Terrain Analysis by Computer

A. Keith Turner and Robert D. Miles, Purdue University

Abstract

The first part of this paper discusses the development of suitable terrain sampling procedures. Reproducible sampling is an important prerequisite to satisfactory terrain analysis. Terrain variability measures have been developed to differentiate between unique terrain types. The results of these analyses are used to develop stratified samples of compound areas covering entire map sheets.

Several computer-oriented terrain analysis methods have been developed and are described in the second part of this paper. Some of these methods are extensions of previously developed or suggested geomorphic techniques in which the computer is used to expedite procedures; other methods are new developments.

Examples of computer-prepared contour maps are included. Such maps are useful for rapid study and comparison of different terrain types.

Introduction

Geomorphology is becoming an increasingly quantitative science. Many geomorphologists now view the landscape as a series of open physical systems tending toward energy equilibrium (1). Such scientists naturally attempt to describe terrain in unambiguous meaningful quantitative terms. Accurate, reproducible, numerical descriptors of terrain have become of great concern to several scientific disciplines, including engineering geology, military geology, soil science, and land locomotion engineering. This paper describes techniques for the analysis of terrain using the computer. Further additions and refinements to the techniques are anticipated.

This research was sponsored by the Indiana State Highway Commission and the Bureau of Public Roads; however, this paper has not had the benefit of their review. The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Bureau of Public Roads or the Indiana State Highway Commission.

The authors also wish to acknowledge the assistance of Dr. V. L. Anderson, Department of Statistics, Purdue University, who reviewed the statistical implications of this study.

Terrain Sampling Procedures

Map digitization, the reduction of graphical to numerical data, is required before computer analysis is possible. Development of simple, logical, and reproducible sampling procedures is an important prerequisite to satisfactory terrain analysis. Sampling can be performed at a series of irregularly spaced points or at a regular spacing forming a sample grid.

Irregularly spaced samples can often describe the terrain better with equal or smaller numbers of sample points than can grid systems which tend to sample repetitiously in uniform areas. Irregularly spaced samples are favored by surface fitting operations, such as trend surface analysis, since the smaller number of sample points minimizes computational complexities.

With gridded samples, the position of a sample in the array automatically denotes its X and Y coordinates. This feature may ease data storage problems within the computer and can be used to advantage when computations require knowledge of adjacent points, since lengthy data search routines are unnecessary. While data are most frequently collected on rectangular coordinate systems, polar coordinates have been used. Stone and Dugundji (11) discuss the merits of various systems.

This research used gridded data to analyze the roughness and variability of homogenous areas. These values in turn allowed the determination of appropriate irregularly spaced sampling distributions of larger areas. Gridded samples were also used to develop hypsometric integrals of various terrain types.

Descriptive Measures of Homogenous Terrain

Two techniques for computer analysis of homogenous terrain types have been evaluated. One of these measures the surface roughness; the other is a technique for estimating the hypsometric integral whereby the labor of calculating this measure is substantially reduced.

Estimation of Surface Roughness

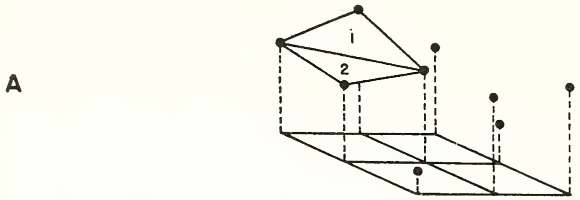
A technique for measuring roughness has been suggested by Hobson (4). This study has used a modified version of his program VECTOR.

Description of Program VECTOR

A rectangular array of elevation readings is the basic input data. A set of interesting triangular planar surfaces are defined by groups of three adjacent elevation readings (Figure 1A). Two different sets of triangles can be obtained from the same data by redefining the triangle corners as shown in Figure 1B. Type 1 triangles have "northeast-southwest" diagonals, while Type 2 triangles have the opposite diagnoals.

Normals to these planes are represented by unit vectors. Mean vector orientation, vector strength, and vector dispersion are computed using methods defined by Fisher (3) and described by Watson (15). Vector strength, R1, is obtained by using the direction cosine method (5). The standardized vector strength, R, (where R equals R1 divided by the number of triangles) ranges in value from zero for no preferred orientation to one for identical orientation. Fisher's dispersion factor, K, (3) indicates the variability or spread of the unit vectors; it takes on small values for highly dispersed distributions and extremely high values for low dispersions.

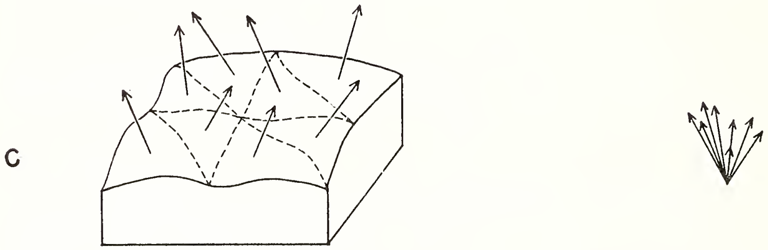
"Smooth" areas generally have high vector strengths and low dispersions (high K values) as shown in Figure 1C. Such areas may be flat or may have a regional tilt. "Rough" areas of non-systematic elevation changes yield low vector strengths and high vector dispersion, thus low K values (Figure 1D).



