Using Spatial Autocorrelation Analysis to Explore the Errors in Maps Generated from Remotely Sensed Data

Russell G. Congalton
Department of Forestry and Resource Management, University of California, Berkeley, CA 94720

ABSTRACT: Three data sets of varying spatial complexity, including an agricultural area, a range area, and a forest area, were chosen for investigation in this study. A difference image was generated for each data set by comparing a Landsat classification with an assumed correct reference classification and noting the agreement and disagreement. Visual inspection and spatial autocorrelation analysis were used to identify and quantify the patterns of error within each difference image. This information is very important when land cover maps generated from remotely sensed data are sampled for accuracy assessment.

INTRODUCTION

IT HAS LONG BEEN RECOGNIZED that, because of sensor resolution and landscape variability, remotely sensed data exhibits a dependency between neighboring pixels. This dependency, called spatial autocorrelation, influences many aspects of remotely sensed data. For example, the entire classification process, especially the selection of training areas, is affected by this relationship between neighboring pixels. In addition, the results of preprocessing algorithms such as smoothing and edge enhancement can be drastically influenced by the amount of autocorrelation exhibited by a particular image. Probably the aspect of remote sensing most affected by spatial autocorrelation is in the analysis of classification error. What are the patterns of error found in remotely sensed data? How are these patterns quantified? How do we properly sample the remotely sensed data to determine the error, given some amount of spatial autocorrelation? These questions must be answered in order to assess the accuracy of maps derived from remotely sensed data. Therefore, it is the objective of this paper to investigate the spatial autocorrelation of remotely sensed data and determine its affects on accuracy assessment.

SPATIAL AUTOCORRELATION ANALYSIS

Spatial autocorrelation analysis is not new. The techniques have long been used by geographers and social scientists. In the 1940s Cruickshank (1940, 1947) investigated the regional variations in the incidence of cancer throughout England. Cox (1969) used spatial autocorrelation analysis to study the regional distribution of democratic voters in presidential elections. More recently, autocorrelation has been applied to ecological studies and the natural sciences. Reed (1982) used autocorrelation analysis to study and model competition in Loblolly pine stands. Campbell (1981) recognized the importance of these techniques in the analysis of remotely sensed data and applied them during the classification process.

As can be seen in the above examples, spatial autocorrelation involves the effect that a certain quality or characteristic at a given location has on that same quality at neighboring locations. In other words, given a group of mutually exclusive units or individuals in a two-dimensional plane, if the presence, absence, or degree of a certain characteristic affects the presence, absence, or degree of the same characteristic in neighboring units, then the phenomenon is said to exhibit spatial autocorrelations (Cliff and Ord, 1973). This technique is directly applicable to error analysis or accuracy assessment of remotely sensed data because each pixel is a mutually exclusive unit and the characteristic of interest is whether or not the pixel has been correctly classified.

The characteristic of interest in spatial autocorrelation analysis can be either continuous or discrete, and the method of analysis varies accordingly. In exploring errors in remotely sensed data, the characteristic of interest is discrete and from a binary distribution (i.e., it is either correct or it is an error). The method of spatial autocorrelation analysis for discrete variables is called the join count statistics and is described by Moran (1948). Let the characteristic of interest, X, be defined such that

$x_i = 1$ if the unit i possesses the characteristic of interest, and

$x_i = 0$ if the unit i does not possess the characteristic of interest.

Also let the join matrix, (δ_{ij}), be defined such that

$\delta_{ij} = 1$ if the unit i and the unit j have a boundary of positive nonzero length in common, and

$\delta_{ij} = 0$ if the unit i and the unit j do not have a boundary of positive nonzero length in common.

The total number of observed 1–1 joins (i.e., the number of pairs of units which both possess the characteristic of interest, X) is given by

$$1\text{–}1 \text{ joins} = 1/2 \sum_{(2)} \delta_{ij}\, x_i\, x_j$$

and the total number of observed 1–0 joins (i.e., the number of pairs of units in which one unit possesses the characteristic of interest, X, and the other does not) is given by

$$1\text{–}0 \text{ joins} = 1/2 \sum_{(2)} \delta_{ij}(x_i - x_j)^2$$

$$\text{where } \sum_{(2)} = \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n}.$$

The total number of observed 0–0 joins can be determined by subtraction because the number of 0–0 joins is equal to the total number of joins minus the 1–1 joins and the 1–0 joins. The total number of joins in a two-dimensional plane is defined as

$$\text{total joins} = 1/2 \sum_{i=1}^{n} L_i$$

where L denotes the number of units joined to the *i*th region (Cliff and Ord, 1973).

