# 10 Spatial statistics and GIS: an integrated approach

*Kenneth L. Kvamme*

## 10.1 INTRODUCTION

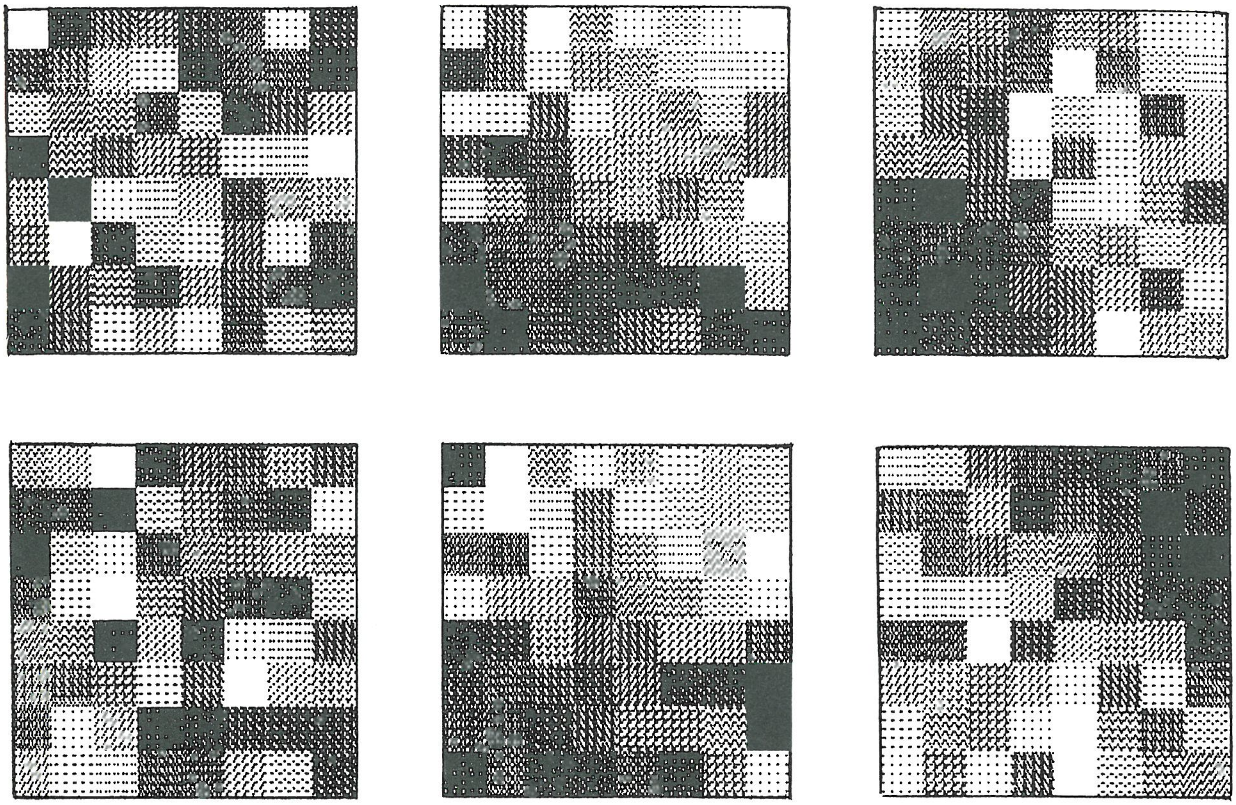
There is a two–fold purpose to this paper. The first is to emphasize the need for spatial statistics. The second is to suggest that Geographic Information Systems (GIS) may be the ideal place to implement these statistical procedures. By "spatial statistics" I do not mean the application of the conventional tests described in most statistics books to data sets that are spatial in origin. Conventional tests usually require independence among observations, something that generally is untrue of spatially distributed information, and these procedures usually are aspatial in nature and design. Spatial statistics, on the other hand, take into account the locational component of the data and, at the same time, allow for the resultant interdependencies that occur between observations (Cliff & Ord 1981).

GIS, of course, are computer software designed for the efficient and rapid retrieval, manipulation, processing, and display of spatially referenced information (Burrough 1986). Consequently, they are a logical place to implement spatial statistical tests. Various GIS components, such as sub–systems for encoding, manipulation, and display, can greatly facilitate the spatial data analysis and interpretation process. This is no trivial matter. Moreover, recent work in spatial analysis (e.g., Harris & Lock 1990) has emphasized the joint importance of the domains of visualization and quantitative appraisal. They complement each other and both are essential to spatial inquiry. The excellent display capabilities of GIS, together with embedded systems for quantitative analysis, can provide an ideal environment for spatial investigations.

