Dennis L. Murphy
*Technicolor Government Services, Inc.*
*U.S. Geological Survey*
*EROS Data Center*
*Sioux Falls, SD 57198*

# Estimating Neighborhood Variability with a Binary Comparison Matrix

A technique was developed to estimate variability of nominal scale data using a 3- by 3-neighborhood function.

## Introduction

Digital spatial data bases are created to support a variety of resource management activities. Sources of spatially referenced data include aerial photography, Landsat multispectral scanners, digital elevation models, and maps of point, line, and region geographic entities. Once established, digital data bases can be analyzed with the functional capabilities of a geographic information system.

The analysis functions of a geographic information system allow resource analysts to derive new variables from the source data stored in the system. Tomlin and Berry (1979) identified four classes of fundamental operations which manipulate geographic data: reclassifying map categories, overlaying maps, measuring cartographic distance, and characterizing cartographic neighborhoods. Quantifying variability is one of many methods to characterize cartographic neighborhoods.

Variables in a geographic information system frequently are of nominal scale; two examples are land-cover categories interpreted from aerial photography and soil classes digitized from a soil survey. Neighborhood variability of nominal scale data is relevant to resource analysis applications of geographic information systems. This paper describes the use of a binary comparison matrix to estimate neighborhood variability of nominal scale data in a raster format data base. The topological characteristics of the binary comparison matrix index are compared with two spatial operators which measure other aspects of neighborhood variability. An example from a land-cover classification of the Kenai National Wildlife Refuge in Alaska demonstrates results produced by the binary comparison matrix.

Abstract: *The analysis of spatial data bases supports resource management, planning, and decisionmaking. Cartographic models of land suitability, for example, incorporate a variety of spatial variables and analytical functions. Neighborhood variability of nominal scale data is a spatial variable relevant to applications of geographic information systems.*

*A technique which utilizes a binary comparison matrix has been developed to implement a neighborhood function for a raster format data base. The technique assigns an index value to the center pixel of 3- by 3-pixel neighborhoods. The binary comparison matrix provides additional information not found in two other neighborhood variability statistics; the function is sensitive to both the number of classes within the neighborhood and the frequency of pixel occurrence in each of the classes. Application of the function to a spatial data base from the Kenai National Wildlife Refuge, Alaska, demonstrates (1) the numerical distribution of the index values, and (2) the spatial patterns exhibited by the numerical values.*

## Method

The binary comparison matrix characterizes 3- by 3-pixel neighborhoods; the technique compares nominal class values and assigns the index value to the center pixel of the neighborhood.

The binary comparison matrix (BCM) is expressed as follows:

$$\text{BCM} = {}^{1}\!/_{2} \sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij}, \quad (1)$$

where $n$ = number of neighborhood elements.

Photogrammetric Engineering and Remote Sensing,
Vol. 51, No. 6, June 1985, pp. 667-674.
0099-1112/85/5106-0667$02.25/0

A boolean operator determines the value of $r_{ij}$ by comparing nominal class values ($C$) of pixel pairs:

$$\text{If } C_i = C_j, \text{ then } r_{ij} = 0, \text{ else } r_{ij} = 1. \tag{2}$$

The algorithm which computes the index uses a part of the 9- by 9-comparison matrix derived from the neighborhood (Figure 1). The binary comparison matrix is similar to a graph connectivity matrix for a network (Abler *et al.*, 1971). The total number of comparisons required is $(n^2 - n)/2$. The matrix size is $n$ by $n$, or $n^2$ (Figure 1b). Because no pixel requires a comparison with itself, $n$ is substracted from $n^2$, representing the omission of the matrix diagonal. One half of the remaining comparisons are redundant, that is, the same comparison is made across the matrix diagonal; the numerator $(n^2 - n)$ is therefore divided by 2. For a 3 by 3 neighborhood, the number of comparisons is $(9^2 - 9)/2$, or 36. The boolean operator in the algorithm compares 36 pixel pairs and sums the resulting values of $r_{ij}$. The sum of the 36 binary comparison matrix values (Figure 1b) produces the same index values as Equation 1 applied to the 81 comparisons of the complete 9 by 9 matrix.

