Conference on

Machine Processing of

Remotely Sensed Data

October 16 - 18, 1973

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

NUMERICAL CLASSIFICATION PROCEDURES IN FLUVIAL GEOMORPHOLOGY[1]

Glen C. Gustafson[2]

University of Munich
Geography Department
Munich, Germany

I. ABSTRACT

A number of up-to-date numerical classification techniques are described. These include the orthogonal and oblique factor analysis methods, and the unweighted pair-group cluster analysis procedure. The techniques are applied to morphometric data from 159 small drainage basins from two geographical regions. Transformation techniques to achieve the normal distribution with respect to symmetry are applied.

II. INTRODUCTION

Numerical classification procedures have a long history of applications in scientific research. For example, factor analysis and principal components analysis have been used in psychological research for over forty years. Similarly, various methods of cluster analysis have found many applications in biology since the 1950's. However, the broader use of these methods in geography and elsewhere has been hindered by disagreement as to specific methods, and also by criticism which is often somewhat out of date. Therefore, one important task at this stage is the identification of appropriate areas of contribution for these methods which have been developed and applied chiefly outside of the earth sciences until recently. A second task is the empirical testing of up-to-date mathematical alternatives in an effort to recognize which methods are optimum for which types of data.

III. DATA ACQUISITION

In the present study, fifteen often-used morphometric variables were determined for a number of small drainage basins from two geologically similar, but widely separated regions. The variables attempt to describe in diverse ways the linear, areal, shape and relief aspects of the basins. The choice of variables was made on the basis of broad application in the geomorphological literature and facility of measurement from orthophotos and topographic maps. For each drainage basin the following parameters were determined: surface area (AREA), total stream length (STRLEN), number of streams (STRNUM), basin perimeter (PERIM), length of primary drainage channel traced to divide (PRIMCH), straight line basin length (BASLEN), slope of primary drainage channel (SLOPE), basin length to basin width ratio (SHAPE 1), ratio of square of length of primary drainage channel to basin area (SHAPE 2), drainage density (DENSTY), channel frequency (FREQCY), relative relief within basin (RELIEF), median elevation of basin (ELEV), ruggedness number (RUGGED), and relative drainage density (RELDEN). Definitions of these

[1]This work was supported by the German Research Association under Project No. Gi 9/17.

[2]Present address: Department of Geology and Geography, Herbert Lehman College, New York City University

parameters are to be found in standard geomorphological references such as Maxwell (1967), Melton (1958b) and Strahler (1968).

The first test area, located in the Far East, was studied by means of orthophotos and drop-line charts at scale 1:12,500. These materials are produced photogrammetrically by means of rectification and enlargement of narrow parallel strips from the aerial negative. Basic information on this new type of photomap and its by-products is available in Blachut (1972), Institut Géographique National (1971) and The Canadian Surveyor (1967).

The second test area, on the southern edge of the Black Forest of Germany, was investigated using the orohydrographic version of the national topographic map (1:25,000), supplimented by orthophotos of scale 1:10,000. From the two test areas a total of 159 drainage basins of second, third and fourth orders were selected, with approximately half in each area. "Basin order" is used here in the sense defined by Strahler (1952) where the first-order basins enclose the smallest unbranched tributaries. Criteria for selection of basins included basin order, wide distribution among the available geological units, and lack of obvious disturbances in the drainage network caused either by mankind or tectonic processes. In this way an input matrix was built up for each of the two study areas. For the Far Eastern and Central European study areas these matrices are 15 x 79 and 15 x 80 respectively.

IV. FACTOR ANALYSIS METHODS EMPLOYED

Despite its apparent objectivity, a factor analytic study requires a number of operational decisions which will effect the results of the analysis. Most basic among these is the choice of the so-called "closed" or "open" model. The former, normally known as principal components analysis, employs unities along the diagonal of the correlation matrix. This expediency has the effect of assuming that the reasons for the dispersion of the measurements are all understood from the limited sample under study. The resulting factors will be speciously inflated, which is considered undesirable (Cattel, 1965). The alternative to the closed model simply leaves a portion of the variance of each variable "open" to a later resolution when a broader sample of input data may be available. In this approach, the diagonal of the correlation matrix is occupied by estimates of the common variance (i.e., shared variance with other variables) of each variable, known in this context as "communality".

Unfortunately, communality estimation has remained a difficult and disputed concept. The range of possible values is, however, generally accepted to extend from an upper limit of unity to a lower limit given by the square of a variable's multiple correlation with all other variables (Cattel, 1965). This latter quantity, know as the Squared Multiple Correlation (S.M.C.), is often recommended as the best available method for communality estimation (Harman, 1968, Kaiser, 1960, Steiner, 1965). Due to these reasons and to its local availability, the S.M.C. method has been used in the present study to estimate communalities. Further detailed description of the alternatives is to be found in Harman, Chapter 5 (1968). It should also be mentioned, that with increasing order of the matrix, that is, number of variables, the diagonal elements have decreasing impact on the results of the factor analysis (Hope, 1968).

The next level of decision in designing the factor analytic model involves the choice of rotation scheme. The unrotated factor matrix often shows loadings on each factor which are rather widely distributed among the variables. The effect of this is to make interpretation of the matrix, and subsequent description of the factors more difficult. Rotation is a methematical method of shifting the frame of reference by means of coordinate transformations. During rotation, the swarm or cluster of points plotted in n-dimensional space remains fixed while the axes rotate to a new position such that the numerical values obtained for the loadings are more distinctly differentiated. The new axes yield new coordinates and loadings, making interpretation clearer.

The amount of rotation is, in general, governed by some concept of "simple structure". This criteria was originally defined by Thurstone. Stated simply, it is a set of rules which seek to maximize loadings of the essential variables on each factor, and to minimize the remaining loadings (Fruchter, 1954). Thus, ideally, only a few variables will have high loadings on any one factor, with the

remaining loadings for that factor being essentially estimates of zero.

Two general types of rotations are available. In the traditional "orthogonal rotation schemes the assumption is made that the factors are not correlated in nature. This is expressed by the fact that all axes remain perpendicular despite rotations around each axis. In some studies, this constraint has been found too restricting to fit nature (Cattel, 1965). This represents one of the most important subjective decisions the factor analyst must reach concerning his particular problem.