A comprehensive overview of spatial statistics is beyond the scope of this paper. What I hope to offer is a demonstration of the weakness and fallibilities of conventional aspatial statistical tests when applied to geographically distributed data. By emphasizing weaknesses, the superiority of procedures that incorporate spatial information will be enhanced. I also hope to show some of the inherent problems that arise when dealing with the quantitative analysis of spatial information, particularly those stemming from interdependencies that occur among observations. This will be accomplished initially through the presentation of some simple example data sets pertaining to the topic of spatial association—the relationship between two variables measured at identical locations over space. The discussion will ultimately conclude, however, with a somewhat more complex example: a modification of Student's t–test that allows for spatial data and exemplifies many of the issues at hand. All of this will be undertaken in contexts that employ the data retrieval, manipulation, and display capabilities of GIS demonstrating, I hope, the importance of GIS to modern spatial analysis.

## 10.2 THE SPATIAL COMPONENT OF INFORMATION

The topic of spatial association offers a convenient way to illustrate many of the issues and difficulties that arise when dealing with geographically distributed information. Spatial association simply refers to the relationship between two variables measured at identical locations over space. Two artificial data sets are utilized, each forming an $8 \times 8$ matrix for $n$=64 observations (Figure 10.1a). The data in each matrix were obtained from a random number generator, sampling without replacement integers in the interval 1–64. The data might represent frequencies for artifact types A versus B or perhaps the



a b c

*Figure 10.1: Spatial association between top and bottom images showing in column a) no association, in column b) positive association, and in column c) negative association.*

transformed results of magnetometer and soil resistivity readings, respectively.

### 10.2.1 Conventional (aspatial) statistics

The point–to–point association between the two matrices can be assessed through a conventional correlation measure. Because the data in each matrix represent a sequence of unique integers they can be treated as ranks and Spearman's $r_S$ or Pearson's $r$ will give an identical result (Siegel 1956:203). Since the data in Figure 10.1a were derived through random number generation we can expect a correlation of about zero; actualy a large number of these data sets were generated until a pair was achieved with a correlation very near zero, in fact $r=r_S=0.002$. Visually, there does not seem to be any spatial association between the images. This is as it should be: visual impressions and quantitative findings should agree.

It should be realized, however, and emphasized, that the foregoing conventional measures of correlation, Pearson's $r$ and Spearman's $r_S$, are aspatial in nature. That is, they measure correlation strictly between common points and fail to consider the association that may exist between nearby or even adjacent locations. Consequently, commonalities that may exist between nearby values cannot be detected making these statistics solely measures of what has been termed "point–to–point" association (Hubert *et al.* 1985). This can be forcefully demonstrated through some simple reshuffling exercises. If the 64 pairs of values in the two images of Figure 10.1a are kept constant and unchanged, but only their ordering or position in the matrix is altered, we can restructure the spatial arrangement of the data yet maintain the same lack of correlation.

In Figure 10.1b one of the 64! possible orderings of the paired data is given that visually points to strong positive spatial association because low, medium, and high values tend to occur in similar regions. Yet, the two values in each grid square are the same 64 pairs of integers that indicate zero correlation. All that is changed is their spatial position. Figure 10.1c shows a spatial arrangement that visually suggests strong negative association with the same uncorrelated data.

These reshuffling examples clearly indicate the unsuitability of conventional statistics that fail to consider the locational component of the data in spatial contexts. The statistics uniformly point to no correlation even though positive or negative association is readily apparent in some of the arrangements (Figure 10.1).

### 10.2.2 Spatial statistics

To illustrate the superiority of statistics that consider spatial information we can turn to one of the more commonly known ones, Moran's *I*, which has, in fact, been employed a great deal in the archaeological literature (e.g., Hodder & Orton 1976; Chadwick 1978; Whitley & Clark 1985; Kvamme 1990a). The reason for selecting this statistic is that it is relatively easy to work with and understand, a simple transformation allows it to detect spatial association, and it is implemented in several common GIS packages.

