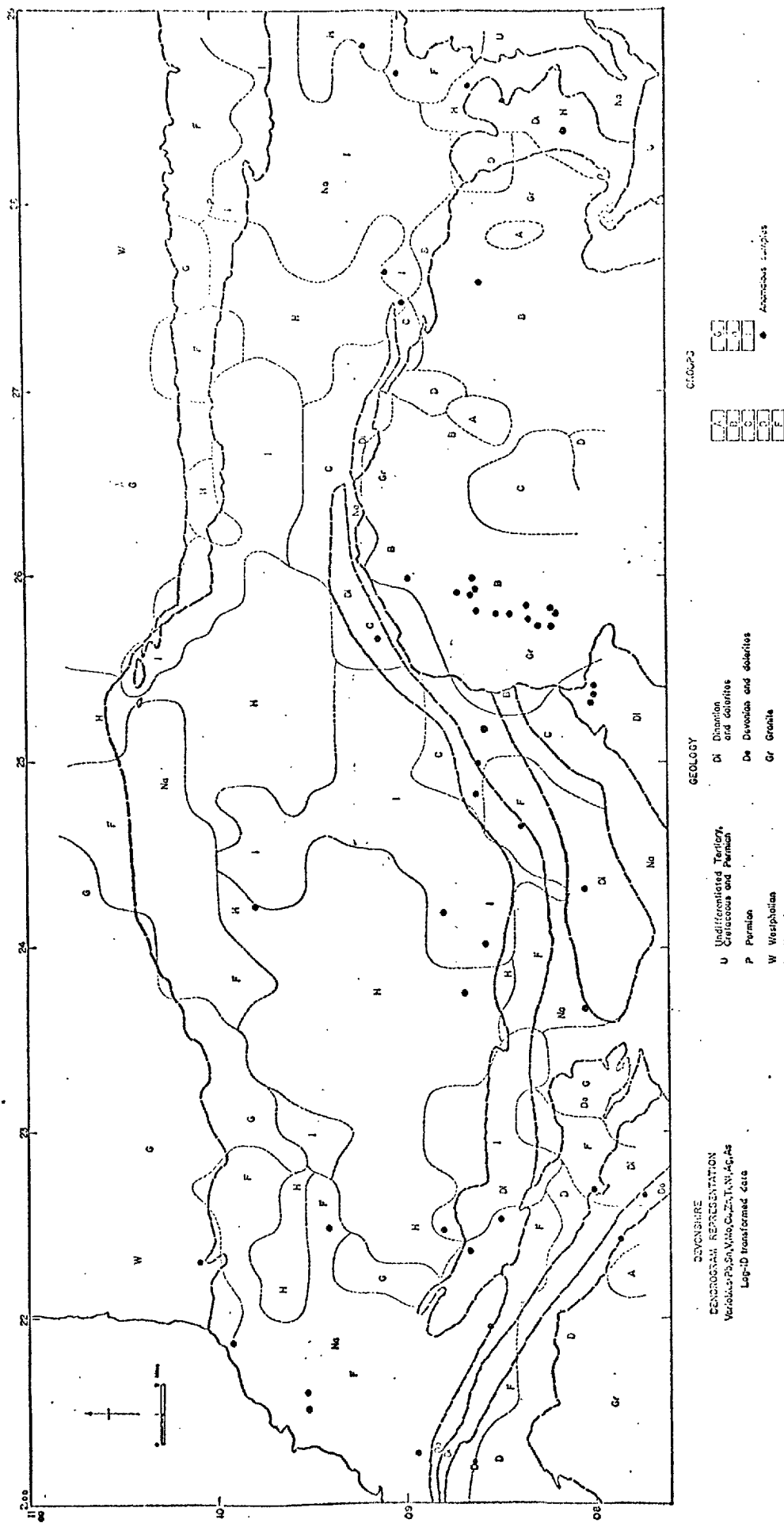
The original classification map can be seen in figure 4-7 and the smoothed results in figure 4-8. The basic dendrogram



DEVONSHIRE
DENDROGRAM REPRESENTATION
ORIGINAL MAP
Variables: Pb, Sn, V, Mo, Cu, Zn, Ti, Ni, Ag, As
Log-10 transformed data

GEOLOGY		GROUPS	
U	Undifferentiated Tertiary, Cretaceous and Permian	X A	△ F
P	Permian	+ B	□ G
W	Westphalian	○ C	○ H
N	Namurian	◊ D	□ I
Di	Dinantian and dolerites	● E	● Anomalous samples
De	Devonian and dolerites		
G	Granite		

FIG. 4-7



is shown in figure 4-6 and the statistical properties of each group are given in table 4-8.

The Westphalian is predominantly represented by group G, and to a lesser extent by group F. The close association between these two groups is confirmed by the EDF analysis (Table 4-9). In the smoothed map group G almost covers the whole of the Westphalian exposures and extends southward onto the Namurian along tributaries of the Tamar river (2250-2375/0950), probably as a response to the influence of Westphalian sediments downstream. The higher proportion of shale intercalations in the eastern half of the Westphalian is not detected by any specific group, in contrast to the untransformed results, probably because of the data compression produced by the log-10 transformation.

Groups H and I and, to a lesser extent, F and G overlie the Namurian in the smoothed map (Fig. 4-8). These groups are, however, mainly irregularly distributed as a composite over Dinantian and Devonian rocks in the south and southwest of the area and east of the Dartmoor granite, but without showing any well defined correlation with the main geological features. A consistent association between groups H and I is well shown by the EDF analysis results (Table 4-9). Only group C seems to show a response to the metamorphosed Dinantian and Namurian rocks along the northern and western margins of the Dartmoor granite and within the granite itself. Group I includes the stream sediments at the head of the Teign Valley and is also irregularly distributed over the Namurian and Dinantian rocks throughout the area.

The Permian rocks were not outlined by any specific group, probably because of the influence of the Namurian stream sediments derived from the south. The occurrence of group F over the Westphalian around the Black Torrington area is probably due to the same cause.

TABLE 4 - 8

Arithmetic means ($\bar{x}$) and standard deviations (s), Devonshire, dendrogram representation log-10 transformed data; total sample size = 664

GROUPS	A	B	C	D	E	F	G	H	I	ANOMAL. SAMPLES
Sample Size	21	31	36	31	16	121	126	98	134	50
Pb $\bar{x}(\log)$	1.7	1.6	1.7	1.7	1.9	1.5	1.5	1.5	1.5	1.7
Pb $s(\log)$	0.1	0.1	0.1	0.2	0.3	0.1	0.1	0.1	0.1	0.4
Sn	2.5	2.3	1.7	1.7	1.4	0.7	0.7	0.7	0.7	1.7
	0.3	0.6	0.3	0.6	0.4	0.3	0.3	0.2	0.2	0.8
V	1.5	1.4	1.9	1.9	2.1	2.0	1.8	2.0	2.0	1.8
	0.1	0.2	0.2	0.2	0.1	0.2	0.2	0.1	0.1	0.4
Mo	0	0.02	0.4	0	0.7	0	0	0.3	0.4	0.4
	0	0.1	0.1	0	0.3	0	0	0.1	0.1	0.3
Cu	1.3	1.1	1.6	1.5	1.8	1.5	1.3	1.5	1.5	1.4
	0.3	0.2	0.2	0.2	0.3	0.2	0.2	0.1	0.2	0.4
Zn	2.0	1.9	2.2	2.3	2.4	2.1	2.0	2.2	2.2	1.2
	0.1	0.1	0.6	0.2	0.2	0.2	0.1	0.1	0.1	1.5
Ti	3.2	3.3	3.5	3.6	3.7	3.5	3.4	3.5	3.6	3.4
	0.2	0.2	0.1	0.2	0.2	0.2	0.1	0.1	0.1	0.3
Ag	-1.0	-0.8	-0.7	-0.8	-0.4	-0.8	-0.7	-0.9	-0.9	-0.9
	0.2	0.4	0.5	0.4	0.5	0.5	0.5	0.3	0.4	0.4
Ni	1.0	0.8	1.6	1.5	1.8	1.7	1.6	1.7	1.7	1.5
	0.2	0.2	0.2	0.3	0.2	0.2	0.1	0.1	0.2	0.6
As	1.4	1.0	1.6	1.8	1.7	1.2	0.6	0.7	1.4	0.8
	0.4	0.4	0.6	0.4	0.4	0.3	0.2	0.3	0.2	1.3

Mean values in p.p.m. (log).

TABLE 4 - 9

Group classification results by EDF analysis, Devonshire, dendrogram representation, log-10 transformed data results, $\sigma = 0.2$.

	ASSIGNED GROUP											% CORRECT
	A	B	C	D	E	F	G	H	I	ANOM.	?	
TRUE GROUP:	A	14	6	1	0	0	0	0	0	0	0	67
	B	7	21	2	1	0	0	0	0	0	1	68
	C	1	2	27	3	1	0	1	0	1	0	75
	D	3	1	1	20	0	3	1	0	2	0	65
	E	0	0	4	0	5	1	0	1	5	0	31
	F	0	0	2	6	1	88	20	1	1	0	73
	G	0	0	0	0	0	7	119	0	0	0	94
	H	0	0	1	0	1	0	1	85	10	0	87
	I	0	0	3	1	2	3	1	8	112	4	84
	Anom.	2	1	3	5	3	1	0	0	10	25	50

Overall correct recognition rate
(Excluding anomalous group) = 80%

Overall correct recognition rate
(For all groups) = 78%

Groups F, G, H and I in general attain intermediate mean metal values, with respect to the rest of the groups (except the "anomalous" samples), in Pb, Mo, Cu, Zn, Ti, Ag and As and higher values in V and Ni (except group E) and lower in Sn; however, group G is characteristically lower in V, Cu, Zn, Ti, Ni and As probably responding to the highly arenaceous nature of the Westphalian rocks.

Groups A and B, and a few samples assigned to groups C and D, occur over the Dartmoor granite. The association of these groups is well shown by the EDF analysis results (Table 4-9). Groups A and B are more representative of the granite and have the highest mean values of Sn probably due to the Sn mineralization; Pb also attains, to a lesser extent, higher values probably because of the proximity of the Pb-Zn mineralization on the metamorphic aureole. These last two groups tend to attain lower metal values for the rest of the elements and could constitute an individual group on the basis of their statistical properties given in table 4-8. A group of "anomalous" samples occurs on the western side over the area of poorly-drained peaty gley soils in which the secondary environment has produced an anomalous increase in the concentration of Mn, Co, Fe and Zn. A group of "unknown" samples (of unusual composition) was detected over the same area with the application of the EDF analysis (Howarth, 1971 b, p.B147).

Groups C and D are more associated with rocks of the metamorphic aureole and could constitute a single group on the basis of their statistical properties (Table 4-8). Both groups attain the highest concentration of Sn, after groups A and B, which suggests the strong granite influence in the stream sediments already confirmed by field investigation. Group E (eliminated during smoothing) appear to be a group with the highest mean metal values (with the exception of Sn and As) and appears to be geochemically more associated with groups C and D, both characteristic of the granite margins. The majority of the E samples are seen distributed on and around the Mn mineralization. The EDF analysis results (Table 4-9) have shown group E

to be closely associated with group C and the group of "anomalous" samples.

In the Bodmin area groups A and D are related to the granite, of these, the stream sediments of group A are exclusively derived from the granite. Samples assigned to group D are distributed on the contact or outside the margin of the granite on streams carrying a mixture of sedimentary and granitic detritus in which the granitic fraction still predominates. The same dilution effect between sedimentary country rock and granitic detritus was also observed in areas near the eastern contact of the Dartmoor granite where group D is present.

Taking into consideration the visually estimated group associations confirmed by the EDF analysis (Table 4-9), a fairly accurate delimitation of the main contacts could be carried out (without a priori geological knowledge): 1) between the Dartmoor granite and the Namurian-Dinantian categories and 2) between the Westphalian (Permian?) and the Namurian-Dinantian.

The proportion of "anomalous" samples is much smaller in comparison with those resulting from analysis of the untransformed data. The majority of the "anomalous" samples, as in the untransformed data results, are located along the mineralized belt between the granitic masses and surrounding most of their margins. This group of "anomalous" samples attains similar high mean values of Pb and Sn to groups A, B, C and D, probably because of its proximity to the Pb-Zn and Sn mineralizations. The rest of the elements showed a diversity of mean values and, the most conspicuous characteristic shown by the group is the very large variability for most of the metal values with respect to the rest of the groups with the exception of group E. Group E also behaves as an "anomalous" group not only because of its high metal concentrations, but also because of the large variability of the values; it is also distributed over areas evidently influenced by mineralization and probably basic igneous rocks.

In general, the "anomalous" samples on the eastern side (2800-2900/0810-0850) detect the Pb-Zn mineralizations within Lower Carboniferous rocks of the Teign Valley. Two "anomalous" samples close to the northeast margin of the Dartmoor granite (2750/0900), and one "anomalous" sample on the northwestern margin (2560/0910) coincide with several Cu deposits. The Mn mineralization within the Dinantian rocks west of the Dartmoor granite (2400-2500/0790-0850) which are relatively higher in Cr, Ni, Ti and V is probably the cause of "anomalous" samples in this area despite the absence of Mn in the variables considered. It is probable that the "anomalous" samples between Launceston and the western side of the Dartmoor granite are a result of the high Mo values in the black pyritic shales within the Dinantian sequence (Fletcher, 1968). Two "anomalous" samples on the northeastern side of the Bodmin Moor granite (2230-2260/0775-0790) are a reflection of the Sn and W deposits located upstream. The remaining "anomalous" samples are of uncertain origin; those samples distributed over the Namurian showed a chemical composition generally closer to the composition of the stream sediments derived from the Namurian bedrock.

Classification of the "anomalous" samples alone yielded seven groups, however, no clear picture reflecting the different mineralizations or mineralized areas resulted, possibly because of the influence of contamination, unknown limitations of the method, or human bias in interpretation of the results.

ii) R-mode results.

The R-mode analysis results were interpreted in the same manner as for the Denbighshire area.

In the case of the untransformed data only the Sn-As association (characteristic of elements present in the granite mineralization and in the metamorphic aureole) was consistent. The remaining elements did not display any well defined association or groups of associations.

With the log-10 transformed data the most conspicuous associations were: Pb-Cu, Sn-As-Mo-Ag, and V-Zn-Ni-Ti. The first

association (although incomplete) would correspond to the mineralization mainly associated with the metamorphic aureole, whereas the second seems to be related to mineralization within the granite (Sn) and within the aureole (As). The third association could apparently be correlated with the basic igneous rocks.

If the results are compared with those of factor analysis obtained by Khaleelee (1969, p.283-287) it is clear that there is no close correlation between the associations found with cluster analysis and those detected with factor analysis. This lack of coincidence is largely due to the smaller number of variables used in this study in contrast with the larger number used by Khaleelee. It is probably also due to the limitations of the dendrogram method which produces only a limited picture of the associations compared with that which results from a factor analysis.

According to the R-mode analysis results, it is considered that the log-10 transformation has given a closer approximation to trace element associations expected on geological grounds especially those forming the geochemical aureole around the granite. Khaleelee (1969) also reported better results with the log-10 transformed data than with the untransformed data results.

iii) Discussion.

As it was pointed out for the Denbighshire area, hierarchical cluster analysis has again proved to be effective for the location of mineralization even in very contaminated areas, whereas it is much more limited for the delineation of the bedrocks, being restricted only to broad geological units (e.g., Westphalian (Permian?), Namurian-Dinantian (Permian?), granite) if no a priori geological knowledge is taken into consideration.

Both the untransformed data and the log-10 transformed data outlined the same number of groups, however, the untransformed results were more successful in the case of groups restricted to the Upper Carboniferous rocks (D) and the Lower Carboniferous and Devonian rocks in the south-central part of the study area. The

Westphalian was better differentiated with the log-10 transformation of the data, whereas the Permian was not detected in either case.

The quantitative evaluation of the performance of the results is difficult because (as was pointed out for Denbighshire) several groups occur simultaneously over a single geological formation with no obvious (detailed) geological features which might explain such classification patterns. Therefore, more detailed geological information and geological field control is desirable in future investigations. Without taking into consideration any a priori geological knowledge, EDF analysis could also be applied using the broad taxonomic groups found by cluster analysis as "training areas" and the results would undoubtedly broadly approach those described by Howarth (1971b).

In the untransformed results 118 samples (14%) of the whole data set were considered "anomalous" samples. 45 of these 118 (38%) were related to known mineral deposits, however, 37 of the 118 (31%) were related to Lower Carboniferous sedimentary rocks (and suspected equivalents at the head of the Teign Valley) intruded by basic igneous rocks with significantly higher Cr, Ni, Ti and V stream sediment contents. The area of poorly-drained peaty gley soils on the western Dartmoor was not detected. The rest of the "anomalous" samples (31%) were apparently unrelated to any source of mineralization or contamination, consequently making the detection of the mineralized areas less evident. However, those "anomalous" samples related to geochemically anomalous areas tended to occur in clusters along the E-W trending mineralized belt, whereas the rest were sparsely distributed over wider (uncontaminated) areas barren of mineralization.

In the log-10 transformed data results only 50 samples (6%) of the whole data set were taken as "anomalous" samples. 20 of these (40%) were related to known mineral deposits and only 6 out of the 50 (12%) were related to the Lower Carboniferous sedimentary rocks (and suspected equivalents at the head of the Teign Valley) intruded by basic igneous rocks; 13(26%) were distributed over an area of poorly-drained peaty gley soils on the western side of the Dartmoor granite; the rest of the "anomalous" samples (22%) were apparently not related to any significant source of mineralization

or contamination. In this case the geochemically anomalous areas were less obscured by areas barren of mineralization and contamination, despite the lower proportion of samples classified as "anomalous" (6%) compared to the untransformed results (14%).

4-4 Non-linear mapping algorithm.

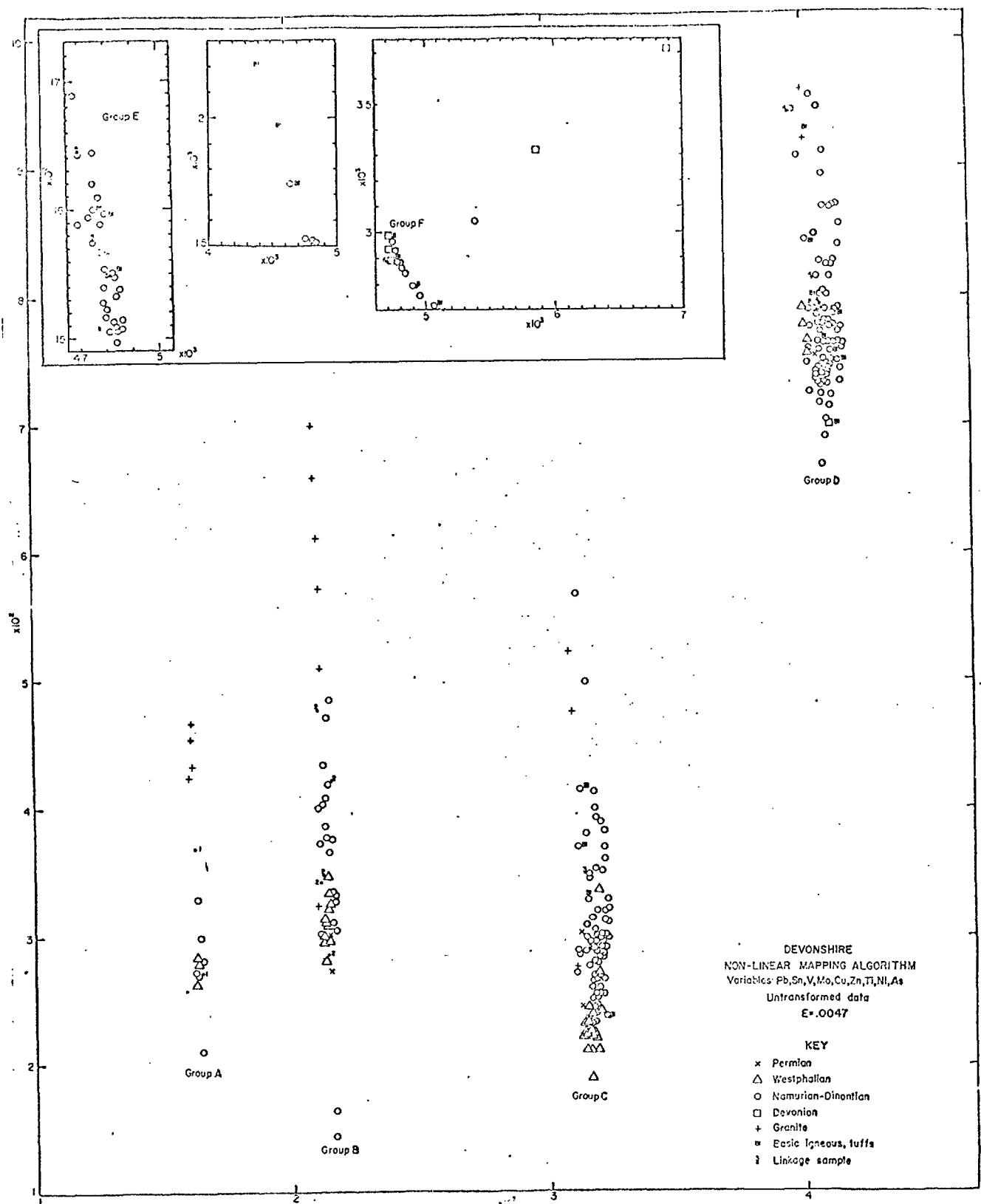
i) Results.

Results are based on the same data set as used for the hierarchical cluster analysis with the exception of silver, which was excluded because the many zero values caused very deformed log-10 transformed NLM mappings which were not representative of the geology of the study area.

a) Classification based on the untransformed data.

Figure 4-9 shows the "basic" groups in a set of 340 samples defined using "linkage samples"; this set of 340 samples showed a small additional group between groups E and F which was not present in the first set of 500 samples, but because of its unimportance it was not plotted on the original map (Fig. 4-10). All the groups, with the exception of D and E, show overlapping among the different geological units. The Namurian and Dinantian were considered as a whole unit for correlation with the bedrock geology because they showed no significant separation in the mapping.

Each of the A, B and C groups (Fig. 4-9) show strong separation of the geological units along the ordinate axis, with the Westphalian having low values, the Namurian-Dinantian intermediate values, and the granite with higher values. The transition between each geological unit in each "basic" group is overlapping, but in general the areas covered by samples representative of each geological feature are clearly outlined. A "majority-vote" smoothing applied to each "basic" group would undoubtedly decrease this



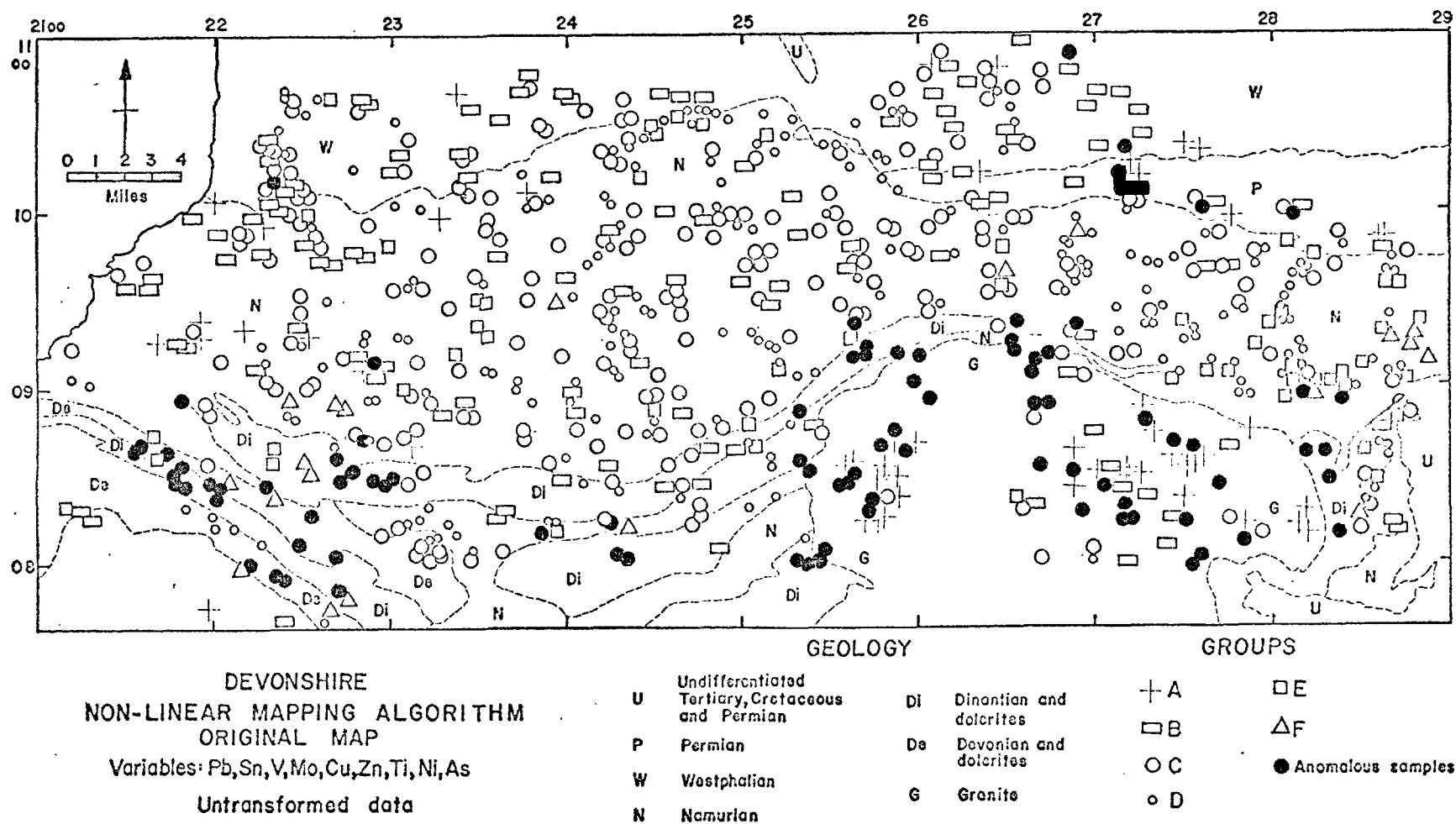


FIG. 4-10

overlap. Group D consists almost entirely of the Namurian-Dinantian category with the exception of a few samples classified as Westphalian, granite or Devonian. Group E is exclusively composed of the Namurian-Dinantian category, and group F by Namurian-Dinantian and Devonian units. The few Permian stream sediment samples occupy erratic positions among groups B, C and D. The stream sediment samples which are more influenced by the basic igneous rocks are scattered from groups C to F and do not show restricted positions in each group. Generally samples which are significantly influenced by basic igneous rocks behaved as "anomalous" samples within the "basic" groups.