The method, then, to test if a phenomenon exhibits spatial

autocorrelation (i.e., the observed join count is significantly different from random) depends on the asymptotic normality of the join count statistics. Because of this normality condition, the first two moments of the join count statistics can be used to specify the location (mean) and the scale (variance) parameters of the normal distribution. These moments can then be evaluated under either of two assumptions (Cliff and Ord, 1973):

- Free sampling (sampling with replacement), where each individual or unit is independently assigned the characteristic of interest, X, with probability p and $q = (1-p)$, respectively, or
- Nonfree sampling (sampling without replacement), where each unit has the same a priori probability of being assigned the characteristic of interest, X, subject to the overall constraint that n_1 units take the value x_i to be one and n_2 units take the value x_i to be zero where $n = n_1 + n_2$.

The nonfree sampling assumption is the correct one for exploring the errors in remotely sensed data because p is usually unknown. However, the values of n_1 and n_2 are known after the entire population has been observed. The first two moments of the join count statistics under the nonfree assumption can be found in Appendix A (Moran, 1948). For a more detailed description of spatial autocorrelation analysis, see Congalton (1984).

DATA

Three data sets of varying spatial complexity were used in this study to explore on a pixel-by-pixel basis the patterns of error in remotely sensed data. A USGS 7 1/2-minute quadrangle-sized area of a forested environment, a rangeland environment, and an agricultural environment was selected. For each quadrangle sized area, a Landsat MSS land-use/land-cover classification and a reference classification were available. Because of the lack of complete ground data for two of the three quadrangles, it was necessary to assume a reference classification. This assumption was valid because ancillary ground data were used in the classification process to improve the assumed correct classification (reference classification). Additional details about how these reference classifications were derived can be found in the paper cited for each data set. Also, it is important to recognize that it is not the specific classification that was being studied here, but rather the pattern of error expressed by the difference of the two classifications. Detailed descriptions of each study area are given so that the pattern of error in the difference image can be related to the vegetation complexity of the area.

The agricultural data set is the Clarke quadrangle, an area of center pivot irrigation in the Umatilla River Basin of north-central Oregon. The Landsat classification was generated from a multi-temporal composite of June, July, and August 1979 data and used a modified clustering approach with a maximum-likelihood algorithm. The resulting classified image was resampled to a 1-acre (63.6-metres square) pixel size and registered to a Universal Transverse Mercator (UTM) map projection (Loveland and Johnson, 1981). The area was classified into nine land-use categories including wheat, alfalfa, potatoes, corn, beans, apples, pasture, nonirrigated land, and unclassified. The reference classification for the agricultural data set was generated from actual ground visits and aerial photography.

The rangeland data set is the Lizard Point quadrangle, an area in the northwest corner of Arizona in Mojave County and is indicative of the southwest desert environment. The land-use/land-cover map was generated from 26 August 1977 Landsat MSS scene using an unsupervised maximum likelihood classifier. The classified image was resampled to a pixel size of 50-metres square and registered to a UTM map projection (Rhode and Miller, 1980). The area was classified into nine land-cover categories, including cropland/pasture, coniferous forest, evergreen woodland, deciduous woodland, Mohave desert shrub, Great Basin desert shrub, mountain shrub, plains grassland,

and barren. The reference classification was generated using ancillary DMA terrain data.

The forest data set is the Garden Point quadrangle, a forested area located in the Lolo Creek section of western Montana. The Landsat MSS data were obtained on 13 July 1974 and classified using a layered maximum likelihood approach. The classified scene was resampled to a 60-metre square pixel and registered to a UTM map projection (Shasby et al., 1981). The classified image consisted of 12 land-use/land-cover categories, including perennial grass, sagebrush/shrub, open Ponderosa pine, healthy short needle conifer, decadent short needle conifer, logging slash/clearcut, barren, nonforest/agriculture, and four road classes. The reference classification was generated using ancillary digital elevation matrix (DEM) terrain data.

METHODS

The exploration of errors in maps generated from remotely sensed data was divided into two steps. The first step was to generate difference images from the original data. The second step was to apply spatial autocorrelation analysis and study the patterns of error.