DEFINITION OF TRIANGULAR PLANES FROM ELEVATION MATRIX

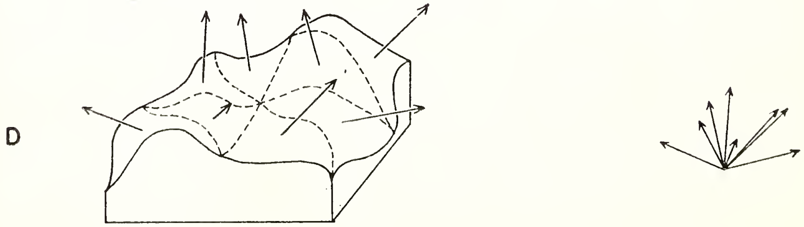
TYPE 1 TRIANGLES

TYPE 2 TRIANGLES



"SMOOTH" TERRAIN

LOW VECTOR DISPERSION



"ROUGH" TERRAIN

HIGH VECTOR DISPERSION

FIGURE I– PROGRAM VECTOR ANALYSIS PROCEDURES
(modified after Hobson)

Description of Test Sites

Twenty-five sample areas were used, twelve of which were located in Indiana. Eleven sites were measured on 1:24,000 scale topographic maps; the twelfth was located on a 1:62,500 scale map and corresponds to one of the 1:24,000 scale test sites. These sites represent several characteristic Indiana landforms including karst plains (2 sites); hills (3 sites), plateaus (1 site), escarpments (2 sites), ground moraine (2 sites), ridge moraine and dunes (1 site), and outwash plains (1 site). A thirteenth site was selected to analyze drumlin topography; part of the Weedsport Quadrangle (1:24,000), New York State, was used for this purpose. Twelve micro-terrain maps were selected for analysis from a series of specially surveyed maps prepared by Stone and Dugundji (11). These maps were prepared with contour intervals ranging between one-tenth and one foot and scales from 1:180 to 1:4800. A variety of California terrain types is represented including flood plains, badlands, and a variety of desert landforms (playas, pediments, wadis, dunes).

Results of Terrain Analysis by Program VECTOR

Analysis of twenty-five test sites, encompassing a variety of landforms and map scales, by program VECTOR suggests that meaningful terrain analysis can be performed on the computer. The following summarize the results of the current analyses:

1) Program VECTOR is equally suitable for analysis of micro and macro terrains, provided that maps of suitable scale and contour intervals are available for the test sites.

2) Fisher's dispersion factor, K, is an excellent descriptor of terrain roughness. Roughness is described as "the presence of nonsystematic elevation changes" (see Figure 1C and 1D). Fisher's dispersion factor is defined as

$$K = \frac{N - 1}{N - R1}.$$

where N is the number of observations (triangles), and R1 is the vector strength (3, 4).

For smooth surfaces, all the unit vectors tend to the same orientation; thus R1 approaches N, and N — R1 tends to zero, and accordingly K approaches infinity. Conversely, for very rough terrains R1 tends to zero, and thus K approaches one. The smallest K value obtained in this study was 31. This was measured for a wadi in the Mohave Desert (11, Figure 29) while the largest K value, 374,773, was measured for a playa surface (11, Figure 55). Indiana sites gave less extreme values. The roughest Indiana site was part of the New Albany Escarpment ($K = 81$) while the smoothest site was the ground moraine just west of Tipton ($K = 30{,}044$). Because of the very large range of K values, the log K transformation is frequently useful.

3) All test sites were analyzed by program VECTOR using Type 1 and Type 2 triangles (Figure 1B). No important differences were observed in any of the measures as a result of type of triangles used in the