The statistical properties of the basic groups are included in table 4-10. The smoothed classification map can be seen in figure 4-11. As in Denbighshire, a significant proportion of the "basic" groups are not representative of a single geological formation and their correlation with the main geological features is therefore limited.

Groups B, C and D predominate over the Westphalian; a minor proportion of samples were classified as group A. However, group B seems to be more representative of the more permeable arenaceous Westphalian rocks, by comparison with the rest of the groups overlying less permeable sedimentary and basic igneous rocks, lavas and tuffs, because of its generally lower content of Pb, V, Mo, Zn, Ti, Ni and As (Table 4-10); its higher concentration of Sn is probably due to some Permian stream sediment samples which have been enriched with cassiterite derived from the Dartmoor granite.

Over the Namurian-Dinantian sequence groups B, C, D, E and F occur throughout. Groups C, D, E and F (representative of bedrock units), overlying sedimentary and basic igneous rocks, lavas and tuffs, have in general higher metal values, except for Sn (Table 4-10). No significant chemical difference exist among the groups but only a gradual increase in concentration for most of the elements is observed from group C up to F. However, groups E and F (which have

TABLE 4 - 10

Arithmetic means ($\bar{x}$) and standard deviations (s), Devonshire, NLM algorithm, untransformed data; total sample size = 726

GROUPS	A	B	C	D	E	F	ANCM. SAMPLES
SAMPLE SIZE	41	101	228	197	53	17	89
Pb $\bar{x}$	41	31	33	43	47	57	62
s	29	11	13	41	17	63	66
Sn	71	23	12	18	15	14	438
	102	70	40	47	33	15	738
V	42	69	89	105	130	132	81
	21	28	29	27	33	29	56
Mo	1.7	1.4	1.7	2.0	2.1	2.3	1.9
	1.6	0.8	0.8	1.6	1.1	1.1	1.6
Cu	18	34	30	36	43	48	48
	11	93	13	13	12	10	104
Zn	74	114	137	157	193	206	154
	55	58	71	56	135	77	111
Ti	1463	2000	2987	4025	4962	6000	4314
	458	141	148	293	191	0	3433
Ni	22	37	45	56	63	71	60
	15	17	21	32	28	32	65
As	11	18	19	23	31	27	104
	14	46	34	33	30	24	207

Mean values in p.p.m.

higher metal values) coincide with areas north of the Bodmin Moor granite underlain by Lower Carboniferous and Devonian rocks with basic igneous intrusions, lavas and tuffs enriched with Mn, Ni, V and Ti. These groups are also seen at the head of the Teign Valley. Groups D and E predominate over the eastern side of the main outcrop and coincide with a sequence formed mainly of shale with few intercalated sandstone beds. Group E, which is basically composed of samples derived from a single geological unit, corresponds to the area at the head of the Teign Valley where Lower Carboniferous, basic igneous and Devonian rocks may be present (I.G.S. pers. comm., Howarth, 1971b, 72). On the central and western side of the main Namurian outcrop the predominance of group C corresponds to a lithology basically composed of shale with a large proportion of intercalated sandstone beds. Immediately south of Black Torrington (2425-2500/1000-1050) in the area of the Westphalian-Namurian contact group D predominates, coinciding with terrain underlain mainly by shales with a low proportion of intercalated sandstone beds, reproducing similar classification results to those obtained over the eastern side of the main Namurian exposures.

Group A predominates over the Dartmoor granite, and the associated groups B and C occur to a lesser extent. Group A seems to be more representative of the granite bedrock and tends to attain the lowest metal values (except for Sn and Pb) (Table 4-10) in contrast with the other groups which are representative of the bedrock units. The highest Sn content (excepting the group of "anomalous" samples) is probably due to the influence of the tin mineralization within the granite. The majority of the samples distributed on the western part of the Dartmoor granite and strongly influenced by the nature of the secondary environment were mainly classified as group A or less often as "anomalous" samples. Over the Bodmin Moor granite the A, B, C association is not evident, probably because of the scarcity of the sampling sites.

The EDF analysis results (Table 4-11) only show the A, B, C association characterizing the Dartmoor granite. The higher correct recognition rates for each group again confirm, as in Denbighshire, the existence of these geochemically "natural" groups (geochemical facies) strictly on the basis of their chemical composition in terms of the untransformed data.

Taking into consideration the visually estimated group association confirmed by the EDF analysis only the granite could be well delineated from the rest of the sedimentary sequence if no a priori geological knowledge is available. No clear differentiation between the Namurian-Dinantian (-Permian?) and the Westphalian (-Permian?) categories could be carried out.

The majority of the "anomalous" samples are distributed along the E-W trending mineralized belt surrounding the granite intrusions and over the Dartmoor granite itself. The group of "anomalous" samples attains the highest concentrations of Pb, Sn and As, and to a lesser extent Cu (Table 4-10), which are most of the elements characterizing the main mineralizations of the area. This group also shows the highest variability for most of the elements with the exception of Mo and Zn. Three "anomalous" samples east of the Dartmoor granite, within Dinantian rocks (2825-2835/0845-0860) are related to Pb-Zn deposits to the north and also to Fe deposits within the granite. The "anomalous" samples on the Dartmoor granite-Dinantian contact around the National Grid coordinates 2650-2690/0900-0925 can be correlated with drainage from nearby As, Cu and Sn deposits. The "anomalous" samples overlying the granite and on the contact with the Namurian southwest of Okehampton (2590-2610/0880-0920) are largely a response to the Sn mineralization within the granite. Few "anomalous" samples coincide with the Mn mineralization east and southeast of Launceston, whilst to the west an appreciable number are distributed over an area strongly influenced by near-by Mn deposits. Three

TABLE 4 - 11

Group classification results by EDF analysis, Devonshire, NLM algorithm, untransformed data results, $\sigma = 10.0$

		ASSIGNED GROUP								
		A	B	C	D	E	F	ANOM.	?	% CORRECT
TRUE GROUP	A	37	2	0	1	0	0	1	0	90
	B	1	95	1	0	0	0	3	1	94
	C	0	7	215	2	0	0	4	0	94
	D	3	0	4	179	6	1	4	4	91
	E	0	0	0	2	49	0	1	1	92
	F	0	0	0	0	0	17	0	0	100
	Anom.	6	6	4	3	2	1	48	19	54

Overall correct recognition rate

(Excluding anomalous group) = 93%

Overall correct recognition rate

(For all groups) = 88%

"anomalous" samples over the Devonian near the northeastern border of the Bodmin Moor granite (2220-2240/0790) are down drainage from Sn deposits within the granite. The agglomeration of "anomalous" samples over Dinantian and Devonian rocks north of the Bodmin Moor granite (2150-2210/0830-0875) and those on the southwestern contact of the Dartmoor granite (2530/0800) show a close correlation with basic igneous rocks. These rocks are significantly higher in Cr, Ni, Ti and V than Dinantian rocks not intruded by basic bodies (Nichol and others, 1971, Table 4). A limited number of "anomalous" samples distributed over Namurian, Westphalian and Permian rocks are of uncertain origin, but their chemical compositions were generally similar to bedrock compositions.

The untransformed data were subsequently processed using only the variables Mn, Ni, Co and Zn. As in Denbighshire the same number of "basic" groups (13) occurred as those obtained with 9 variables. In the first three groups there was an evident separation between the granite and the Namurian-Dinantian categories; the Westphalian and the Namurian-Dinantian categories were similarly distributed from groups 4 to 10 as in the 9 variable mapping (Fig. 4-9). The Namurian-Dinantian category predominated in the remaining groups. Stream sediments classified as Devonian and basic igneous were irregularly distributed in the 4-variable mapping from groups 6 to 13. Subsequently, the "basic" groups were plotted on one inch to a mile geological map but the resulting picture was too complex to carry out efficient interpretation.

b) Classification based on the log-10 transformed data.

Part of the results obtained with the log-10 transformed data are shown in figure 4-12. This illustrates the NLM mapping based on 500 stream sediment samples. A similar structure was

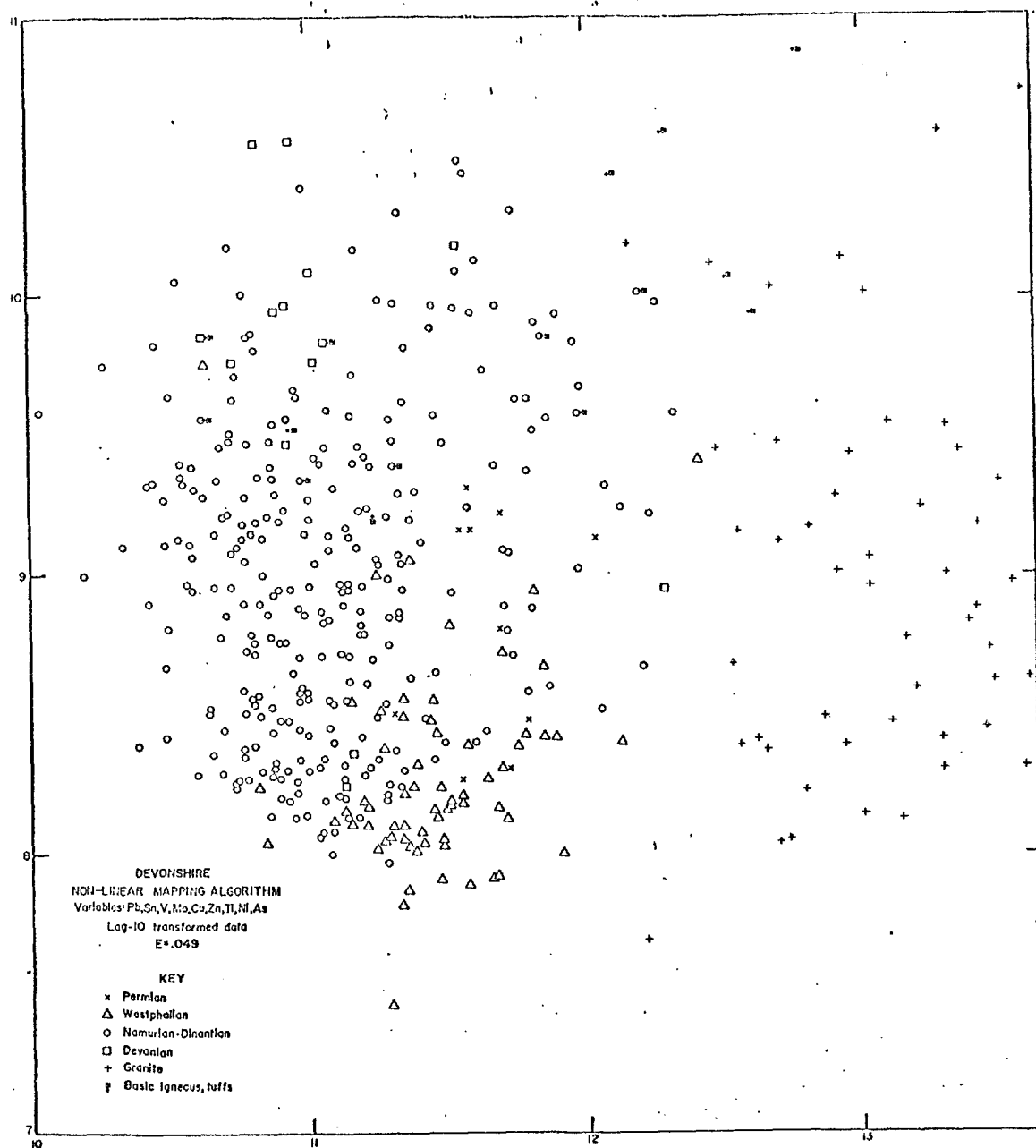


FIG. 4-12

obtained with the remaining 340 samples. The Westphalian, Namurian-Dinantian and granite categories are perfectly separated with a minimum of overlap. However, the Devonian and Permian show more overlapping with the Namurian-Dinantian and the Namurian-Dinantian-Westphalian, respectively. The Dinantian and Namurian were grouped together because of the complete overlap between both units. Stream sediments completely or partially influenced by basic igneous rocks are consequently irregularly distributed within the Namurian-Dinantian and Devonian groups. Stream sediments connected with anomalous geochemical environments (e.g. mineralization, contamination and secondary environment) were generally distributed around the periphery of the main cluster.

As in Denbighshire, the log-10 transformed data were also normalized but no improvement resulted.

The "anomalous" samples derived from the untransformed data were subsequently re-classified following the log-10 transformation of the data (Fig. 4-13). The Mn mineralization within the Dinantian is characterized not only by Mn itself but also by higher concentrations of Cr, Ni, Ti and V and is perfectly separated from the Sn mineralization within the granite. The Pb-Zn, Cu and As mineralizations formed an intermediate mixed group. Stream sediments influenced by the peaty gley soils over the western side of the Dartmoor granite, in which the secondary environment has caused an anomalous increase of some metals, are completely isolated by the mapping from the main cluster and form a striking group of "anomalous" samples unconnected with mineralization and contamination.

The log-10 transformed data were subsequently processed using only the variables: Mn, Ni, Co and Zn. The separation of the groups after processing the first 500 samples underwent a noticeable improvement because the granite category was then perfectly separated from the other categories. The Permian

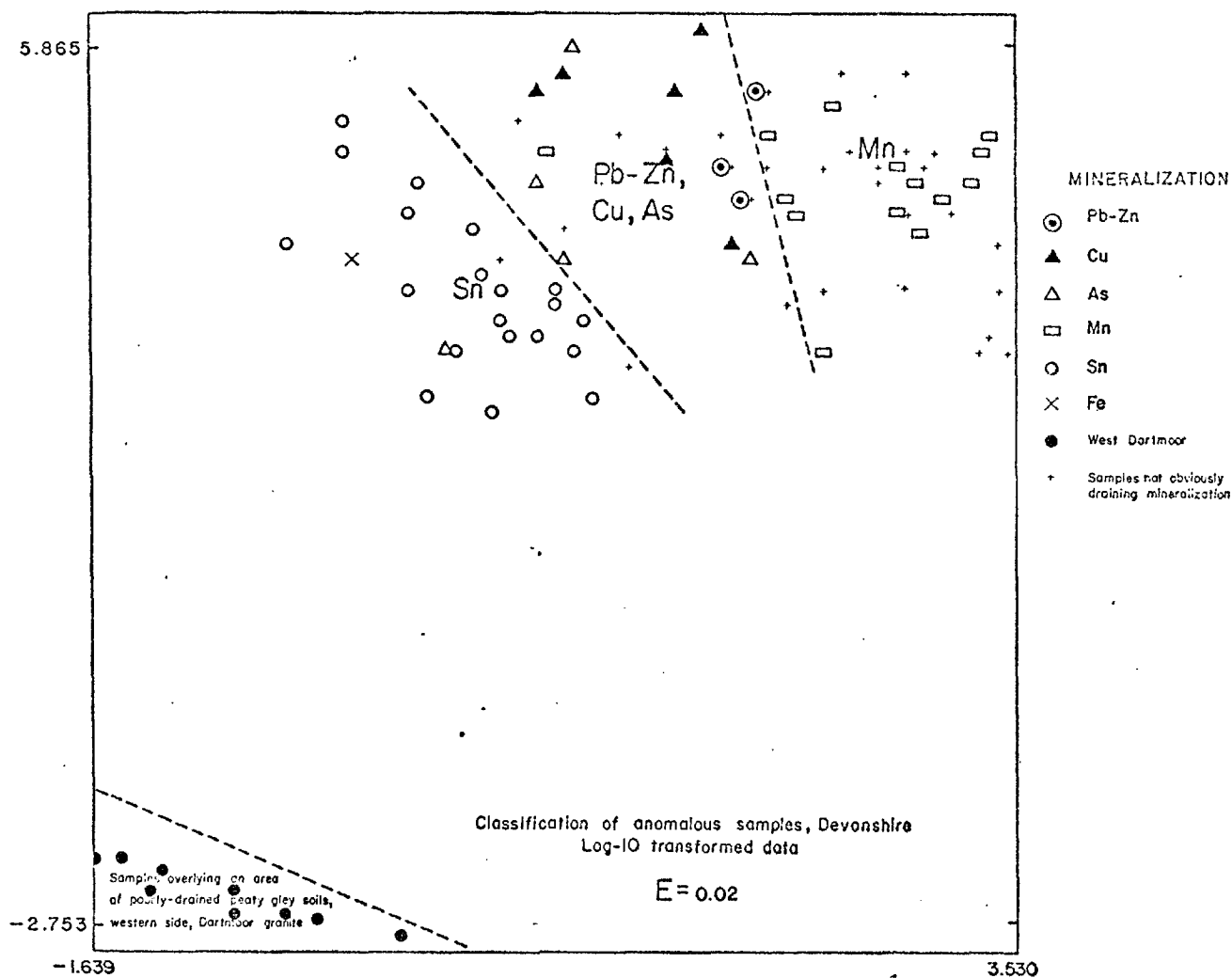


FIG. 4-13

group was also better separated from the Namurian and Westphalian. The rest of the groups were similar to figure 4-12. When the remaining 340 samples were processed the granite was again poorly separated, but this time the Permian was irregularly distributed over the Westphalian and the Namurian-Dinantian groups. The remaining geological units displayed the same general distribution observed in figure 4-12.

c) Classification based on the normalized data.

Part of the normalized raw data results are shown in figure 4-14. This illustrates the NLM mapping based on 500 stream sediment samples. A similar structure was obtained with the remaining 340 samples. The Westphalian, Namurian-Dinantian and granite categories are perfectly separated with the minimum of overlap. The Devonian and Permian show more overlap with the Namurian-Dinantian and the Westphalian categories, respectively, probably because of their chemical similarity. Stream sediments, as in the log-10 results, completely or partially influenced by basic igneous rocks are (consequently) irregularly distributed within Namurian-Dinantian and Devonian groups. These classification results repeat those obtained with the log-10 transformed data (Fig. 4-12). Those samples distributed along the periphery of the mappings, and evidently related with geochemically anomalous environments were considered as "anomalous" samples for further evaluation of mineralization (contamination) and secondary environment conditions.

ii) Discussion

The difficulties of overlapping among bedrock units and problem of correlation between the classification patterns (groups) and the main geological formations, were similar to those pointed out for Denbighshire, with the untransformed data. The log-10 transformed data, as in Denbighshire, produced much improved

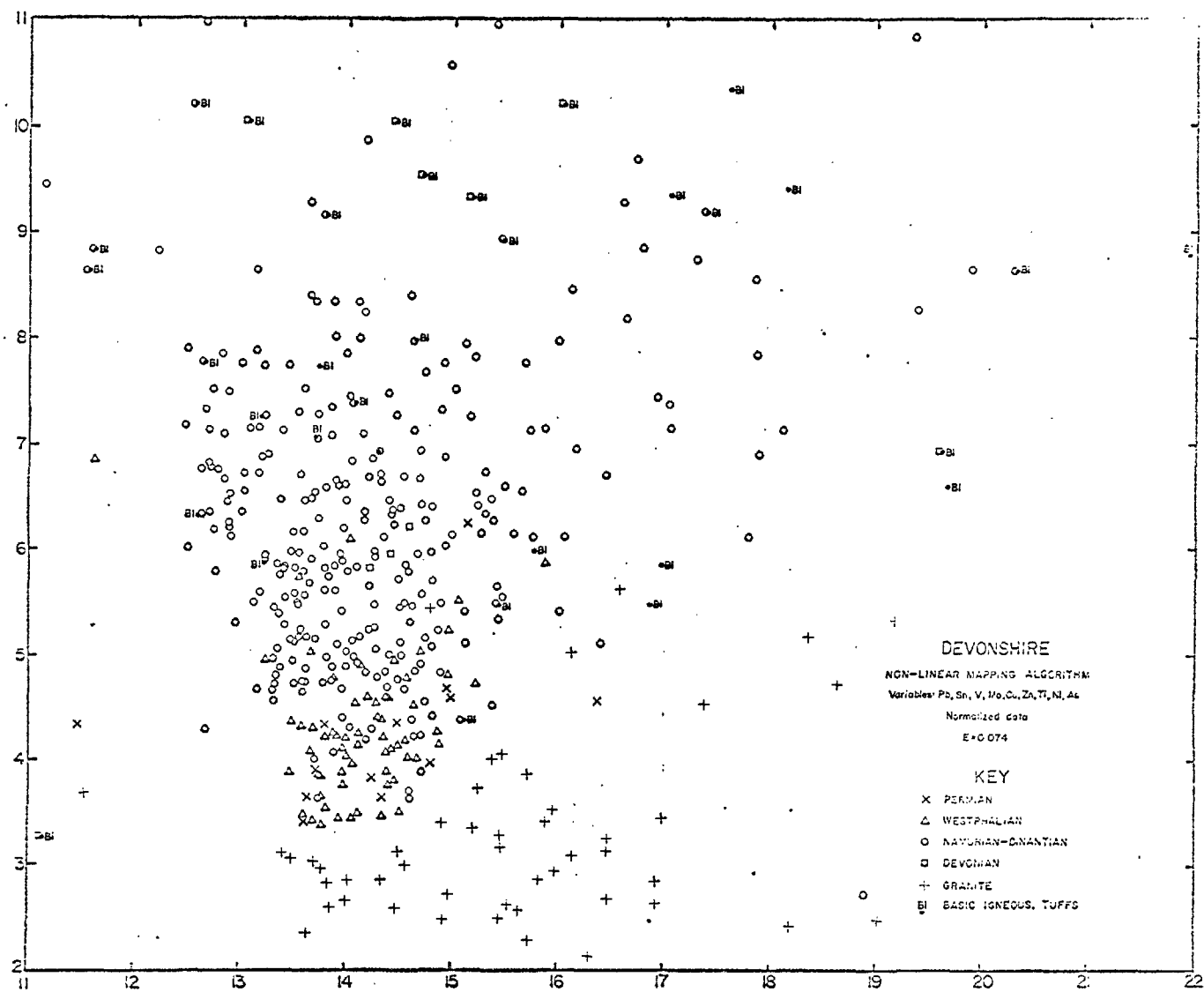


FIG 4-14

delineation of the main bedrock units with a minimum of overlapping. More detailed geological information and field control would again be necessary in order to test the performance of the method more exhaustively.

For practical purposes the NIM algorithm has again proved in this particular area to be a highly sensitive tool for the delineation of geochemically anomalous areas.

In the untransformed data results (9 variables) 89 samples (11%) of the whole data set were considered as "anomalous". 56 of these 89 (63%) were found to be related to areas more influenced by mineralization (contamination) and secondary environment (area of poorly-drained peaty soils of west Dartmoor); of these, 8 samples (9%) were found related with the nature of the secondary environment. 27 of 89 (30%) "anomalous" samples were estimated to be related to the Lower Carboniferous (Devonian?) sedimentary rocks without reported mineralization (and suspected equivalents at the head of the Teign Valley) intruded by basic igneous rocks very high in ~~Mn~~, Cr, Ni, Ti and V. The rest of the "anomalous" samples (7%) were scattered over areas without any noticeable source of mineralization or derived contamination.

In the log-10 transformed data results, 173 samples (21%) of the whole data set were considered as "anomalous". 99 of these 173 samples (57%) were estimated to be related to areas more influenced by mineralization (contamination) and secondary environment connected with the poorly-drained peaty gley soils of west Dartmoor; of these, 17 samples (10%) were found related with the nature of the secondary environment. 27 of 173 (16%) "anomalous" samples were estimated to be related to the Lower Carboniferous (Devonian?) sedimentary rocks without reported mineralization (and suspected equivalents at the head of the Teign Valley) intruded by basic igneous rocks. The rest of the

"anomalous" samples (27%) were not attributable to any obvious cause.

In the normalized raw data results, 173 samples (21%) of the whole data set were taken as "anomalous" samples. 108 of 173 (63%) were found to be related to areas more influenced by mineralization (contamination) and the secondary environment conditions in west Dartmoor; of these, 16 samples (9%) were found connected with the nature of the secondary environment. 44 of 173 (25%) "anomalous" samples were estimated to be related to the Lower Carboniferous (Devonian?) sedimentary rocks without reported mineralization (and suspected equivalents at the head of the Teign Valley) intruded by basic igneous rocks. The rest of the "anomalous" samples (12%) were not attributable to any obvious cause.

In general, the "anomalous"-classified stream sediments which were not attributable to any obvious cause, have chemical compositions which resemble the compositions of those stream sediments derived from relatively unmineralized (uncontaminated) bedrock sources. In this case, good control of pH and Eh would probably assist in the further discrimination of mineralization from secondary environment conditions.

The "anomalous" samples related to geochemically anomalous areas showed a denser agglomeration along the E-W trending mineralized belt and on the Dartmoor granite itself (in some cases), whereas those unrelated to any obvious source of mineralization, contamination or secondary environment conditions were only sparsely distributed.

4-5 Conclusions: Hierarchical cluster analysis and NLM algorithm.

1) The bedrock geology was better detected by the dendrogram representation (untransformed and log-10 transformed data) than with the NLM algorithm, untransformed data. The NLM algorithm,

log-10 transformed data, achieved comparable results to those of the dendrogram representation, nevertheless, these results appear to be more consistent because of the minimum amount of distortion introduced to the final two-dimensional mapping.

2) The mineralization (contamination) and the nature of the secondary environment on west Dartmoor were better detected by the "anomalous" samples classified with the log-10 transformed dendrogram representation than with the untransformed data (Fig. 4-12). Tin mineralization was better detected by the untransformed dendrogram representation than with the log-10 transformation.