Moran's *I* attempts to measure the strength of correlation of a single variable with itself over space, or what is known as spatial autocorrelation (Cliff & Ord 1973). Consequently, it differs from the previous correlation measures because it does not assess the relationship between two variables, and it explicitly incorporates spatial information. It is defined as:

$$[1] \quad I = \left( \frac{n}{\sum_i \sum_j W_{ij}} \right) \frac{\sum_i \sum_j W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where $\bar{x}$ is the observed mean of the variate of interest. Thus, each observation $x_i$ is compared with each other observation $x_j$ considering the spatial relationship between them expressed by $W_{ij}$. The latter can be regarded as a weight and can take on many forms. Perhaps the simplest is a binary scheme where $W_{ij} = 1$ for adjacent locations; $W_{ij} = 0$ otherwise. Another common weighting scheme is $W_{ij} = 1/d_{ij}$, where $d_{ij}$ is the Euclidean distance between locations *i* and *j*. In the latter nearby observations receive much more weight than those widely separated, agreeing with what Tobler (1970:234) has referred to as «the first law of geography: everything is related to everything else, but near things are more related than distant things.» Many other weighting schemes are possible and can be based on explicit knowledge of spatial relationships or interaction (Cliff & Ord 1973; Hodder & Orton 1976:178). Importantly, the sampling distribution of *I* is known and well–documented, allowing construction of significance tests (Cliff & Ord 1973, 1981).

What has a measure of spatial autocorrelation, the relationship of a single variable with itself over a region, have to do with our problem of the spatial association between two variables? Quite simply, we can define a single new variable, *z*, that characterizes the covariation between the two original variables, *x* and *y*, location–by–location in our three example data set pairs of Figure 10.1. If there is true (positive or negative) spatial association between any of the image pairs then similar values of *z* will tend to cluster, Moran's *I* should indicate significant spatial autocorrelation in *z*, and therefore spatial association between *x* and *y*. Several indices of covariation are possible candidates such as $z_i = x_i y_i$, $z_i = |x_i - y_i|$ or $z_i = (x_i - y_i)^2$. In order to bear some similarity with the previous correlation measures, the following is employed $z_i = (x_i - \bar{x})(y_i - \bar{y})$.

With GIS it is quite easy to generate this new variable using techniques often referred to as "Map Algebra" (Berry 1987). The following steps might be undertaken:

1) the mean of each map held within a GIS, representing the *x* and *y* variables, is first computed;
2) the means are subtracted from the respective values of each location in the maps;
3) a new map representing *z* is generated by taking the product of the previous outcomes.

This process was undertaken for each of the map pairs shown in Figure 10.1. The results, given in Figure 10.2, suggest the success of this tactic. Figure 10.2a, derived from the initial random order of Figure 10.1a, appears equally random with no indication of spatial pattern in *z*. Yet, Figures 10.2b and 10.2c, representing the spatial distribution of z resulting from apparent positive (Figure 10.1b) and negative (Figure 10.1c) spatial association, respectively, clearly indicate pattern. That is, high values of *z*, indicating large same–direction deviations from the respective means of *x* and *y*, cluster in similar regions, as do low values which represent opposite–direction mean deviations in *x* and *y*.

These impressions can be tested statistically with Moran's *I*. One sampling distribution for *I* assumes that the variate of interest is normally distributed; a second does not and is based on a randomization assumption (Cliff & Ord 1973, 1981; see Kvamme 1990a for a worked example). The latter is chosen here because the data are treated as ranks. The results, using reciprocals of Euclidean distances between cell centers as weights, are pleasing. The *z* data in Figure 10.2a, resulting from random data in random order,