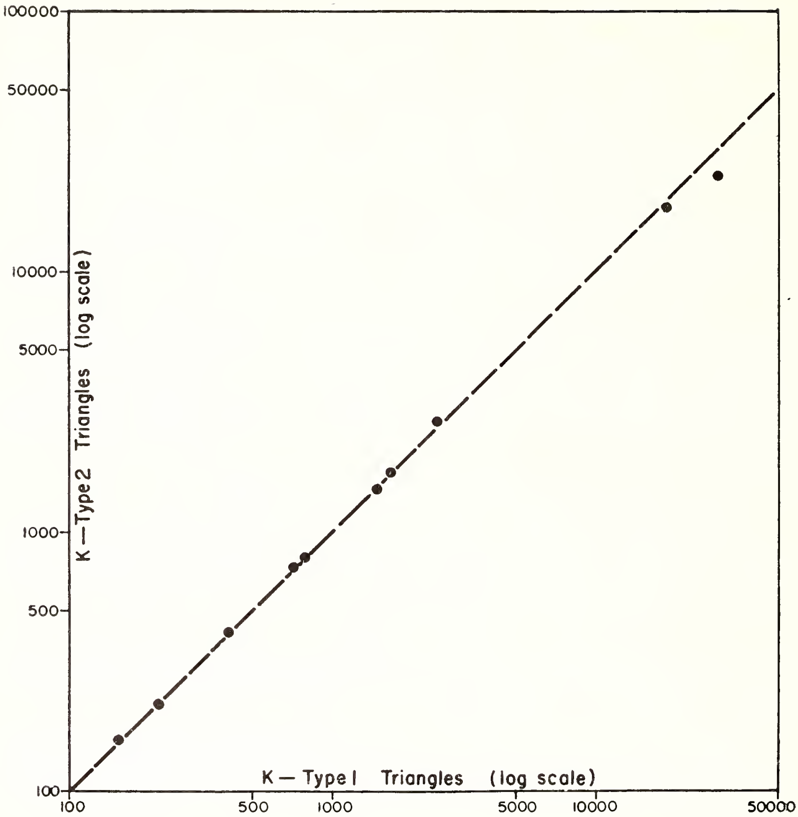


FIGURE 2- EFFECT OF TRIANGLE TYPE ON 'K'

analysis. Figure 2 shows the values of log K for Type 1 triangles plotted versus log K for Type 2 triangles; all points lie close to the 45° line.

4) Since K is a measure of vector dispersion it may be correlated with measures of the variability of attitudes of the triangular planes. The log (variance of dips) was plotted against log K (Figure 3B) and a least squares linear regression line fitted. The linear correlation coefficient, r = —0.964, indicates a strong correlation between log (variance of dips) and log K.

If log (maximum dip) is plotted against log K (Figure 3A) and a least squares regression line fitted, r = —0.950. The difference between the two correlation coefficients is not statistically significant. A strong correlation between log (maximum dip) and log K is geomorphically reasonable, since steep slopes are normally associated with rough topography. However, given two areas with the same maximum slope value,