3) Mineralization (contamination) and the nature of the secondary environment in west Dartmoor were better detected by the "anomalous" samples classified with the untransformed and normalized NLM mapping than with the log-10 transformed data (Fig. 4-12). Both untransformed and normalized data results achieved similar values. These figures were higher than those from the untransformed dendrogram representation and similar to the log-10 dendrogram. A better outline of the Sn mineralization was achieved with the untransformed NLM mapping than with the untransformed dendrogram representation.

4) The exclusive classification of the "anomalous" samples alone with the dendrogram representation was poorly performed. The best classification was obtained with the NLM algorithm, log-10 transformed data.

TABLE 4-12

Comparative successful recognition rates (%) of the "anomalous" samples in relation to different geochemical environments, Devonshire.

METHOD	DATA OPTION	MINERALIZATION (CONTAMINATION) AND SECONDARY ENVIR. WEST DARTMOOR.	SECONDARY ENVIRON- MENT, WEST DARTMOOR.	BASIC IGNEOUS (LOWER CARBON- IFEROUS/ DEVON- IAN?)	AREAS BARREN OF MINERAL- IZATION.	NO. OF "ANOMALOUS" SAMPLES.
		%	%	%	%	
NLM	UNTRANSFORMED (RAW DATA)	63	9	30	7	89
ALGORITHM	LOG-10	57	10	16	27	173
	NORMALIZED	63	9	25	12	173
DENDRO- GRAM	UNTRANSFORMED (RAW DATA)	38	0	31	31	118
	LOG-10	66	26	12	22	50

CHAPTER 5

DISCRIMINANT AND CLUSTER ANALYSIS OF DERBYSHIRE DATA

5-1 Location and description of the study area.

i) Geography.

The study area (Figs. 5-1, 5-2) is situated principally in ~~the~~ southern Derbyshire but includes adjoining parts of north Staffordshire, Cheshire and Yorkshire and is bounded by the National Grid coordinates 3400 (south), 3900 (north), 3900 (west) and 4400 (east) covering a total surface area of 960 sq. miles. Some of the principal towns are: Sheffield, Chesterfield, Ashover, Crich, Derby, Ashbourne, Cheadle, Ecton, Buxton and Stoke-on-Trent.

The Carboniferous Limestone (Visean) dome in the central part has developed an undulating plateau ranging in elevation between 900 to 1000 feet. This elevated plateau has been poorly dissected superficially, and tributary drainage is well developed and normally deeply incised where shales and sandstones are interbedded. The drainage (Fig. 5-1) developed over the Namurian-Visean Shales separates the Visean limestone from the prominent scarps of the Main Namurian Sandstones in the north, east and part of the west. The scarps of the Main Namurian Sandstones attain elevations of up to 2000 feet decreasing southward. The direction of the drainage in this part and in most of the study area is predominantly to the south and east. The west is mainly occupied by the Staffordshire hills formed mainly by the Westphalian and Namurian-Visean Shales. The hills attain elevations of more than 1000 feet descending westward into the undulating Cheshire plain which is composed mainly of Permo-Triassic sediments. Lower elevations (<800 feet) of an undulating topography mainly occur south of the Visean limestone where Namurian-Visean Shales and Permo-Triassic sediments outcrop. The direction of drainage in this part is also to the south. The drainage at the western side is only cutting overburden derived from the Newer Drift. The main rivers are usually floored by recent and extensive alluvial plains.

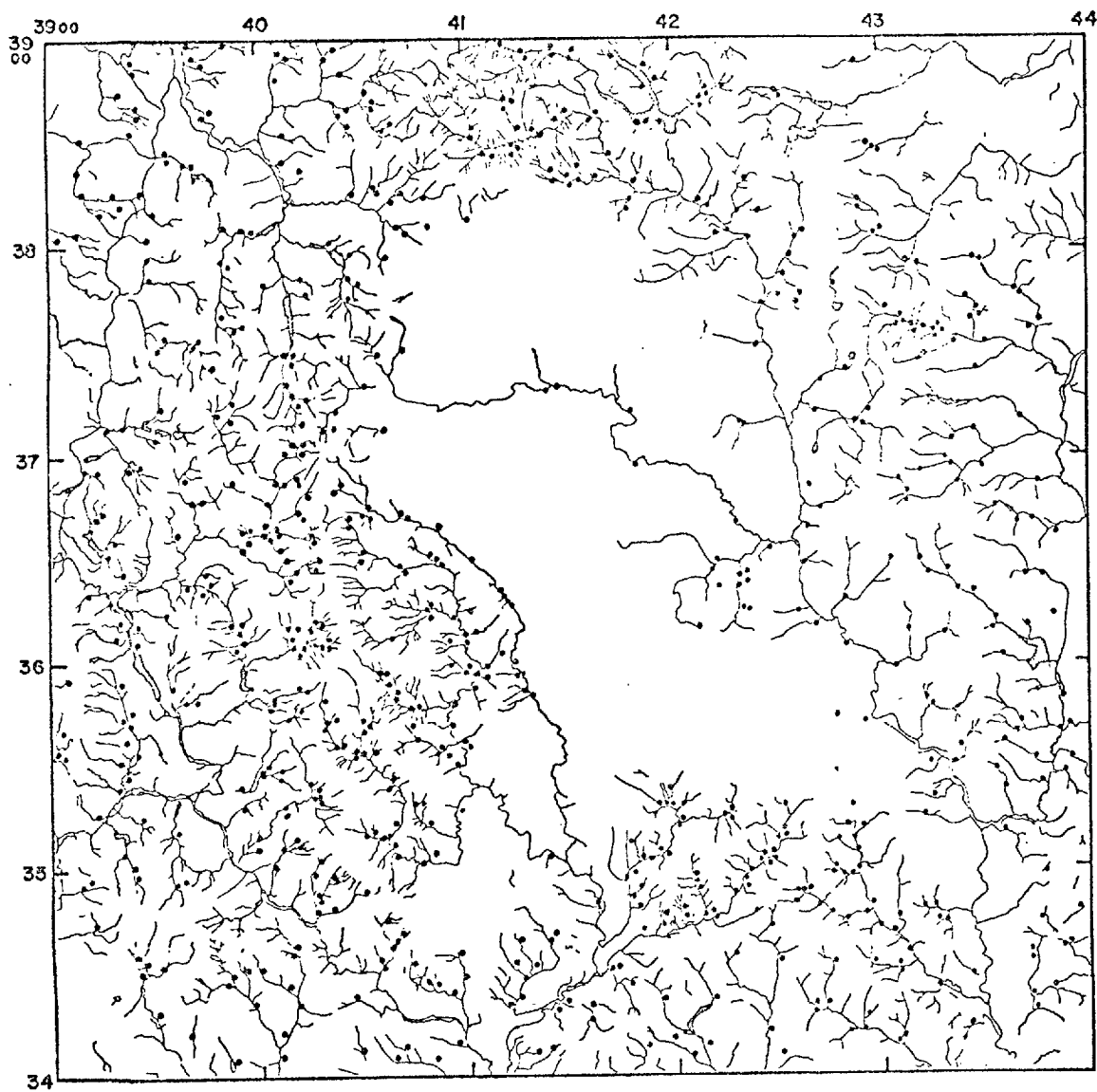


FIG. 5-1
DERBYSHIRE
DRAINAGE SYSTEM AND SAMPLING LOCALITIES
(After Khaleel, 1969)

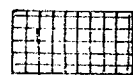
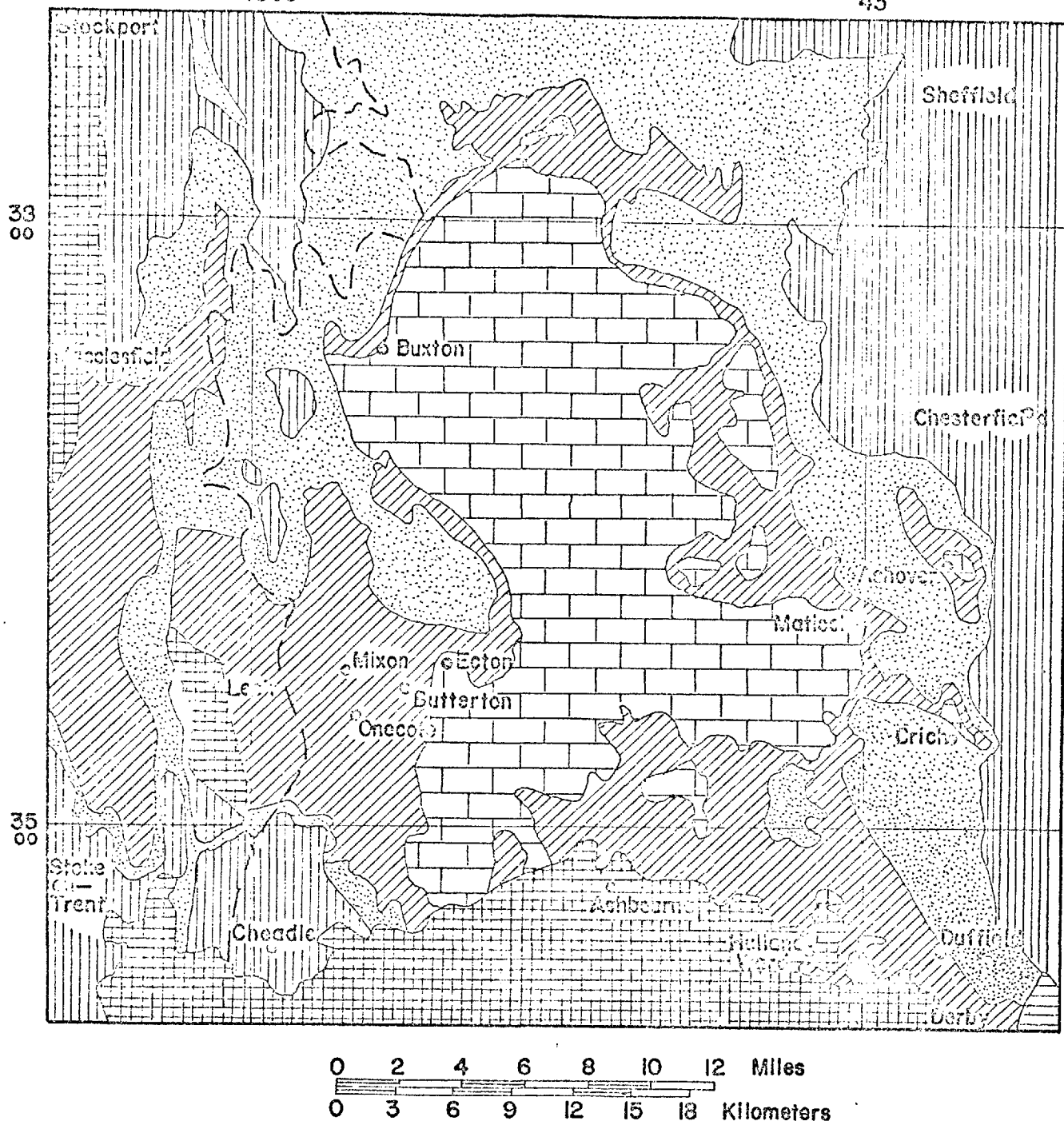
The climate is northwest European and maritime, strongly modified by altitude. The maximum rate of annual rainfall attains more than 50 inches to the north and west of the plateau and falls to below 35 inches over the lower terrains to the south-east. The mean rainfall is 30-40 inches.

A large number of soil groups have been recognised in the study area. The soils of an area to the north and east of Derby have been mapped in detail (Bridges, 1963); the soils of Staffordshire have also been surveyed in a broad sense by Mackney and Burnham (1964) and around Hartington on the western margin of the limestone plateau (Cazalet, 1968). Acid freely-drained brown earths have been developed over the Carboniferous Limestone dome and lime-rich rendzina soils have originated over the steeper, well-drained slopes of the same dome. Gleyed soils are commonly distributed over the Namurian and Westphalian rocks and over the western cover of glacial drift. Peat deposits, peaty gleys and peaty gleyed podzols are exposed over the higher wetter uplands whilst in the steeper slopes of these uplands brown earths have been developed. Well-drained, sandy brown earths are exposed over Carboniferous and Permo-Triassic sandstones and gritstones.

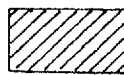
ii) Geology and Regional Geochemistry.

The general description of the geology follows that of Edward and Trotter (1954) and the regional geochemistry is a brief discussion of Nichol and others (1970a). Rocks of Visean, Namurian, Westphalian, Permo-Triassic and Quaternary age have been mapped in the study area. A geologic map is given in figure 5-2. The mean chemical values of stream sediments derived from the principal geological formations are given in table 5-1.

The basal Carboniferous Limestone Series (Visean) exposed at the centre of the area contains thick bedded massive limestone as its most noticeable rock type, which is also folded as a gentle dome trending N-S. Mudstones and sandstones are the other strata of the sequence. Intercalations of cherts and spilitic lavas,



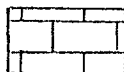
Permo-Triassic



Namurian-Visean Shales



Westphalian (Coal Measures)

Carboniferous Limestone
(Visean)

Main Namurian Sandstones

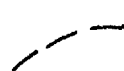
East^{ern} limit of Newer Drift

FIG. 5-2

DISTRIBUTION OF THE PRINCIPAL GEOLOGICAL UNITS, DERBYSHIRE
(After I.G.S., Nichol and others, 1970a)

TABLE 5 - 1

Range and mean* metal content of stream sediments derived from the principal geological formations. (After Nichol and others, 1970a, p.10). Values in p.p.m. unless otherwise stated.

ELEMENT	NAMURIAN- VISEAN SHALES	MAIN NAMURIAN SANDSTONES	WESTPHALIAN (COAL MEASURES)	TRIAS (NEW RED SANDSTONE)
Mo	6.5 2-25	<2	<2	<2
As	11 3.5 - 35	<5	<5	<5
Se	2.8 0.2 - 9.0	-	-	-
Ni	70 40 - 130	35 20-60	50 30-90	25 15-35
V	80 40 - 160	45 30-75	65 40-100	30 20-45
Cu	60 35 - 85	30 20-45	45 25-75	30 20-45
Fe ₂ O ₃ %	2.5 1.4 - 4.5	1.5 0.7-3.4	2.5 1.6-3.9	1.0 0.6 - 1.7
Cr	55 35 - 85	55 30 - 95	60 40 - 90	45 25 - 75
Co	25 15 - 45	15 7 - 30	25 10 - 50	10 5 - 20
Mn	490 170 - 1400	290 85 - 980	800 300 - 2100	225 115 - 450
Ti	1900 1250-3000	1600 940 - 2700	2400 1600-3600	1500 990-2300
Sn	<5	<5	6 2.5 - 15	5 1.5 - 15
No. OF SAMPLES	62	69	65	27

* Geometric mean; range = geometric mean $\pm$ log standard deviation

Pb and Zn omitted in view of the extensive contamination from old mines and smelting activity.

agglomerates, and tuffs ("toadstones") occur. Geochemical characteristics are not reported because of the absence of sampling over the limestone owing to the lack of drainage.

The Millstone Grit Series (Namurian) conformably overlies and surrounds the Visean. It consists of a group of mudstones and shales with beds of sandstones and grit. Marine shales predominate below the Main Namurian Sandstones and in places are underlain by lithologically identical shales of Visean age contemporaneous with the Carboniferous Limestone. For the convenience of this study the Visean Shales have been grouped with the Namurian Shales using the general term Namurian-Visean Shales. The upper rocks of the Namurian have been called for convenience the Main Namurian Sandstones. Stream sediments derived from the Namurian-Visean Shales show a striking enrichment of Mo, As, V, Ni, and Cu, and to a lesser degree, Co and Fe. Rock analysis shows conclusively that the coincident stream sediment patterns of high Mo, As, V, Cu (and Se) reflect the concentration of these metals in the predominant argillaceous facies of the Namurian-Visean Shales.

The Coal Measures (Westphalian) form a sequence consisting of interstratified mudstones, shales, siltstones, sandstones, and few coal seams and conformably overlies the Namurian sediments. The formation shows a generally higher content of Cu, V, Ni, Co, Mn, Cr, Ti, and Fe in the stream sediments similar to the argillaceous Namurian-Visean Shales compared to the predominantly arenaceous Main Namurian Sandstones and Permo-Triassic. Relatively high Mn is the only characteristic feature of sediments in the Westphalian. There is a direct relationship between the Mn content of the three principal Carboniferous formations and their associated stream sediments. However, the relative enrichment of Mn characterizing stream sediments derived from the Westphalian may be due in part to the influence of variations in the secondary environment. On a smaller scale, there is an association of Ti and Fe with minor Ni, Cu, and Co immediately south of Sheffield. A similar composite pattern accompanied by V and minor As occurs north of Cheadle in the southwest. There is also a close correlation of Zn with Cu in the vicinity of the town.

The Permo-Triassic sediments (New Red Sandstone) consist of marls, breccias, sandstones, and pebble beds which are predominantly red in colour. The Permian is restricted to a small area in the north-west corner of the study area, the Trias unconformably overlies the Carboniferous in the south. Stream sediments derived from Triassic rocks on the southern part of the study area show a relative impoverishment of Co, Fe, Mn, and possibly, Ni, V, and Cr where most elements tend to be present in low amounts particularly towards the west. The general tendency for low metal values in stream sediments derived from the Trias is confirmed by rock data.

Deposits of Newer Drift are exposed in the eastern part of the study area and cover the Cheshire plain, wedging out along the 1250 feet contour as the ground rises eastward toward the hills in northeast Staffordshire. The drift related to the Weichsel glaciation is reddish-brown sandy clayed, and is believed to be derived from the Trias. Most stream sediment samples overlying terrains covered by the Newer Drift were collected from tributaries incised in the thick cover of the drift. The stream sediment composition is uniform irrespective of the underlying bedrock and compared with drift-free areas is intermediate in position but is more similar to the trace-element content in stream sediments derived from the more arenaceous formations than to that of the argillaceous type.

Eastward of the Newer Drift, remanent patches of Older Drift are not uncommon on the Namurian and Permo-Triassic in the southern part of the study area. Soil sampling has shown that the composition of the drift is not necessarily related to the underlying bedrock. This drift generally covers the higher ground and consequently has had little or no influence in the geochemistry of the stream sediments which are mainly derived from the residual overburden of the valley slopes.

Variations in the secondary environment (weathering) do not appear to have had any great influence on the major stream sediment patterns characterizing the different geological formations.

The distribution of mineralization and the location of smelter sites (etc.) are shown in figure 5-3. The mineralization is mainly restricted to the eastern side of the study area approximately following the contact between the Carboniferous Limestone (Visean) and the Namurian sediments. The limestone inliers at Crich and Ashover are also included. The richest ore has been developed immediately below the contact with the Namurian-Visean Shales and beneath the tuff ("toadstone") horizons in the limestone. Less prominent mineralization also occurs on the western side of the limestone around Ecton including the Mixon inlier. The mineralization is mainly composed of galena with generally subordinate, although locally important, sphalerite; chalcopyrite only attains ore grade in the vicinity of Ecton and the Mixon inliers. The mineralization occurs as veins and replacements in steeply dipping fracture systems. Barite, fluorite, calcite and occasionally subordinate quartz occur as gangue minerals. Contamination from past mining and smelting activity is generally widespread and locally intense mainly in the eastern part of the study area over Namurian and Westphalian sediments. The stream sediment geochemistry shows a prominent zone of high Pb, Zn, and Ba values flanking the limestone to the east and south-east, generally correlating with the distribution of known mineralization and past mining and smelting activity. This zone is up to 6 to 7 miles wide. The pattern is extremely heterogeneous, with anomalous values ranging from 150 to greater than 10,000 ppm Pb, 300 to greater than 10,000 ppm Zn, and 1500 to greater than 10,000 ppm Ba. The metal ratios also vary markedly in individual streams, a few of which also contain anomalous Cu up to 300 ppm.

5-2 Empirical discriminant analysis.

i) Introduction.

As in Denbighshire it was necessary to use all the available geological information in evaluating the performance of the EDF analysis applied to 705 stream sediment samples.

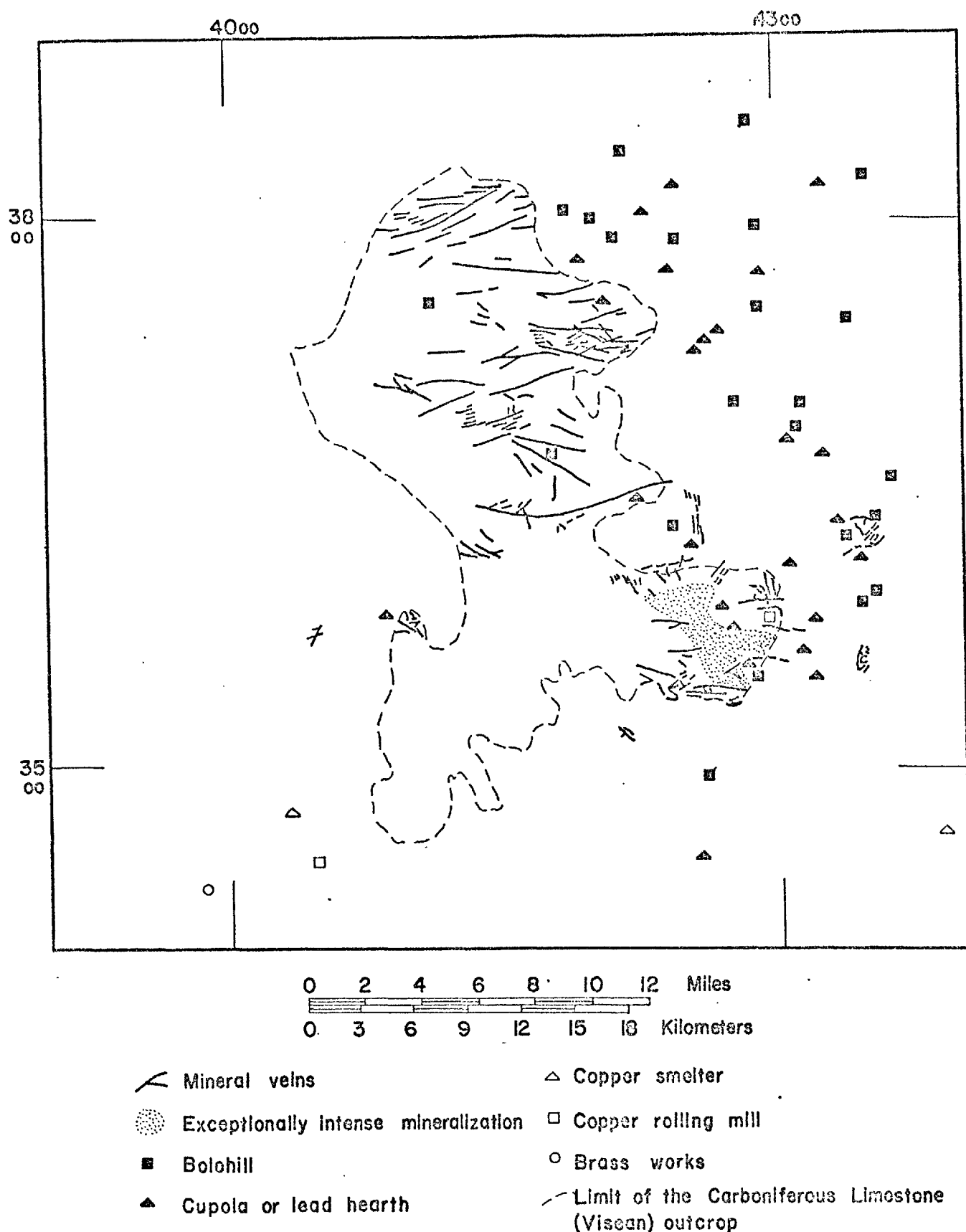


FIG. 5-3

DISTRIBUTION OF PRINCIPAL MINERAL VEINS AND KNOWN SMELTING SITES, DERBYSHIRE
(After I.G.S., Nichol and others, 1970a)

The geochemistry of this area is particularly complex because of contamination by past mining and smelting activity, the presence of glacial drift, and a drainage system which mainly transports sediments which are not representative of a single geological unit.

In order to avoid as far as possible the influence of the glacial materials and to obtain results more representative of the bedrock geology and the mineralization (contamination) the "training areas" were located eastwards of the Newer Drift over areas in which the drainage system mainly transports sediments derived from a single geological formation.

The four training areas (Table 5-2) were located using a 1/4 inch to a mile geological map of the study area and were initially processed without the log-10 transformation of the data. Lithologically and geochemically homogeneous geological formations were selected as representative of a specific bedrock or mineralization type. A class for the mineralization was not included in the untransformed EDF analysis. Only the variables: Pb, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, As and Fe_2O_3 were considered "good" variables (Table 5-3) in the classification procedure. Ga, Sn and Ag were disregarded because of their erratic concentration values over all classes, combined with the virtual absence of Ag.

The loss values and a priori probabilities for the occurrence of the classes were made equal for reasons discussed earlier.

The quantitative evaluation of the bedrock geology EDF results was carried out by selecting stream sediment samples collected over areas, although isolated, strictly representative of a specific lithology (e.g. Namurian-Visean Shales) or an association of related lithologies (e.g. Westphalian). The evaluation of the mineralization EDF (also the cluster analysis and NLM) results was carried out on the basis of those stream sediment samples strongly influenced by the presence of mineralization in the southeastern part of the

TABLE 5 - 2

Training area boundaries for bedrock classes

CLASS	WEST	EAST	SOUTH	NORTH	Size of Training Set
NAMURIAN-UISEAN SHALES	4030	4075	3515	3600	25
MAIN NAMURIAN SAND- STONES	4040	4100	3800	3900	17
WESTPHALIAN	4000	4035	3425	3500	14
PERMO-TRIASSIC	4100	4200	3400	3460	19

TABLE 5 - 3

Average training area compositions, Derbyshire

TRAINING AREA	p.p.m.												SAMPLE SIZE
	Pb	V	Mo	Cu	Zn	Ti	Ni	Co	Mn	Cr	As	Fe ₂ O ₃ %	
NAMURIAN- VISEAN SHALES	63	148	21	134	505	2184	120	31	632	79	33	3.7	25
MAIN NAMURIAN SANDSTONES	156	35	3.7	40	170	1259	41	18	302	51	3.4	1.2	17
WESTPHALIAN (COAL MEAS.)	45	74	2.3	68	222	2789	52	27	1432	68	7.8	3.2	14
PERMO-TRIASSIC (NEW RED SAND- STONE)	124	30	1.4	31	110	1487	20	8	154	42	6.0	0.7	19

Carboniferous Limestone (Visean) dome and over the Ashover, Crich and Ecton inliers; the mineralized areas in the northern half of the Visean dome and adjoining rocks to the east were also subsequently included in this evaluation.

