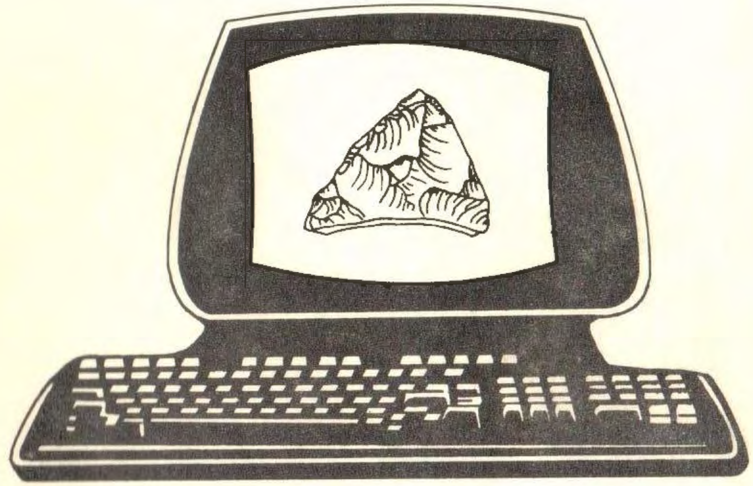
RETURN POSTAGE GUARANTEED

100169
DEPARTMENT OF ANTHROPOLOGY
ARIZONA STATE UNIVERSITY
TEMPE, ARIZONA
85287

ADVANCES IN COMPUTER ARCHAEOLOGY



VOL. 1
FALL 1984

TABLE OF CONTENTS

QUADRAT ANALYSIS IN ARCHAEOLOGY: A PROGRAM FOR TWO-TERM LOCAL VARIANCE
Donald Howes 1

THE PLOT THICKENS: COMPUTER MAPPING AND ARTIFACT PATTERNING
Mary Lou Larson and Lucy Chronic 13

ARCHAEOSTAT 2.0 - APPLESOFT PROGRAMS FOR DATA ANALYSIS
Peter L. Eidenbach 26

CUE 36

QUICK QUERIES 39

QUADRAT ANALYSIS IN ARCHAEOLOGY:

A PROGRAM FOR TWO-TERM LOCAL VARIANCE

Donald Howes
Department of Anthropology
Washington State University
Pullman, Washington

ABSTRACT

Dimensional Analysis of Variance (DAV) was first introduced into archaeology in 1973 and has seen limited use in the field of spatial analysis. Since its introduction, a body of experimental data has been developed by ecologists, dealing with methodological problems inherent in DAV. As an outgrowth of this work, a modification of the basic procedure, termed two-term local variance, has been developed.

This technique allows the examination of all multivariate combinations of starting positions within a grid, eliminating known problems associated with starting position. Two-term local variance can examine any size grid or transect and provides a close spacing of points when the mean square/block size values are plotted.

A BASIC program is presented which calculates two-term local variance scores on the raw data and on data transformed to detect large-scale amplitude pattern. This program, running on a BASIC interpreter, should be adequate for grid or transect sizes of 128 units or less. If a large number of units are to be used in calculations, it is recommended that the program be compiled to enhance program execution.

Introduction

Quadrat analysis techniques for the analysis of spatial patterns were first introduced to the archaeological community by Whallon (1973), who used a technique that he termed Dimensional Analysis of Variance (DAV) to analyze a living surface in Guila Naquitz Cave, Oaxaca. This technique was first developed by Grieg-Smith (1952) and was quickly applied to the analysis of spatial pattern by ecologists (see Ripley [1981:109] for reference to some field studies; Grieg-Smith [1979] provides a detailed overview of plant animal studies up to that date). Outside the field of ecology, however, this and other quadrat techniques have seen only limited use, with some applications in geography (Getis 1964; Moellering and Tobler 1972; Rogers 1974) and archaeology (Brose and Scarry 1976; Burley 1976; Jermann 1978; Dohm 1981; Howes 1982).

In the archaeological community, criticism of quadrat analysis has focused on DAV, where most of the comments have dealt with methodological problems inherent in the original technique as developed by Grieg-Smith (Hodder and Orton 1976:32-38; Kimmel 1978). A body of experimental data dealing with the methodology of DAV is now available in the ecological literature and these problems have mainly been explained, if not solved. (An overview of these developments, while important for the application and interpretation of quadrat analyses, is outside the scope of this paper; I would recommend that the interested reader consult Ripley [1981] who cites most of the relevant literature).