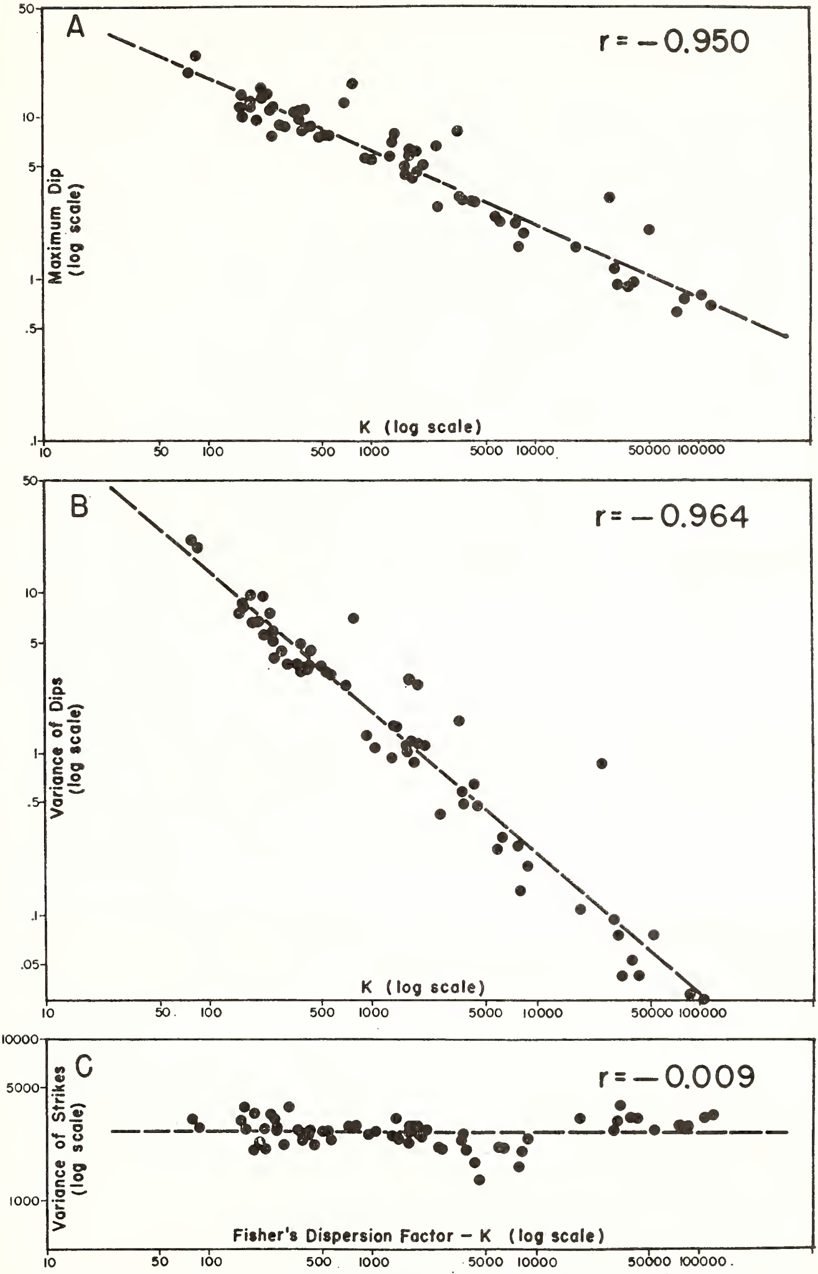


FIGURE 3— RELATIONSHIPS BETWEEN 'K' AND STRIKE AND DIP VARIABILITY

one might have a much greater range in slope values, and hence a larger variance of dip and greater roughness. Thus maximum dip may not be quite as good a descriptor of terrain roughness as variance of dip.

Initially it was believed that a similar correlation might exist between log (variance of strikes) and log K. Figure 3C is a plot of log (variance of strikes) versus log K. The linear correlation coefficient r = 0.009 indicates essentially no linear correlation exists between log (variance of strikes) and log K. This may be due in part to the test sites all having highly variable strike values. Careful examination of the results suggests that K may be related to a function of the variabilities of both strikes and dips. At the present time, however, a satisfactory model has not been discovered, although several have been tried. 5) The orientations of the triangular planar elements become progressively poorer estimators of the true terrain roughness as the sample spacing increases, as shown in Figure 4. A filtering action takes place so that smoothing of the terrain occurs; statistics, such as K, derived from the triangle orientations reflect this smoothing.

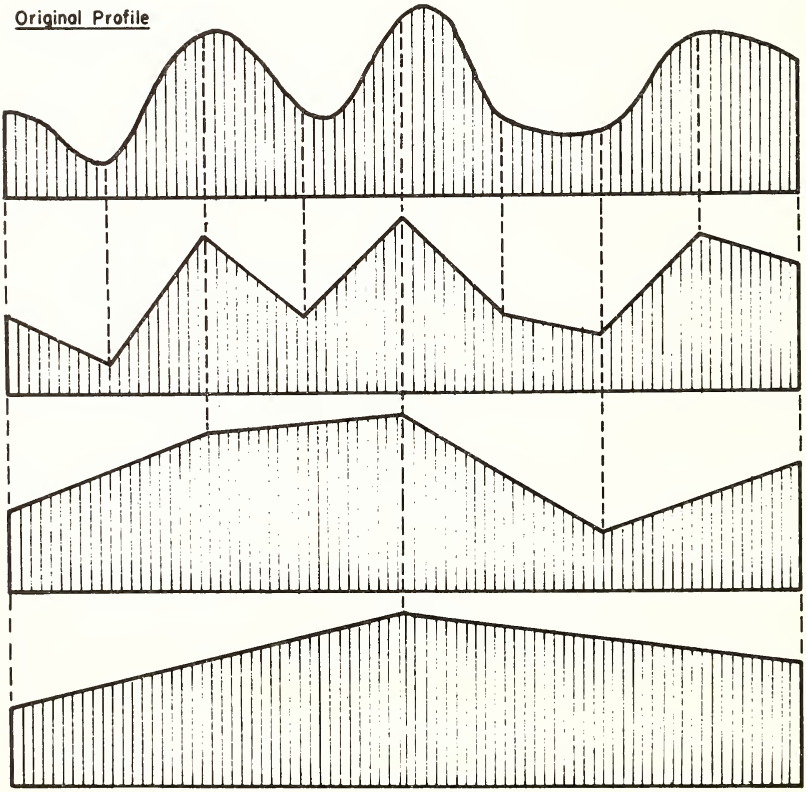


FIGURE 4—EFFECT OF SAMPLE SPACING ON TERRAIN ROUGHNESS ESTIMATION

Photographic enlargements up to four diameters of several test sites were produced, sampled, and analyzed using program VECTOR. Since all these data sets were rectangular arrays, it was also possible to analyze alternate columns and rows. These procedures allowed the examination of the effects of grid size (sample spacing) on results of program VECTOR analysis.

Testing of these comparisons for significance requires the use of non-parametric statistics. Strahler (12, 14) showed that slopes have normal distributions. In contrast, the distributtions of dips of the triangular planes are not normal, but vary widely from skewed to almost exponential distributions. Similar uncertainties surround the distribution of K and of strikes of these planes. Transformation of these distributions to more nearly normal form would allow the use of parametric statistics. The Walsh test, described by Siegel (10), is a suitable non-parametric test, and was used to evaluate K values measured for Indiana sites. The test showed that samples collected on a 500 foot spacing gave significantly smoother K values (at the 95% confidence level) than did samples collected on a 125 foot spacing. The investigator must consider the effect of sample spacing on roughness estimators when determining the roughness of any area and set his grid size accordingly.