DIFFERENCE IMAGE GENERATION

An image called a difference image was generated for each of the three quadrangle-size data sets. Each image consists of a matrix of zeros and ones in which the ones indicate disagreement (error) between the classified image and the reference image and the zeros indicate agreement. All further analysis was based on these difference images. Each image was then displayed in order to visually examine the extent and pattern of error in each. A similar type technique was employed very effectively by Hudson (1987).

SPATIAL AUTOCORRELATION ANALYSIS

The characteristic of interest in analyzing each difference image is error. Because this characteristic is discrete and binary, the proper method of analysis is the join count statistics under the nonfree sampling assumption. In terms of relative efficiency, it is better to use the 1–0 join statistic over the 0–0 or the 1–1 join statistics. It is also recommended that the proportion of the characteristic of interest be as close to 0.5 as possible. Because this situation is not the usual case with a characteristic of interest such as error, then it is better to use the join count statistic for the proportion greater that 0.5 than the join count statistic for the proportion less than 0.5. For testing purposes using the normal distribution, it is better to use the join count statistic for the 1–1 joins rather than the 0–0 joins (Cliff and Ord, 1973).

The null hypothesis tested using spatial autocorrelation analysis is that the characteristic of interest, error, has an equal probability of occurrence in every cell and the error in one cell is independent of the error in every other cell. In other words, the error is randomly distributed throughout the difference image. The first two moments of the join count statistics (see Appendix A) are used to compute a measure of autocorrelation and to test the null hypothesis in each difference image. The test statistic used is the standard normal deviate. The normal assumption is valid in the case of the difference images because of the central limit theorem and the very large sample size. The analysis was performed by a FORTRAN computer program written by the author. If the null hypothesis is rejected at some given confidence level, then the error is not random but, rather, exhibits spatial autocorrelation. In addition, if the 1–0 join count standard normal deviate is positive, then the error is uniformly distributed throughout the difference image. On the other hand, if the null hypothesis is rejected and the 1–0 join count standard normal deviate is negative, this implies that the characteristic of interest, error, is clustered (Cliff and Ord, 1973).

A test of spatial autocorrelation can be performed for each successive lag (see Figure 1) until a lag is found in which no autocorrelation exists. Each lag is defined by a new join matrix. For example, lag 1 consists of the join matrix between the central pixel and its nearest surrounding pixels, while lag 2 consists of the join matrix between the central pixel and the next pixels outward from the surrounding pixels, and so on. Therefore, the first lag in which no autocorrelation exists defines the distance between independent error pixels.

In addition to testing for autocorrelation, the direction of the dependency can also be tested. The join counts are classified into three possible cases as described by the directions moved by the corresponding chess pieces. In the rook's case, the joins occur in the north-south or east-west direction while the bishop's case defines joins that occur diagonally. The queen's case is a combination of the rook's and bishop's case. Observation of the joins for the three cases can lead to identifying the preferred direction of autocorrelation in each difference image.

RESULTS

A difference image was generated for each of the three study areas by comparing pixel by pixel the Landsat classification with the assumed correct reference data. In every pixel that the Landsat classification agreed with the reference data (i.e., no error), a zero was placed in the difference image. In each pixel where the two data sets disagreed, a one was placed in the difference image. Table 1 contains the population parameters for the three difference images. The image sizes vary slightly because the agricultural data set was somewhat larger than the other two data sets and also because the pixel size varies for each image. Visual display of the difference images was achieved by assigning the ones (error) to black and the zeros (correct) to white.

Figure 2 shows the agricultural difference image. Notice that the errors occur in large blocks. This result is expected because agricultural crops are grown in large homogeneous areas. Therefore, if a crop type is misclassified, a large block of error

TABLE 1. POPULATION PARAMETERS FOR THE THREE DIFFERENCE IMAGES

	Difference Image		
	Agriculture	Range	Forest
Image shape (pixels)	245 × 176	240 × 160	241 × 144
Image size (pixels)	43,120	38,400	34,704
Pixel size (metres square)	63.6	50	60
Population mean (proportion incorrect)	0.2411	0.3243	0.2180
Population variance	0.1830	0.2191	0.1705

FIG. 2. The agriculture difference image.

results. For this reason, the agricultural data set was considered the simplest or least spatially diverse of the three data sets.