Two-Term Local Variance

As an outgrowth of some of this experimental work Hill (1973) developed a modification of the normal variance analysis technique that he termed "two-term local variance". This adaptation was designed to eliminate the effect that starting position had upon the location of peaks within the mean square/block size graph, which had been noted by Usher (1969).

Unlike a normal variance analysis procedure, two-term local variance examines all multivariate combinations of starting positions for quadrat scores in the grid, in an attempt to eliminate the problems associated with starting position. For a block size of one, the formula $mean(1/2[x_1-x_2]^2, 1/2[x_2-x_3]^2, \ldots 1/2[x_{n-1}-x_n]^2)$ where,

x_i = sample size in each quadrat

will provide the value for the variance at that block size. The formula $mean(1/4[x_1+x_2-x_3-x_4]^2, \ldots 1/4[x_{n-3}+x_{n-2}-x_{n-1}-x_n]^2)$ will provide the variance for block size two, and so on. The use of a two-term local variance eliminates the need for the block size grouping to be of the series 2^n, as it is for DAV, and is applicable for any block size less than 1/2 N (Hill 1973:228).

Since this technique examines all starting position combinations at each block size, the score that is produced for each block size represents the overall average for the block size, resulting in a smoothing of the data. Also, some of the peaks that are produced on the graph may be artifacts of the technique and are analagous to harmonic peaks (the peaks are periodic in nature and are proportional to the height of the previous peak). It has been suggested (Usher 1975:573) that a peak should be considered as representative of a scale of pattern only if there are no larger peaks to its left. Finally, the calculated variances are not independent, and so cannot be tested for

significance (this is a common problem with all quadrat analysis techniques that are based on a null hypothesis of a poisson distribution, and is not restricted to two-term local variance [Mead 1974]). On the plus side, two-term local variance can examine any size of grid or transect and there is a close spacing (every block size) of points on the mean square/block size graph. In this way, two of the major problems inherent to DAV have been resolved.

Using this technique, analysis can be performed on both the raw data and data which have been transformed to examine two different types of pattern. Usher (1975) has distinguished between what he terms abundance and amplitude pattern. Abundance pattern refers to the overall size of the concentration (its diameter or length), while amplitude pattern is the degree of variability between the concentration value and the value for the spaces between the concentrations (Usher 1975:570). Using artificial distributions to examine the differences between abundance and amplitude pattern, Usher found that two-term local variance was unable to detect amplitude pattern at large block sizes (Usher 1975:575). To counteract this problem, Usher suggests that amplitude data be transformed to abundance data using the formula

$$x_j = x_i - X$$

where,

x_j = transformed datum

x_i = original datum

X = mean of original data

This transformation will suppress small scale amplitude variation and accentuate any large-scale amplitude variation that may be present in the data.

Program Implementation

The program is written in Microsoft BASIC, version 5 (Table 1). An attempt was made to make the program as generalized as possible, to enhance the transportability of the program for other systems. Any place where it was necessary to use any machine or version specific statements will be pointed out and the function of those statements will be explained.

Line 210 clears the screen and homes the cursor; this command is the same as CLS or HOME in other versions of BASIC. Lines 250-280 ask for the names of the input and output files and open those files for reading and writing data. It was decided to use a file read/write system rather than DATA statements for input since this would allow the user greater flexibility in inputting data and would not require the modification of the program to run more than one data set. Data can be input to a file through a text editing program or by the use of a simple data entry program which can be written in BASIC. Lines 250-260 ask for the file names in valid CP/M format, giving the drive the program is resident on, the file name and the data type extension. This information should be modified to fit the requirements of your system. The input file should be structured so that the first value in the file is the number of quadrats used in the analysis, followed by the data scores for each quadrat.