Most terrain types have a characteristic slope length. Statistics derived from grid spacings equal to or less than the equivalent map distance will probably show much less dependence on grid size than will values obtained from larger grids. One method of selecting a grid size would be to estimate the characteristic slope lengths for the terrain to be studied and use a grid size slightly smaller than the equivalent map distance.

Another method is possible if the primary purpose of determining the roughness is to compare areas so as to design suitable stratified sampling procedures, as described in a following section. In such a case a grid size equal to the minimum sampling distance needed for the purposes of analysis, applied to all subareas, will allow their comparison on equal terms.

Estimation of the Hypsometric Integral

A second method of comparing two or more homogenous areas is to compare their hypsometric integrals. The hypsometric integral was devised by Strahler (13). It is a dimensionless measure of subsurface volume of a drainage basin. Strahler suggested that, under certain conditions, this integral would provide a quantitative expression of stage of basin denudation and Schumm (9) successfully used it for this purpose.

Strahler's method of measuring the hypsometric integral is somewhat tedious and this may explain why this morphometric measure has been little used. Chorley and Morley (2) suggested a simplification. They approximated drainage basins by a regular geometrical form, a lemniscate, and were able to more easily calculate the hypsometric integral. However, their technique resulted in values which varied from Strahler's values. They suggested a transformation to correct their values.

Description of Program GCON

The availability of a computer program GCON, originally written in the MAD language by Professor W. R. Tobler, University of Michigan, Ann Arbor, suggested a second method of estimating the hypsometric integral. This program accepts gridded elevation values and performs linear interpolation on them to produce contour maps on the normal computer line printer. Various scaling and contour interval specification options are available.

The program defines a map as an array of printing posittions, N characters per line wide and M lines long. Thus there are N x M character spaces in the entire map. Each character position is evaluated in turn and the appropriate interval list is incremented by one. Thus a frequency table is readily produced of percent area within each contour interval.

Accuracy of this integration procedure is controlled by the fixed size of the unit incremental area, a single character. Thus larger map sizes will increase the accuracy of the area determinations. Visual comparison of the computer-developed contour map with the original topographic map allows the investigator to estimate the reliability of his estimated hypsometric integral.

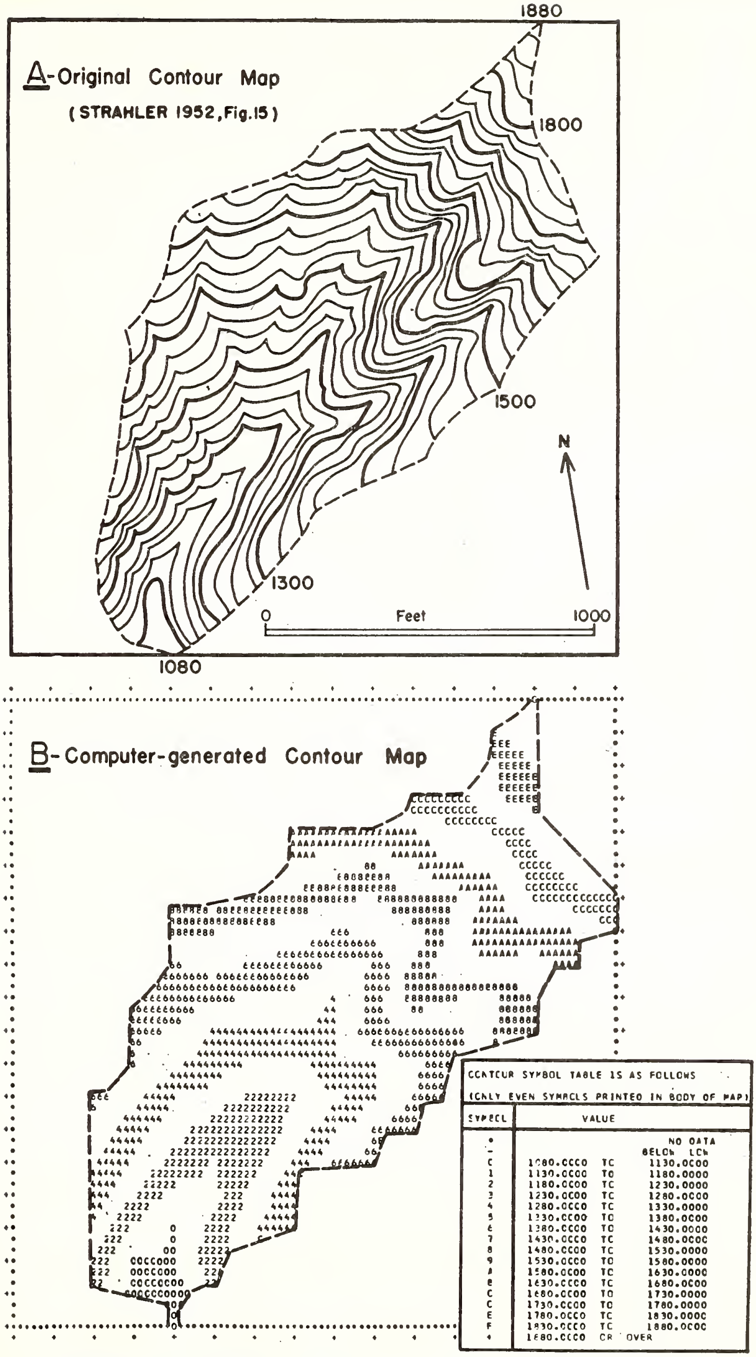
While the program requires rectangular map areas, analysis of irregular watersheds can be accomplished by setting all areas outside the watershed to arbitrarily very high values. The maximum and minimum contour levels corresponding to the highest and lowest values for the watershed are included as data. The program calculates the elevation range and excludes all elevations lying more than one range above the watershed's highest point, or one range below the lowest point. The frequency table will accordingly show the number of character spaces within each contour level in the basin. These values can be quickly transformed for plotting the hypsometric curve. A new subroutine has been developed to allow the computer to make this transformation, plot the curve, and estimate the area below the curve—the hypsometric integral.