The alternative to the first approach involves making the assumption that the factors operating in nature may be correlated. Here the reference axes must no longer be orthogonal. The method is therefore called oblique rotation. The degree of obliqueness may be controlled if desired, and subsequently checked in the factor correlation matrix. This is a small summary matrix giving the correlation coefficient between each factor and every other factor. If, however, the input data involves factors which are truly uncorrelated, then uncorrelated factors will be produced (Harbaugh and Merriam, 1968). The primary davantages of oblique factors are summarized in Fruchter (1954, p. 196) and Cattel (1965, p. 405). In the present study, both orthogonal and oblique methods have been utilized and compared as far as possible. These methods are respectively the "Varimax" method of Kaiser (1958) and the 'Direct Biquartimin" or "Simple Loadings" procedure of Jennrich and Sampson (1966).

Having selected the factor analytic model, with an appropriate communality estimation procedure, and either orthogonal or oblique axes, finally, the analyst must have a criteria to decide when to stop factoring. Operationally this decision is often considered together with the communality issue, since one affects the other. The purely mathematical process of extracting factors can continue until as many factors are produced as there were original variables. Since the process "begins" with the most important factor and continues to more and more minor influences, at some point, the so-called error factors will begin appearing. These may be recognizable as mixtures or slightly altered versions of previous factors. Unfortunately ther is no sampling theory which will allow one to locate this threshold with precision.

The most common method to date has been simply to include only factors having an eigenvalue of 1.0 or more. This criteria is henceforth referred to as the eigenvalue threshold. The eigenvalue, loosely defined, is an expression of the relative information content of the associated factor. It is calculated from the correlation matrix. Geometrically it can be interpreted as the length of any particular axis of the point swarm. For the first factor, the eigenvalue represents the length of the longest axis, and so on for the remaining factors. In evaluating this method of finding the acceptance threshold for factors, it is important to note that any factor with an eigenvalue in excess of 1.0 supplies more information than an unmodified variable alone (Carey, 1969). This criteria, although not ideal or final in any way, has been found by several workers to be the best interim approach available (Harman, 1968, Cattel and Dickman, 1962, Kaiser, 1960). It should be emphasized that this cut-off point has been found empirically useful but is, in effect, quite arbitrary. Therefore various methods of shifting it slightly have been proposed where this is scientifically meaningful (King, 1969). Cattel, in a longer discussion on this topic, emphasizes that taking too many factors is far less serious than taking too few (1958).

In this connection it should be mentioned that several attempts have been made at designing statistical significance tests for the number of factors (e.g., Bartlet, 1950, or Lawley and Maxwell, 1963). These tests, however are under continuing debate by statisticians and others. Kaiser summs up the issue thusly: the significance tests produce "statistically correct but scientifically issue-confusing" factors and require enormous additional calculations (1960). Harbaugh and Merriam emphasize that test of significance for the original correlation coefficients are not either necessary or desirable since factor analysis merges the influences of many minor correlations (1968).

In conclusion, in the design of a factor analytic model, the researcher must pass through several levels of decision which will influence the outcome of the analysis. He is guided in these decisions by the computer program alternatives available to him, the experience with these alternatives as reported in the literature, and his own knowledge of the relationships being studied. The specific

techniques employed in the present study are indicated in Fig. 1 along with other options available locally. The program employed is the "BMD-X-72"package of the UCLA Medical Facility.

V. CLUSTER ANALYSIS METHODS EMPLOYED

In the present study, classification of variables and classification of objects have been attempted. These represent respectively the R- and Q-techniques of analysis and are illustrated in the left-hand plane of Fig. 2. The cluster techniques are, in general, much simpler to carry out and interpret than those of factor analysis. In essential terms, cluster analysis is a method of searching a large symmetrical correlation matrix for the highest relationships between units (i.e., variables or objects) and then listing the associated units as they are found. In this sense, it is not too different from the conventional, by-hand interpretation of a correlation or similarity matrix. By one mathematical method or another, the cluster analyst seeks to discover what internal structure, if any, exists within a given set of data. Some of the specific techniques are summarized in an introductory paper by Sokal (1966)and more thoroughly, in textbook form by Sokal and Sneath (1963). Additional introductory material of much value is to be found in Harbough and Merriam (1968).

Although there are many clustering methods, the most common approach, and that used in the present study, involves building up groups agglomeratively, starting from small nuclei. The specific technique employed here is one of the average linkeage methods, namely the unweighted pair-group method (UWPG) of Sokal and Sneath (1963). Detailed information on the various alternative classification methods is contained in Spence and Taylor (1970), Johnston (1968), and Sokal and Rohlf (1962).

As with factor analysis, the researcher must make several operational decisions which will affect, to some degree, the results. The most important of these is the choice of the clustering method itself. After this is the choice of a similarity measure. There are three important types often used in cluster analysis. The most common are perhaps the correlation coefficients. A second type of similarity coefficient is the distance measure, which expresses association as the distance between two sample points plotted in n-dimensional space. Angular measures represent the third type. Here, similarity is expressed as the angle between two standardized vectors representing two units to be classified. Spence and Taylor (1970) provide a broad discussion of these and other coefficients. In the present study, the Pearson correlation coefficient has been used for all factor analyses, and also the R-mode cluster analyses. The Q-mode cluster analyses have employed the Cosine Theta Coefficient of Imbrie and Purdy (1962). This is an angular measure, calculated from standardized data.

The final product of a cluster analysis is normally a cluster diagram. This is ideally a simple tree-like drawing which shows the internal structure of the data. The smallest branches are the individual "operational taxonomic units" (OTU's) which join at various similarity levels. In this fashion, larger and larger subgroups are built up which eventually encompass all OTU's. In the past, several types of such diagrams have been applied. The newest of these, and the clearest to interpret is the "Dendrograph" (McCammon, 1968). Along one axis, the abscissa, is the similarity level. The ordinate, however is also scaled to show the relationship between individual OTU's. Thus, both the within-group and the between-group similarities are readily apparent.