Lines 290-350 read in the total number of quadrats used (290), dimension the arrays used later in the program (300), input the value for each quadrat into its array element (310-340) and close the input data file (350). Line 320 is an error trapping check to see if there is an attempt to read past the end of the input file. If this occurs, the error flag is set and no attempt is made to read further. Once the data have been input, the data are transformed and

Table 1. Source Code for Two-Term Local Variance Program.

```
10 ' ********************************************************************
20 ' *                                                                  *
30 ' *              TWO-TERM LOCAL VARIANCE (HILL 1973)                 *
40 ' *                                                                  *
50 ' *              PROGRAM INPLEMENTATION: DONALD HOWES                *
60 ' *                                                                  *
70 ' *              JULY 25, 1984                                       *
80 ' *              RELEASE 1.0, VERSION 2                              *
90 ' *                                                                  *
100 '********************************************************************
110 '
120 ' THIS PROGRAM CALCULATES A TWO-TERM LOCAL VARIANCE.
130 ' THE TECHNIQUE IS SIMILAR TO DIMENSIONAL ANALYSIS OF VARIANCE,
140 ' IN THAT DATA VALUES GROUPED BY QUADRAT ARE USED.
150 ' UNLIKE DAV, WHICH GROUPS QUADRATS BY A POWER OF 2, A TWO-TERM
160 ' LOCAL VARIANCE CAN GROUP QUADRATS IN ANY BLOCK SIZE, UP TO 1/2
170 ' THE TOTAL NUMBER OF QUADRATS.
180 ' THIS ALLOWS FOR A MUCH FINER EXAMINATION OF THE SCALE AT WHICH
190 ' PATTERNS OCCUR.
200 '
210 PRINT CHR$(4)
220 '
230 ' OPEN THE INPUT DATA FILE ON DISK AND READ IN DATA VALUES
240 '
250 INPUT "NAME OF INPUT FILE (INCLUDING DRIVE AND EXT.): ",IN$
260 INPUT "NAME OF OUTPUT FILE (INCLUDING DRIVE AND EXT.): ",OT$
270 OPEN "I",#1,IN$
280 OPEN "O",#2,OT$
290 INPUT#1,TOT
300 DIM POINT(TOT),CORREC(TOT),TTL(TOT),TTCL(TOT)
310 FOR I=1 TO TOT
320 IF EOF(1) THEN GOTO 340
330 INPUT#1,POINT(I)
340 NEXT I
350 CLOSE#1
360 AVER=0.0
370 FOR I=1 TO TOT
380 AVER=AVER+POINT(I)
390 NEXT I
400 MEAN=AVER/TOT
410 FOR J=1 TO TOT
420 CORREC(J)=ABS(POINT(J)-MEAN)
430 NEXT J
440 '
450 ' INITIAL LOOP FOR DATA COMPILATION
460 '
470 PRINT CHR$(4)
480 PRINT "PROCESSING DATA"
490 FOR ALT=1 TO 2
500 COUNT=1.0
510 FOR L=1 TO INT(TOT/2)
```

```
520 INCREM=L-1
530 TERMIN=(TOT+1)-(2*L)
540 Z=TOT+1.0
550 NUM=Z-(2.0*COUNT)
560 INTER=0.0
570 '
580 ' VARIABLES "A" TO "D" ARE USED AS INNER LOOP DELIMITERS
590 '
600 FOR K=1 TO TERMIN
610 A=K
620 B=INCREM+A
630 C=B+1
640 D=INCREM+C
650 SUM=0.0
660 DIF=0.0
670 '
680 ' THE INNER TWO LOOPS CALCULATE THE VARYING SIZE GROUPS OF DATA
690 ' USED IN THE CALCULATION OF THE TWO-TERM LOCAL VARIANCE
700 '
710 FOR I=A TO B
720 IF ALT=1 THEN SUM=SUM+POINT(I) ELSE IF ALT=2 THEN SUM=SUM+CORREC(I)
730 NEXT I
740 FOR J=C TO D
750 IF ALT=1 THEN DIF=DIF+POINT(J) ELSE IF ALT=2 THEN DIF=DIF+CORREC(J)
760 NEXT J
770 INTER=(((SUM-DIF)^2)/(2.0*COUNT))+INTER
780 NEXT K
790 '
800 ' "TSUM" IS THE VALUE FOR THE TWO-TERM LOCAL VARIANCE AT THE
810 ' BLOCK SIZE UNDER INVESTIGATION
820 '
830 TSUM=INTER/NUM
840 IF ALT=1 THEN TTL(L)=TSUM ELSE IF ALT=2 THEN TTCL(L)=TSUM
850 COUNT=COUNT+1.0
860 PRINT ".";:
870 NEXT L
880 NEXT ALT
890 FOR I=1 TO INT(TOT/2)
900 PRINT #2,TTL(I);",";TTCL(I)
910 NEXT I
920 CLOSE#2
930 GOSUB 1000
940 END
```


read into a second array in lines 360-430. This array will be used for the investigation of amplitude pattern.