Figure 3 shows the rangeland difference image. The pattern of error here is less blocky than that in the agricultural difference image but more blocky than in the forest difference image. The rangeland area has a mixture of both large homogeneous grass areas and small heterogeneous woody vegetation areas. It is this mixture that causes confusion along boundaries (i.e., edge effects) resulting in a more linear pattern of error in the difference image. The rangeland data set is considered to be in between the agricultural and forest data sets in terms of spatial diversity.

The most spatially complex data set is the forest data set, as indicated by the difference image (Figure 4). The pattern of error is the least blocky and most linear of the three difference images. This linear pattern is caused by confusion along bound-
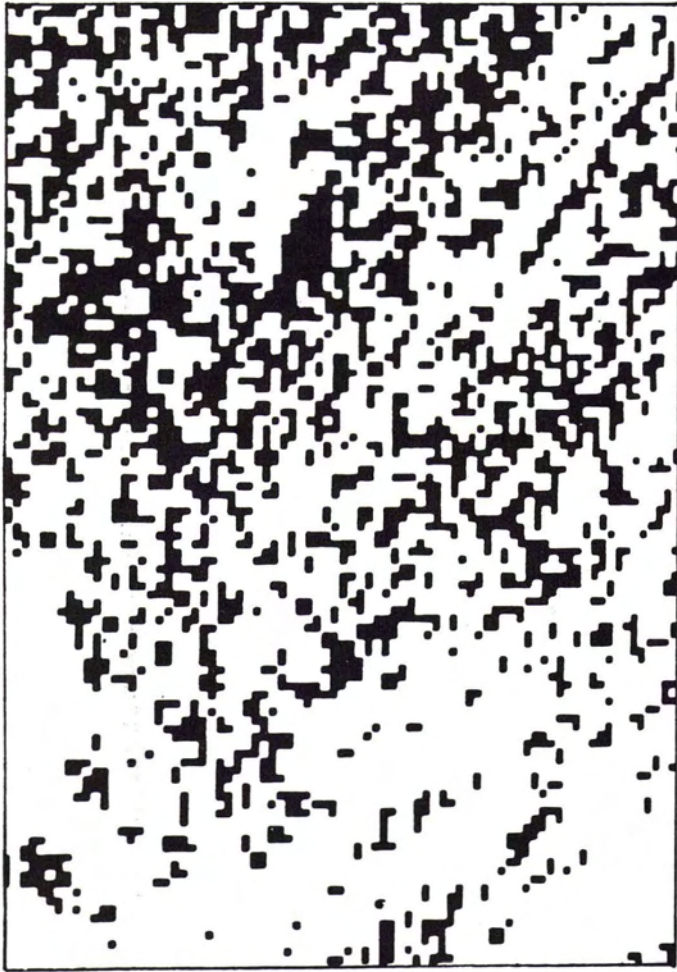

C = central pixel
L1 = joins for the first lag with the central pixel
L2 = joins for the second lag with the central pixel
⋮
LN = joins for the nth lag with the central pixel

FIG. 1. Diagram showing the lags used in spatial autocorrelation.

FIG. 3. The range difference image.

FIG. 4. The forest difference image.