VI. PREPROCESSING OF INPUT DATA

Factor and cluster analysis, using the Pearson correlation coefficient, require no representative sample and no assumptions regarding the frequency distribution of the variables under study (Parks, 1966, Thurston, 1945). However, it has been common practice to submit only normally distributed data (Kendall, 1965, Fruchter, 1954). Reasons for this include, the possible intermediate use of significance tests, and for greater consistency from one study to anaother. Normalizing transformations are therefore often used to achieve this condition. Various methods.for assessing the normality of a distribution are available. Among these are: arithmetic probability paper, skew and kurtosis evaluation, Chi-Square tests, and Kolmogorov-Smirnov tests. In the present work, skew calculations have

been used for this purpose. This ratio of moments of the distribution describes succinctly and precisely the type and degree of departure from normality. Skew is defined here as:

$$B_1 = \frac{m_3}{\sqrt{m_2^3}}$$

where m_2 = second moment of the distribution

m_3 = third moment of the distribution.

By this definition, a logarithmic distribution, for example, is said to have positive skew. For normally distributed data, skew is equal to zero.

Since moderate departure from normality is not crucial, many geographic ap-aplications in the past have applied no correcting transformations. In other studies, only the most skewed variables have been modified by the use, in most cases, of logarithmic transformations. In the present work it was found that such logarithmic transformations often tend to overcompensate. That is, a variable X with high positive skew is transformed into a variable log X with moderate to high negative skew. The potential use of several other transforming functions to improve this situation is referred to by Mather (1968) and also by Miller and Kahn (1962). In addition, Dixon provides programming instructions for around thirty such transformations (1968). This diversified transformation approach has been applied to the present data. The list of transformations tested on each variable for both sets of data is given in Table 1. The number of variables for which each transformation was optimum with respect to skew is also indicated. "Optimum" in this sense is defined as that transformation which yields the lowest skew value for each variable individually. The result of these procedures is the production of a new data matrix in which the distribution of each variable is transformed according to its needs and very closely approximates the normal distribution with respect to symmetry.

If the data is to be used as input for subsequent analysis based on distance coefficients or angular measures of similarity, some form of standardization is also required. This is not a necessity with the Pearson correlation coefficient, used in most factor analyses, since its calculation involves division by the standard deviation, which is a form of standardization (Mather, 1968). By far the most common standardization procedure is to convert all measurements to standard units. This involves subtracting the mean of a variable from each of the observed values and dividing this by the standard deviation. This results in all variables having a mean of 0.0 and a standard deviation of 1.0. All objects in the sample can now be represented as vectors of similar length, despite the units of measurement used for each. Since applications of many of these correcting procedures are not common in the geographic literature, an effort was made to compare their effects. Factor analyses were carried out based on: 1) raw data, no transformations, 2) log-normalization in extreme cases, and 3) optimum transformation of each variable for normal skew characteristics, followed by standardization.

VII. FACTOR AND CLUSTER ANALYSIS RESULTS

The problem of defining the primary dimensions of third-order drainage basins by means of statistical analysis has been attempted in several previous studies. In an effort to facilitate comparison of results, the present study was first carried through for the third-order basins alone. These results are summarized in Table 3. The input variables, however, differ somewhat between the studies. This is, of course, a subjective decision based on experience and the purposes of the study. It should be mentioned that deleting all variables relating to a particular prime factor simply has the effect of removing the factor altogether, the less important factors shifting upwards to replace the lost dimension.

Factors identified for the 53 third-order basins in the Asian study area included basin size, basin shape, relief characteristics, dissection intensity, and relative dissection intensity. Those from the 45 basins in the European study area were essentially the same but with the relief and shape factors reversed in order. This means simply that for the European basins more of the total variance is concentrated in the relief measurements than in the shape measurements. The stream number variable has been clearly absorbed into the basin size factor, as

might be expected in a homogeneous region. There is a moderate tendency for relief aspects to be loaded on the basin size factor also. However, in general these loadings are around 0.5 or less, and so often have little influence on a factor which is already heavily loaded. The existence of dissection characteristics as fully independent factors may also relate at least partly to the rather homogeneous nature of the two test regions. The reinterpretation of the Melton data (Mather and Doornkamp, 1970) from arid basins in the southwest United States, for example, also yielded an independent dissection factor. However, in the more heterogeneous data from Mather and Doornkamp (1970), it was absorbed into the basin size factor.

For purposes of verification, the above data was also subjected to R-mode cluster analysis. For the third-order basin data, as shown in the upper half of Fig. 3, three primary clusters were produced: basin size and relief, basin shape, and dissection intensity. Variables joining these clusters independently included elevation, slope, and relative drainage density. Ruggedness, which is drainage density times relief, clusters with density for the Wies area in Central Europe, and with relief for the Asian area. This is typical for the results of cluster studies. The simplification forces an OTU into the single cluster where it is most similar, rather than splitting the effects as factor analysis does. The clusters could easily be given descriptive names to facilitate comparison with the factored results shown in Table 3.

When the entire data for each of the two study areas is included in the analysis, the factors change moderately. The most noticeable change, perhaps, is an increase in the relative importance of the dissection variables. As the fourth factor in the earlier analyses, they carried approximately 10% of the total variance. In the present case, as factor two, they contribute around 17% of the variance. One can conclude from this that with increasing heterogeneity of drainage basins, the diagnostic importance of dissection intensity also grows. Drainage density is thought to reflect very sensitively the overall balance of physiographic processes in the landscape. Only one other significant change in the loadings is apparent. The slope and relief characteristics, in addition to loading on a unique factor, also load moderately on the first factor. This is again, likely a result of the increasing heterogeneity of the full data for each region. Therefore, large fourth-order basins have greater overall size, including relief, than small second-order basins.