Results of Analysis by Program GCON

The above concepts were tested by re-analyzing seven drainage basins whose hypsometric integrals were measured by Strahler (13). Figure 5A is a copy of Strahler's map of one of these basins; Figure 5B is a copy of a computer-generated contour map of the same basin. Figure 6 is a plot of the hypsometric curve for the same basin. Strahler's values are also plotted showing the close agreement between the two methods. Table 1 tabulates the results for all seven drainage basins, Chorley and Morley's uncorrected values for the same basins are also listed for comparison.

Methods of Stratified Sampling of Map Sheets

Most topographic quadrangle maps cover a variety of terrain types. Such areas may be termed "compound" as opposed to "simple" terrains



```
CCNTCUR SYPBOL TABLE IS AS FOLLOWS
(CNLY EVEN SYMBCLS PRINTED IN BODY OF MAP)

SYPBCL          VALUE

  .                           NO DATA
  -                         BELOW  LCW
  C      1080.CCCO   TC    1130.0C00
  1      1130.0C00   TO    1180.0000
  2      1180.0C00   TC    1230.0000
  3      1230.0C00   TC    1280.0C00
  4      1280.0CCO   TC    1330.0000
  5      1330.CC00   TO    1380.0C00
  6      1380.CCC0   TO    1430.0000
  7      1430.0CCO   TC    1480.0C0C
  8      1480.CC00   TC    1530.0000
  9      1530.0CCO   TO    1580.0000
  A      1580.0C00   TC    1630.0000
  B      1630.CCCO   TC    1680.0C00
  C      1680.0C00   TO    1730.0000
  C      1730.CC00   TO    1780.0000
  E      1780.0CCO   TC    1830.000C
  F      1830.0CCO   TC    1880.0C0C
  +      1880.CCCO   CR  OVER
```

FIGURE 5-DRAINAGE BASIN, VERDUGO HILLS, CALIFORNIA.

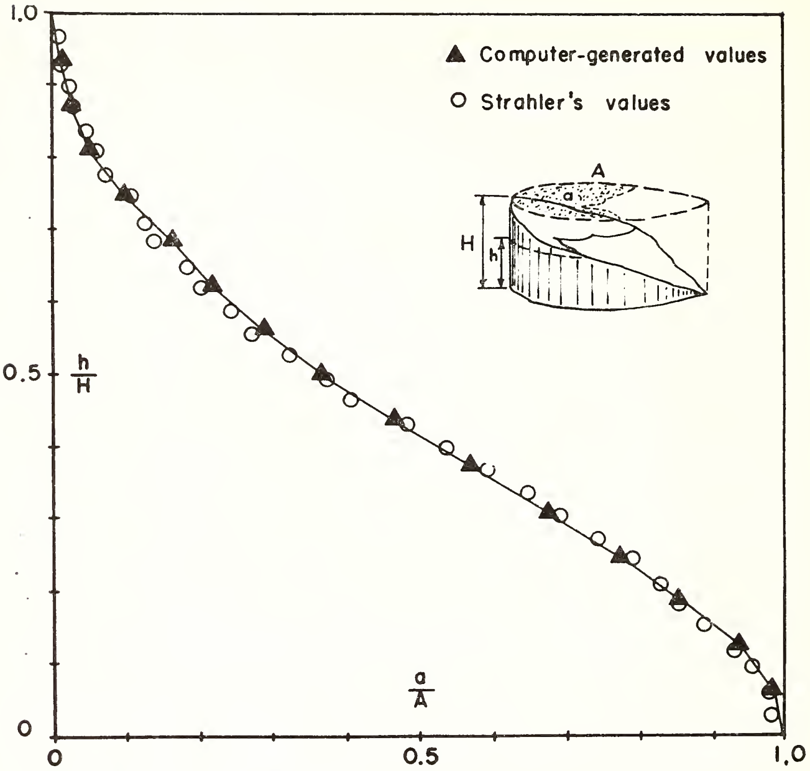


FIGURE 6— HYPSOMETRIC CURVE, DRAINAGE BASIN IN VERDUGO HILLS (see Figure 5)