The application of the EDF analysis to the Derbyshire regional geochemical data was carried out in the same manner as explained before for the Denbighshire area.

ii) Results.

a) Classification based on untransformed data.

The optimum smoothing parameter (σ) was chosen on the basis of the behaviour of the testing set classification results as the smoothing parameter was iteratively increased (Fig. 5-4). A rapid rise in the number of samples assigned to one of the training set categories, rather to the "unknown" category, occurs as σ increases from 1.0 to 10.0. An exception to this is the Westphalian category which shows a sudden increase in the number of samples assigned to it at about 9.0, 18.0, 31.0 and 36.0 of smoothing parameter suggesting the presence of different populations closely allied to the Westphalian. The classification result begins to reach a plateau for all bedrock classes from about $\sigma = 40.0$ as a result of excessive overlap of the estimated probability density functions at the larger σ values. According to a previous estimation of the percentage of samples to be considered as "unknown" by the classifier in order to test both bedrock and mineralization classes, an optimum result appears to be achieved with σ in the region of 45.0.

The map based on the results of the first classification is shown in figure 5-5, and the percentage of correct classifications or misclassifications for the testing set samples assigned to each class are given in table 5-4.

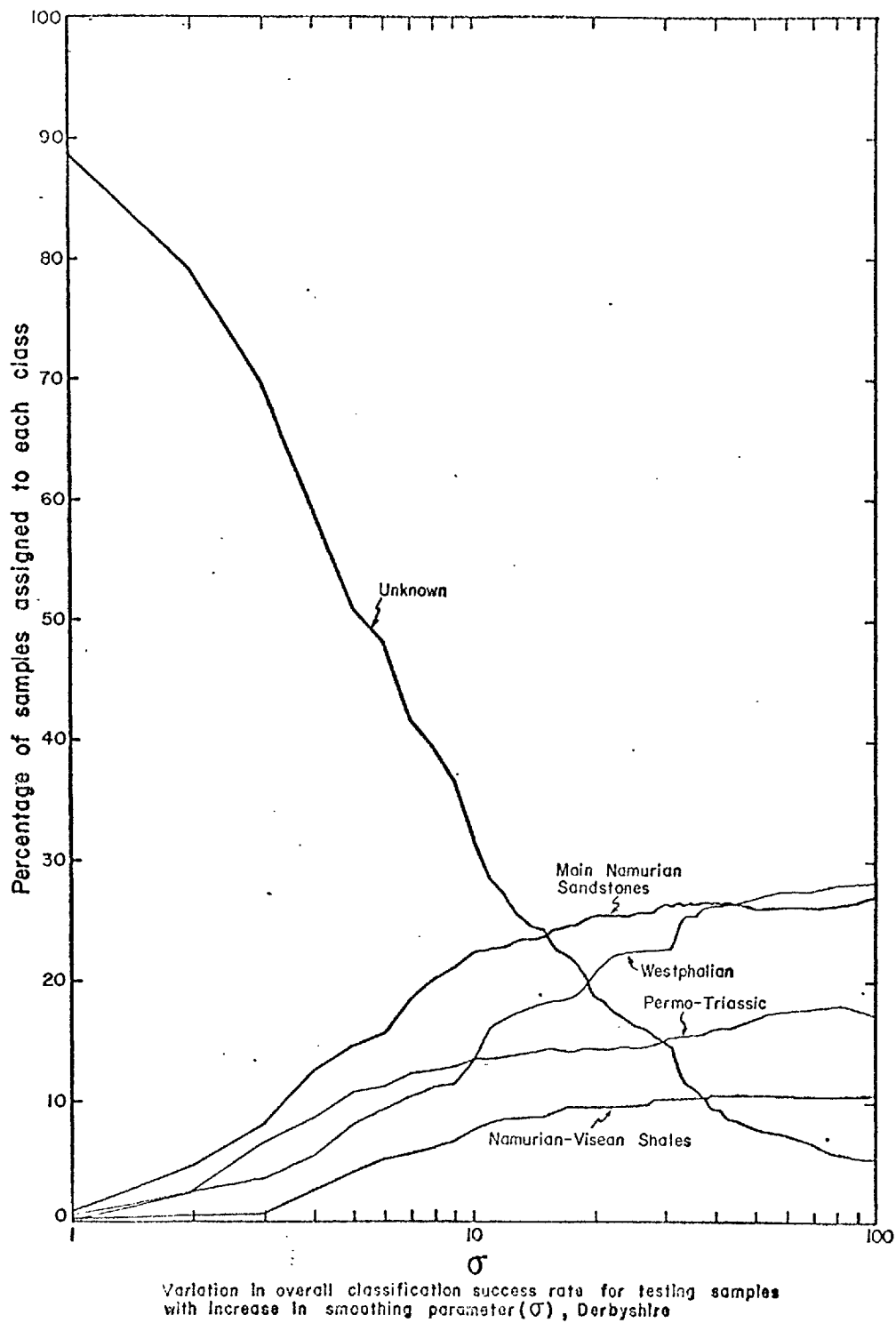
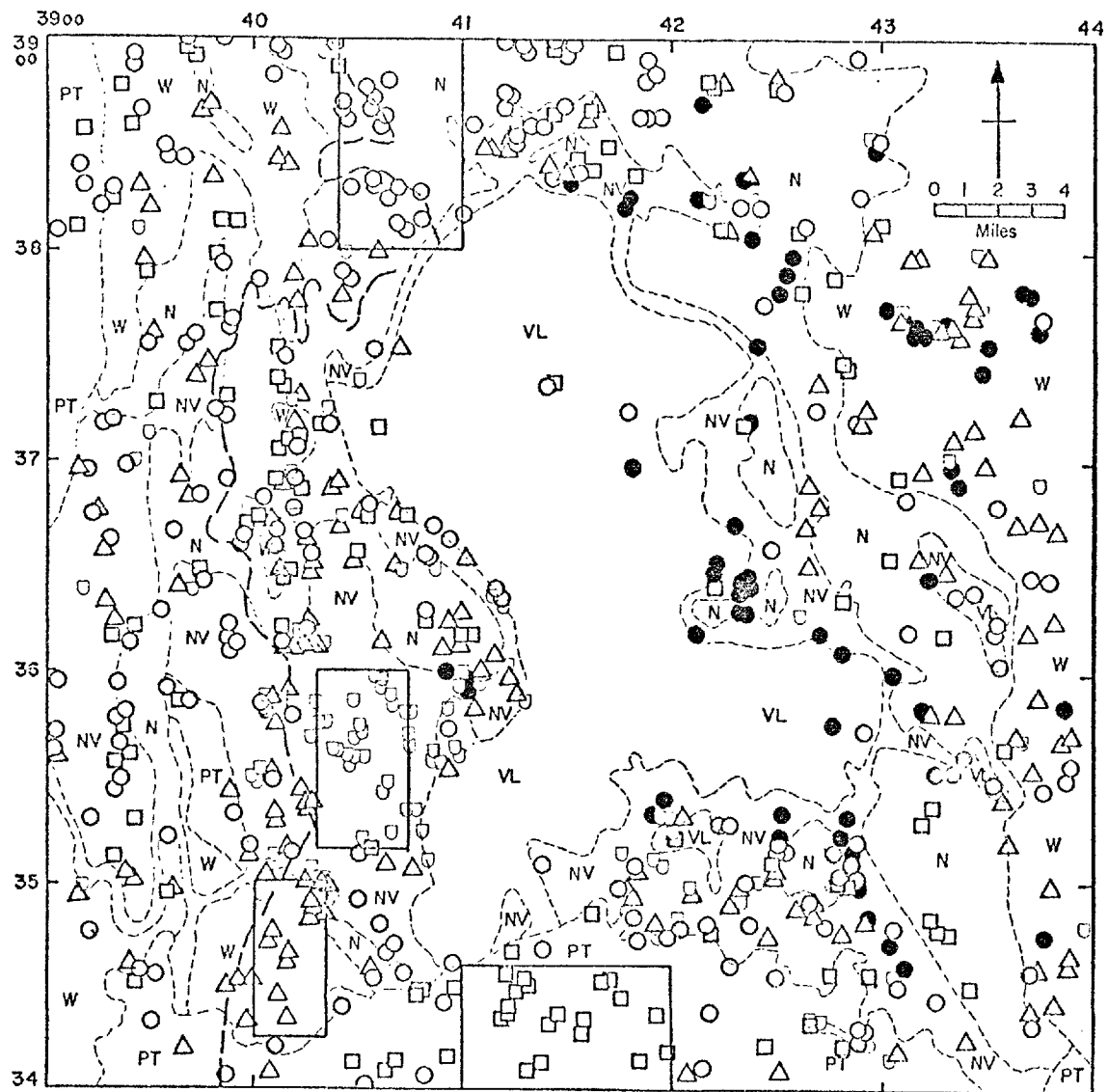


FIG. 5-4



DERBYSHIRE
DISCRIMINANT ANALYSIS
 Variables: Pb, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr,
 As, Fe₂O₃
 Untransformed data

GEOLOGY
 PT Permo-Triassic
 W Westphalian
 N Main Namurian
 Sandstones
 NV Namurian-Visean Shales
 VL Visean (Carb. Limestone)
 --- Newer Drift's eastern
 limit

CLASSES
 □ Permo-Triassic
 △ Westphalian
 ○ Main Namurian
 Sandstones
 □ Namurian-Visean
 Shales
 ● Unknown
 Training area

FIG. 5-5

TABLE 5 - 4

Classification of 173 stream-sediment samples in bedrock category testing set, EDF transformed data.

ASSIGNED CLASS							
TRUE CLASS	NAMURIAN VISEAN SHALES	MAIN NAMURIAN SANDSTONES	WESTPHALIAN	PERMO- TRIASSIC	UNKNOWN	SAMPLE SIZE	% CORRECT
Namurian Visean Shales	18	11	23	3	-	55	33
Main Namurian Sandstones	1	21	10	12	9	53	40
Westphalian	7	4	27	-	15	53	51
Permo-Triassic	2	4	2	4	-	12	33

Minimum overall correct recognition rate = 40%

It can be seen that the classification of the bedrock geology is somewhat confused, probably due to the influence of the glacial drift (Newer Drift) cover in the west, the contamination over the eastern part and the drainage system dissecting different geological formations.

The classification of the Namurian-Visean Shales is poor (Table 5-4). Recognition of the outcrop over the southeastern part of the area is not very successful probably because of the vicinity of the mineralization and the geochemical influence of other geological formations (Carboniferous Limestone (Visean) and Main Namurian Sandstones) outcropping as inliers and outliers, respectively, within the formation. The majority of the samples northwest, west and south of the training area over the same formation were classified either as Westphalian, Main Namurian Sandstones or "unknown" probably due to the influence of the Newer Glacial drift. Most of the stream sediment samples east of the training area, less affected by the drift, were correctly classified.

The Main Namurian Sandstones are also poorly classified (Table 5-4). Only over the northeasternmost outcrop of the formation were a large number of samples classified as Main Namurian Sandstones. However, most of the misclassified samples were assigned to the Permo-Triassic category, probably because of the chemical similarity between both geological formations (Table 5-1). The classification results are erratic over the remaining outcrops.

The eastern Westphalian outcrop attained the highest correct recognition percentage (Table 5-4) for all classes probably because of its continuous exposure with a drainage system favouring the transportation of Westphalian stream sediments. Over the west-central side (N.G.R., 3990-4030/3630-3730) of the study area, the samples derived from two small Westphalian outcrops do not show any relationship with the underlying bedrock probably because of

the drainage system is also dissecting Main Namurian Sandstones and Namurian-Visean Shales.

The EDF classification over the Permo-Triassic is also poor (Table 5-4). The frequent appearance of Main Namurian Sandstones-classified samples over the Permo-Triassic is observed as a response to the close geochemical similarity of the two formations (Table 5-1).

A great majority of samples over the Newer Drift were classified as Main Namurian Sandstones and Permo-Triassic probably as a response to stream sediment compositions intermediate between those of the two formations. When Nichol and others (1970 a, p.11) compared the composition of stream sediments-derived from the Newer Drift with drift-free areas, the trace element content in stream-sediments on the drift was more similar to the trace elements derived from the more arenaceous formations than to that of the argillaceous lithologies.

Since no "mineralization" classes had been included in the training set, around 9 per cent of the stream sediments were correctly recognized as not resembling any of the bedrock-derived categories, and were consequently classified as "unknown". 31 of 62 "unknown" samples (50%) show a close correlation with the mineralization. The rest of the "unknown" samples are related to those areas strongly influenced by contamination.

b) Classification based on log-10 transformed data.

In order to try to improve the results obtained using untransformed data log-10 transformation was applied to the data using the same training areas as before, but with the addition of a new training area for the mineralization category. This new training area was established on the southeast side of the Carboniferous Limestone (Visean) dome where the significance of the lead-zinc mineralization is known (the coordinates are 4190 west, 4300 east, 3510 south, 3550 north; sample size 30).

A smoothing parameter (σ) of 0.04 was iteratively computed using all classes, applying the same criteria as before, and the

corresponding proportion of samples assigned to each class is shown in table 5-5.

TABLE 5 - 5

Percentage of testing samples assigned to each class

σ	N-Visean Shales	M.Namur. Sand- stones.	Westphalian	Permo-Trias.	Mineral- ized	Unknown
0.04	13.1	26.2	20.7	12.8	10.2	3.1

The classification results are shown in figure 5-6 and the success rates for the testing set classifications for each class are given in table 5-6. Compared with the untransformed results (Table 5-4) the classification of the bedrock geology shows an increase in the successful recognition rate among the bedrock types for the testing set for each geological formation, except the Westphalian and Main Namurian Sandstones. A significant proportion of the samples formerly classified as "unknown" were now classified as mineralized (contaminated) coinciding with the main mineralized areas and with the widespread influence of mining and industrial contamination mainly over the eastern side. The "unknown" classified samples, although reduced in number (Table 5-6), do not show any clear relationship with the main mineralized areas, instead they are distributed erratically throughout the study area.

In order to test the geochemical affinity of the Main Namurian Sandstones with the other geological units and to investigate the extent to which the classification patterns are affected by not choosing all the appropriate training areas, the Main Namurian Sandstone training area was eliminated. The correct recognition rates for the testing set classes are shown in table 5-7. The majority of the samples (52%) formerly classified as Main Namurian Sandstones within the Main Namurian Sandstones exposures, were now mainly assigned to the Permo-Triassic category. The chemical

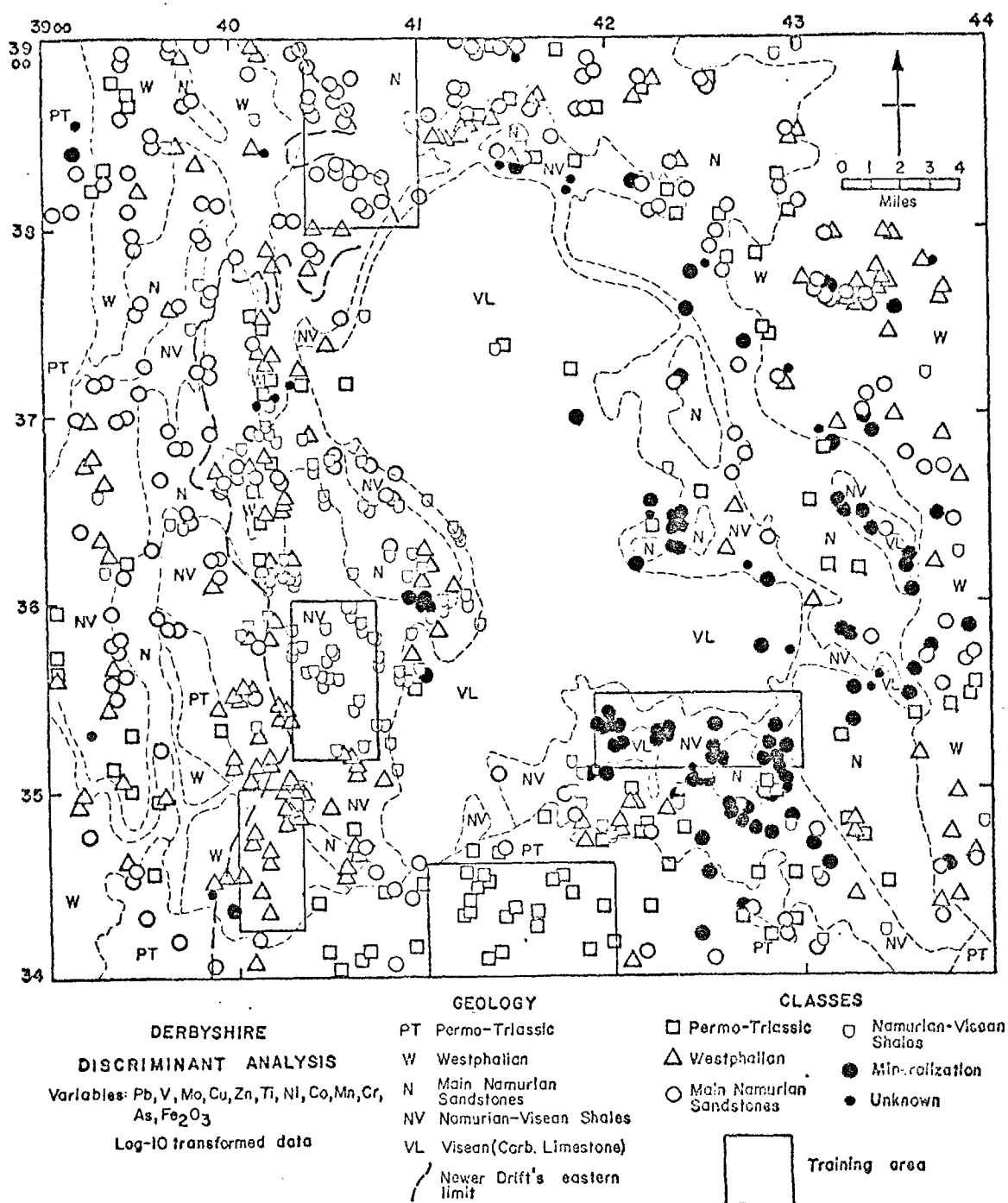


TABLE 5 - 6

Classification of 262 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data.

ASSIGNED CLASS								
TRUE CLASS	NAMURIAN VISEAN SHALES	MAIN NAM- URIAN SAND- STONES	WESTPHALIAN	PERMO- TRIASSIC	MINERAL-- IZATION	UNKNOWN	SAMPLE SIZE	% CORRECT
NAMURIAN VISEAN SHALES	25	3	15	4	7	1	55	45
MAIN NAMURIAN SANDSTONES	3	20	7	13	8	2	53	38
WESTPHALIAN	3	16	24	2	7	1	53	45
PERMO-TRIASSIC	-	3	1	6	1	-	11	55
MINERALIZATION	17	7	7	11	39	9	90	43

Minimum overall correct recognition rate (bedrock) = 44%
 (bedrock and mineralization) = 44%

TABLE 5 - 7

Classification of 276 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data.

ASSIGNED CLASS								
TRUE CLASS	NAMURIAN- VISEAN SHALES	MAIN NAM- URIAN SANDSTONES	WESTPHALIAN	PERMO- TRIASSIC	MINERAL- IZATION	UNKNOWN	SAMPLE SIZE	% CORRECT
NAMURIAN VISEAN SHALES	25	-	17	5	6	2	55	45
MAIN NAMURIAN SANDSTONES	4	-	17	35	9	2	67	-
WESTPHALIAN	3	-	35	6	8	1	53	66
PERMO-TRIASSIC	-	-	2	8	1	-	11	73
MINERALIZATION	17	-	7	16	41	9	90	46

Minimum overall correct recognition rate (bedrock) = 37%
 (bedrock and mineralization) = 39%

similarity between the Main Namurian Sandstones and the Permo-Triassic stream sediments is again evident with this new classification. The rest of the classification patterns (bedrock and mineralization classes) also showed an improved spatial distribution and an increase in the correct recognition rate for each class, with the exception of the Namurian-Visean Shales class.

In a second experiment the training area over the Main Namurian Sandstones was restored but the Namurian-Visean Shales training area eliminated. The classification patterns over the Namurian-Visean Shales now showed significant changes because the majority of the stream sediment samples over the (largest) western and, to a lesser extent, the southeastern outcrops were now assigned to the "mineralization" class, probably because the "mineralization" training samples are stream sediments draining not only mineralized rocks but also extensive non-mineralized Namurian-Visean Shale exposures on the southeast side of the area. The correct recognition rates for the remaining bedrock classes, given in table 5-8, are very similar to those of table 5-6. The distribution of the classification patterns over the bedrock and mineralization units were similar to those in figure 5-6, with the exception of the Namurian-Visean Shales as discussed above.

It was subsequently decided to keep the same Namurian-Visean Shales training area and to relocate the Main Namurian Sandstones training area because it is apparently influenced by the Newer Drift detritus. The coordinates for this new training area are: 4100 west, 4230 east, 3875 south, 3900 north with a sample size of 13. Since some stream sediment samples within the Permo-Triassic training area might also be influenced by upstream mainly Namurian-Visean Shale detritus and, to a lesser extent, by Carboniferous Limestone (Visean) and Main Namurian Sandstone materials it was decided to reduce, as much as possible, the training samples to

TABLE 5 - 8

Classification of 295 stream-sediment samples based on bedrock and mineralization training sets, EDF log-10 transformed data.

ASSIGNED CLASS								
TRUE CLASS	NAMURIAN- VISEAN SHALES	MAIN NAM- URIAN SANDSTONES	WESTPHALIAN	PERMO- TRIASSIC	MINERAL- IZATION	UNKNOWN	SAMPLE SIZE	% CORRECT
NAMURIAN VISEAN SHALES	-	2	29	6	48	3	88	-
MAIN NAMURIAN SANDSTONES	-	21	11	13	7	1	53	40
WESTPHALIAN	-	17	25	2	8	1	53	47
PERMO-TRIASSIC	-	3	1	6	1	-	11	55
MINERALIZATION	-	7	8	14	52	9	90	58

Minimum overall correct recognition rate (bedrock = 25%
(bedrock and mineralization) = 35%

a set solely draining Permo-Triassic rocks. The coordinates for this new training area are: 4100 west, 4145 east, 3400 south, 3460 north; sample size 11.

The resulting testing set classifications are shown in table 5-9. It was immediately obvious that by avoiding or minimizing the influence of possible sources of contamination of the training samples the distribution of the classification patterns were very similar to those of figure 5-6, however, the correct recognition rates for each class were improved (Table 5-9) and the minimum overall correct recognition rates for the bedrock categories alone, and the combined bedrock and mineralization categories attained their highest values compared with previous results.

iii) Discussion

The results of the stream sediment classification for the four bedrock classes, based on the untransformed data (Table 5-4), indicate a minimum overall correct recognition rate of 40 per cent. This low value is a consequence of the non-transformation of the data and of the widespread influence of mineralization, contamination and glacial drift which have produced a high variability in the data values. With the log-10 transformation the variability of the data values is reduced and an increase in the minimum overall correct recognition rate for the combined bedrock and mineralization categories to 44 per cent (Table 5-6) was achieved.

In the case of the untransformed data (Table 5-4) in which no "mineralization" class was included, the minimum overall successful classification rate is increased to 47 per cent if the Namurian-Visean Shales and the Main Namurian Sandstones are taken together as the integrated Millstone Grit Series, while with the log-10 transformed data (Table 5-6) the minimum overall correct classification rate attains the same 47 per cent; when both the bedrock and mineralization categories are considered the minimum overall

TABLE 5 - 9

Classification of 270 stream-sediment samples based on bedrock and mineralization training sets, EDF. log-10 transformed data.

ASSIGNED CLASS								
TRUE CLASS	NAMURIAN- VISEAN SHALES	MAIN NAM- URIAN SANDSTONES	WESTPHALIAN	PERMO- TRIASSIC	MINERAL- IZATION	UNKNOWN	SAMPLE SIZE	% CORRECT
NAMURIAN VISEAN SHALES	29	4	13	-	8	1	55	53
MAIN NAMURIAN SANDSTONES	2	28	7	6	9	1	53	53
WESTPHALIAN	3	1	37	1	10	1	53	70
PERMO-TRIASSIC	-	4	1	11	3	-	19	58
MINERALIZATION	16	3	8	8	46	9	90	51

Minimum overall correct recognition rate (bedrock) = 58%
(bedrock & mineralization) = 56%

overall correct classification rate attains only 46 per cent. However, if contamination influencing the training set is minimized the minimum overall correct classification rate strictly for the bedrock geology (Table 5-9) increases to 61 per cent, and with both bedrock and mineralization categories it is 58 per cent.

In the case of the untransformed data (Fig. 5-5), 17 per cent of the sediments draining known mineralization southeast of the Carboniferous Limestone (Visean) dome and the inliers of Ashover, Crich and Ecton, were classified as "unknown". When the mineralized veins of the northern half of the Visean dome and adjoining rocks along the east margin were considered then the overall classification rate rose to 28 per cent of "unknown"-classified samples as indicators of mineralization in very contaminated areas.

In the case of the log-10 transformed data an increase to 43 per cent correct classifications (Table 5-6) was achieved for the sediments in streams draining the whole mineralization when a training area for the mineralization was considered.