Based on the most sophisticated set of input data, which has undergone optimum normalization and standardization, and by means of the Varimax rotation scheme, five significant factors can be defined (see Table 4). These agree in general for both study areas and can be interpreted as follows: basin size (and partially relief), dissection intensity, basin shape, relief characteristics, and dissection completeness. The use of the Simple Loadings oblique rotation on this data produces factor loadings which are slightly easier to interpret. In the European area, the factors themselves remain unchanged, despite minor changes in the loadings. For the Asian basins, dissection intensity reverts to the fourth factor, being replaced by relief. The correlations between the factors themselves, shown in the factor correlation matrix, are generally 0.25 or less. For comparison purposes, the corresponding results for the cluster analyses are given in the bottom half of Fig. 3.

An important use of the present findings is in data reduction. First, in subsequent geomorphological studies in this type of landscape, it may be possible to greatly reduce the number of terrain characteristics measured, without important loss of information. This possibility could be checked by means of a random sample of basins from the area to be studied. In the present study, based on fifteen input characteristics, five important factors have been shown to exist in nature. The substitution of, for example, one heavily loaded variable for each factor, represents one type of data reduction. Criteria for the selection of such a variable may relate to the factor loading, or ease of measurement or other considerations. The second type of data reduction is the use of the factor scores matrix in the place of the original measurements. The relationship between each case of the input data and each factor is shown by these factor scores. For example, the very large basins and the very small basins influence the size factor heavily. Therefore, their loadings in the factor scores matrix are large on factor one.

It has been found in the present study that this artificial data matrix for five factors may be substituted for the original data matrix with practically no

loss of information concerning each individual basin. This type of data reduction is useful if subsequent statistical work is to be carried out. Here, for example, the size of the matrix has been reduced by two-thirds. If an orthogonal rotation scheme has been used to produce the factor scores, they have the additional advantage, for some purposes, of being totally uncorrelated.

One possible use for such a reduced data matrix is in Q-mode cluster analysis for regionalization purposes. As an illustration and essentially a by-product of the present study, such an analysis has also been carried out, although the data is not ideal for this purpose. Nevertheless, using standardized data, the basins from the two regions seperate from each other completely into two clusters at the 50-phenon level (i.e., level of overall similarity). Figure 4 is a highly reduced reproduction of the associated dendrograph. Interpretation at the 65-phenon level indicates three clusters in the European area (upper half of Fig. 4), and three in the Asian area (lower half of Fig. 4). Within each of the two regions the clusters, however, are not highly differentiated from each other. This may be taken as further evidence of the internal homogeneity of each of the two test areas. Thus, meaningful geographical interpretation within each of the two test areas proved difficult. In addition, unfortunately, little detailed physiographic information is available on the two areas. Several conclusions could nevertheless be made. First, basins from the two regions could be reliably seperated from each otherer despite the particular operational decisions, such as number of variables considered, similarity coefficient used, and so on. Second, the use of the factor scores matrix yields a nearly identical classification to that produced by the use of all original variables associated with those factor scores. Finally, a large number of the basins within each cluster form contiguous units in the landscape which have a moderate north/south orientation tendency.

VIII. CONCLUSION

Classifications with two different goals have been undertaken in the present study. The classification of variables (i.e., measurements) resulted in the identification of the primary geometric dimensions of two groups of low order drainage basins in crystalline rocks. These dimensions are: basin size, dissection intensity, basin shape, basin relief, and dissection completeness. The relationships between these primary dimensions are best abstracted in the factor correlation matrix. The factor analysis of variables also made possible an important data reduction. One type is the substitution of the factor scores matrix for the complete input matrix in subsequent work. Another type is the use of only the most important variable, or variables, for each factor.

The effort to classify basins was not entirely successful. Basins from the two test areas could be seperated from eachother with reliability. However, the internal grouping within each study area proved difficult to judge due to lack of detailed terrain information. Thus, if a sample contains essentially different physiographic subunits, these differences should be reflected in some of the morphometric characteristics, and thus allow discrimination by cluster analysis.

The factor analytic model, due to its sophistication and flexibility, is able to incisively identify underlying influences in a complex set of multivariate data. The cluster analytic model, a much simpler procedure, is a useful compliment to the former. It may be used to verify and somewhat generalize the factored results. Its greatest benefit lies in its simplicity. In the Q-mode, based on a large number of input characteristics for each of a group of objects, it unequivically allocates each object to the one group where it is most similar. Depending on the clustering method chosen, this may however vary slightly. Testing procedures using discriminant functions have, in the past, been used for mathematically evaluating and refining such classifications.

REFERENCES

Bartlett, M.S. (1950) "Tests of Significance in Factor Analysis". British Journal of Statistical Psychology. Vol. 3, pp. 77-85.

Berry, B.J.L. (1971) "Comparative Factorial Ecology". Suppliment Issue of Economic Geography. Vol. 47, No. 2.

Blachut, T.J. (1972) "Orthophoto Technique: Basic Instruments and Methods". World Cartography. Vol. 12, pp. 80-92.

Canadian Surveyor (1967) International Symposium on Photo Maps and Orthophoto Maps. Ottawa.

Carey, G.W. (1969) "Principle Componant Factor Analysis and its Application to Geography". in Quantitative Methods in Geography: A Symposium. Washington: American Geographical Society, pp. 6-20.

Cattell, R.B. (1965) "Factor Analysis: An Introduction to Essentials". Biometrics Vol. 21, pp. 190-215 and 405-435.

Cattell, R.B. (1958) "Extracting the Correct Number of Factors in Factor Analysis" Educational and Psychological Measurement. Vol. 18, pp. 791-838.

Cattell, R.B. and K. Dickman (1962) "A Dynamic Model of Physical Influences Demonstrating the Necessity of Oblique Simple Structure". Psychological Bulletin. Vol. 59, pp. 389-400.

Dixon, W.J. (1968) BMD Biomedical Computer Programs. Berkeley: Univ. Calif.Press.

Fruchter, B. (1954) Introduction to Factor Analysis. Princeton: van Nostrand.

Gustafson, G.C. (1973) Quantitative Investigation of the Morphology of Drainage Basins using Orthophotography."Münchener Geographische Abhandlungen". Vol. 11. Selbstverlag des Geographischen Institutes der Universität München.

Harbough, J.W. and D.F. Merriam (1968) Computer Applications in Stratigraphic Analysis. New York: John Wiley and Sons.