An alternative expression for the BCM index is the following:

$$\text{BCM} = \tfrac{1}{2}\left[n^2 - \sum_{i=1}^{K} f_i^2\right], \tag{3}$$

where $n$ = number of neighborhood elements,
$f_i$ = frequency of elements in class $i$, and
$K$ = number of classes in the neighborhood.

| | | |
|---|---|---|
| 1 A | 2 A | 3 B |
| 4 B | 5 B | 6 B |
| 7 B | 8 B | 9 B |

(a)

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | - | | | | | | | | |
| 2 | 0 | - | | | | | | | |
| 3 | 1 | 1 | - | | | | | | |
| 4 | 1 | 1 | 0 | - | | | | | |
| 5 | 1 | 1 | 0 | 0 | - | | | | |
| 6 | 1 | 1 | 0 | 0 | 0 | - | | | |
| 7 | 1 | 1 | 0 | 0 | 0 | 0 | - | | |
| 8 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | - | |
| 9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | - |

(b)

FIG. 1. (a) Example of a 3 by 3 neighborhood. (b) Binary comparison matrix illustrating each of the relevant $r_{ij}$ values for the 3 by 3 neighborhood shown in Figure 1a. The BCM index for the neighborhood is 14.

For a 3 by 3 neighborhood, $n^2$ is a constant (81). The sum of the squares of class frequencies is equal to the count of comparisons ($r_{ij}$) which have the same class value. Subtracting this sum from $n^2$ produces a count of pixel comparisons which have different classes, but the count includes the redundant information in the full 9 by 9 matrix. Multiplication by one-half eliminates the redundant pixel comparisons. For a 3 by 3 neighborhood, with $K = 2$, $f_1 = 2$, $f_2 = 7$:

$$\begin{aligned}\text{BCM} &= \tfrac{1}{2}\,[81 - (4 + 49)]\\ &= 14\end{aligned}$$

Equation 3 demonstrates that the index incorporates both the number of classes occurring in the neighborhood ($K$) and the frequency of occurrence in each class ($f_i$). Thus the index value is sensitive to changes in either $K$ or $f_i$.

As $K$ increases, the index will also increase. For most values of $K$, however, varying the class frequencies will produce different results for the BCM index (Figure 2). For a specified $K$, a concentration of neighborhood elements in a single class (a high $f_i$) will result in a low BCM index. Conversely, relatively low class frequencies across all classes will

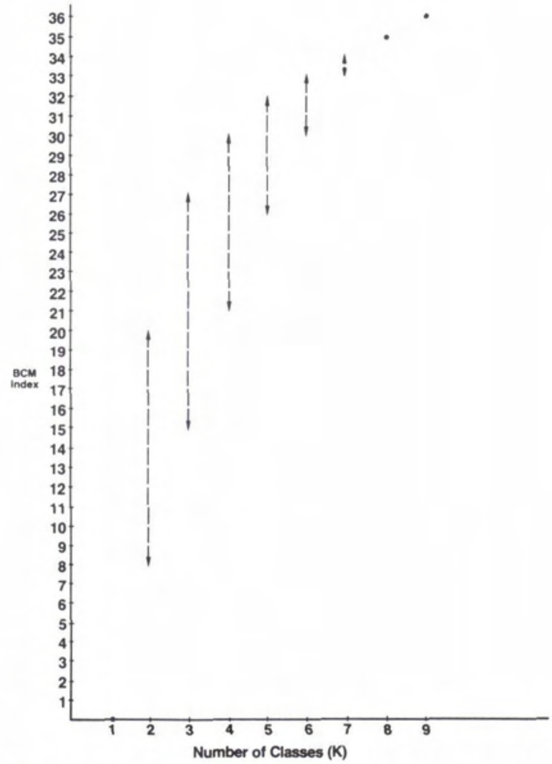



FIG. 2. Minimum-maximum BCM index values as a function of the number of classes in a neighborhood.

produce a high BCM index. For example, consider two extremes for neighborhoods with $K = 2$:

Case One — $f_1 = 1, f_2 = 8$
BCM $= \frac{1}{2}[81 - (1 + 64)] = 8$
Case Two — $f_1 = 4, f_2 = 5$
BCM $= \frac{1}{2}[81 - (16 + 25)] = 20$