TABLE 1

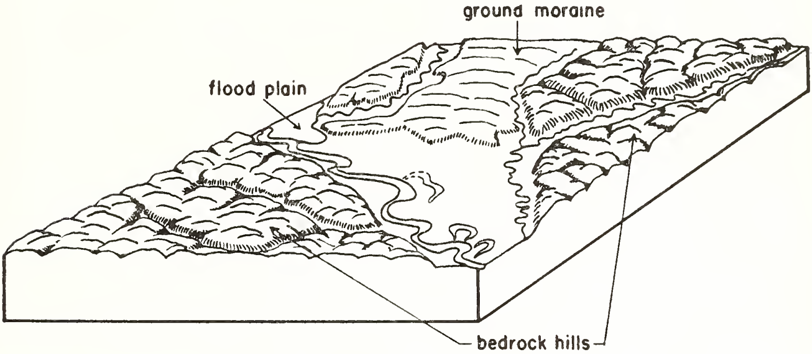
Comparison of Hypsometric Integrals

Values from Strahler (13)					
		Composite Values			
Figure Number	Single-Basin Hypsometric Integral	Mean Hypsometric Integral	Estimated Population Standard Deviation	Computer Estimated Hypsometric Integral	Approximated Hypsometric Integral from Chorley and Morley (2)
14	79.5%			78.7%	86.0%
15	43.0%			43.5%	41.0%
16	17.6%			16.3%	12.7%
17-1		59.7%	6.55%	62.5%	59.6%
17-2		54.2%	5.20%	54.4%	53.2%
17-4		46.8%	4.58%	43.6%	44.5%
17-5		40.8%	5.88%	46.2%	35.7%

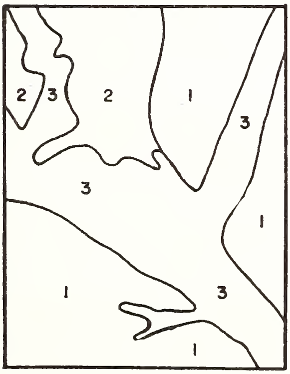
which show only one type of landform. In statistical terminology such compound map areas contain more than one population.

Suppose a researcher wishes to digitize a map area as shown in Figure 7A. He might partition the map into three "simple" areas as shown in Figure 7B. The investigator may decide on a total number of sample locations for the entire map, or he may decide on some sampling density for any one of the simple areas. In either case, if he wishes to sample all areas with equal reliability, he must distribute the number of sample points for each subarea according to its relative area and its relative variability.

The percent of the total map area contained in any subarea is easily estimated. It can be obtained by a planimeter if greater accuracy



A. BLOCK DIAGRAM OF MAP AREA (after Lobeck)



AREA TABLE

SUBAREA	% AREA
1. bedrock hills	47
2. ground moraine	17
3. flood plain	36

B. MAP SUBDIVIDED FOR STRATIFIED SAMPLING

FIGURE 7– STRATIFIED SAMPLING PROCEDURES

is desired. Thus, the problem resolves into determining a suitable numerical descriptor of the variability of the various terrain types.

Terrain Variability Measures

A perfectly smooth, planar, horizontal surface can be adequately described by a single elevation value; any planar inclined surface requires three elevation values, and associated X and Y coordinates, to describe it. Gently curved surfaces require still more data points to satisfactorily describe their overall orientation. The overall orientation of subareas might be described in a number of ways. The range of the elevations is the simplest value to obtain, requiring no prior sampling procedure.

A particular subarea's roughness will also affect the number of sampling points required to describe it. It has already been shown that K, Fisher's dispersion factor, is an ideal numerical descriptor of roughness. Since K values can range from one to extremely large numbers the variability factor, v, is defined as

$$v = \frac{\text{Range of elevations (feet)}}{\log (K)}$$

v can range from zero for completely flat surfaces which must by definition be smooth, to the range in elevations for extremely rough areas ($K = 10$, thus $\log (K) = 1$). Table 2 lists values of elevation range, K, log (K), and the resulting v for all twenty-five test sites.

If the variability factor and the percent area belonging to each subarea on a map sheet, such as Figure 7B, are known and the total number of samples to be distributed on the map is decided on, the following equation will distribute the points among the various subareas

$$n_i = \frac{v_i a_i N}{\Sigma v_i a_i}$$

where i is the subarea number, v_i is the variability of the ith subarea, a_i is the percent area of the ith subarea, and N is the total number of samples on the map. Such a distribution of sample points among the various subareas, will result in approximately equal sampling reliability over an entire map area.

Conclusions

Homogenous terrain types can be compared using measures of roughness, variability, and the hypsometric integral. The digital computer can aid in the calculation of these measures. In addition to their geomorphic applications these measures could be used as predictors of several engineering construction costs, such as highway grading costs, since they are associated with terrain geometry.