Harman, H.H. (1968) Modern Factor Analysis. Chicago: Univ. Chicago Press.

Hope, K. (1968) Methods of Multivariate Analysis. London: London Univ. Press.

Imbrie, J. and E.G. Purdy (1962) "Classification of Modern Bahamian Sediments". Amer. Assoc. of Petroleum Geologists Memorandum. Vol.1, pp. 253-272.

Institut Géographique National (1971) Proceedings of the National Symposium on Orthophotographs and Orthophotomaps. Paris.

Jennrich, R.I. and P.F. Sampson (1966) "Rotation for Simple Loadings". Psychometrika. Vol. 31, pp. 313-323.

Johnston, R.J. (1968) "Choice in Classification: the Subjectivity of Objective Methods". Annals, Assoc. Am. Geographers, Vol. 58, pp. 575-589.

Kaiser, H.F. (1958) "The Varimax Criterion for Analytic Rotation in Factor Analysis". Psychometrika. Vol. 23, pp. 187-200.

Kaiser, H.F. (1960) "The Application of Electronic Computers to Factor Analysis". Educational and Psychological Measurement. Vol. 20, pp. 141-151.

Kendall, M.G. (1965) A Course in Multivariate Analysis. London: Charles Griffin

King, L.J. (1969) Statistical Analaysis in Geography. New Jersey: Prentice-Hall.

Krumbein, W.C. and F.A. Graybill (1965) Introduction to Statistical Models in Geology. New York: Mcgraw-Hill.

Lawley, D.N. and A.E. Maxwell (1963) Factor Analysis as a Statistical Method. London: Butterworth and Co.

Mather, P.M. (1968) "Numerical Classification in Geomorphology". in The Use of Computers in Geomorphological Research, Edited by J.C. Doornkamp. British Geomorphological Research Group.

Mather, P.M. and J.C. Doornkamp (1970) "Multivariate Analysis in Geography, with Particular Reference to Drainage Basin Morphometry". Transactions of the Institute of British Geographers. No. 51, pp. 163-187.

Maxwell, J.C. (1967) "Quantitative Geomorphology of some Mountain Chaparral Watersheds in Southern California", in Quantitative Geography edited by W.L. Garrison and D.F. Marble, pp. 108-226.

McCammon, R.B. (1968) "The Dendrograph: A New Tool for Correlation". Bulletin Geological Soc. of America, Vol. 79, pp. 1663-1670.

Melton, M.A. (1957) "An Analysis of the Relations among Elements of Climate, Surface Properties, and Geomorphology". Ph.D. Dissertation, Columbia Univ.

Melton, M.A. (1958) "Correlation Structure of Morphometric Properties of Drainage Systems and their Controlling Agents". Journal of Geology, Vol.66, pp.442-460.

Miller, R.L. and J.S. Kahn (1962) Statistical Analysis in the Geological Sciences. New York: John Wiley and Sons.

Muehrcke, P. (1972) Thematic Cartography. Washington: Assoc. Am. Geographers.

Parks, J.M. (1966) "Cluster Analysis Applied to Multivariate Geological Problems". Journal of Geology. Vol. 74, pp.703-715.

Sokal, R.R. (1966) "Numerical Taxonomy". Scientific American. Vol. 215, pp.106-16.

Sokal, R.R. and F.J. Rohlf (1962) "The Comparison of Dendrograms by Objective Means". Taxonomy. Vol. 11, pp. 33-40.

Sokal, R.R. and P.H.A. Sneath (1963) Principles of Numerical Taxonomy. San Francisco: W.H. Freeman and Co.

Spence, N.A. and P.J. Taylor (1970) "Quantitative Methods in Regional Taxonomy". in Progress in Geography, Vol. II. London: Edward Arnold.

Steiner, D. (1965) "die Faktorenanalyse—ein Modernes Statistisches Hilfsmittel des Geographen für die Objektive Raumgliederung und Typenbildung". Geographica Helvetia. Vol. 20, pp. 20-34.

Strahler, A.N. (1968) "Quantitative Geomorphology" in Encyclopedia of Geomorphology, Edited by R.W. Fairbridge. New York: Reinhold.

Strahler, A.N. (1952) " Hypsometric Analysis of Erosional Topography". Bulletin of the Geological Society of America. Vol. 63, pp. 1117-1142.

Thurstone, L.L. (1945) "The Effects of Selection in Factor Analysis". Psychometrika. Vol. 10, pp. 165-198.

Überla, K. (1968) Faktorenanalyse. Berlin: Springer Verlag.

Table 1. Transformations tested for each variable and usage of each to achieve optimum normality with respect to symmetry.

Transformation	Utilization for each:	
	2nd, 3rd and 4th order basins	3rd order basins alone
1. $y = x$	2	4
2. $y = \sqrt{x}$	3	2
3. $y = 1/x$	0	1
4. $y = 1/x^2$	0	0
5. $y = \log_e x$	11	11
6. $y = x^2$	0	0
7. $y = x^{1.5}$	0	0
8. $y = \sqrt{x} + \sqrt{x+1}$	5	5
9. $y = \sqrt[3]{x}$	9	7
10. $y = 1/x^3$	0	0

Table 2. Example factor matrix with associated data; corresponds with Line 3 in Table 4.

ROTATED FACTOR MATRIX (simple loadings) S.M.C. Communality Estimation
INPUT DATA: Standardized variables with optimum normality, Asia (n=79)