Because the index is sensitive to both $K$ and $f_i$, a higher value of $K$ in a neighborhood does not necessarily produce a higher result for the BCM index. Compare the following with Case Two above:

Case Three — $K = 3, f_1 = 7, f_2 = 1, f_3 = 1$
BCM $= \frac{1}{2}[81 - (49 + 1 + 1)]$
$= 15$

Although Case Three contains three classes, the BCM measure of neighborhood variability is lower than that obtained for Case Two, which contains two classes. The lower index for Case Three is due to the concentration of neighborhood elements in one of the classes, $f_1$. Note, however, that the index for Case Three is greater than that obtained for Case One.

### Comparison with Other Methods

Two other raster format variability measures that were compared with the binary comparison matrix technique are

- Number of different classes (NDC) (C. Dana Tomlin, Yale University, unpublished manual for the Map Analysis Package, 1980; Environmental Systems Research Institute, Geographic Information Software Descriptions), and
- Center versus neighbors (CVN) (Mead *et al.*, 1981).

Four characteristics of the estimation methods were compared:

- The set upon which a topology is generated, that is, the topological space;
- The method of generating the topology;
- The method of calculating the index from the topology; and
- The range of values for the index.

The following definitions from Munkres (1975) were applied in the comparisons:

- A topology on a set $X$ is a collection $T$ of subsets of $X$ having the following properties:
  - The empty set $\emptyset$ and $X$ are in $T$,
  - The union of the elements of any subcollection of $T$ is in $T$, and
  - The intersection of the elements of any finite subcollection of $T$ is in $T$.
- The set $X$ for which a topology has been specified is called a topological space.

### Number of Different Classes (NDC)

The NDC method utilizes the nine-element neighborhood, $W$, as the topological space. The topology on $W$ is generated by partitioning $W$ into subsets of nominal class values.

For example, the algorithm would implicitly partition the neighborhood shown in Figure 1a into two subsets:

$$W_1 = [1,2]$$
$$W_2 = [3,4,5,6,7,8,9]$$

The topology on $W$ includes: $\emptyset$, $W_1$, $W_2$, and $W$. The estimation of neighborhood variability is a count of the subsets, $W_i$; for the example in Figure 1a, the index value is 2. In general, the NDC index value ranges from 1 to 9.

### Center Versus Neighbors (CVN)

The CVN method compares the neighborhood center with the other eight elements. The topological space $V$ is a set of eight ordered pairs representing the comparisons of class values between neighborhood elements. The eight ordered pairs are partitioned into two sets: $V_0$, the ordered pairs of elements with the same nominal class value, and $V_1$, the ordered pairs with different class values. The method counts the elements in $V_1$ to estimate neighborhood variability. The index ranges from 0 to 8.

For the example (Figure 1a), the subsets of the topology are

$$V_0 = [(5,3), (5,4), (5,6)\ (5,7), (5,8), (5,9)]$$
and
$$V_1 = [(5,1), (5,2)].$$

The count of elements in $V_1$ is 2.

### Binary Comparison Matrix (BCM)

The topological space, $R$, for the BCM technique is 36 ordered pairs representing the paired comparison of neighborhood elements (Figure 1b). The binary operator implicitly partitions $R$ into two sets of neighborhood element relations: $R_0$, the ordered pairs with identical class values, and $R_1$, the ordered pairs with different class values. The index is a count of ordered pairs in $R_1$; the index values range from 0 to 36. For the example (Figure 1), the two sets are

$$R_0 = [(1,2), (3,4), (3,5), (3,6), (3,7), (3,8), (3,9), (4,5), (4,6), (4,7), (4,8), (4,9), (5,6), (5,7), (5,8), (5,9), (6,7), (6,8), (6,9), (7,8), (7,9), (8,9)], \text{ and}$$
$$R_1 = [(1,3), (1,4), (1,5), (1,6), (1,7), (1,8), (1,9), (2,3), (2,4), (2,5), (2,6), (2,7), (2,8), (2,9)].$$

The resulting index value is 14.