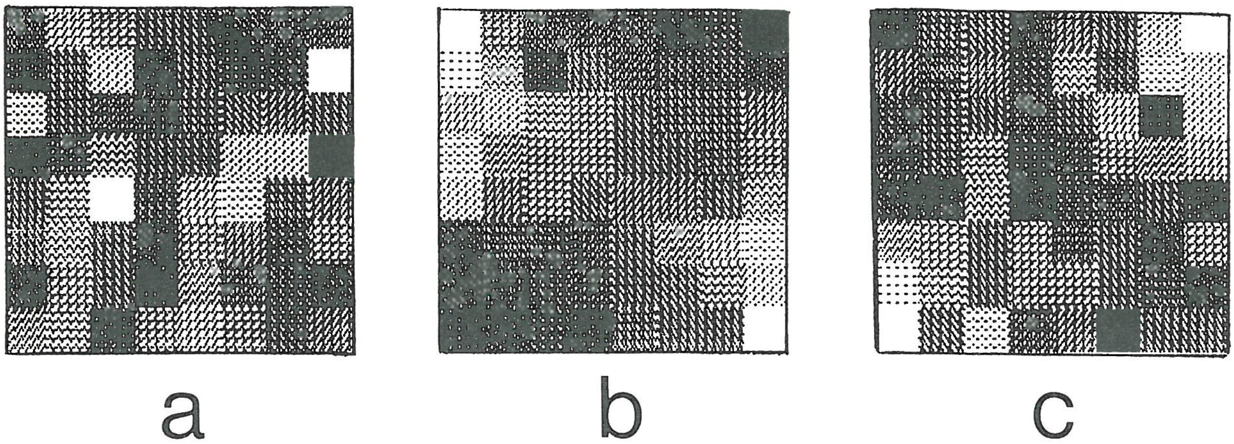



*Figure 10.2: Bivariate transformation of the images in each column of Figure 10.1 showing a) no pattern and b,c) pattern, or spatial autocorrelation.*

exhibit no significant spatial autocorrelation ($I = -.029$; $p=.366$). On the other hand, the $z$ data in Figures 10.2b and 10.2c, representing the strong spatial association shown in Figures 10.1b and 10.1c, yield highly significant levels of autocorrelation ($I_{pos} =.097$; $p <.001$; $I_{neg} =.028$; $p =.003$). Thus, a simple transformation of two variables, coupled with a proper statistic that incorporates spatial information, Moran's $I$, is able to correctly detect spatial association when it is present, as well as indicate when it is not. This contrasts quite dramatically with the conventional aspatial correlation measures, Pearson's $r$ and Spearman's $r_S$, which completely failed to detect any association when it was present, yielding in all cases values near zero.

The approach summarized here for detecting spatial association, although not described previously in the literature, was presented largely to illustrate properties of spatial statistics using a relatively well–known and simple spatial autocorrelation measure, Moran's $I$. Much more rigorous and insightful procedures exist for confronting the spatial association problem which may be found in the works of Hubert *et al.* (1985), Hubert & Golledge (1982), Tjostheim (1978), and Gorenflo & Gale (1986).

## 10.3 EFFECTIVE SAMPLE SIZE

A second important and related problem when dealing with the analysis of spatially distributed information is the effective sample size. This issue is best illustrated by returning to the spatial association problem and a conventional measure of correlation. Suppose one is working with a single hectare, 100 × 100 m square, and is able to make systematic observations of two variables. The variables might be slope angle, elevation, depth to bedrock, or a magnetometer reading, for example. A grid spacing of 50 m might be initially chosen, yielding a 3 × 3 matrix of observations for each variable. This situation is depicted in Figure 10.3a where each variable illustrates a weak north–south trend (reflected by the gray levels). Pearson's $r$ is a modest $r=.32$ that is nowhere near significant with a sample size of only $n=9$ ($p \approx .40$).

Now suppose that it is important to show that the two variables illustrate statistically significant correlation through use of Pearson's $r$. Even though the relationship seems to be rather weak, with $r=.32$, it is common knowledge that significance is a function not only of $r$, but also of sample size, $n$. In fact, assuming bivariate normality and independent observations, the statistic

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

which is clearly a function of $r$ and $n$, follows a $t$–distribution with $n$–2 degrees of freedom.

These relationships can easily be exploited to achieve the appearance of statistical significance. Suppose that we return to the same 100 × 100 m parcel, but decrease the grid spacing between observations from 50 m to 25 m, yielding 5 × 5 matrices and $n=25$ observations (Figure 10.3b). If the correlation stays about the same, at $r=.32$, the significance of the relationship increases markedly, with $p=.12$. By most standards this still would not be regarded as a statistically significant outcome, however. Consequently, the process might be taken one step further by sampling the same region at 12.5 m intervals yielding 9 × 9 matrices