Variable	Factor 1	2	3	4	5
1 AREA	-0.96079	0.01516	0.16533	0.19145	-0.01725
2 STRLEN	-1.02453	0.02421	0.17335	-0.01329	-0.06588
3 STRNUM	-1.01831	0.02745	0.18781	-0.21174	0.17495
4 PERIM	-0.93618	0.04412	-0.04036	0.16664	-0.01731
5 PRIMCH	-0.84278	0.02451	-0.31293	0.12401	-0.04021
6 BASLEN	-0.80081	0.00692	-0.33611	0.15016	-0.03119
7 SLOPE	0.47820	-0.90045	0.20608	0.02875	-0.06812
8 SHAPE 1	0.12448	0.00055	-0.91719	0.04775	-0.00440
9 SHAPE 2	0.04861	0.03803	-1.00222	-0.09508	-0.01083
10 DENSTY	0.04948	0.03547	-0.02177	-0.98980	-0.21311
11 FREQCY	0.03185	0.01333	-0.03183	-0.84160	0.42637
12 RELIEF	-0.56130	-0.58017	-0.20721	0.15403	-0.06886
13 ELEV	-0.19712	-0.40119	-0.13161	0.21583	0.03315
14 RUGGED	-0.59330	-0.62147	-0.23959	-0.24537	-0.16793
15 RELDEN	-0.01644	-0.01480	-0.00972	0.03796	1.00269
Identification of important variables for each factor:	STRLEN STRNUM AREA PERIM PRIMCH BASLEN (RUGGED) (RELIEF) (SLOPE)	SLOPE RUGGED RELIEF	SHAPE 2 SHAPE 1	DENSTY FREQCY	RELDEN
Eigenvalue for each factor:	7.22	2.58	1.86	1.42	0.96
Cumulative proportion of total variance:	0.48	0.65	0.78	0.87	0.94

Table 3. Summary of factor analysis results for third-order basins from various studies.

Input Data					Primary Dimensions or Factors and Associated Eigenvalues				
Source	Method	Location	Objects	Variables	F_1	F_2	F_3	F_4	F_5
1971 Doornkamp and King; Melton data, 1957	Factor Analysis: Matrix Diagonal	Southwest United States	156 3rd order basins	12	size	number of streams	dissection intensity	basin relief	-
1970 Mather and Doornkamp	Factor Analysis: Varimax	Southern Uganda	130 3rd order basins	18	size and dissection (8.67)	number of streams (3.85)	Stream length ratio (1.70)	basin relief (1.47)	bifurcation ratio (0.84)
1973 Gustafson	Factor Analysis: Varimax	Far East	53 3rd order basins	15	size (7.37)	shape (2.25)	relief (1.82)	dissection intensity (1.46)	relative dissection intensity (1.09)
1973 Gustafson	Factor Analysis: Varimax	Black Forest, Germany	45 3rd order basins	15	size (7.18)	relief (3.21)	shape (1.74)	dissection intensity (1.10)	relative dissection intensity (0.68)

Table 4. Factor analysis results for all basins under various operational conditions.

Input Data			Type of Rotation	Factors and Associated Eigenvalues				
Area	Objects	Transformations		F_1	F_2	F_3	F_4	F_5
Far East	79	none	Simple Loadings	size (relief) (7.06)	shape (2.58)	dissection intensity (1.86)	relief (1.34)	relative dissection intensity (0.95)
Far East	79	Log in extreme cases	Varimax	size (relief) (7.24)	dissection intensity (2.57)	shape (1.87)	relief (1.39)	rel. dissec. intensity (0.95)
Far East	79	optimum (see Text)	Simple Loadings	size (relief) (7.22)	relief (2.58)	shape (1.86)	dissection intensity (1.42)	rel. dissec. intensity (0.96)
Central Europe	80	none	Varimax	size, slope (6.84)	dissection intensity (2.54)	shape (1.77)	relief (1.33)	rel. dissec. intensity (0.88)
Central Europe	80	optimum (see Text)	Simple Loadings	size, slope (6.97)	dissection intensity (2.57)	shape (1.84)	relief (1.44)	rel. dissec. intensity (0.93)

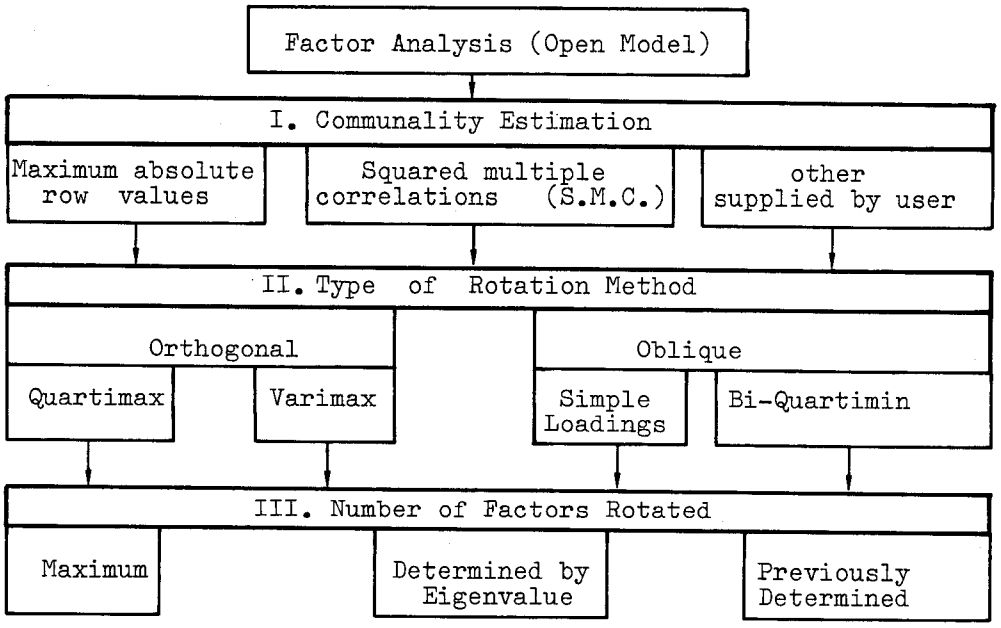

Figure 1. Schematic diagram of the levels of decision in designing the factor analytic model, and some alternatives.

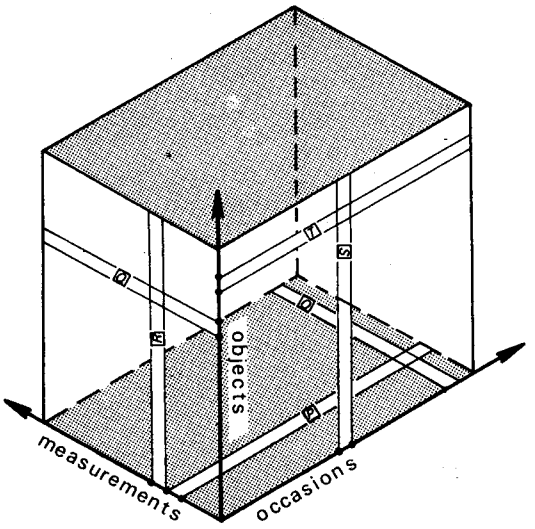

Figure 2. The covariation chart showing modes of analysis (modified from Cattel, 1965).