### Summary of Topological Characteristics

The topological space distinguishes the three variability estimation techniques (Table 1). The NDC technique incorporates the set $W$ as the topological space; the CVN and BCM methods use a relation on $W$ for the topological space (a relation is a subset of the cartesian product of $W \times W$). Because NDC does not use a relation for the topological space, there

TABLE 1. COMPARISON OF THREE NEIGHBORHOOD VARIABILITY ESTIMATION TECHNIQUES

| | Number of Different Classes (NDC) | Center versus Neighbors (CVN) | Binary Comparison Matrix (BCM) |
|---|---|---|---|
| Topological Space | $W = [1,2,3,4,5,6,7,8,9]$ | $V = [(5,1),(5,2),(5,3),(5,4),(5,6),(5,7),(5,8),(5,9)]$ | $R = [(1,2),(1,3),(2,3),(1,4),(2,4),(3,4),(1,5),(2,5),(3,5),(4,5),(1,6),(2,6),(3,6),(4,6),(5,6),(1,7),(2,7),(3,7),(4,7),(5,7),(6,7),(1,8),(2,8),(3,8),(4,8),(5,8),(6,8),(7,8),(1,9),(2,9),(3,9),(4,9),(5,9),(6,9),(7,9),(8,9)]$ |
| Topology | $\emptyset$, $W$, $W_1 \ldots W_k$ – Class subsets | $\emptyset$, $V$, $V_0$ – Comparisons similar, $V_1$ – Comparisons different | $\emptyset$, $R$, $R_0$ – Comparisons similar, $R_1$ – Comparisons different |
| Index Range | Count of $W_i$'s 1–9 | Count of elements in $V_1$ 0–8 | Count of elements in $R_1$ 0–36 |

are no explicit comparisons of neighborhood elements. For this reason, NDC is not sensitive to changes in class frequency. The CVN method does use a relation on $W$ for the topological space; however, the relation is a subset of the topological space $R$ used in the BCM method. The relation $R$ is the only topological space which represents the complete, paired comparison of neighborhood elements.

## EXAMPLES OF THE THREE METHODS

Figure 3 compares the index values derived from each of the three methods. The examples represent the four possible class frequencies in a 3 by 3 neighborhood when $K$ equals 2. The NDC method produces the same value for all four examples, demonstrating that the NDC measure of variability is not sensitive to changes in class frequency.

The index derived from the CVN method is different for each of the four examples; however, two results should be noted. First, the value of the center pixel has a major impact on the index. For example, in Figures 3c and 3d, the value decreases from 6 to 5 because one additional pixel in Figure 3d is identical with the center pixel. Additional examples (Figures 4a and 4b) demonstrate that the spatial distribution of neighborhood elements changes the CVN index, while the index values from the NDC and BCM methods are unchanged. Second, the CVN method is not sensitive to the number of classes in the neighborhood. Although the neighborhood in Figure 4c contains seven classes, the CVN index is the same as the neighborhood which contains two classes (Figure 4b).

The BCM index increases with increasing complexity of the neighborhood (Figures 3 and 4). The lowest index value when $K$ equals 2 is 8; the highest value is 20. The BCM technique is the only method of the three compared which is sensitive to both the number of classes occurring in the neighborhood and the frequency of elements in each class.

## SPATIAL SENSITIVITY

The BCM technique is not sensitive to changes in the spatial distribution of class values; the method produces identical index values for neighborhoods with equal values of $K$ and $f_i$, regardless of the spatial distribution of classes within the neighborhood. Figures 5a and 5b document this result with an example of two neighborhoods, each containing two classes with frequencies of 3 and 6. A second index is required which is sensitive to changes in the spatial distribution of nominal scale data.

One measure of the spatial distribution of the classes within a neighborhood is an index of edge between neighborhood elements. Mead *et al.* (1981) suggested two spatial distribution measures in 3 by 3 neighborhoods: interspersion and juxtaposition.

| | | |
|---|---|---|
| A | B | B |
| B | B | B |
| B | B | B |

a

| | | |
|---|---|---|
| A | A | B |
| B | B | B |
| B | B | B |

b

| | | |
|---|---|---|
| A | A | B |
| B | A | B |
| B | B | B |

c

| | | |
|---|---|---|
| A | A | B |
| A | A | B |
| B | B | B |

d

| Example | $f_1$ | $f_2$ | NDC | CVN | BCM |
|---|---|---|---|---|---|
| a | 1 | 8 | 2 | 1 | 8 |
| b | 2 | 7 | 2 | 2 | 14 |
| c | 3 | 6 | 2 | 6 | 18 |
| d | 4 | 5 | 2 | 5 | 20 |

FIG. 3. A comparison of variability indices for three estimation methods, applied to four class frequency possibilities when $K$ (number of classes) equals 2.