The variability factor can be used to determine appropriate irregularly spaced sampling distributions of larger compound areas covering one or many map sheets. The importance of these techniques should not be underestimated. McIntyre (7) states that irregularly spaced random samples may better digitize topographic maps than gridded data under at least some conditions. Computer analysis of these

TABLE 2

Results of Program VECTOR Analyses

Test Site Number	Terrain Type	Elevation Range (ft)	Roughness K	log K	Variability v (ft)
A. Turner Test Sites					
02393-1	escarpment	315	85	1.929	163.5
02393-3	escarpment	278	82	1.914	145.1
03022-1	hills	340	397	2.599	131.0
02487-1	hills	277	220	2.342	118.5
02039-2	hills	212	158	2.199	96.5
W-1	drumlins	183	222	2.346	78.3
02039-1	karst plain	166	723	2.859	57.2
02039-3	karst plain	170	1329	3.124	54.6
02393	plateau	81	1759	3.245	25.0
02393-2	ridge moraine and dunes	53	2892	3.461	15.3
02329-2	ground moraine	60	24664	4.392	13.7
02329-1	outwash plain	28	18024	4.256	6.6
02625-1	ground moraine	26	30044	4.478	5.8
B. Stone and Dugundji Sites (subarea numbers are figure numbers)					
M-29	Boulder-free wadi	25.0	31	1.491	16.8
M-57	Pleistocene lake terrace	10.8	51	1.708	6.3
M-33	Micro-badlands	9.0	117	2.068	4.4
M-45	Sand-sheet	10.5	309	2.490	4.2
M-49	Complex dunes	8.7	161	2.207	3.9
M-53	Turret dunes	7.0	133	2.124	3.3
M-41	Playa drainage channels	5.3	233	2.367	2.2
M-37	Floodplain mounds	3.0	269	2.430	1.2
M-39	Floodplain ridges	2.3	189	2.276	1.0
M-35	Salt ploygons	1.4	102	2.009	0.7
M-47	"Devil's Cornfield"	1.15	975	2.989	0.4
M-55	Playa	0.47	374773	5.573	0.084

compound areas by techniques such as trend-surface analysis will open the way to a number of interesting studies.

Map comparison procedures, as suggested by Miller (8), are one obvious application. Geometric analysis of at least more regular terrain types, such as drumlinized topography, by the methods developed by Whitten (16) and Loudon (6) could have useful applications to glacial geology, geomorphology, and civil engineering. The authors are currently examining the applicability of trend surface analysis to the highway location problem.

Literature Cited

1. CHORLEY, R. J. 1956. Some neglected source material in quantitative geomorphology. Jour. Geol. **64**:422-423.

2. CHORLEY, R. J. and L. S. D. MORLEY. 1959. A simplified approximation for the hypsometric integral. Jour. Geol. **67**:566-571.

3. FISHER, R. A. 1953. Dispersion on a sphere. Proc. Royal Society of London, series A. **217**:295-305.

4. HOBSON, R. D. 1967. FORTRAN IV programs to determine surface roughness in topography for the CDC 3400 computer. Kansas Geol. Survey Computer Contrib. **14**:28p.

5. JOHNSON, R. E. and F. L. KIOKEMEISTER. 1957. Calculus with analytic geometry. Allyn and Bacon Inc. 709p.

6. LOUDON, T. V. 1964. Computer analysis of orientation data in structural geology. Office of Naval Research, Geography Branch, Tech. Rept. 13, ONR Task No. 389-135:129p.

7. MCINTYRE, D. B. 1967. Trend-surface analysis of noisy data. Kansas Geol. Survey Commputer Contrib. **12**:45-56.

8. MILLER, R. L. 1964. Comparison-analysis of trend maps. Computers in the Mineral Industries, Stanford Univ. Publ., Geol. Sci. **9**:669-685.

9. SCHUMM, S. A. 1956. Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. Geol. Soc. American Bull. **67**:597-646.

10. SIEGAL, S. 1956. Nonparametric statistics. McGraw-Hill.

11. STONE, R. O. and J. DUGUNDJI. 1965. A study of microrelief—its mapping, classification, and quantification by means of a Fourier analysis. Engineering Geology **1**:89-187.

12. STRAHLER, A. N. 1950. Equilibrium theory of erosional slopes approached by frequency distribution analysis. Amer. Jour. Science **248**:673-696, 800-814.

13. STRAHLER, A. N. 1952. Hypsometric (area-altitude) analysis of erosional topography. Geol. Soc. American Bull. **63**:1117-1142.

14. STRAHLER, A. N. 1956. Quantitative slope analysis. Geol. Soc. America Bull. **67**:571-596.

15. WATSON, G. S. 1957. Analysis of dispersion on a sphere. Royal Astron. Soc. Monthly Notices, Geophysics Supp. **7**:153-159.

16. WHITTEN, E. H. T. 1967. Fourier trend-surface analysis in the geometrical analysis of subsurface folds of the Michigan Basin. Kansas Geol. Survey Computer Contrib. **12**:10-11.