It is highly probable that a large proportion of the misclassification are, in fact, correct because of the geochemical "spectrum" of contamination of the stream sediments from the glacial drift and the mining and smelting activities.

iv) Conclusions

The widespread influence of contamination by past mining and smelting activities, the presence of glacial drift, and a drainage system which mainly transports sediments which are not representative of a single geological formation has caused a very complex geochemistry with very erratic individual element patterns and consequently very poor classification results regarding both bedrock geology and mineralization, however:

1) the bedrock geology was more correctly classified with the log-10 transformed data than with the untransformed data when all the appropriate training areas were considered and the influence of contamination minimized,

2) the mineralization was generally detected by the "unknown"-classified samples in the untransformed data results; it was better detected with the log-10 transformed data when a training ^{area} for mineralization was considered.

5-3 Hierarchical cluster analysis

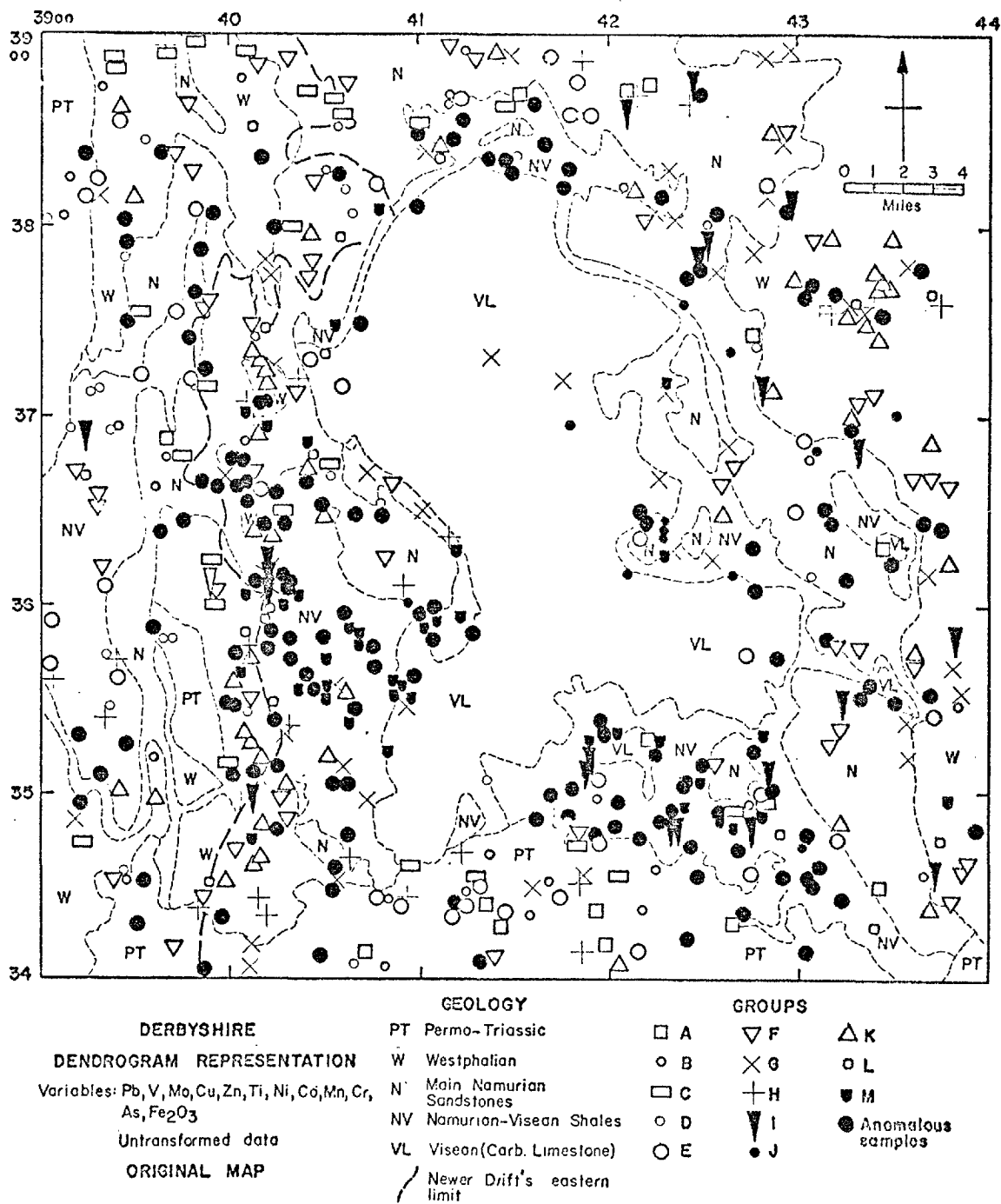
i) Q-mode results.

The results from both untransformed and log-10 transformed data are based on 640 samples and on the 12 variables: Pb, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, As and Fe_2O_3 . The distance coefficient in the "basic" dendrograms for the bedrock units ("basic" groups), with the exception of the "anomalous" samples, generally ranged between 0.82 and 0.96. Coefficients ranging between 0.67 and 0.98 were found in the "subsequent" dendrograms.

Smoothed maps are not included for this particular area because the extreme variability of the original classification maps is such that it did not assist in their interpretation.

The original untransformed results can be seen in figures 5-7 and 5-8. The statistical properties for each group are included in table 5-10.

The classification patterns in untransformed results are very confused, probably due to the complicated distribution of the main geological units. The drainage system commonly does not dissect a single geological formation and consequently transports composite or diluted stream sediments. There is also widespread natural (glacial drift) and artificial (mining and industrial) contamination. The common lower rates of correct recognition shown by the EDF analysis results (Table 5-11) and the irregularity in the misclassification for each group (with the exception of



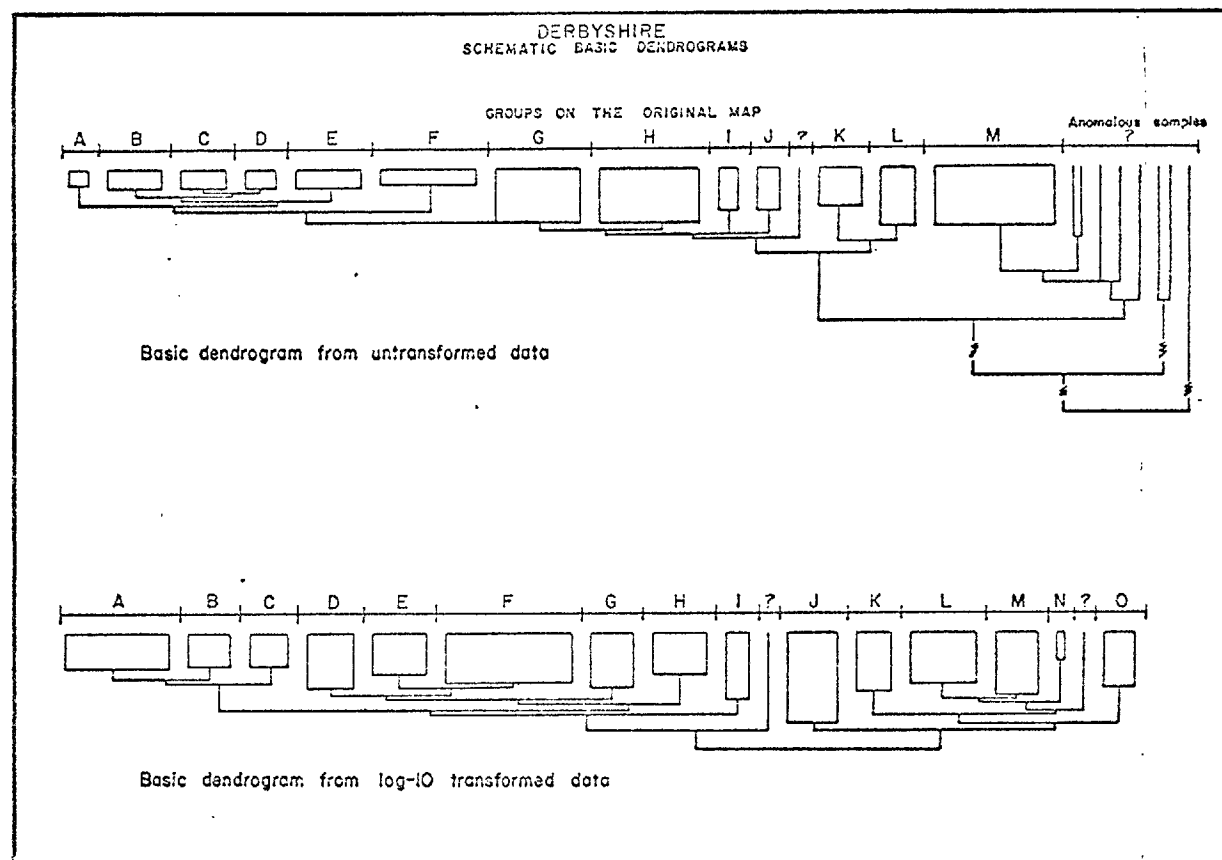


FIG. 5-8

TABLE 5 - 10

Arithmetic means ($\bar{x}$) and standard deviations (s), Derbyshire, dendrogram representation, untransformed data; total sample size = 593

GROUPS	A	B	C	D	E	F	G	H	I	J	K	L	M	ANOMALOUS SAMPLES
SAMPLE SIZE	14	40	27	26	36	60	50	40	21	13	52	16	63	135
Pb $\bar{x}$	149	104	51	46	75	99	139	102	1445	3885	100	85	542	640
s	244	179	55	73	96	144	236	174	2256	2332	220	87	1427	1760
V	32	38	45	33	28	55	61	72	58	59	100	114	143	67
	9	22	8	20	8	16	27	39	26	24	33	53	47	45
Mo	1	1.9	1.7	1.5	1.1	1.7	4.0	4.0	3.5	8	1.9	8	19	6
	0	2.4	1.2	2.3	0.7	1.2	3.6	6.5	2.8	10	2.5	9	10	8
Cu	28	31	39	36	22	42	48	54	42	43	48	63	82	61
	4	9	8	13	7	9	23	32	16	17	14	24	42	31
Zn	131	117	171	126	129	148	225	191	380	1645	183	300	507	426
	70	83	81	57	131	55	229	123	361	995	115	102	487	627
Ti	2000	1226	1307	965	867	1920	2298	1829	1633	1762	3531	3375	2405	1676
	0	458	273	243	191	173	1204	751	698	783	858	857	767	511
Ni	26	34	39	33	22	41	56	44	70	48	52	105	100	68
	10	30	8	15	8	8	26	21	38	20	27	29	40	72
Co	13	13	17	14	9	19	25	20	41	26	24	35	31	36
	9	9	5	7	5	5	12	11	26	15	9	11	12	93
Mn	230	288	306	272	155	481	697	642	2768	919	838	1988	990	1105
	135	278	154	272	99	365	641	645	2024	834	634	1427	986	2377
Cr	45	75	39	44	39	48	69	56	56	50	81	84	82	135
	7	60	9	30	48	14	75	26	30	25	25	60	33	777
As	3.6	3.1	4.6	4.4	3.1	4	6.7	13.0	11	17	9	7	34	14
	1.8	0.8	2.8	4.2	0.5	2.9	5.4	13.0	12	19	11	8	30	18
Fe ₂ O ₃	1.2	1.2	1.4	1.0	0.6	1.9	2.1	2.8	2.7	2.0	2.9	4.2	3.7	2.2
	0.4	0.7	0.3	0.3	0.3	0.6	1.0	1.9	1.0	1.0	1.6	1.1	1.5	2.0

Mean values in p.p.m., except Fe₂O₃ (%)

TABLE 5 - 11

Group classification results by EDF analysis, Derbyshire, dendrogram representation, untransformed data results, $\sigma = 45.0$.

ASSIGNED GROUP																	
	A	B	C	D	E	F	G	H	I	J	K	L	M	ANOM.	?	% CORRECT	
T R U E G R O U P	A	10	0	0	0	3	0	0	0	0	0	0	0	1	0	71	
	B	0	4	6	7	2	1	3	0	0	0	0	2	3	0	10	
	C	0	5	13	3	1	1	1	0	0	0	0	0	3	0	48	
	D	0	1	2	5	15	0	1	1	0	0	0	1	0	0	19	
	E	0	2	2	16	9	0	2	1	0	0	0	0	4	0	25	
	F	17	0	4	0	0	29	2	0	1	0	0	0	7	0	48	
	G	2	2	4	2	5	7	3	1	0	10	1	2	9	0	14	
	H	4	3	5	1	4	5	4	0	0	3	0	4	5	0	10	
	I	0	0	0	0	1	1	0	9	1	0	1	5	3	0	43	
	J	0	0	0	0	0	0	0	3	4	0	1	1	2	2	31	
	K	0	0	0	0	1	5	1	0	0	36	2	5	2	0	69	
	L	0	0	0	0	0	1	1	1	1	4	5	1	2	0	31	
	M	0	3	1	0	0	2	3	9	3	1	6	3	22	8	35	
	Anom.	3	4	9	8	8	21	13	7	4	3	6	5	13	22	9	16

Overall correct recognition rate
(Excluding anomalous group) = 34 %

Overall correct recognition rate
(For all groups) = 30%

group A) would probably explain the disturbing effect of the drainage system and contamination on the stream sediments; however, groups A and F, and D and E constitute closely associated groups that if taken together increase the correct recognition rate in each case to 80% and 73%, respectively. None of these group associations showed significant correlation with a specific geological feature. Regarding the bedrock geology, only one group (M) occurs predominantly over one unit (Namurian-Visean Shales on and around Mixon); this particular response is because the exposures, which are not strongly affected by contamination, have a drainage system mainly transporting Namurian-Visean Shales which are very high in Mo. The high Pb, Zn and Cu mean values and standard deviations shown by group M (Table 5-10), are comparable with similar values in the group of "anomalous" samples, and suggest the influence of near-by geological sources high in these elements, perhaps the galena sphalerite and chalcopyrite mineralization in the vicinity of Mixon and Ecton. Practically all groups show a diversity of mean values making the task of correlation difficult or impossible. This is confirmed as well by the EDF analysis results.

Samples considered as "anomalous" are irregularly distributed throughout the study area and detect the mineralization (contamination) well on the eastern side of the area, but such a response is confused by a widespread and dense pattern of "anomalous" samples overlying rocks on the western side on and around the eastern limit of the Newer Drift. The group of "anomalous" samples is relatively lower in mean Pb and Zn values (Table 5-10) and appears to be less representative of mineralization and contamination. This is probably due to the lowering effect on the overall mean values produced by those samples classified as "anomalous" on the western side which are derived from a region relatively barren in mineralization and contamination (smelting sites), with the exception of those areas near Ecton and Mixon. Groups I and J, distributed mainly on the mineralized and contaminated areas, appear to be representative of geochemically anomalous areas (e.g. mineralization and contamination) because of their very high mean values of Pb and Zn, the majority of which are higher

than those values shown by the group of "anomalous" samples.

The original log-10 transformed results can be seen in figure 5-9 and 5-8. The statistical properties for each group are included in table 5-12.

The log-10 transformed results display less variability than those derived from the untransformed data. Over the Namurian-Visean Shales outcrop on an around Mixon, group L predominates and it is also evident over the Namurian-Visean Shales in the southeast of the area. Group L attains the highest Mo content (Table 5-12) as with group M of the untransformed data results, and also higher mean values of Cu and Zn suggesting the additional influence of the mineralization in the vicinity of Mixon and Ecton. Stream sediments over the Permo-Triassic rocks on the southern part of the area and Main Namurian Sandstones on the northern portion were classified within a relatively well defined association (mainly composed of groups H, A and C) which has similar lower mean metal values (Table 5-12). Stream sediments derived mainly from Westphalian rocks were also classified in another relatively well defined association of groups (mainly E and F) also generally similar in intermediate mean metal values (Table 5-12). The first association is probably due to similarity in chemical composition between the Permo-Triassic, and Main Namurian Sandstones (Table 5-1). Both associations are confirmed by the EDF analysis results (Table 5-13). The correct recognition rate for the H, A and C association taken together increased to 76% and the rate for the E and F association, together with group A which is also closely allied to both groups, rose to 87%. Again, generally lower correct recognition rates and a higher degree of irregularity in the misclassification results are observed, probably as a response to the disturbing effect produced in the stream sediments by the drainage system and contamination. Those samples assigned to group E south of Sheffield coincide with the association of Ti and Fe with minor Ni, Cu and Co over the Westphalian reported by Nichol and others (1970a, p.8).

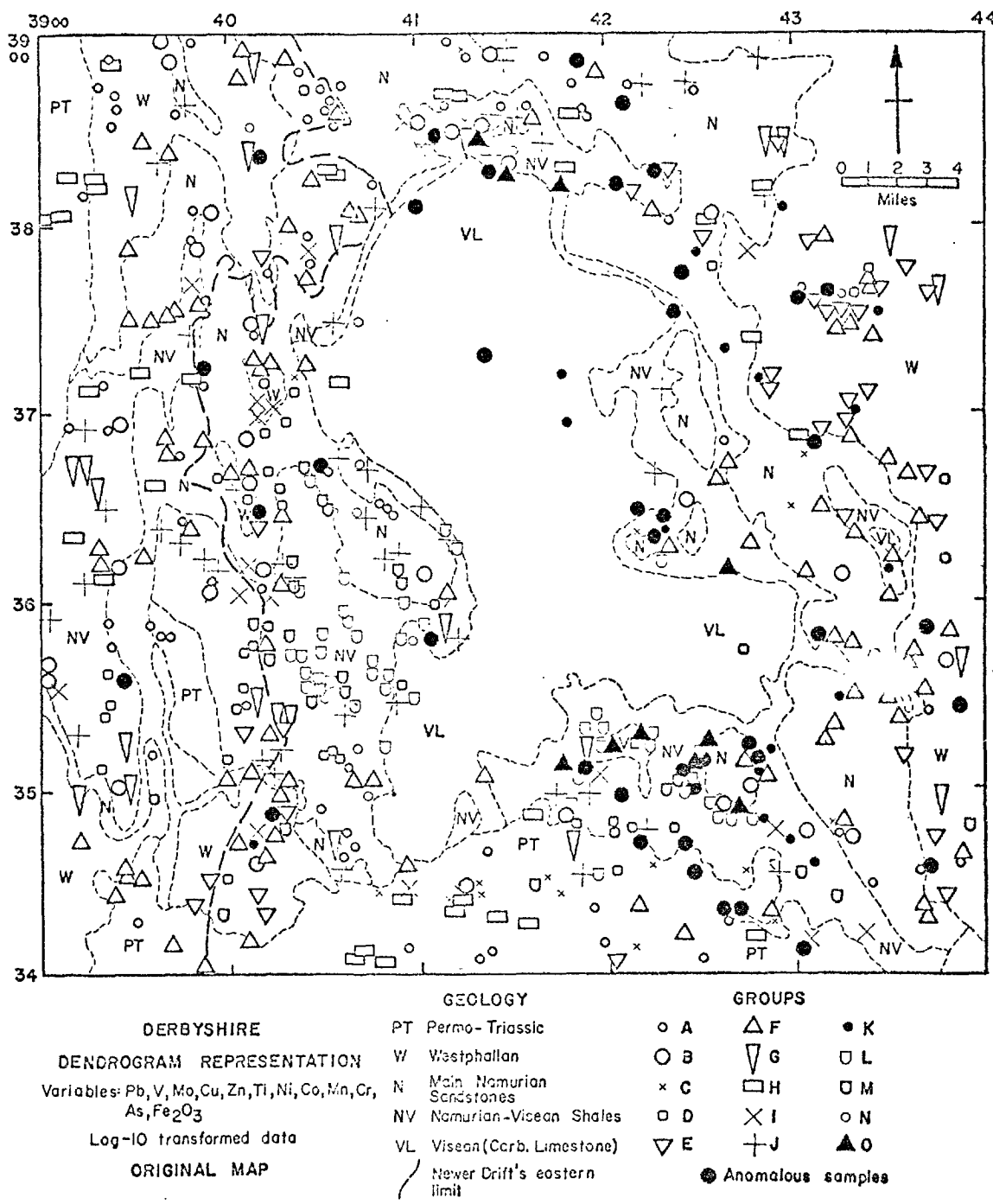


FIG. 5-9

TABLE 5 - 12

Arithmetic means ($\bar{x}$) and standard deviations (s), Derbyshire, dendrogram representation, log-10 transformed data; total sample size = 606

GROUPS	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	ANOMALOUS SAMPLES
SAMPLE SIZE	79	29	16	39	36	107	25	32	8	51	22	50	32	20	15	45
Pb $\bar{x}(\log)$	1.6	1.6	1.9	1.6	1.8	1.9	1.8	1.5	1.4	1.6	3.2	2.2	1.7	1.7	3.0	2.5
Pb s(log)	0.5	0.4	0.3	0.4	0.4	0.6	0.3	0.3	0.3	0.3	0.5	0.6	0.4	0.3	0.9	0.9
V $\bar{x}$	1.6	1.7	1.4	1.7	1.8	1.7	1.9	1.4	1.6	1.8	1.7	2.2	1.9	2.1	2.0	1.8
V s	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.1	0.4	0.2	0.1	0.1	0.2	0.2	0.2	0.3
Mo $\bar{x}$	0	0.2	0.02	0.2	0.04	0.02	0.4	0	0	0.8	0.5	1.3	0.9	1.2	1.2	0.6
Mo s	0	0.1	0.07	0.2	0.1	0.1	0.2	0	0	0.2	0.2	0.2	0.3	0.2	0.2	0.5
Cu $\bar{x}$	1.5	1.6	1.4	1.6	1.7	1.6	1.8	1.4	1.3	1.7	1.6	2.0	1.7	1.8	1.8	1.6
Cu s	0.1	0.1	0.1	0.2	0.2	0.1	0.2	0.2	0.2	0.1	0.2	0.3	0.2	0.1	0.1	0.2
Zn $\bar{x}$	2.0	2.1	2.1	2.2	2.2	2.2	2.3	1.8	1.7	2.2	2.6	2.7	2.5	2.3	3.0	2.5
Zn s	0.4	0.2	0.3	0.2	0.6	0.2	0.2	0.3	0.2	0.2	0.4	0.4	0.3	0.2	0.5	0.5
Ti $\bar{x}$	3.2	3.2	3.0	3.3	3.4	3.2	3.4	3.0	3.0	3.2	3.2	3.4	3.3	3.4	3.3	3.2
Ti s	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.1	0.1	0.2	0.2
Ni $\bar{x}$	1.5	1.5	1.3	1.6	1.7	1.6	1.8	1.3	1.3	1.7	1.8	2.0	1.8	2.0	1.8	1.7
Ni s	0.1	0.1	0.1	0.1	0.2	0.1	0.2	0.2	0.3	0.2	0.2	0.2	0.3	0.2	0.1	0.3
Co $\bar{x}$	1.2	1.2	0.7	1.3	1.4	1.3	1.5	0.8	0.9	1.4	1.4	1.5	1.4	1.5	1.3	1.3
Co s	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.3
Mn $\bar{x}$	2.3	2.4	2.1	2.8	3.1	2.6	2.9	1.9	1.8	2.8	2.9	2.8	2.9	2.8	2.7	2.7
Mn s	0.2	0.2	0.3	0.4	0.3	0.3	0.3	0.3	0.4	0.5	0.3	0.4	0.4	0.4	0.2	0.6
Cr $\bar{x}$	1.7	1.8	1.6	1.8	1.8	1.7	1.8	1.4	1.6	1.7	1.6	1.9	1.8	1.9	1.7	1.8
Cr s	0.2	0.3	0.3	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.5
As $\bar{x}$	0.5	0.5	0.5	1.3	0.6	0.6	0.6	0.5	1.6	0.7	0.8	1.6	1.4	0.7	1.2	0.8
As s	0.2	0.09	0	0.2	0.1	0.2	0.2	0	0.2	0.3	0.4	0.2	0.2	0.3	0.4	0.4
Fe ₂ O ₃ $\bar{x}$	0.03	0.01	-0.3	0.2	0.4	0.2	0.3	-0.1	0.05	0.2	0.3	0.6	0.3	0.4	0.3	0.3
Fe ₂ O ₃ s	0.2	0.4	0.4	0.5	0.3	0.3	0.5	0.3	0.2	0.4	0.2	0.3	0.5	0.5	0.2	0.5

Mean values in p.p.m. (log); except Fe₂O₃ (%)

TABLE 5 - 13

Group classification results by EDF analysis, Derbyshire, dendrogram representation, log-10 transformed data results, $\alpha = 0.04$.

		ASSIGNED GROUP																	
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	ANOM.	?	% CORRECT
T R I B U T E G R O U P	A	44	3	0	0	0	23	1	7	0	0	0	0	0	0	0	0	1	56
	B	8	10	0	0	0	6	2	0	0	3	0	0	0	0	0	0	0	34
	C	2	0	8	0	0	2	0	4	0	0	0	0	0	0	0	0	0	50
	D	0	2	0	26	2	4	1	0	0	1	0	0	1	0	0	2	0	67
	E	0	1	0	1	17	11	1	0	0	2	1	0	0	0	0	1	1	47
	F	13	2	0	1	12	73	2	0	0	0	1	0	0	0	0	3	0	68
	G	0	1	0	0	5	3	12	0	0	2	1	0	0	1	0	0	0	48
	H	5	0	4	0	0	0	0	22	0	0	0	0	0	0	0	1	0	69
	I	0	0	0	3	0	0	0	0	4	0	0	0	0	0	0	1	0	50
	J	0	2	0	0	4	0	4	0	0	29	0	0	3	5	0	3	1	57
	K	0	0	0	1	1	3	0	0	0	0	8	0	1	0	1	7	0	36
	L	0	0	0	1	0	0	0	1	0	0	1	43	3	0	1	0	0	86
	M	0	0	0	2	0	0	0	0	0	2	0	7	17	2	1	1	0	53
	N	0	0	0	0	0	0	1	0	0	3	0	1	2	12	1	0	0	60
	O	0	0	0	0	0	0	0	0	0	2	1	2	1	0	7	2	0	47
	Anom.	2	3	0	1	2	7	0	1	1	4	5	1	2	3	1	9	3	20
Overall correct recognition rate (Excluding anomalous group)		= 59%																	
Overall correct recognition rate (For all groups)		= 56%																	

Samples considered as "anomalous" occur in smaller amounts than those obtained with the untransformed data and do not show as clear a pattern in the east. Some of the mineralized areas such as Ashover and Crich are not detected by any "anomalous" sample. On the western side only a few scattered samples were classified as "anomalous", thus indirectly enhancing the "anomalous" pattern on the eastern side, an effect opposite to that obtained with the untransformed data. The group of "anomalous" samples is relatively high in Pb and Zn (Table 5-12) and is to a certain extent less representative of mineralization (contamination) than groups K and O which have still higher values for Pb and Zn.