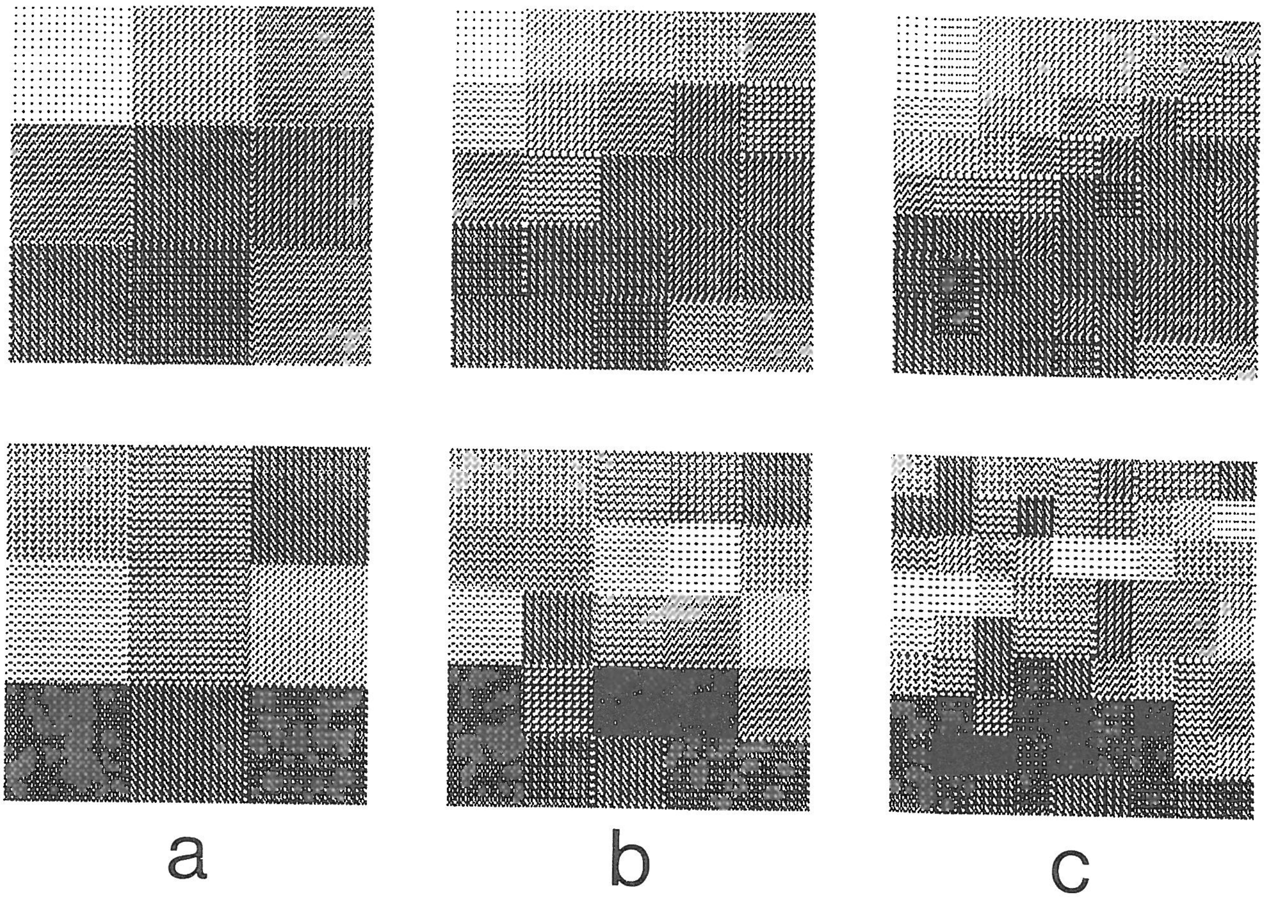



*Figure 10.3: Sampling a finite region where two variables (top and bottom) are measured at increasingly higher resolutions (a–c) to inflate sample size.*

and $n$=81 observations (Figure 10.3c). Now, with the correlation still at $r$=.32, high statistical significance is finally reached with $p$<.001.

What has been accomplished here? It should be clear that in spatial contexts we can continue to sample a region indefinitely at ever increasing levels of resolution. In so doing it is an easy matter to inflate the apparent sample size, $n$. It therefore is possible to inflate the apparent significance of conventional statistics in spatial contexts. In fact, Pearsonian correlations of nearly zero can always be made to appear significant if $n$ is made arbitrarily large. This type of abuse is made all the more easy today with GIS technology since with raster data structures it is a trivial matter to interpolate or resample a grid at virtually any level of resolution thereby achieving thousands, if not millions, of observations.

Yet, is new information really generated by sampling at increasingly higher levels of spatial resolution? Interdependencies in spatial data sets mean that by taking a new observation one does not necessarily increase the effective sample size by one. The new observation might be closely related to others nearby making much of the information it carries redundant on its neighbors. This is particularly true when one continues to sample a finite region. New observations become more–and–more dependent on previous ones until virtually no new information is generated. A good analogy for this phenomenon may be found in time series analysis where data generally are temporally autocorrelated and quarterly observations are not four times as informative as yearly observations over the same period.