Agriculture Autocorrelation

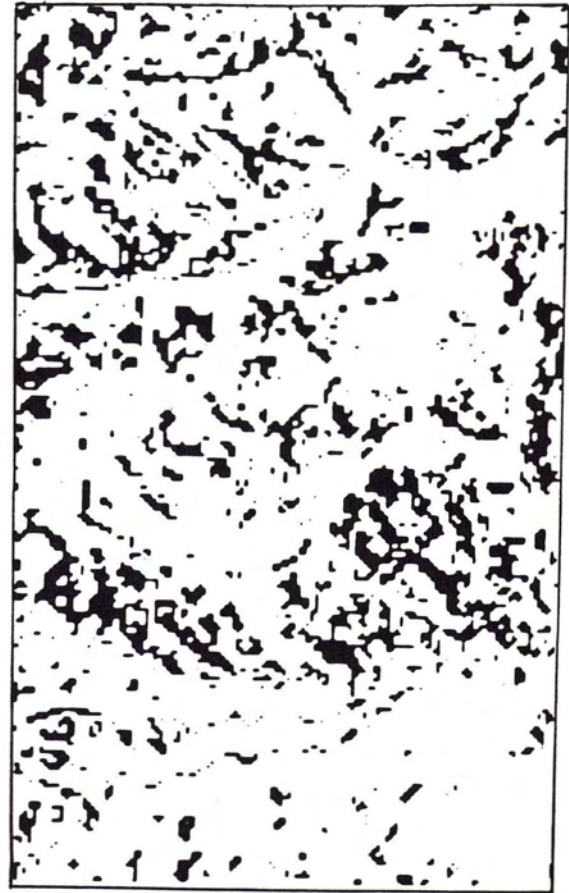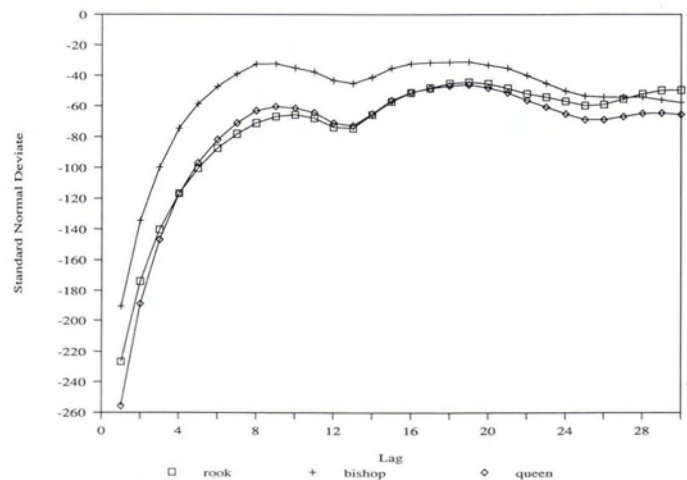FIG. 5. A plot of the spatial autocorrelation analysis for the 1–0 join counts on 30 lags of the agriculture difference image.

aries of various land-cover categories. This edge effect may also be confounded by topography, namely, the shadowing caused by the linear arrangement of mountains in the scene.

The spatial autocorrelation analysis was performed on each difference image using the 1–0 join count statistic. The analysis was run for 30 lags in order to find the distance at which the errors became independent (i.e., exhibited no spatial autocorrelation). The standard normal deviate, the test of the null hypothesis, and the resulting spatial distribution (random, clustered, or uniform) were determined for each successive lag and for the three possible directional cases (rook, bishop, and queen). The results of the analysis are presented graphically for ease of understanding.

Figure 5 shows the results of the spatial autocorrelation analysis for the 1–0 join count statistic on 30 lags of the agricultural difference image. In all lags the null hypothesis was rejected and, therefore, the error is said to exhibit spatial autocorrelation. The negative standard normal deviate (i.e., test statistic) indicates a positive autocorrelation and, therefore, a clumped or clustered distribution. The graph shows the standard normal deviate increasing (getting closer to zero) in the first five lags and then fluctuating over a narrow range for the remaining 25 lags. It is not unexpected that an error in one location in the difference could have a positive influence on a pixel 30 pixels or less away. In fact, visual observation of he blockiness of the distribution of error in the difference image would lead one to agree with this result.

The results of the spatial autocorrelation analysis using the

1–0 join count statistic on 30 lags of the rangeland difference image are given in Figure 6. Here again all the results are significant (reject the null hypothesis) and all the standard normal deviates are negative, indicating a positive autocorrelation. Therefore, the distribution of error in the range difference image is also clumped. Note that the graph shows the standard normal
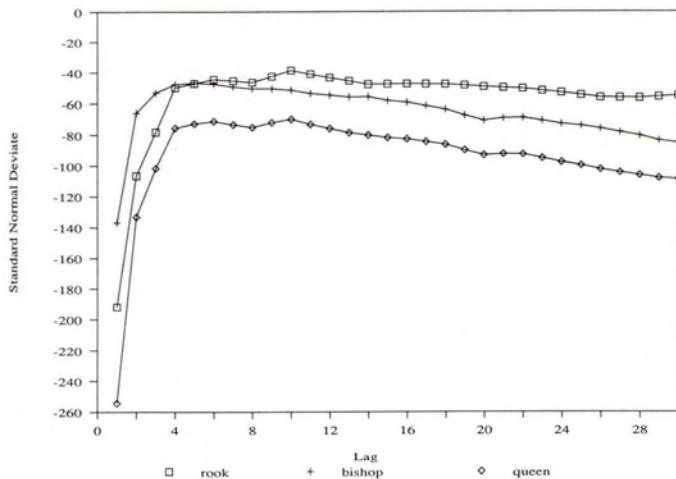
Range Autocorrelation