No attempt was made to classify mineralization types because of the presence of a basically single type of mineralization (Pb-Zn).

ii) R-mode results.

The R-mode results have been interpreted in a similar manner as before.

With the untransformed data the most stable associations were: Pb-Zn, Mo-As and, less consistently, the V-Ti-Ni-Cr-Fe association. Mn was irregularly linked to this last association. No other important associations occurred.

These associations did not consistently occur with the log-10 transformed data and instead a high variability was observed. The former Mo-As association of the untransformed data was this time absent.

Comparing the results with the R-mode factor analysis results obtained by Khaleelee (1969, p.318): 1) there was an agreement with the Pb-Zn association representative of mineralization and contamination defined with the untransformed data; in the log-10 transformed data the same association was present in less important factors combined with other elements, coinciding with the inconsistent log-10 transformed cluster analysis results; 2) the closer

agreement between the Mo-As association and the As-Cu-Mo-V association found by Khaleelee in the untransformed data results and related with the black marine shales (Namurian-Visean Shales); as in the case of the log-10 transformed cluster analysis the association was not very well detected and did not appear until the lower factors, and 3) there was no clear correspondence between the less consistent association of the untransformed cluster analysis results and the rest of the associations reported by Khaleelee. The Cr-Fe-Ga-Ti-V association was only detected with the untransformed factor analysis results and correlated with similar association in rocks of the Lower Palaeozoic in Denbighshire. It was decided that, as in factor analysis, the untransformed cluster analysis results provided on balance the best agreement with part of the element associations reported by earlier workers.

iii) Discussion

In the application of hierarchical cluster analysis to the classification of stream sediments in which the majority are not derived from a single geological formation and are strongly influenced by mineralization and contamination, it has been demonstrated that it gives confused and erratic results in relation to the bedrock units. Neither were the most conspicuous mineralized areas detected by the "anomalous" samples in an area obviously influenced by widespread contamination due mainly to past mining and smelting activity although they were located by some sample groups.

It is not therefore possible to expect good results for the bedrock to be obtained with the method in regional stream sediment surveys in areas similar to Derbyshire. It is highly probable that undeveloped or virgin areas with this general geological configuration presently exist throughout the world (with the exception of the artificial contamination) and the results would always tend to be confused on the basis of the composite nature of the stream sediments. It is believed that for areas of this type the practicability of the method would be greater if soil and rock geochemical surveys are carried out because each sample becomes a more reliable in situ representative of the main geological features in the absence of natural contamination

at the sampling sites.

In the untransformed data results 135 of the 640 samples (21%) were considered as "anomalous". 50 of these 135 (37%) were found to be connected with mineralization and contamination and about 40 (30%) were more related with mineralization; 65 (48%) were mainly distributed on and around the Newer Drift; 20 "anomalous" samples (15%) were not apparently related to any major source of mineralization or associated contamination. These quantitative results illustrate the misleading effect that the "anomalous" samples on the western side have had in the detection of mineralized and contaminated areas, and the apparently lower masking effect of areas mainly exclusively contaminated by smelting activity over those areas strongly influenced by mineralized veins as well as mining and smelting activities.

In the log-10 transformed data results 41 samples (7%) were considered as "anomalous". 23 of these 41 (56%) were found to be related to mineralization and contamination, and 14 (34%) were more related to mineralization; 8 (20%) were mainly distributed over the western side on and around the Newer Drift; 10 (24%) were not apparently related to any known major source of mineralization or associated contamination. In this case the log-10 transformed results gave an obviously better classification of mineralization and contamination because the misleading influence of the "anomalous" samples over the western side is minimised.

5-4 Non-linear mapping algorithm (NLM)

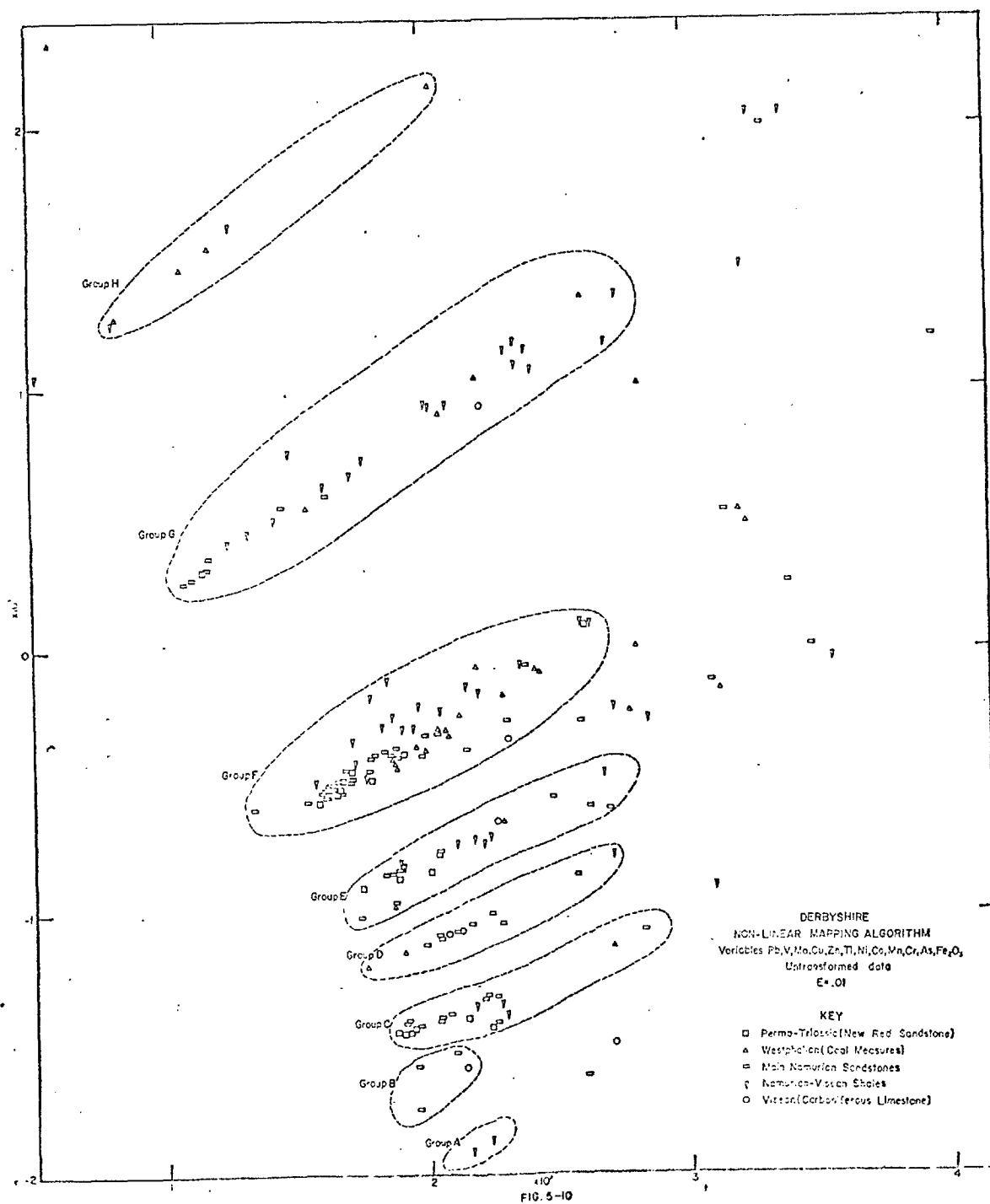
i) Results

Results are based on 705 samples and on the combination of the 12 variables: Pb, V, Mo, Cu, Zn, Ti, Ni, Co, Mn, Cr, As and Fe_2O_3 , and on the 4 variables: Mn, Ni, Co and Zn (except when the data was normalized).

Figure 5-10 shows the "basic" groups defined when the first 500 samples of the whole data set were processed without normalization or log-10 transformation of the data. The remaining samples of the set were tested in the same manner with the "linkage samples" and similar results were obtained. All groups displayed a marked overlap between different bedrock units probably for the reason discussed in the previous section (5-3).

The problems of correlation with the main bedrock units discussed in the dendrogram interpretation also apply here, however, the very complex erratic distribution of the different group patterns in this area gave the worst difficulties of the interpretation. Only a gradational separation was possible between the combined Permo-Triassic-Main Namurian Sandstones category, because of the high degree of overlap between both units, and the Namurian-Visean Shales. The erratically distributed Westphalian samples are more associated with the Namurian-Visean Shales. Those samples distributed around the periphery of the diagram and isolated from the main clusters ("basic" groups) were also considered "anomalous" samples.

The "basic" groups and the "anomalous" samples were plotted as a classification map (Fig. 5-11) and the statistical properties for each group are given in table 5-14. No important correlation with the bedrock units may be carried out because of the high variability in the distribution of the groups as a response to the disturbed constitution of the stream sediments. No well defined associations between groups are shown by the EDF analysis results (Table 5-15). The higher rates of correct recognition for each group confirm again, as in previous areas, the strictly "natural" response of each group to the mainly trace element geochemistry on the basis of the untransformed data. Samples classified as "anomalous" outlined the mineralized and contaminated areas well because they are present in conspicuously higher density



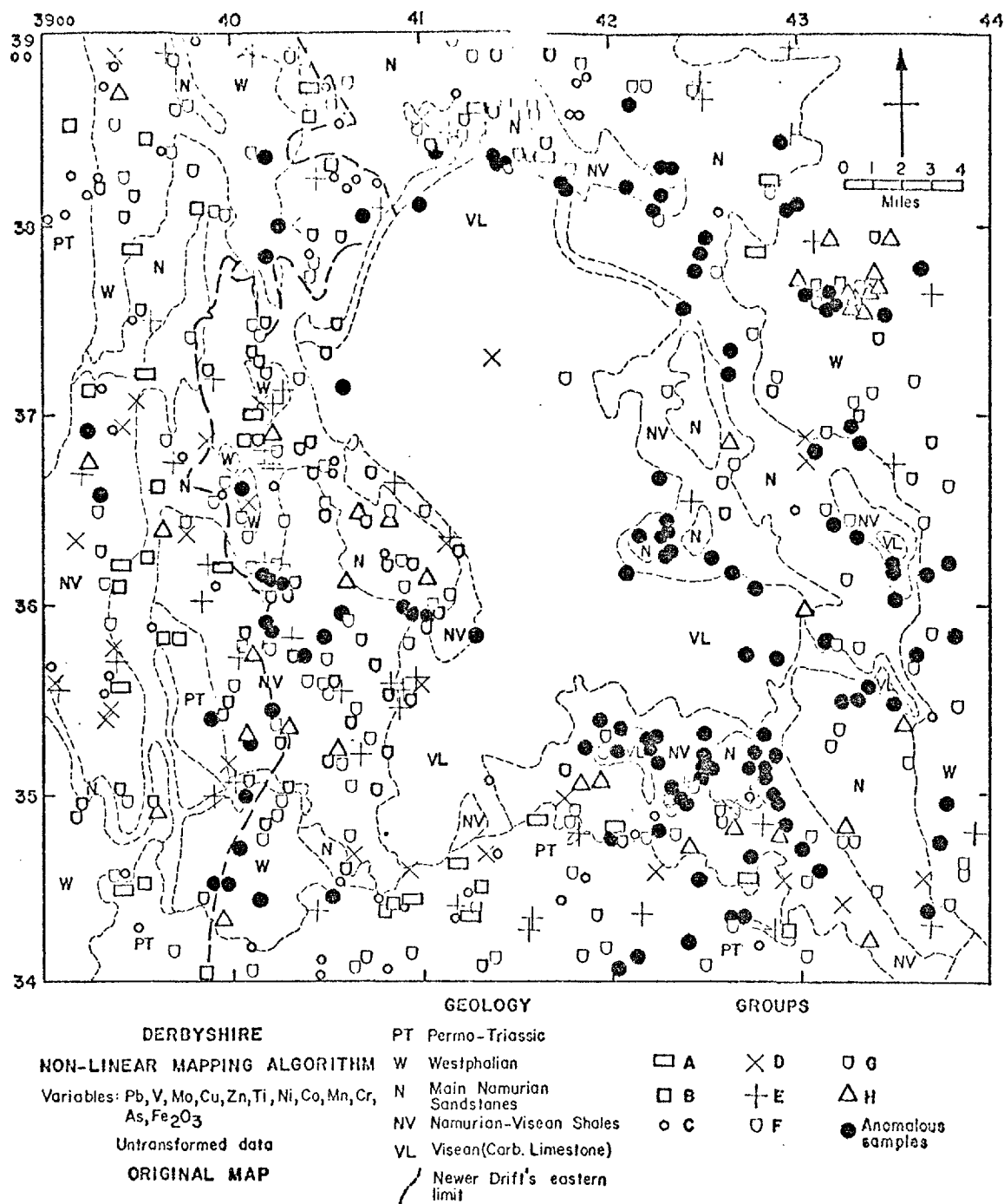


FIG. 5-11

TABLE 5 - 14

Arithmetic means ($\bar{x}$) and standard deviations (s), Derbyshire, NLM algorithm, untransformed data; total sample size = 638.

GROUPS	A	B	C	D	E	F	G	H	ANOMALOUS SAMPLES
SAMPLE SIZE	19	37	51	44	70	172	81	36	128
Pb $\bar{x}$	46	49	57	55	93	95	78	139	1874
Pb s	39	64	51	64	116	121	121	237	2727
V $\bar{x}$	26	33	39	47	64	72	103	116	73
V s	8	19	24	25	36	42	48	53	51
Mo $\bar{x}$	1.2	1.3	1.8	2.1	4.1	5.8	8.3	5.9	7.0
Mo s	0.5	0.8	2.0	2.0	5.0	8.5	9.9	8.1	8.6
Cu $\bar{x}$	25	28	42	37	44	51	61	54	57
Cu s	9	10	25	15	25	29	38	20	40
Zn $\bar{x}$	92	115	150	169	194	203	273	257	845
Zn s	79	76	91	114	138	151	258	211	1432
Ti $\bar{x}$	603	908	1059	1300	1661	2015	2963	3917	2058
Ti s	73	350	416	0	369	173	155	363	1319
Ni $\bar{x}$	23	27	35	39	53	52	77	68	78
Ni s	9	8	19	18	35	34	45	34	74
Co $\bar{x}$	8	12	14	17	22	22	27	26	42
Co s	5	6	6	8	11	11	10	8	96
Mn $\bar{x}$	152	211	318	357	490	512	764	818	1977
Mn s	104	201	256	227	523	442	548	524	2572
Cr $\bar{x}$	37	43	51	50	55	60	77	77	144
Cr s	33	38	54	28	27	31	40	30	798
As $\bar{x}$	15	5	7	9	11	11	15	9	15
As s	25	7	15	10	15	15	23	9	18
Fe ₂ O ₃ $\bar{x}$	0.8	1.0	1.5	1.5	1.9	2.3	4.3	3.3	2.7
Fe ₂ O ₃ s	0.2	1.2	1.4	0.7	1.2	1.2	10.9	1.7	2.0

Mean values in p.p.m., except Fe₂O₃ (%)

TABLE 5 - 15

Group classification results by EDF analysis, Derbyshire, NLM algorithm, untransformed data results, $\sigma = 45.0$

ASSIGNED GROUP											
	A	B	C	D	E	F	G	H	ANOM.	?	% CORRECT
TRUE GROUP	A	18	1	0	0	0	0	0	0	0	95
	B	1	33	1	0	0	1	0	1	0	89
	C	0	0	50	0	0	0	1	0	0	98
	D	0	0	0	44	0	0	0	0	0	100
	E	0	0	0	0	63	0	2	4	0	90
	F	0	0	1	0	2	163	3	3	0	95
	G	0	1	0	0	1	2	76	0	1	94
	H	0	0	0	0	1	1	0	34	0	94
	Anom.	2	4	3	0	8	2	1	0	93	73

Overall correct recognition rate
(Excluding anomalous group) = 94%

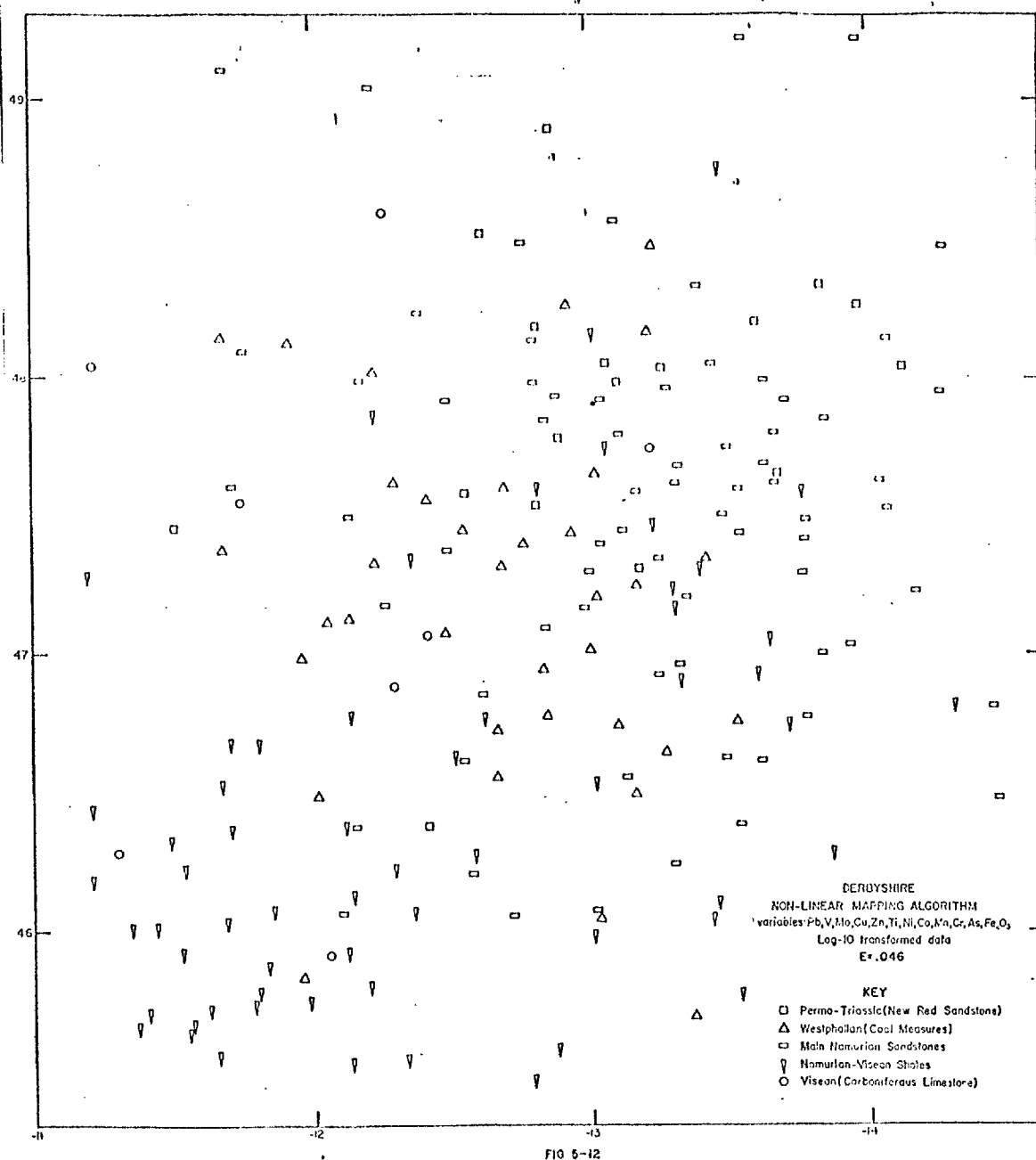
Overall correct recognition rate
(For all groups) = 90%

over the eastern half of the area in contrast with the western half. The group of "anomalous" samples attains the highest concentration (and variability) of Pb and Zn (table 5-14) which are the elements characterizing the mineralization and contamination.

Part of the results obtained with the log-10 transformed data are shown in figure 5-12. This illustrates the mapping obtained with the same 500 samples; a similar structure was obtained with the remaining 205 samples. The figure shows an extreme degree of overlap, mainly between the Westphalian (as an intermediate group on the structure) and the remaining categories, and between the Permo-Triassic and the Main Namurian Sandstones. Disregarding the Westphalian category, the same gradational separation between the combined category Permo-Triassic-Main Namurian Sandstones and the Namurian-Visean Shales obtained with the untransformed data can again be observed.

Part of the results obtained with the normalized data are presented in figure 5-13. This shows the mapping obtained with some 500 samples tested with the log-10 transformed data; a similar structure was obtained with the remaining 205 samples. The mapping shows a similar structure to that obtained with the log-10 transformed data (Fig. 5-12), however, a higher proportion of Namurian-Visean Shales are seen scattered throughout the remaining categories.

The distribution of the main geological bedrock units in the untransformed, log-10 and normalized NIM mappings would also serve to corroborate the already established chemical similarity (or proximity) between the stream sediments derived from the Permo-Triassic and the Main Namurian Sandstones, and between those stream sediments derived from the Westphalian and the Namurian-Visean Shales (Table 5-1). The intermediate chemical composition



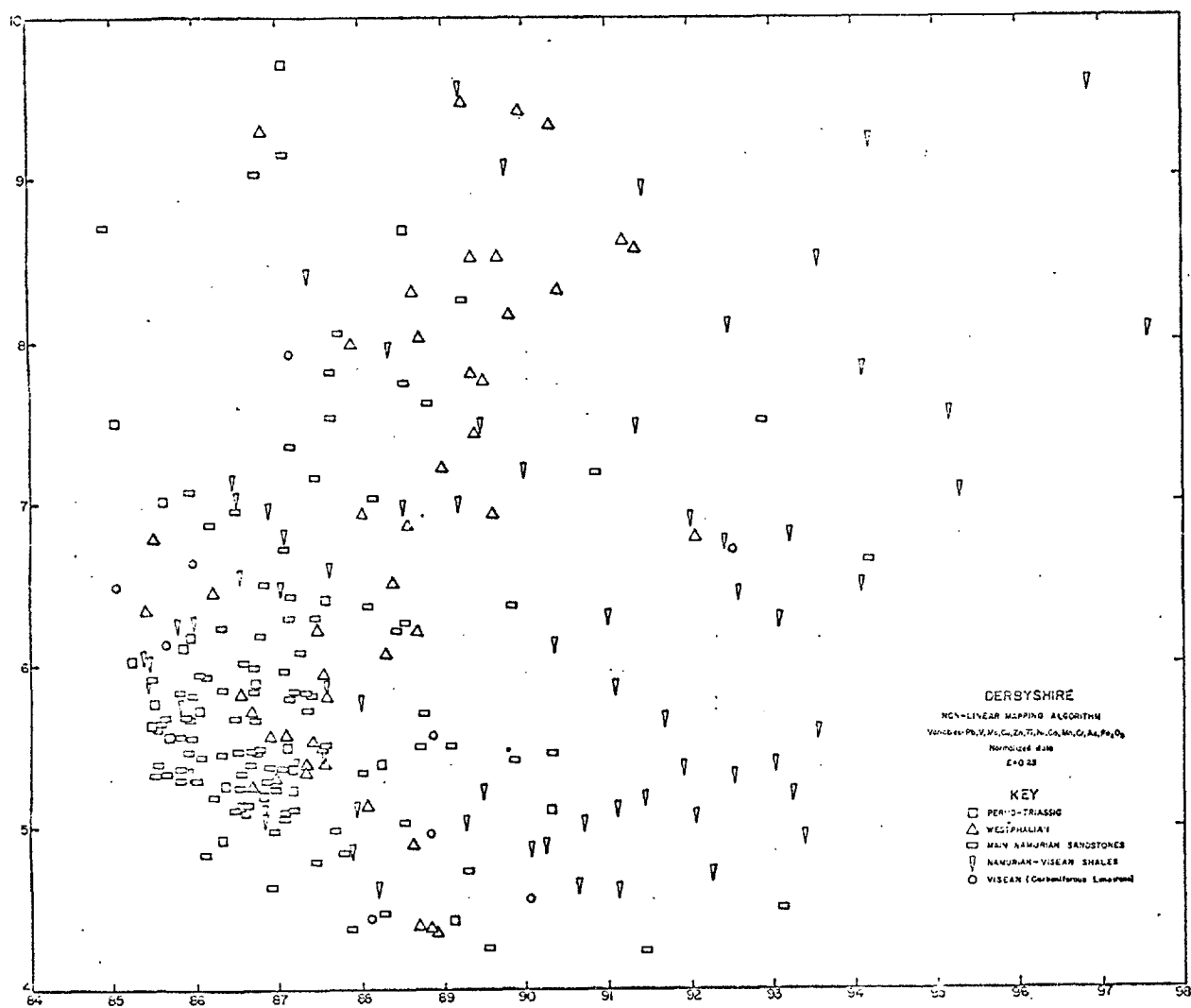


FIG. 5-13

of the Westphalian was very well shown by all mappings because the general intermediate (with high overlap) distribution occupied by the Westphalian with respect to other categories. Stream sediments related to anomalous geochemical environments (e.g. mineralization and contamination) were again generally distributed around the periphery of the diagram.

The "anomalous" samples alone derived from the original untransformed data were subsequently classified implementing the log-10 transformation (Fig. 5-14). A clear linear separation between those "anomalous" samples distributed on the eastern side, mainly influenced by mineralization and contamination, and those overlying the western half which is relatively unmineralized (with the exception of Ecton and Mixon areas) was obtained. The "anomalous" samples over the eastern half, which are more influenced by mineralization, did not show separation from samples mainly influenced by contamination (smelting sites), instead both types of geochemical environments are seen intermingled. The Ecton mineralization is clearly classified within the western relatively unmineralized area.