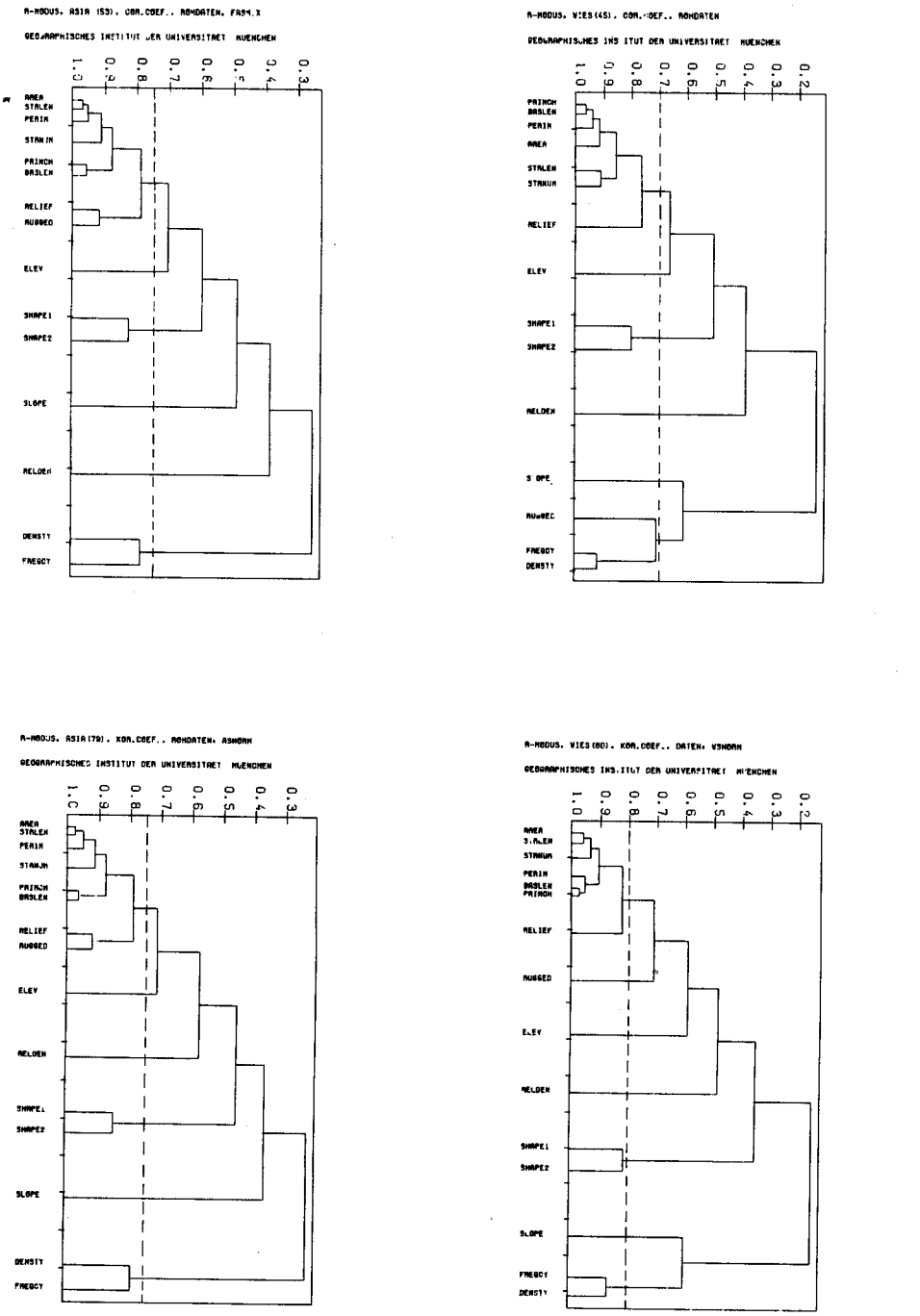

Figure 3. Dendrographs for four different sets of data in the present study; Identification is as follows: left side, Asian study area, right side, European study area; upper half, third order basins, lower half, second, third and fourth order basins; interpretation level, discussed in text, is indicated with a dashed line.