FIG. 6. A plot of the spatial autocorrelation analysis for the 1–0 join counts on 30 lags of the range difference image.
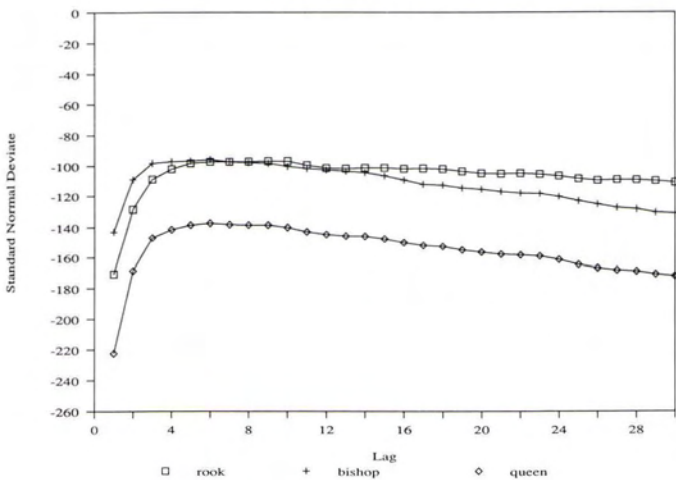
Forest Autocorrelation

FIG. 7. A plot of spatial autocorrelation analysis for the 1–0 join counts on 30 lags of the forest difference image.

deviate increasing until about lag 10 where it slowly begins to decrease again. The standard normal deviates for the rook's case are relativley constant after the fourth lag. This result indicates a constant autocorrelation effect in the horizontal and vertical directions while the diagonal (bishop's case) direction is much more variable.

Figure 7 presents the results of the spatial autocorrelation analysis for the 1–0 join count statistic on 30 lags of the forest difference image. Again, all the lags were shown to be significant and all the standard normal deviates are negative, resulting in positive spatial autocorrelation. It is somewhat unexpected that the error in the forest difference image would be clustered, especially noting the linear trend of the error in the difference image. However, it appears that this edge effect error has more of a clumped nature than a uniform one. While errors 30 pixels apart are still positively correlated, they are not correlated to the same magnitude as in the agricultural difference image. The forest difference image exhibits much of the same behavior as the range difference image. The standard normal deviates tend

to increase until lag 9 or 10 and then slowly decrease again. Also the standard normal deviates for the rook's case are somewhat constant after lag 3.

The spatial complexity of the three data sets, as observed in the visual display of the difference images, is also evident in the spatial autocorrelation results. In lag 1, the standard normal deviate is most negative for the agriculture difference image, indicating that it is most clumped, while the standard normal deviate for the forest difference image is least negative. These results complement the original notion that the agricultural data set is the least spatially complex while the forest data set is most complex.

CONCLUSIONS

The creation of difference images and the use of spatial autocorrelation analysis can lead to a better understanding of the patterns of error in land-cover maps generated from remotely sensed data. This knowledge is useful for many reasons. The pattern of error may be influenced by ground features such as topography or by computer processing techniques such as resampling. It is important in assessing the accuracy of such maps that some idea of the pattern of error be known so that the correct sampling scheme will be used (Congalton, 1988).

It is expected that, with the increased spatial resolution of the Thematic Mapper (TM) and SPOT satellite data, spatial autocorrelation techniques will play an even greater role in understanding the pattern of error. Continued research in this area is necessary in order to fully understand and make use of the information contained in remotely sensed data.

REFERENCES

Campbell, James, B., 1981. Spatial Autocorrelation Effects upon the Accuracy of Supervised Classification of Land Cover. *Photogrammetric Engineering and Remote Sensing*, Vol. 47, No. 3, pp. 355–363.

Cliff, A.D., and J.K. Ord, 1973. *Spatial Autocorrelation*. Pion Limited. London, England. 178 p.

Congalton, Russell G., 1984. *A Comparison of Five Sampling Schemes Used in Assessing the Accuracy of Land Cover-Land Use Maps Derived from Remotely Sensed Data.* Ph. D. Dissertation. Virginia Polytechnic Institute and State University. 147 p.