The group of "anomalous" samples (69 samples) influenced by mineralization and contamination alone were subsequently classified with the same 12 variables; a rather noisy structure composed of several isolated small clusters was obtained (Fig. 5-15). It was then decided to find out the best subset of elements by examination of all possible combinations of the 1 to 12 variables in order to obtain a better discrimination of samples representing mineralization from contamination.

In this problem, the highest classification success rate (85.5%) was attained with the combination: Mo-Cu-Zn-Co-Cr-As-Fe, and the combination: V-Cu-Zn-Ti-Co-As-Fe was also found to be an equally successful "best feature set". Fig. 5-16 shows the classification obtained with the first set of variables. Those samples

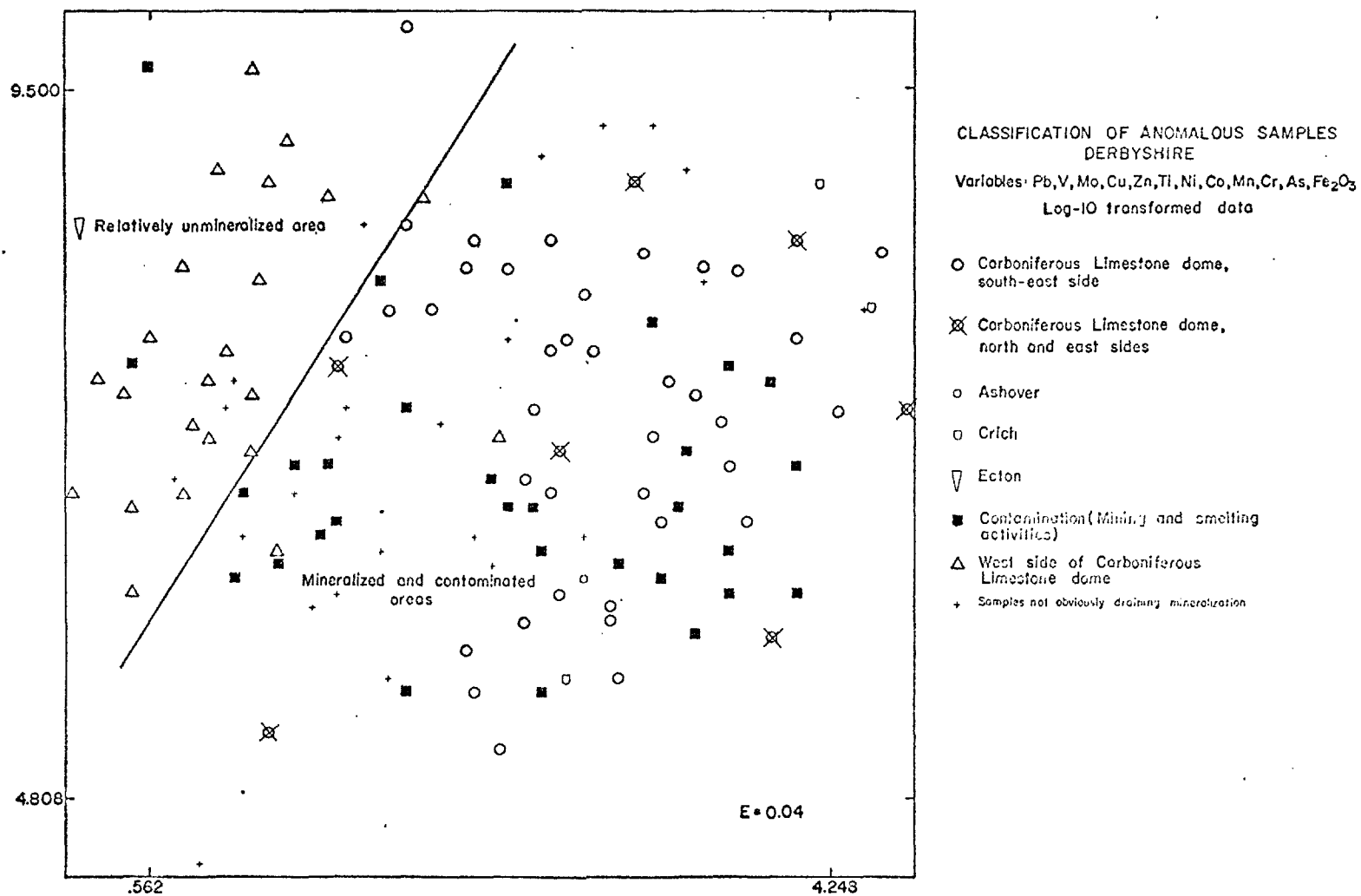


FIG. 5-14

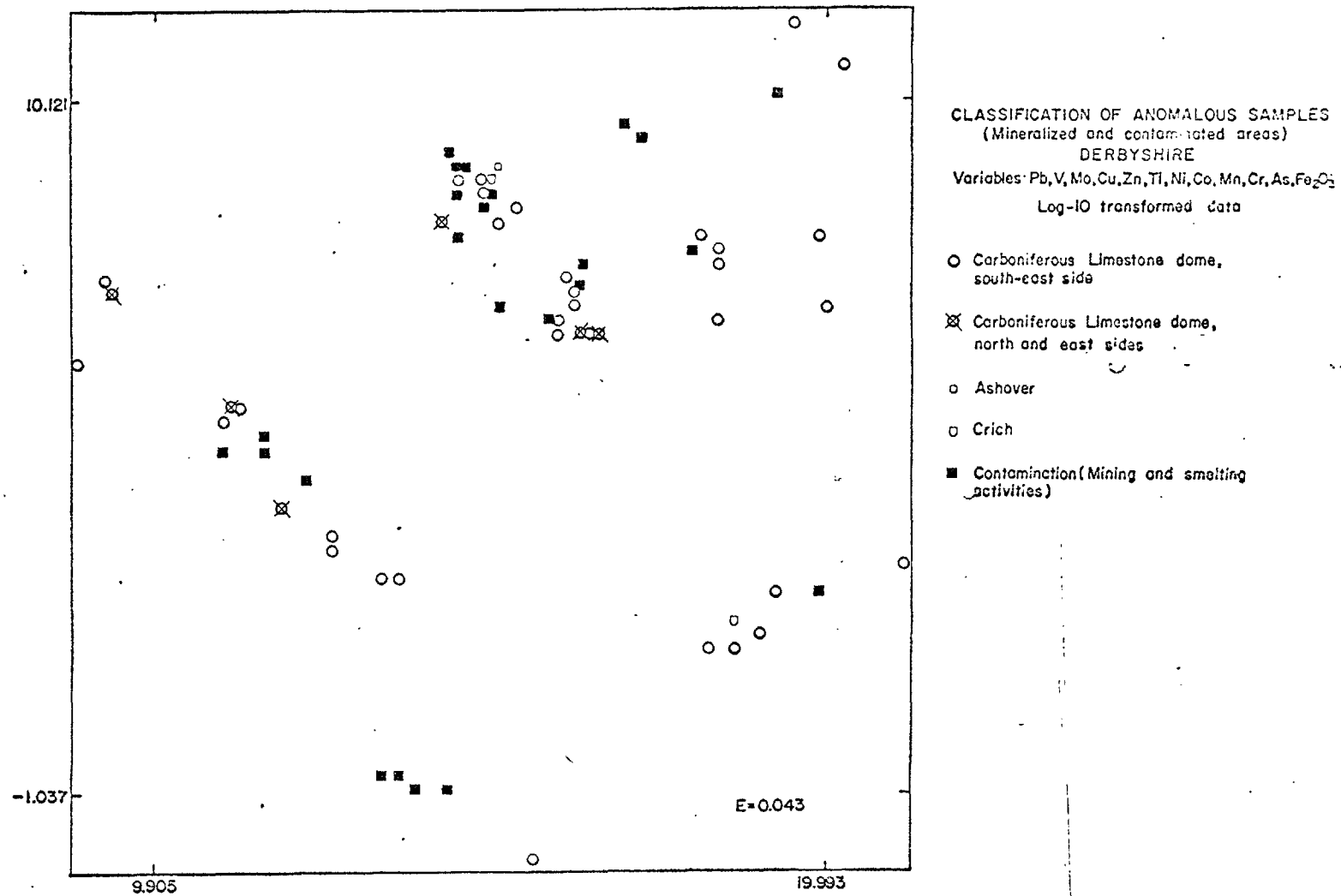


FIG. 5-15

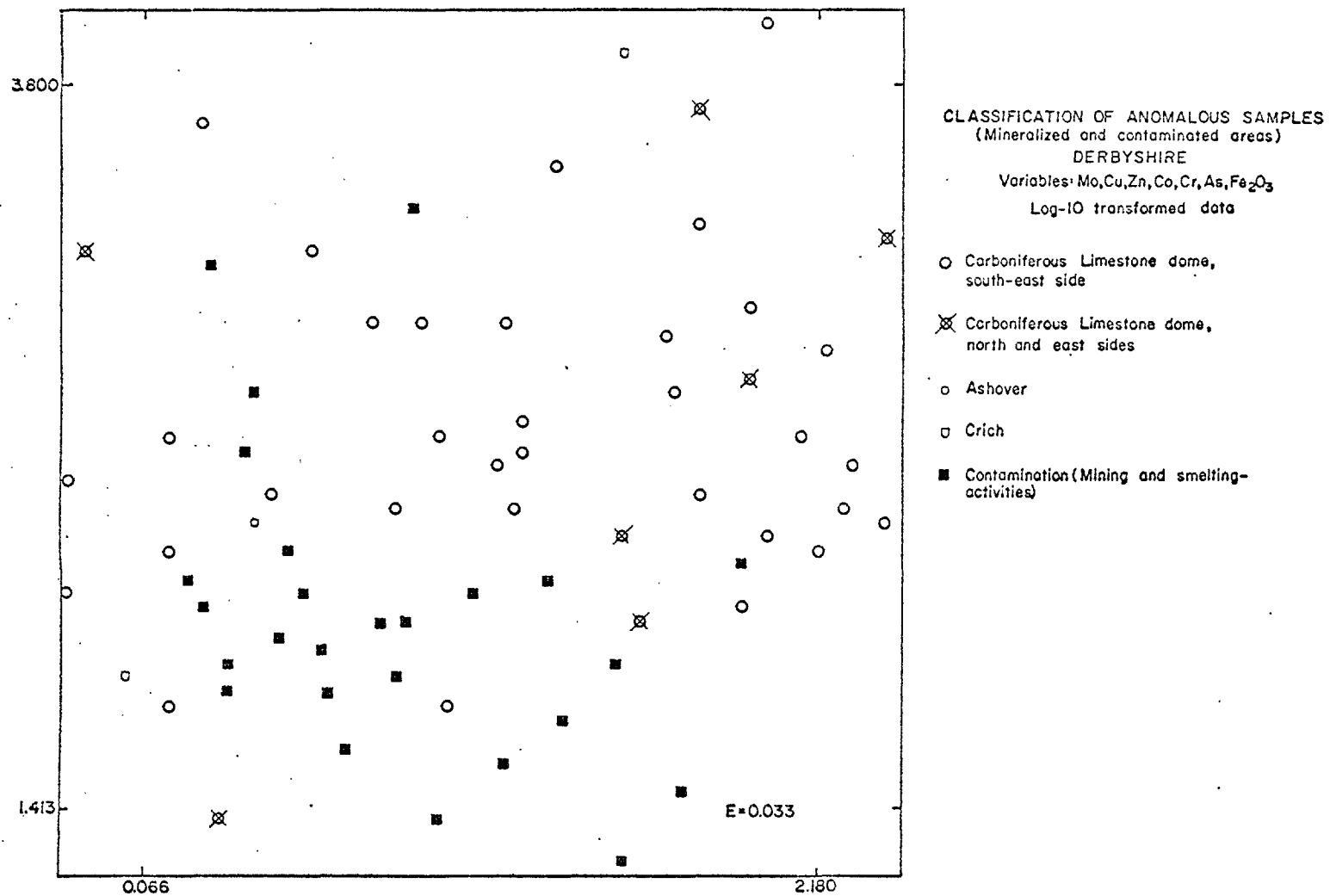


FIG.5-16

mainly influenced by contamination are now almost perfectly well differentiated from those samples mainly influenced by mineralization. Mineralization types were probably not separated because of the occurrence of only one conspicuous mineralization (Pb-Zn) in the study area and the set of variables considered. Similar results to those of figure 5-16 were obtained with the second set of variables.

The untransformed and log-10 transformed data were subsequently processed with only the variables Mn, Ni, Co, Zn but no significant results were obtained instead generally confused noisy mappings were obtained.

ii) Discussion

The application of the NLM algorithm to the composite and strongly contaminated stream sediments of Derbyshire has produced the most erratic results and therefore made the correlation of the different classification patterns with principal units of the bed-rock very difficult.

In the untransformed results 128 samples out of 705 (18%) were considered as "anomalous". 86 of these 128 (67%) were related to mineralization and contamination and 65 (51%) were estimated to be more related with mineralization. 26 (20%) were mainly distributed over the western side in the vicinity of the Newer Drift; 16 (13%) were apparently unrelated with any known major source of mineralization or associated contamination. In this case the relatively smaller confusing influence of the "anomalous" samples on the western side of the area made the pattern corresponding to the mineralization and contamination on the eastern side clearer.

In the log-10 transformed results 117 samples (17%) were considered "anomalous". 68 of these 117 (58%) were related to mineralization and contamination and about 31% of these were more associated with mineralization; 41 (35%) were mainly distributed over the western side and 8 (7%) were apparently unrelated with

any known source of major mineralization and associated contamination throughout the study area.

In the normalized results 117 samples (17%) were considered "anomalous". 52 of these 117 (44%) were related to mineralization and contamination and about 23% of these were more associated with mineralization; 55 (47%) were mainly distributed over the western side and 10 (9 %) were apparently unrelated with any known source of major mineralization and associated contamination.

In both log-10 and normalized results the relatively greater influence of the "anomalous" samples on the western side of the area (mainly with the normalized data) made the detection of the mineralized and contaminated areas on the eastern side more difficult.

For practical purposes the classification of the "anomalous" samples alone (log-10 transformed data) distributed over mineralized, contaminated and relatively unmineralized areas has served to illustrate the importance of a successive classification of "anomalous" samples in order to discriminate (without additional geological field information) those areas most likely to bear mineral deposits (e.g. the eastern half of the study area) in contrast with those relatively unmineralized areas (e.g. the western side) which obscure the results and their interpretation. In the classification of the "anomalous" samples of Devonshire (Fig. 4-13) the stream sediments over the western side of the Dartmoor granite (highly influenced by the nature of the secondary environment) were perfectly separated from those mainly influenced by mineralization. In new and relatively (geologically) unknown areas a good control of the secondary environment distribution (e.g. pH and Eh measurements) and any source of contamination would undoubtedly assist in a better interpretation of the results.

5-5 Conclusions: Hierarchical cluster analysis and NLM algorithm.

1) The classification results of the bedrock geology were in general confused in both the dendrogram representation and the untransformed NLM mapping, however, better results were achieved with the log-10 dendrogram representation.

2) The best NLM classification results for both the bedrock geology, and the group^{of} "anomalous" samples derived from the untransformed NLM mapping, were obtained with the log-10 and normalized (only bedrock) NLM mappings. Separation of mineralized and contaminated areas on the eastern side from relatively unmineralized (uncontaminated) areas on the eastern side was perfectly well obtained with the log-10 NLM mapping using the same original 12 variables. Subsequently differentiation of mineralization from contamination was obtained with the "best" set of variables selected according to a feature selection procedure.

3) Both mineralization and contamination together, and mineralization alone, were better detected by the "anomalous" samples classified with the dendrogram log-10 transformed data than with the untransformed data (Table 5-16). The smaller confusing influence of the "anomalous" samples on the western side made the detection of mineralization and contamination on the eastern side more conspicuous.

4) Both mineralization and contamination together, and the mineralization alone, were better detected by the "anomalous" samples classified with both untransformed and log-10 NLM mappings than with the normalized mapping (Table 5-16). The untransformed NLM mapping achieved better results than the log-10 NLM mapping.

5) Mineralization and contamination were together better detected by the untransformed and log-10 transformed NLM mappings than with the dendrogram representation or the normalized NLM mapping (Table 5-16). Mineralization and contamination together, and mineralization alone, were better detected with the log-10 dendrogram representation than with the normalized NLM mapping. Mineralization alone was better detected with the untransformed and log-10 transformed NLM mappings than with the untransformed dendrogram, however, the log-10 transformed dendrogram attained higher classification rates than the untransformed dendrogram and the log-10 and normalized NLM mappings. Normalization of the raw data shows poorer detection rates regarding geochemically anomalous areas with respect to the other methods and transformations.

TABLE 5-16

Comparative successful recognition rates (%) of the "anomalous" samples in relation to different geochemical environments, Derbyshire.

METHOD	DATA OPTION	MINERALIZATION & CONTAMINATION %	MINERALIZATION %	WESTERN SIDE %	AREAS BARREN OF MINERAL & CONT. %	No. OF "ANOMALOUS" SAMPLES
NLM ALGORITHM	Untransformed (Raw Data)	67	51	20	13	128
	Log-10	58	31	35	7	117
	Normalized	44	23	47	9	117
DENDRO- GRAM	Untransformed	37	30	48	15	135
	Log-10	56	34	20	24	41

CHAPTER 6

GENERAL CONCLUSIONS AND RECOMMENDATIONS

6-1 Introduction

On the basis of the nature of the regional geochemical data of Denbighshire, Devonshire and Derbyshire (mainly trace-element geochemistry) and the available geological information, a short discussion with general conclusions follows to present the achievements of the classification methods and the limitations found. The recommendations subsequently made have the purpose of suggesting an orientation path to follow in future research on the basis of these results.

6-2 General conclusions

i) Empirical discriminant analysis.

The EDF analysis has shown very encouraging results in the Denbighshire and Derbyshire areas which can be complemented with the results previously obtained in Devonshire (Howarth, 1971b). The quantitative results for the correct classification performance of the method were less successful than those obtained with the Devonshire data, but the evaluation has been largely successful for illustrating the promising application of EDF analysis both in regional geological mapping and in routine mineral exploration surveys in new prospect areas. On the basis of the general evaluation of the results in both areas of the United Kingdom the following general conclusions can be summarized:

1) The bedrock geology was better classified with the log-10 transformed data than with the untransformed data. Similar results were obtained in Devonshire (Howarth, 1971b). The classification results were improved when the most appropriate "training areas" were established and the influence of contamination minimized.

2) The mineralization (contamination) was well detected by the "unknown" -classified samples with the untransformed data without the establishment of "mineralization" training areas.

Classification tests using different percentages of "unknown" samples (derived from the testing set) were carried out by increasing or decreasing the "optimum" smoothing parameter (σ). The mineralization was, in all cases, well detected. When the number of "unknown" samples attained higher proportions (e.g., 30% ; low σ) an appreciable number of "unknown" samples were distributed over non-anomalous bedrock features that obscured the detection of the geochemically ^{really} ~~isotopically~~ anomalous areas. Where the number of "unknown" samples attained lower proportions (e.g., 5% ; high σ) the number of "unknown" samples, distributed over non-anomalous bedrock features significantly decreased, but the detection of the geochemically anomalous areas was, although less clear, always covered by a reduced, but persistent, number of "anomalous" sample patterns. Therefore, by increasing or decreasing σ around the optimum (iteratively determined) value, those "unknown" samples which appear persistently over very localized areas would be those which would have the highest probability of indicating possible anomalous areas. This will have great importance in virgin areas due to the lack of knowledge about mineralization types present; in such areas the detection of mineral target areas would mainly be carried out based on the "unknown"-classified samples.

3) The mineralization types were better classified with the log-10 transformed data when the appropriate "mineralization" training areas were established.

ii) Hierarchical cluster analysis (dendrogram representation) and the non-linear mapping algorithm.

Regional geochemical reconnaissance surveys in new prospect areas, principally in tropical (virgin) areas with dense exuberant vegetation, a thick cover of deeply weathered overburden masking the "fresh" outcrops along flanks and crests,

and with difficult penetration, would undoubtedly provide ideal material to test the practicability of the EDF analysis provided that the minimum required geological information for the establishment of the most appropriate "training areas", as much as possible representative of the main geological features, is available. In this preliminary phase, the hierarchical cluster analysis (dendrogram representation) and the NLM algorithm would hopefully supply valuable information in order to delineate those areas that geochemically would correspond to the main mappable geological features (e.g. bedrock and mineralization).

The results of the dendrogram representation and the NLM algorithm in Denbighshire, Devonshire and Derbyshire have served to test the practicability of the cluster-seeking techniques in order to carry out the delimitation of the groups (or associations of groups) that generally showed correlation with the main features of the bedrock geology, the mineralization (contamination) or the secondary environment, principally in areas with drainage systems mainly transporting stream sediments derived from a single geological formation. The following conclusions can be summarized regarding both cluster-seeking techniques, these are:

- 1) On the basis of the qualitative evaluation of the "basic" groups (with the exception of the "anomalous" samples), the best classification results of the bedrock geology and features of the secondary environment were obtained with the log-10 NLM mapping and to a lesser extent with the normalized NLM mapping. Good results were also obtained regarding mineralization and secondary environment conditions; these results would be significantly improved with better field information.

- 2) On the basis of the qualitative evaluation of the "basic" groups (with the exception of the "anomalous" samples), better classification results of the bedrock geology (except in Devonshire) and features of the secondary environment were obtained with the log-10 dendrogram representation than with the un-

transformed versions of both the dendrogram representation and the NLM mapping. Despite the larger problems of distortion of the dendrogram structure, compared with the NLM structure, the group (or associations of groups) outlined, either with the untransformed or log-10 transformed data were found to be generally related with the main (broad) geological features of the bedrock and the secondary environment.

3) Reliable and consistent correlations with bedrock features were not obtained with the untransformed NLM mapping because of the high overlap between different geological features with the same "basic" groups. However, the EDF analysis (Table 6-1) has shown the highest overall correct recognition rates for all study areas with the application of this version of the NLM algorithm, probably because the defined groups are more representative of real-world "natural" conditions (geochemical facies) in terms of chemical composition than of mapped geological units or lithological facies.

4) On the basis of the quantitative evaluation of the "anomalous" samples, the best detection of mineralization (contamination) and features of the secondary environment was achieved with untransformed and normalized (except in Derbyshire) NLM mappings, however, good detections were also given by the rest of the methods and their data transformations. Better results were obtained with the log-10 dendrogram representation than with the untransformed version.

5) When different groups of "anomalous" samples alone were classified, better results were obtained with the log-10 NLM mapping than with the dendrogram representation.

6) The groups of "anomalous" samples generally showed a high rate of misclassification probably because of the diversity of different geological sources (geological formations) influencing the "anomalous"-classified samples. The groups of

TABLE 6 - 1

Overall correct recognition rates (%) of the EDF analysis applied to the cluster-seeking results.

DATA	GROUPS	DENBIGHSHIRE		DEVONSHIRE		DERBYSHIRE	
		DENDROG.	NLM	DENDROG.	NLM	DENDROG.	NLM
UNTRANSFORMED DATA	NON-ANOMALOUS	61	93	49	93	34	94
	ANOMALOUS	23	49	28	54	16	73
LOG-10 TRANSFORMED DATA	NON-ANOMALOUS	63	*	80	*	59	*
	ANOMALOUS	22	(52)**	50	(57)**	20	(58)**
NORMALIZED DATA	NON-ANOMALOUS	-	*	-	*	-	*
	ANOMALOUS	-	(80)**	-	(63)**	-	(44)**

* Effectively single cluster structure without a priori information.

** Estimated from additional information on mineralization (contamination) and secondary environment conditions (except for Derbyshire) distribution.

"anomalous" samples seem to be composed of some apparently independent small sub-groups of "anomalous" samples for each geological formation, sometimes not strongly reflecting mineralization (contamination). No obvious causes are known to produce the samples classified as "anomalous" over relatively unmineralized areas, confusing then the "anomalous" patterns over mineralized (contaminated) areas.

7) In the dendrogram representation (untransformed and log-10 transformed data) some "basic" groups (with the exception of the groups of "anomalous" samples) showed strong correlation with mineralized (contaminated) areas.

8) Terrains conspicuously influenced by the unusual nature of the secondary environment (Mn induced conditions) in the study areas were not detected by any specific group pattern(s) when using the variables: Mn, Ni, Co, and Zn which are conspicuously enhanced in the secondary environment. Higher overlap and variability between all important geological features was observed.

9) Normalization of the data did not significantly eliminate overlap between groups in the log-10 NIM mappings; results were generally similar to those performed without normalization.

10) The R-mode analysis yielded some consistent variable associations correlated to a certain extent with some of the previous factor analysis results (Khaleelee, 1969), but the persistent variability of other variable associations suggests inconsistency of these associations throughout the different data transformations and apparently they could be artificial products of the sampling, and may not therefore be significant in the factor analysis.

11) EDF analysis subsequently applied to the cluster-seeking technique results assisted the more "objective" definition of the degree of association or similarity between geochem-

ically similar "basic" groups. With this procedure, broad geological features were very well delimited by "associations" of "basic" groups, (with the exception of Derbyshire) principally with the results derived from the dendrogram representation. A summary regarding the overall correct classification rates for independent non-anomalous and anomalous groups, for each transformation and study area, is included in table 6-1.

With the log-10 and normalized NLM mappings the number of samples considered "anomalous" (peripheral samples on the single-cluster two-dimensional structure) can, to a certain extent, be arbitrarily decided by the geochemist within wide maximum limits and advantageously made uniform in number in order to achieve better comparative quantitative evaluations. With the untransformed NLM mapping the number of "anomalous" samples to be selected showed more stable maximum limits highly independent of human bias.

The relatively advantageous freedom of arbitrarily choosing larger sets of "anomalous" samples, with the log-10 and normalized NLM mappings, had the disadvantage that a good proportion of "anomalous" samples showed higher correlation with non-anomalous bedrock features. However, it is believed this difficulty could be minimized by selecting a smaller number of peripheral ("anomalous") samples mainly over more isolated regions of the mapping farther away from the main cluster. With the untransformed NLM mapping those samples considered "anomalous" showed generally lower correlation with non-anomalous bedrock features (except in the "unfavourable" Derbyshire area).