Statements 490-880 constitute the heart of the program, where the actual value for the two-term local variance at each block size is calculated. There are five FOR-NEXT loops nested at four levels. The outer loop (490) determines whether the value is calculated on the raw or transformed data. The second loop (510) sets the values for the counter variables used in the calculation of values from each starting position and those variables that are needed for the expansion of the added and subtracted terms at each block size. The third loop (600) uses the variables calculated in the previous loop to start calculation of two-term local variance at each block size. The inner two loops (710-760) run sequentially, rather than nested, and are used to calculate the added and subtracted terms at each starting position. In these two loops, the value for ALT is checked and either the raw or transformed data matrix is used, depending on the value. In line 770 the scores for all the different starting positions are totalled, and the actual score for the two-term local variance is calculated in line 830. Again, depending on the value for ALT, this score is output to one of two different matrices.

Once all the calculations are completed, the two output data matrices are written to disk in lines 890-910 and output file is closed (920). The output statement in line 900 writes the data to disk with the score for the raw value followed by the score for the transformed value as a data pair on a single line. This can be changed to suit the wishes of the user.

The GOSUB statement in line 930 was used to call a plotting routine, which produced Figure 1 as an illustration of a use of the output file. The plotting routine is not reproduced here, since it is device specific to an Amdek DXY-100 plotter and would not be transportable.

Figure 1 was produced using simulated data for an 8x8 area, giving a total of 64 units for which data are available. The plot for raw data scores (solid line) shows a single peak present at a block size of 29, indicating a large-scale concentration with an area of 29 square meters. Transforming the data to search for amplitude pattern (dashed line) shows a minor peak at a block size of 19, although the height of the peak is small enough that it probably is not a significant concentration.

Like all BASIC interpreter programs, this one runs slowly, especially since there is a large amount of computation taking place in the loops. Timing the program on a Vector Graphic 4, using an Intel 8088 microprocessor running at 5.1 MHz, the program takes 11:52 to complete its calculations. This is an indication that, if a large number of units are contemplated (i.e. more than 128), it would be best to compile the program using Microsoft's BASCOM compiler to speed program execution. This procedure will entail some minor modification of the source code, so that compilation can be accomplished.

REFERENCES

Brose, D.S. and J.F. Scarry

1976 The Boston Ledges Shelter: comparative spatial analysis of Early Late Woodland occupations in Summit County, Ohio. Midcontinental Journal of Archaeology 1:179-228.

Burley, D.V.

1976 On the spatial analysis of a plowed site at Bartibog. Canadian Archaeological Association, Bulletin 8:82-105.

Dohm, K.M.

1981 Similarities in spatial characteristics of several Basketmaker II sites on Cedar Mesa, Utah. Unpublished MA Thesis, Department of Anthropology, Washington State University.

Getis, A.

1964 Temporal land-use pattern analysis with the use of nearest-neighbour and quadrat methods. Annals of the Association of American Geographers 54:391-399.

Greig-Smith, P.

1952 The use of random and contiguous quadrats in the study of the structure of plant communities. Annals of Botany N.S. 16:293-316.

1979 Pattern in vegetation. Journal of Ecology 67:755-779.

Hill, M.O.

1973 The intensity of spatial pattern in plant communities. Journal of Ecology 61:225-235.

Hodder, I.R. and C. Orton

1976 Spatial analysis in archaeology. Cambridge University Press, Cambridge.

Howes, D.

1982 Spatial analysis at a Northwest Coast fishing camp: the Hoko River site. Unpublished MA Thesis, Department of Anthropology, Washington State University.

Jermann, J.V.

1978 Archaeological investigations at Home Valley, Skamania County, Washington. University of Washington, Office of Public Archaeology, Institute of Environmental Studies, Reconnaissance Report 18.