Both methods are based on the CVN approach. The interspersion measure is the same as CVN; juxtaposition expands upon CVN by incorporating two weights:

- Spatial weighting of edge by pixel location, that is, the edges between element 5 and elements 2, 4, 6, and 8 in Figure 1a (the edges orthogonal to the center pixel) are each counted as two edges. The edges between the center and diagonal elements (elements 1, 3, 7, and 9) are weighted with value 1.
- A relative importance value that assigns a rank between 0 and 1 to edges between each possible pair of classes.

It is suggested here that an alternative estimate of neighborhood edge for use with the BCM index is a measure of class value changes between all adjacent neighborhood elements. The edge index need not be restricted to comparisons between the center and other eight neighborhood elements. Changes in nominal class values along rows and columns will estimate the total edge occurring within a neighborhood. A 3 by 3 neighborhood has 12 edges; therefore, the edge index ranges from 0 to 12. The edges are a subset of the relation $R$ in the binary comparison matrix (Figure 6).

In the example shown in Figure 5, Case One has three row changes and two column changes, for a total edge index of 5. Case Two changes four times across the rows and four times down the columns for a total edge index of 8. Although the two cases have identical values for $K$, $f_1$, $f_2$, and BCM, they have different edge index values. The edge index is sensitive to the spatial distribution of the elements in the neighborhood.

| | | |
|---|---|---|
| A | A | B |
| A | B | B |
| B | B | B |

a

| | | |
|---|---|---|
| A | B | B |
| B | A | B |
| B | B | A |

b

| | | |
|---|---|---|
| A | A | C |
| B | A | D |
| G | F | E |

c

| Example | K | $f_1$ | $f_2$ | $f_3$ to $f_7$ | NDC | CVN | BCM |
|---|---|---|---|---|---|---|---|
| a | 2 | 3 | 6 | – | 2 | 3 | 18 |
| b | 2 | 3 | 6 | – | 2 | 6 | 18 |
| c | 7 | 3 | 1 | 1 | 7 | 6 | 33 |

FIG. 4. Comparison of variability indices for three estimation methods applied to selected examples of 3 by 3 neighborhoods.

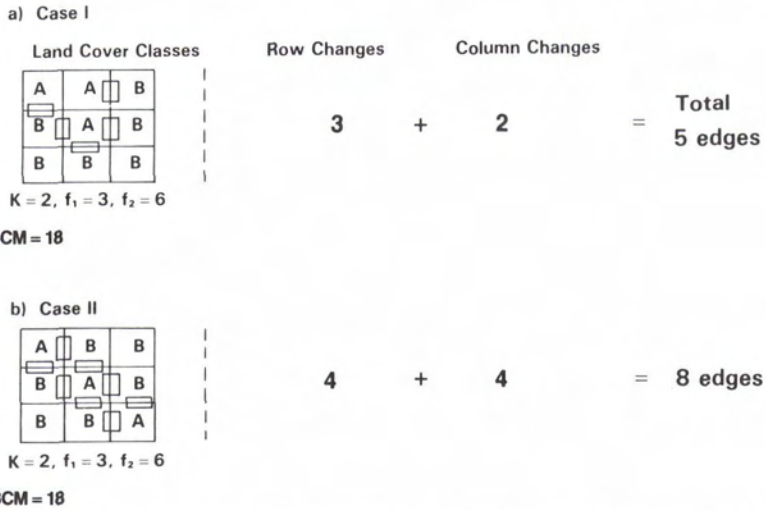



FIG. 5. Estimates of orthogonal edge for two examples of 3 by 3 neighborhoods with two classes, frequencies of 3 and 6.