With the untransformed and log-10 transformed dendrogram results, different numbers of "anomalous" samples were selected according to the possibilities given by the structure. The more compact (compressed in the distance scale) log-10 dendrogram structure, with a lower degree of correlation variability,

produced a more limited number of "anomalous"-classified samples in contrast with almost 3 times more obtained with the untransformed dendrogram due to a wider (less compressed) structure and then with a higher correlation variability between samples and groups of samples. The more compact groups defined with the log-10 dendrogram structure apparently seem to be more likely responding to the influence of the non-anomalous bedrock features. In the study areas, mainly in those considered favourable to test the present classification techniques, such as Denbighshire and Devonshire, the "anomalous" samples derived from the log-10 dendrogram showed a lower response to the non-anomalous bedrock features, then, a better outline of the geochemically anomalous areas was obtained. The higher variability of the correlation levels in the untransformed dendrogram structure produced a larger proportion of "anomalous"-classified samples over the non-anomalous bedrock, therefore, obscuring the detection of the true anomalous areas. It is also believed that by choosing those samples (mainly in the untransformed dendrogram) linked at lower distance correlation levels, the possibility of misclassifying non-anomalous bedrock features would be appreciably reduced.

It must be clear that some of the previous conclusions and arguments do not have to be considered as "generalized" rules of immediate application in all cases, because, as it has been discussed in earlier chapters, differences in the results using, for example, different types of variables are to be expected. One example is shown by the Mount Shoobridge granite data (major element geochemistry) discussed in section 2-3. In this case, the perfect delimitation of the main petrological groups was achieved with the untransformed NIM mapping, whereas with the log-10 NIM mapping the classification was poorly carried out. In contrast, the regional geochemical data (mainly trace element geochemistry) considered in this investigation was, instead, much better classified with the log-10 NIM mapping.

In Devon neither the cluster-seeking techniques nor the EDF analysis (Howarth, 1971 b) showed significant classification patterns responding to the sandstone strata interbedded with

the Namurian shales. Instead, the results were confused, suggesting that the sampled material (stream sediments) and/or the sampling density (one sample per square mile) were not suitable to detect acute differences of lithology and chemical composition like those characterizing the interbedded sandstones and shales of the Namurian Formation (Table 6-2).

The individual classification patterns produced by the EDF analysis and the cluster-seeking techniques (as a response to the chemical composition) corresponded in a large number of cases with the bedrock and mineralization features. In the three areas under study the relationship between well defined classification patterns and features of the bedrock was not found in all cases because of the lack of more detailed geological information (e.g. lithological facies). It became obvious from the field investigations carried out that because of the difficulty in relating changes in classification patterns derived from the stream sediment geochemistry to observed variations in bedrock geology, mainly because of the extreme lack of bedrock exposures (in streams and road cuttings, e.g., in cultivated areas such as Devon), the limitations of the stream sediments as a sampling medium and the sampling density, more detailed correlation between the classification patterns and detailed lithological variations in the bedrock would be practically impossible to obtain over the majority of each area (for the reasons mentioned above). In terms of a field survey in virgin territory this would involve such an expenditure of effort that it would almost certainly not be cost-effective in the reconnaissance phase of the survey with which we are concerned here.

It is believed that the solution to this type of problem, from a more academic than practical point of view, could be better investigated carrying out a specific research project combining mapping of lithologic facies with sampling of different geological materials (e.g. stream sediments, soils and rocks) and an increase of the sampling density.

TABLE 6 - 2

Mean metal values of 6 shale and 13 sandstone samples of the Namurian (Crackington) Formation, Devonshire (Spectrographic analysis).

ELEMENT	SHALE $\bar{x}$ (p.p.m.)	SANDSTONE $\bar{x}$ (p.p.m.)
Pb	37	19
Ga	27	16
V	283	66
Cu	41	20
Zn	233	192
Ti	2933	1662
Ni	57	40
Co	30	14
Mn	350	248
Cr	127	53
Sn	5	5
Ag	0.2	0.2
As	500	500

Rock sample locations and descriptions:(Table 6-2).

No.24:	N.G.R., 2130/0953.	Shale, black somewhat weathered.
20:	2145/0978.	Shale, idem as above.
22:	2138/0968.	Shale, idem as above.
18:	2155/0983.	Shale, light black and more weathered.
7:	2550/0930.	Shale, light black, limonite along fractures and deeply weathered.
2:	2440/1005.	Shale, black, somewhat weathered.
23:	2130/0953.	Sandstone, grey, medium-grained, poorly weathered.
16a:	2180/1000.	Sandstone, light grey, medium-grained deeply weathered.
19:	2145/0978.	Sandstone, light grey, medium-grained, deeply weathered with limonite coatings mainly along fractures.
17:	2155/0983.	Sandstone, light grey, medium-grained, deeply weathered with limonite coatings mainly along microfractures.
15:	2118/0988.	Sandstone, dark grey, relatively unweathered, with limonitic coatings along fractures.
14a:	2225/0980.	Sandstone, grey, medium-grained, relatively unweathered with limonite along fractures.
8:	2415/1118.	Sandstone, generally fresh, with turbidity structure, medium to fine-grained grading to shale.
9:	2420/1117.	Sandstone, idem as above.
6:	2550/0930.	Sandstone, grey, medium-grained, generally fresh, limonitic coatings along microfractures.
4:	2460/0985.	Sandstone, grey, medium-grained, microfractures filled with silica, medium-weathered.
5:	2535/0940.	Sandstone, grey, medium-grained, fairly fresh, with some limonitic coatings along fractures.
3:	2440/1005.	Sandstone, same as above.
2a:	2440/1005.	Sandstone, same as above.

6-3 Recommendations.

The recommendations below suggest some of the steps that should be considered in future investigations regarding the application of discriminant and cluster analysis in regional geochemical surveys as complementary routine tools for regional geological mapping and mineral exploration surveys, these are:

1) Both techniques should be simultaneously applied to poorly known or unknown (geologically) new areas. Cluster analysis should firstly be carried out seeking the most significant groups that assist in the definition of the most representative "training areas" for the subsequent application of the discriminant analysis. Overlapping, as far as possible, in both cluster and discriminant results should be looked for as a logical response to the most conspicuous geological and geochemical ("natural") environments representative of the mappable regional geological units, the mineralization, and/or any other important geological feature.

For the subsequent practical applications of these techniques to similar regional geochemical data (stream-sediment samples and trace-element geochemistry), the following suggestions are given:

a) Preferably apply the log-10 NLM mapping (or the normalized NLM mapping) to stream-sediment samples properly identified with sample number and codified lithological qualities (e.g. A = sandstone; B = granite; C = poorly drained peaty gley soils; D = relatively flat topography) or other interesting geological attributes recorded at the sampling sites. For example, a sample identification 345BCD would indicate that the sample was collected over relatively flat granite exposures covered with poorly-drained peaty gley soils. With this sort of identification procedures the geochemist will be able to determine on the scatter

plot whether samples with similar identifications constitute well defined groups that could be used for geological mapping and as "training" areas for subsequent application of the EDF analysis in order to corroborate the "natural existence of well defined mappable geological (and geochemical) features. Mineralization and secondary environment features could also be detected simultaneously with these techniques by choosing a small proportion of "anomalous" samples (e.g. 10% of the total (large) data set; e.g. 800-1000 samples) such as those more isolated from the main cluster(s) scattered round the periphery of the diagram.

b) In order to discriminate mineralization (and unusual secondary environment conditions) from normal bedrock conditions preferably the untransformed NLM mapping, or perhaps the log-10 dendrogram representation, are recommended. The untransformed NLM mapping is less influenced by human bias during the selection of the "anomalous" samples because most of the samples related to geochemically anomalous environments were clearly ("naturally") separated from the main clusters. The more compact structure of the log-10 dendrogram, and the higher degree of linkage independence shown by the samples considered as "anomalous" with respect to the groups related with the bedrock, makes them more effective to detect geochemically anomalous areas (mineralized and influenced by unusual conditions of the secondary environment). With the application of both methods the classification rates related with the geochemically anomalous areas should (on the basis of this work) generally attain the highest values of correct classification. The detection of mineralization would better be achieved by controlling the influence of unusual conditions of the secondary environment. In this case pH and Eh measurements (and some other additional information, e.g. presence of poorly-drained peaty soils; frequent occurrence of crusts of Mn and Fe oxides) should assist to discriminate both anomalous environments. pH and Eh measurements should be carried out either contemporaneous with the regional sampling or, subsequently, after the data has been processed, on those possible mineral target (and adjoining) areas detected

by the "anomalous" samples as financial economy of the project and the geochemist's criteria suggest.

c) The log-10 EDF analysis should be applied mainly to classify bedrock features and mineralization types either on the basis of a priori knowledge or after the cluster-seeking results have been exhaustively evaluated in virgin areas. Mineralization (and/or secondary environment) alone should be well detected by the "unknown"-classified samples with the untransformed EDF analysis.

2) Both techniques should be applied to rock-(in situ) and soil-(residual) derived samples, consequently more representative of the geology and geochemistry at the sampling site, avoiding as much as possible any cause of contamination. The artificial contamination in rock and soils samples along highway sources does not reflect trace element contamination if the rock samples are trimmed in the manner that is customary in geological sampling (U.S. Geol. Surv., 1969) and the soil samples are collected far enough from the roadway or any other source of artificial contamination. It is then obvious that the rate of artificial contamination in rock and soil materials of new prospect areas will practically be negligible.

3) Different spacing between rock soil and stream-sediment samples should be studied with both techniques in order to know more about the behaviour of the techniques and the results obtained under different sampling densities at the scale of regional geochemical surveys.

4) A relatively "good" geological and geochemical control should be carried out at each sample site (e.g. photogeological control and/or geomorphological field observations; detailed lithological and soil descriptions; Eh and pH measurements, and any other important geological observation). Some of the

difficulties found during the delimitation of the groups, without detailed a priori geological knowledge, principally regarding the bedrock geology, were due to the virtual absence of additional geological information at the sampling sites.

According to the results obtained with 69 mineralized and contaminated "anomalous" samples from Derbyshire (Fig. 5-16), after a feature selection procedure, detailed criteria for the differentiation of areas exclusively influenced by mineralization, from those with combined mineralization and contamination, or contamination alone, could also be further considered as a subject of undoubtedly practical importance. Favourable results would be of immediate application in areas of past mining and industrial activities as the United Kingdom.

APPENDIX 1

Empirical discriminant function (EDF).

Let us assume that for each class Ω_j , $j = 1, 2, \dots, k$, we have observations on a p -feature vector $X = \{x_1, x_2, \dots, x_p\}$, and that the a priori probabilities h_j , $j = 1, 2, \dots, k$, of occurrence of each class are known. Let the multivariate probability density function for the j th category be $f_j(X)$, that is, the probability that X belongs to category j . These functions may be of any form provided that they are everywhere non-negative, integrable, and that their integrals over all space equal unity. The classifier must perform the classification on the basis of this information with a minimum of misrecognition.

Defining a decision function $d(X)$, where $d(X) = d_i$ means that X is assigned to Ω_i , let ℓ_i be the loss (penalty) incurred if $d(X) = d_j$ when X is a member of Ω_i . It is assumed that the loss is zero for a correct decision. The problem is now to choose a decision criterion such that the average loss over all classes is minimized. Fu (1968) has shown that under the optimal rule, in the sense of minimizing the average loss, $\sum_i \ell_i h_i f_i(X)$ is smaller than under any other decision rule. Using a symmetrical loss function

$$d(X) = d_i; \ell_i = 0$$

$$d(X) = d_j, j \neq i; \ell_i = \text{const}$$

when X is a member of Ω_i , then the Bayes decision rule is to assign X to the category for which $h_r \ell_r f_r(X)$ is a minimum.

While it may be possible to estimate the a priori probabilities and loss-function values to the correct order of magnitude for each category, it may be impossible to know the probability density functions.

Again following Specht, we will estimate the probability densities for each category as a sum of exponentials based upon the set of training samples, which all have a positive probability of occurring, and we will assume that samples not in the training set but near a given sample point (in p space) will have about the same probability of occurring as the training samples.

Assuming that the estimated probability density function for a category is smooth and continuous, and that the first partial derivatives are small, Specht proposed that an interpolation function $g(X, X_i)$ be found such that

$$f(X) = \frac{1}{m} \sum_i g(X, X_i)$$

where m is the number of training patterns available and $g(X, X_i)$ is the contribution of the i th training pattern to the estimated density. If it is assumed that each training pattern contributes independently to the overall density distribution, and that $g(X, X_i)$ is a function of the Euclidean distance of X from the i th pattern point in p space, following Specht we write

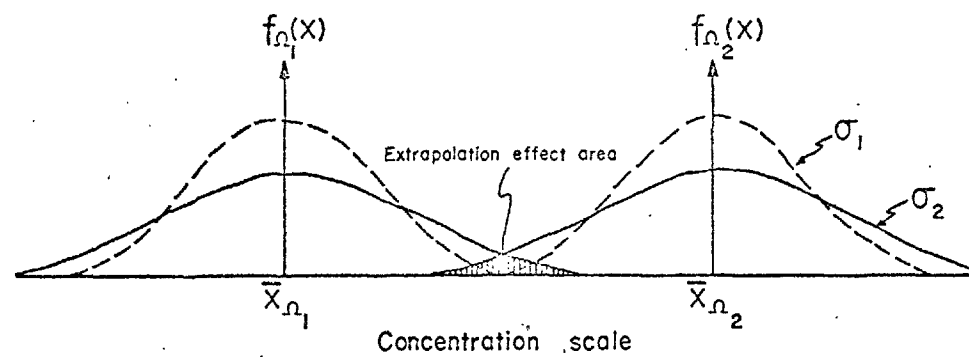
$$g(X, X_i) = \frac{1}{(2\pi)^{p/2} \sigma^p} \exp \left[-\frac{(X-X_i)'(X-X_i)}{2\sigma^2} \right]$$

where σ is a "smoothing parameter" (see page 19). The estimated density function for the Ω th category is then

$$f_{\Omega}(X) = \frac{1}{(2\pi)^{p/2} \sigma^p} \cdot \frac{1}{m} \sum_{i=1}^m \exp \left[-\frac{(X-X_{\Omega i})'(X-X_{\Omega i})}{2\sigma^2} \right]$$

where $X_{\Omega i}$ is the i th training pattern from category Ω .

The probability density function (for each category of the training set) obtained with the increase of the smoothing parameter may in some cases overlap other functions along the concentration scale (Fig. 1). Extrapolation from the observed data values



Extrapolation effect of two unimodal probability density functions for categories Ω_1 and Ω_2 with increase of σ ($\sigma_1 < \sigma_2$)

FIG. 1

can take place in both directions along the concentration scale. If a relatively large smoothing parameter has to be used in order to extrapolate satisfactorily between data points, points on the ends of the different populations (e.g. Ω_1 and Ω_2 , Fig. 1) could be assigned to rather different categories.

APPENDIX 2

The non-linear mapping (NLM) algorithm.

The mathematical model is explained by Sammon (1969, p.401) as follows: suppose that we have n vectors in a p -space designated X_i , $i = 1, \dots, n$ and corresponding to these we define n vectors in a q -space ($q=2$) designated Y_i , $i = 1, \dots, n$. Set the distance¹ between the vectors X_i and X_j in the p -space be defined by $d_{ij}^* \equiv \text{dist} \mid X_i, X_j \mid$ and the distance between the corresponding vectors Y_i and Y_j in the q -space be defined by $d_{ij} \equiv \text{dist} \mid Y_i, Y_j \mid$. Let us now randomly² choose an initial q -space configuration for the Y vectors and denote the configuration as follows:

$$Y_1 = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1q} \end{bmatrix} \quad Y_2 = \begin{bmatrix} y_{21} \\ \vdots \\ y_{2q} \end{bmatrix} \quad \dots \quad Y_n = \begin{bmatrix} y_{n1} \\ \vdots \\ y_{nq} \end{bmatrix}$$

Next we compute all the q -space interpoint distances d_{ij} , which are then used to define an error E , which represents how well the present configuration of n points in the q -space fits the n points in the p -space, i.e.:

$$E = \frac{1}{\sum_{i < j}^n [d_{ij}^*]} \sum_{i < j}^n \frac{[d_{ij}^* - d_{ij}]^2}{d_{ij}^*}$$

1. Any distance measure could be used; however, if we have no a priori knowledge concerning the data, we would have no reason to prefer any metric over the Euclidean metric. Thus this algorithm uses the Euclidean distance measure.

2. For the purpose of this discussion it is convenient to think of starting configuration as being selected randomly; however, in practice the initial configuration for the vectors is found by projecting the p -dimensional data orthogonally onto a q -space spanned by the q -original coordinates with the largest variances.

Note that the error is a function of the $q \times n$ variables y_{ij} , $i = 1, \dots, n$ and $j = 1, \dots, q$. The next step in the NIM algorithm is to adjust the y_{ij} variables or equivalently change the q -space configuration so as to decrease the error. We use a steepest descent procedure to search for a possible local minimum of the error.

REFERENCES¹

- ARCHER, A.A., 1959, The distribution of non-ferrous ores in the Lower Palaeozoic rocks of North Wales: Symposium on the future of non-ferrous mining in Great Britain and Ireland, Inst. Min. Metall: Gt. Brit., p.259-276.
- BALL, D.F., 1960, The soils and land use of the district around Rhyl and Denbigh (Sheets 95 and 108): Mem. Soil Surv. Gt. Brit; Agricult. Res. Council, Harpenden, 139 p.
- _____ 1963, The soils and land use of district around Bangor and Beaumaris (Sheets 94 and 107): Mem. Soil Surv. Gt. Brit., Agricult. Res. Council, 182 p.
- BRIDGES, E.M., 1963, The soils and land use of the district north of Derby: Mem. Soil Surv. Gt. Brit., 112 p.
- CAMERON, E.M., SIDDELEY, G. and DURHAM, C.C., 1971, Distribution of Ore Elements in rocks for evaluating ore potential: Nickel, Copper, Cobalt and Sulphur: The Canadian Inst. of Min. and Metall., special vol. No.11, p.298-313.
- CAZALET, P., 1968, Investigation of the profiles characteristics, development and distribution of the soils of part of the southern Pennines: Ph.D. thesis, University of London.
- CLAYDEN, B., 1964, Soils of the Middle Teign Valley district of Devon: Bull. Soil Surv. Gt. Brit., No.1, Agricult. Res. Council, Harpenden, 111p.
- CLAYDEN and MANLEY, D.J.R., 1964, The soils of the Dartmoor granite: In Dartmoor Essay edited by I.G. Simmons, Devon Ass. Advnt. Sci. Lit and Art., p.117-140.
- DANSER, B.H., 1950, A theory of systematics: Bible Biotheor., v.4, p.113-180.
- DAVOINE, D., 1969, La distinction geochemique ortho-para des leptynites: Bull. Soc. Fr. Cristallogr., v.92, p.59-75.
- DEARMAN, W.R. and EL SHARKAWI, M.A.H., 1962, The shape of the mineral deposits in the Lower Culm Measures of north-west Dartmoor: Trans. Royal Geol. Soc. Cornwall, v.9, p.286-296.
- DINES, H.G., 1956, The metalliferous mining region of south-west England: Mem. Geol. Surv. Gt. Brit., 2 Vols., 795 p.

1) References with bibliographies of cluster and discriminant applications in other fields of geology are indicated thus *.

- EDMONDS, E.A., McKEOWN, M.C. and WILLIAMS, M., 1969, British regional geology, south-west England: Natural Environ. Res. Council, I.G.S., 130p.
- EDWARDS, W. and TROTTER, F.M., 1954, The Pennines and adjacent areas (based on previous edition by D.A. Wray): The Geol. Surv. Gt. Brit., 86p.
- FLETCHER, W.K., 1968, Geochemical reconnaissance in relation to copper deficiency in livestock in the Pennines and Devon: Ph.D. thesis, University of London.
- FORGY, E.W., 1963, Detecting "natural" clusters of individuals: Western Psychological Ass., Santa Monica, Calif., U.S.A.
- _____ 1965, Cluster analysis of multivariate data: Efficiency versus interpretability of classifications: AAS - Biometric Soc. (WNAR) Meetings, Riverside, Calif., U.S.A.
- FU, K.S., 1968, Sequential methods in pattern recognition and machine learning: Academic Press, London 227p.
- GILMOUR, J.S.L., 1937, A taxonomic problem: Nature, v.139, p.1040-1042.
- _____ 1940, Taxonomy and philosophy: In J.S. Huxley (ed.), The New Systematics, Clarendon Press, Oxford, p. 461-474.
- _____ 1951, The development of taxonomic theory since 1851: Nature, v.168, p.400-402.
- _____ 1961 b, Taxonomy: In A.N. McLeod and L.S. Cobley (eds.), Contemporary Botanical Thought, Oliver and Boyd, Edinburgh, and Quadrangle Books, Chicago, p.27-45.
- GOWER, J.C., 1967, A comparison of some methods of cluster analysis: Biometrics, v.23, p.623-637.
- * GRIFFITHS, J.C., 1966, Application of discriminant functions as a classification tool in geosciences; Kansas State Geol. Surv. Computer Contribution 7, p.48-52.
- * HARBAUGH, J.W. and MERRIAM, D.F., 1968, Computer applications in stratigraphic analysis: John Wiley and Sons, Inc., 282p.
- HORSNAIL, R.F., 1968, The significance of some regional geochemical patterns in North Wales and south-west England: Ph.D. thesis, University of London.
- HOWARTH, R.J., 1971 a, An empirical discriminant method applied to sedimentary-rock classification from major-element geochemistry: Math. Geology, v.3, No.1, p.51-60.
- _____ 1971 b, Empirical discriminant classification of regional stream sediment geochemistry in Devon and east Cornwall: Trans. Inst. Min. Metall. (Section B: Appl. Earth Sci.), v.80, p.B 142-149.

- _____ 1972, Empirical discriminant classification of regional stream sediment geochemistry in Devon and east Cornwall (Reply): Trans. Inst. Min. Metall. (Section B. Appl. Earth Sci.), v.81, p. B 115-119.
- * _____ 1973, Pattern recognition problem in applied geochemistry: Proceedings of the Fourth International Geochemical Exploration Symposium 1972, Inst. Min. and Metall., London, England.
- KHALEEL, J., 1969, The application of some data processing techniques in the interpretation of geochemical data: Ph.D. thesis, University of London.
- MACKNEY, D. and BURNHAM, C.P., 1964, The soils of the West Midlands: Mem. Soil Surv. Gt. Brit.
- NAGY, G., 1968, State of the art in pattern recognition: Proc. IEEE, v.56, p.836-862.
- NICHOL, I. and others, 1970a, Regional geochemical reconnaissance of the Derbyshire area: Rep. Inst. Geol. Sci., 37p.
- _____ 1970 b, Regional geochemical reconnaissance of the Denbighshire area: Rep. Inst. Geol. Sci., 40p.
- _____ 1971, Regional geochemical reconnaissance of part of Devon and North Cornwall: Rep. Inst. Geol. Sci., 26p.
- NICHOL, I. and LARSON, J.O., 1973, Comparison of multivariate techniques in the interpretation of geochemical data: Proceedings Fourth International Geochemical Exploration Symposium 1972, Inst. Min. and Metall. London, England.
- OBIAL, R.C., 1970, Cluster analysis as an aid in the interpretation of multi-element geochemical data: Trans. Inst. Min. Metall. (Section B: Applied Earth Sci.), v.79, p. B 175-180.
- OBIAL, R.C. and JAMES, C.H., 1973, The use of cluster analysis in geochemical prospecting with particular reference to southern Derbyshire: Proceedings Fourth Int. Geoch. Expl. Symp. 1972, Inst. Min. and Metall., London, England.
- PARKS, J.M., 1970, Fortran IV program for Q-mode cluster analysis on distance function with printed dendrogram: Kansas Geol. Surv. Comp. Centre, v.46, 32p.
- RHODES, J.M., 1969, The application of cluster and discriminant analysis in mapping granite intrusions: Geology Depart., Australian National University, Canberra, A.C.T., Lithos 2, p.223-237.

- SAMMON, J.W., 1969, A nonlinear mapping for data structure analysis: Trans IEEE, v. C-18, p.401-409.
- SHERLOCK, R.L., 1918, The geology and genesis of the Trefriw pyrites deposits: Quart. Jour. Geol. Soc., v.74, p.106-115.
- SIMPSON, G.G., 1961, Principles of animal taxonomy: Columbia University Press, New York, 247p.
- SMITH, B. and GEORGE, T.N., 1961, British regional geology, North Wales: Department of Scientific and Industrial Research, Geol. Surv. and Mus., 97p.
- SNEATH, P.H.A., 1958, Some aspects of Adansonian classification and of the taxonomic theory of correlated features: Ann. Microbiol. Enzimol, v.8, p.261-268.
- SOKAL, R.R. and MICHENER, C.D., 1958, A statistical method for evaluating systematics relationships: University of Kansas Sci. Bull., v.38, P.1409-1438.
- SOKAL, R.R. and SNEATH, D.H.A., 1963, Principles of Numerical Taxonomy: W.H. Freeman and Co., San Francisco, 359p.
- SPEATCH, D.F., 1967, Generation of polynomial discriminant functions for pattern recognition: IEEE Trans. Electr. Computers, v. 16, p.308-319.
- TAYLOR, D., 1968, The interpretation of lead and zinc stream sediment anomalies related to contamination and mineralization in the United Kingdom: Ph.D. thesis, University of London.
- U.S. GEOLOGICAL SURVEY, 1969, Environmental geochemistry, geochemical Survey of Missouri, Plans and progress for first six-month period (July-December, 1969), Denver, Colo., 49p.
- VALENTINE, J.W. and PEDDICORD, R.G., 1967, Evaluation of fossil assemblages by cluster analysis: Journ. Palaeontology, v.4(2), p.502-507.
- WIGNALL, T.K., 1969, Generalized Bayesian classification functions: K classes: Economic Geology Bull., v.64, p.571-574.
- WILKINSON, P., 1970, Automatic data handling in geochemistry and allied fields: Systematics Assoc. Special Volume No.3, Data processing in Biology and Geology, ed. J.L. Cutbill, p.205-234.