Kimmel, R.H.

1978 Dimensional analysis of variance: notes on a statistical approach for the identification of artifact association. Unpublished MA Thesis, Department of Anthropology, University of North Carolina.

Mead, R.
1974 A test of spatial pattern at several scales using data from a grid of contiguous quadrats. Biometrics 30:295-307.

Moellering, H. and W.R. Tobler
1972 Geographical variances. Geographical Analysis 4:34-50.

Ripley, B.D.
1981 Spatial statistics. John Wiley & Sons, New York.

Rogers, A.
1974 Statistical analysis of spatial dispersion: the quadrat method. Pion Limited, London.

Usher, M.B.
1969 The relation between mean square and block size in the analysis of similar patterns. Journal of Ecology 57:505-514.

1975 Analysis of pattern in real and artificial plant populations. Journal of Ecology 63:569-586.

Whallon, R., Jr.
1973 Spatial analysis of occupation floors I: application of dimensional analysis of variance. American Antiquity 38:266-278.

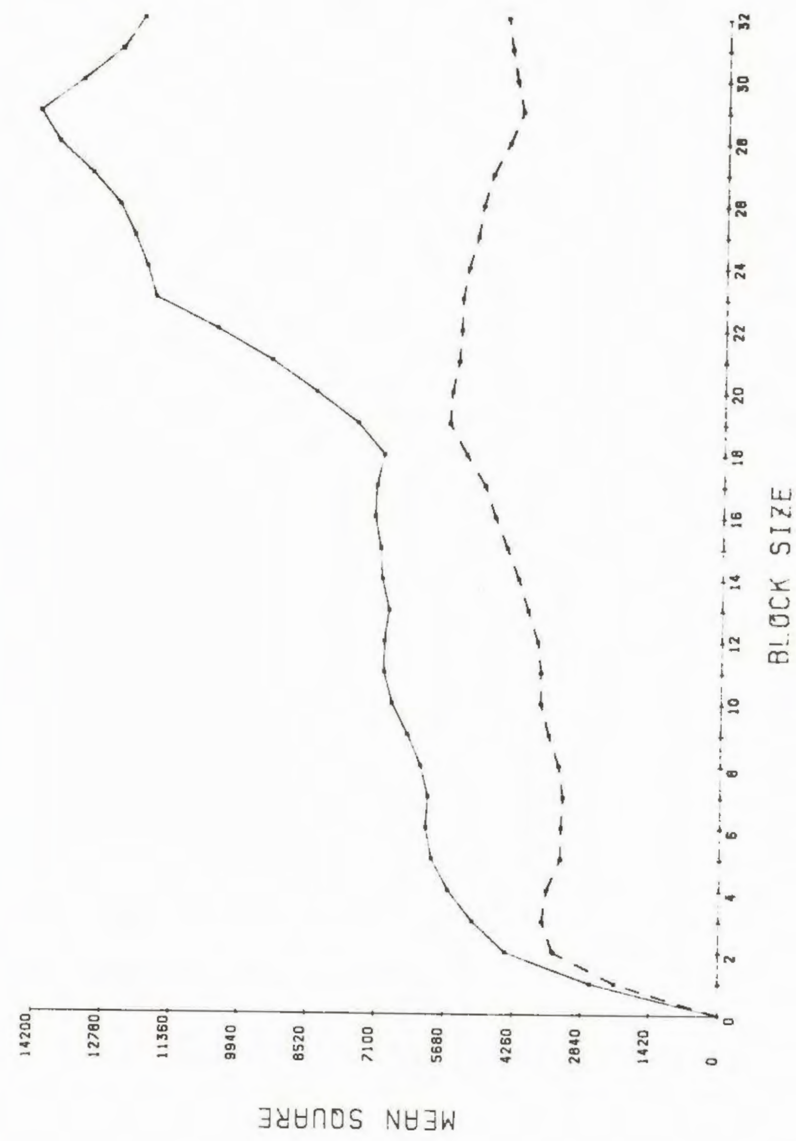


Figure 1. Bivariate plot using simulated data. A single peak is present at a block size of 29, indicated by the raw data (solid line). One additional amplitude peak is found at a block size of 19 (dashed line).