## Application of BCM

The BCM technique was applied to a portion of a digital land-cover classification of the Kenai National Wildlife Refuge, Alaska. First, the land-cover classes were aggregated into major groups approximating physiognomic categories (Plate 1a) and then the BCM index was computed using the aggregated classes.

The index values present a spatial pattern (Plate 1b). The pattern contains areas of homogeneity (purple), edges between major land-cover classes (blue-green), and focal points of neighborhoods with high land-cover variability (yellow). Examples of homogeneous regions are the forested areas (dark green) and the lakes (blue) in the upper left, as well as the peatlands/wetlands (yellow) in the lower right, of Plate 1a. The transition zones between major land-cover types appear as linear features (blue-green) in Plate 1b. The upper right portion of Plate 1a contains many small land-cover regions; this condition produces the points of high variability (yellow) shown in Plate 1b.

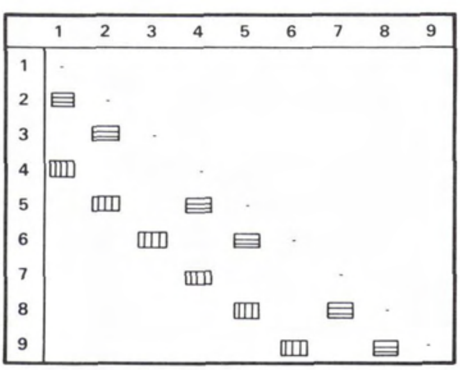


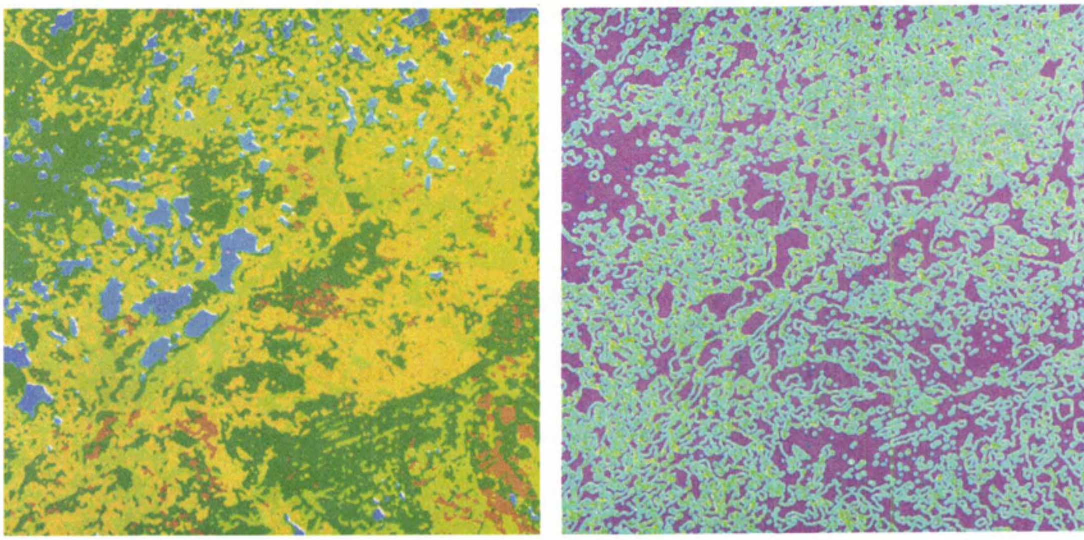

FIG. 6. Location of orthogonal edge relations in the binary comparison matrix.

Figure 7 is a histogram of the numerical distribution of index values. About 40 percent of the pixels occur in homogeneous neighborhoods (BCM = 0). Ten percent of the pixels are located in neighborhoods with index values greater than 23. The maximum index value in this example is 34. High values of BCM represent increasing neighborhood complexity as a function of the total number of classes and the distribution of the neighborhood elements between the classes.

## Conclusions

The binary comparison matrix is a technique to estimate neighborhood variability of nominal scale data in a raster format spatial data base. The method implements a binary operator in a 3- by 3-neighborhood function. The index is sensitive to both the number of classes occurring in a neighborhood and the frequency of neighborhood elements in each class. An examination of the topological characteristics of BCM, compared with the characteristics of the NDC and CVN indices, demonstrates the different aspects of variability measured by the three methods.