———, 1988. A Comparison of Sampling Schemes Used in Generating Error Matrices for Assessing the Accuracy of Maps Generated from Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 5, pp. 593–600.

Cox, K.R., 1969. The Voting Decision in a Spatial Context. *Progress in Geography* (L.C. Board, R.J. Charley, P. Haggett, and D.R. Stoddart, eds.). Arnold, London. pp. 81–117.

Cruickshank, D.B., 1940. A Contribution Towards the Rational Study of Regional Influences: Group Information under Random Conditions. *Papworth Research Bulletin*, Vol. 5, pp. 36–81.

———, 1947. Regional Influences in Cancer. *British Jour. of Cancer*, Vol. 1, pp. 109–128.

Hudson, W.D., 1987. Evaluating Landsat Classification Accuracy from Forest Cover Type Maps. *Canadian Journal of Remote Sensing*, Vol. 13, No. 1, pp. 39–42.

Loveland, T.R., and G.E. Johnson, 1983. The Role of Remotely Sensed and other Spatial Data for Predictive Modeling—The Umatilla, Or-

egon Example. *Photogrammetric Engineering and Remote Sensing*, Vol. 49, No. 8, pp. 1183–1197.

Moran, P.A.P., 1948. The Interpretation of Statistical Maps. *Jour. Royal Stat. Soc.*, Series B, Vol. 10, pp. 243–251.

Reed, David D., 1982. *The Spatial Autocorrelation of Individual Tree Characteristics in Loblolly Pine Stands*. Ph.D. dissertation. Virginia Polytechnic Institute and State University. 126 p.

Rhode, W.G., and W.A. Miller, 1980. *Arizona Vegetation Resource Inventory (AVRI) Project Final Report*. USGS EROS Data Center internal report to BLM. Sioux Falls, S.D. 200 p.

Shasby, M.B., R.E. Burgan, and G.R. Johnson, 1981. Broad Area Forest Fuels and Topography Mapping Using Digital Landsat and Terrain Data. *Proceedings of the Seventh International Symposium on Machine Processing of Remotely Sensed Data*. Purdue University, West Lafayette, Indiana. pp. 529–538.

APPENDIX A
JOIN COUNT STATISTICS

The first two moments of the join count statistics under the assumption of nonfree sampling (Moran 1948).

Moments of the 1-1 join count statistics:

$$u_1'(1-1) = \frac{An_1^{(2)}}{n^{(2)}}$$

$$u_2(1-1) = \frac{An_1^{(2)}}{n^{(2)}} + \frac{2Dn_1^{(2)}}{n^{(3)}}$$
$$+ \frac{[A(A-1)-2D]n_1^{(4)}}{n^{(4)}} - \left[\frac{An_1^{(2)}}{n^{(2)}}\right]^2$$

Moments of the 0-0 join count statistics:

$$u_1'(0-0) = \frac{An_2^{(2)}}{n^{(2)}}$$

$$u_2(0-0) = \frac{An_2^{(2)}}{n^{(2)}} + \frac{2Dn_2^{(3)}}{n^{(3)}} + \frac{[A(A-1)-2D]n_2^{(4)}}{n^{(4)}} - \left[\frac{An^{(2)}}{n^{(2)}}\right]^2$$

Moments of the 1-0 join count statistic:

$$u_1'(1-0) = \frac{2An_1n_2}{n^{(2)}}$$

$$u_2(1-0) = \frac{2An_1n_2}{n^{(2)}} + \frac{4[A(A-1)-2D]n_1^{(2)}\,n_2^{(2)}}{n^{(4)}}$$
$$+ \frac{2Dn_1n_2(n_1+n_2-2)}{n^{(3)}} - 4\left[\frac{An_1n_2}{n^{(2)}}\right]^2$$

Notation:

$u_1'(x)$ = the first moment of x about the origin (i.e., the expected value of x).

$u_2(x)$ = the second moment of x about the mean (i.e., the variance of x).

$n^{(i)}$ = $n(n-1)\dots(n-i-1)$.

n = the total number of regions in the population.

n_1 = the total number of regions in the population with the characteristic of interest.

n_2 = the total number of regions in the population without the characteristic of interest.

A = $1/2 \sum_{i=1}^{n} L_i$.

D = $1/2 \sum_{i=1}^{n} L_i(L_i - 1)$.

L_i = the number of regions joined to the ith region.