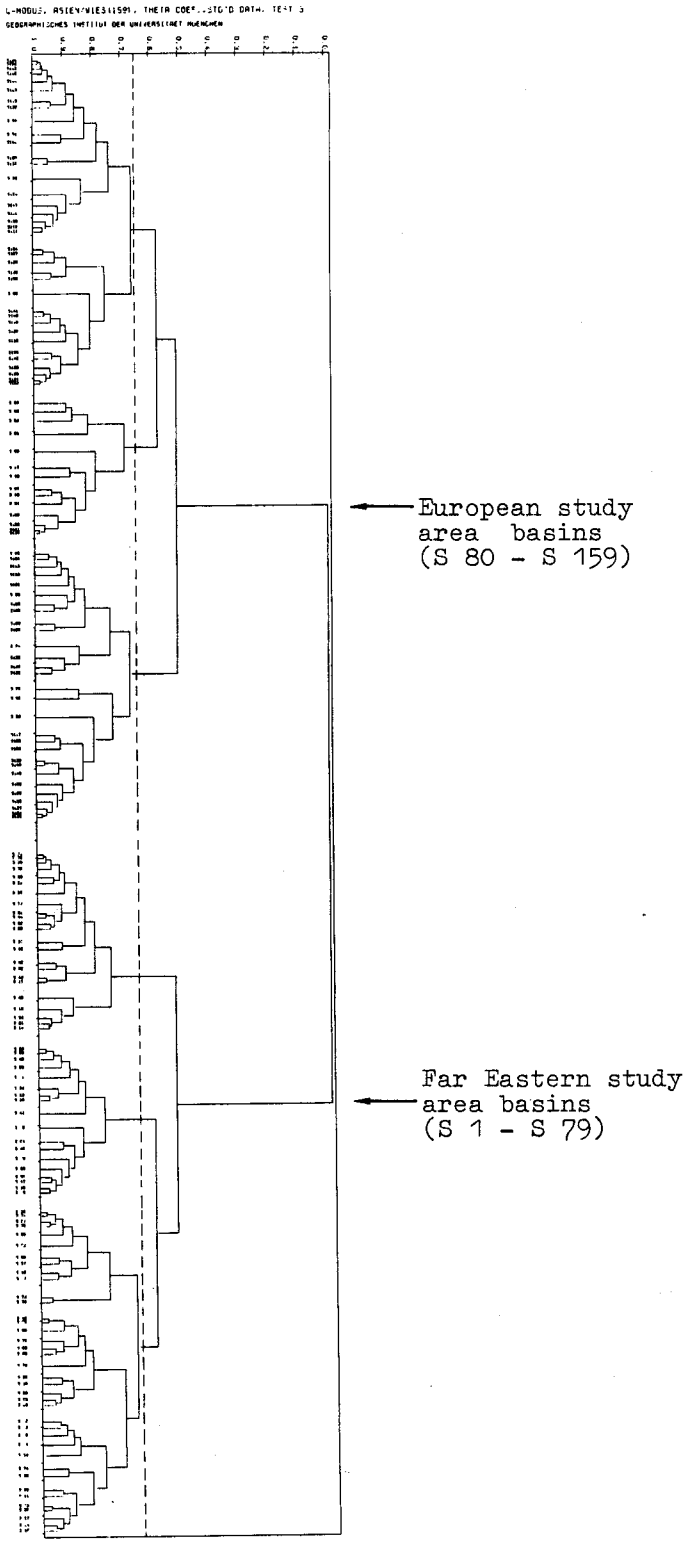

Figure 4. Dendrograph showing Q-mode classification of 159 drainage basins from two test areas; 65-Phenon interpretation yields three subgroups in each region.

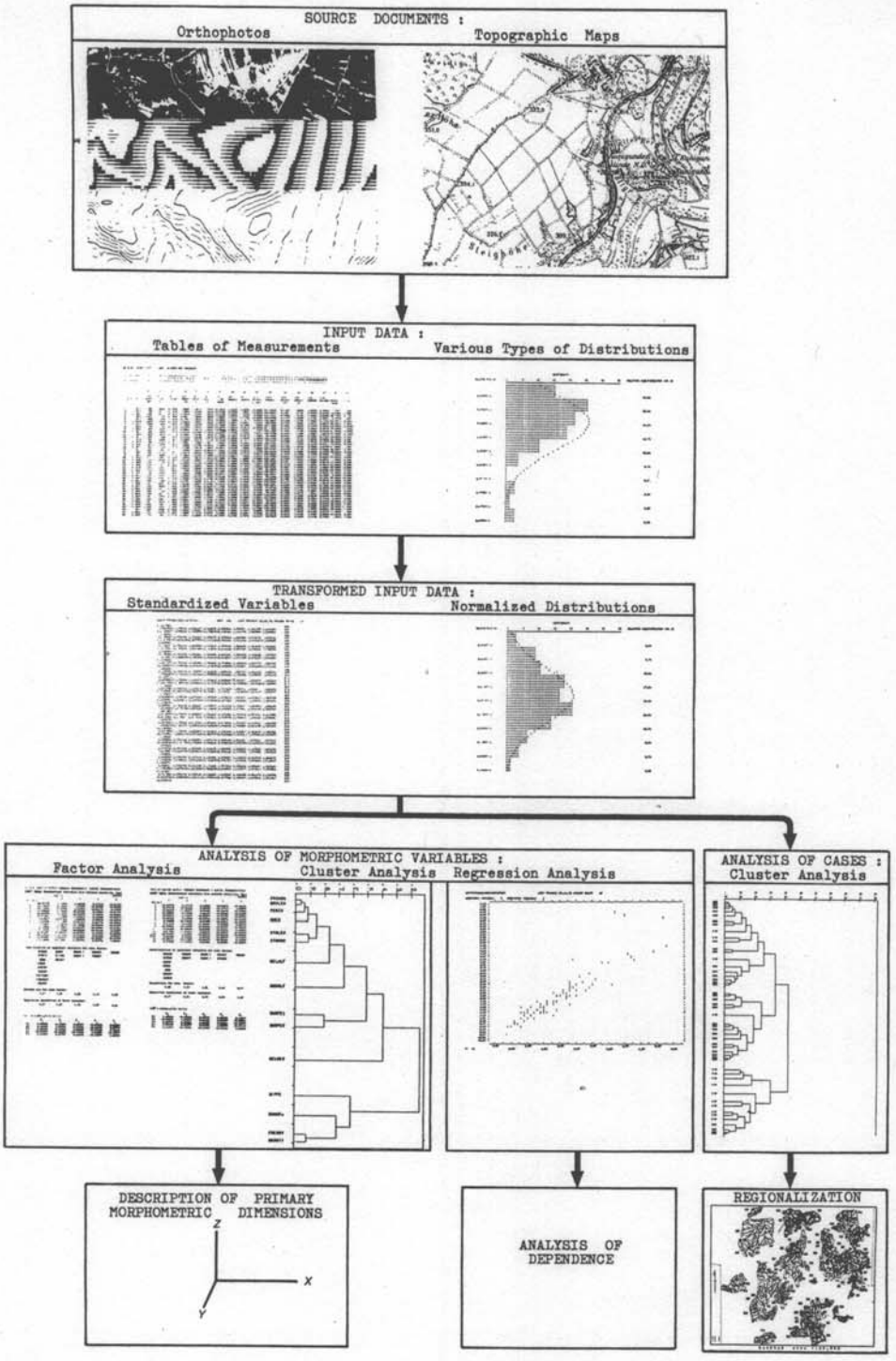

Figure 5. Schematic flow diagram summarizing the sequence of statistical operations in the present study.