The index is not sensitive to the spatial distribution of land-cover class values; it does not measure edge. However, a subset of the relations in the binary comparison matrix can be extracted to estimate edges internal to the neighborhood. The edge index



(a) (b)

PLATE. 1 (a) Aggregated land-cover classes for a portion of the Kenai National Wildlife Refuge, Alaska. Dark green = forest, light green = shrub, yellow = peatlands/wetlands, brown = grasses and disturbed areas, blue = water. (b) Spatial pattern of BCM index values for a portion of the Kenai National Wildlife Refuge, Alaska. Purple = homogeneous areas, blue − green = edges between areas, yellow = points representing neighborhoods with high land-cover variability.

values are sensitive to the spatial distribution of class values in the neighborhood.

The technique has been implemented in two raster systems—the Interactive Digital Image Manipulation System (IDIMS) and the Remote Information Processing System (RIPS). Results of the spatial operator have been presented as maps to resource analysts interested in portraying landscape variability as a part of a land management plan. In addition, digital images of the BCM index have been included in a data base containing telemetry data, land-cover categories, and terrain variables to characterize wildlife utilization regions.

Neighborhood variability is a parameter impor-

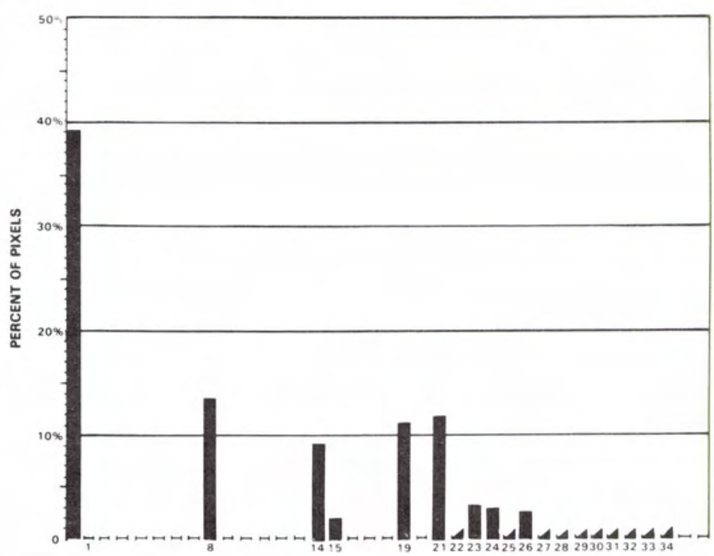



FIG. 7. Distribution of the BCM index derived from a classification covering a portion of the Kenai National Wildlife Refuge, Alaska.

tant to resource analysis applications of geographic information systems. Variability of nominal scale data can be measured in a geographic information system with neighborhood operators. The binary comparison matrix provides a variability statistic which quantifies the complexity of a neighborhood as a function of the number of classes and the distribution of neighborhood elements between the classes. The technique has utility both as a map product and as a variable in resource modeling applications.

Acknowledgments

The author acknowledges the assistance of Jan W. van Roessel to formally derive Equation 1, and the contribution made by Brian Dealey to implement the algorithm within the Remote Information Processing System.

References

Abler, R. J., J. S. Adams, and P. Gould, 1971. Spatial Organization. *The Geographer's View of the World*, Prentice-Hall, Inc., Englewood Cliffs, N.J., pp. 258-261.

Mead, R. A., T. L. Sharik, S. P. Prisley, and J. T. Heinen, 1981. A Computerized Spatial Analysis System for Assessing Wildlife Habitat from Vegetation Maps, *Canadian Journal of Remote Sensing*, Vol. 7, No. 1, pp. 34-40.

Munkres, J. R., 1975. *Topology: A First Course*, Prentice-Hall, Inc., Englewood Cliffs, N.J., p. 76.

Tomlin, C. D., and J. K. Berry, 1979. A Mathematical Structure for Cartographic Modeling in Environmental Analysis, *Proceedings, American Congress on Surveying and Mapping*, 39th Annual Meeting, Washington, D.C., pp. 269-283.

(Received 5 November 1982; revised and accepted 11 March 1985)