The interdependencies generally present in spatial data sets have the effect of causing a reduction in the apparent sample size to what might be termed an effective sample size, or an equivalent number of independent observations (Cliff & Ord 1975). One problem in the analysis of spatially autocorrelated data, then, lies in working out what the effects of this reduction might be. For Pearson's $r$ applied to spatially autocorrelated data Bivand (1980), for example, has offered some results based on computer simulations.

The remaining sections of this paper provide a further focus on this issue of effective sample size by presenting a spatial statistical test that actually estimates it from empirical data. The test does so through a simple transformation of the apparent sample size, *n*, and the strength of the observed spatial autocorrelation, as measured by Moran's *I*. Consequently, it also provides additional insight into the other spatial–statistical topic of this paper, tests that incorporate the locational component of the data.

## 10.4 A MODIFIED *T*–TEST FOR SPATIAL DATA

A modified Student's *t*–test, appropriate for the comparison of means when samples consist of spatially autocorrelated observations, was presented by Cliff & Ord in 1975. This test illustrates many of the issues and concerns discussed previously and has achieved some attention as a point of methodological interest (e.g., Haggett *et al.* 1977; Haining 1980; Cliff & Ord 1981), but little in the way of actual application to empirical data sets. The latter is probably true because it is somewhat more complex than the conventional (aspatial) *t*–test and, in any case, it is difficult to implement without appropriate computer software. There is also some question about the kinds of spatial contexts to which it may be applied (Haining 1980). The following material is derived primarily from Cliff & Ord's (1975) initial presentation, and is greatly abridged. My intent in the remainder of this paper is to:

1) present this relatively obscure test to a wider and contemporary (archaeological) audience because it is important in itself, but also as a means to illustrate spatial autocorrelation issues and problems;
2) clarify its role given several possible limitations and shortcomings;
3) give an example application of the test in a GIS context showing a possible archaeological use; and
4) illustrate a recently created software implementation compatible with a popular GIS package.

### 10.4.1 The conventional student's *t*–test

Under a null hypothesis, $H_0$, of no difference between population means, $\mu_1$ and $\mu_2$, Student's *t*–statistic:

[2] $$t = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{\bar{x}_1-\bar{x}_2}}$$

follows a *t*–distribution with $n_1 + n_2 - 2$ degrees of freedom, when certain assumptions can be met. These assumptions include normal populations, homogeneity of variance, and independent observations. $\bar{x}_1$ and $\bar{x}_2$ are sample means from populations 1 and 2 and $\hat{\sigma}_{\bar{x}_1-\bar{x}_2}$ is the estimated standard error of the difference between sample means given by

[3] $$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{S_p^2\left(\frac{1}{n_1}\right)+\left(\frac{1}{n_2}\right)}$$

The common variance assumption requires a pooling of the individual sample variances ($S_1^2$ and $S_2^2$)

[4] $$S_p^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{n_1+n_2-2}$$

The independence assumption is what causes problems in spatial contexts, since geographically distributed data often are spatially autocorrelated, as noted in previous sections.

Student's *t*–test (Eqs. 2–4) generally is regarded as quite robust in the face of mild departures from normality, and corrections exist for the case when the sample variances are assumed unequal (Hays 1988). However, the severe consequences that can arise when the last assumption, independence, is violated have been demonstrated by Cliff & Ord (1975) through computer simulation.

Cliff & Ord (1975) took artificial data sets spatially autocorrelated by a known amount, $\rho$, and through Monte Carlo methods obtained repeated samples against which the performance of *t* (Eqs. 2–4) could be evaluated. Specifically, they assumed a quadrilateral, or first–order, spatial autoregressive process, usually regarded as the simplest start–point in studies of this nature (Cliff & Ord 1975,1981; Griffith 1987:28). That is, the value of a variate $x_i$, distributed in a lattice in space, is determined by its four neighbors and the strength of $\rho$, plus a random error term

[5] $$x_i = \rho\sum_j W_{ij}^* x_j + e_i$$

If we recall that $W_{ij}$ is a weight between the $i^{th}$ and $j^{th}$ observations, then $W_{ij}^*$ is a scaled or standardized weight

[6] $$W_{ij}^* = \frac{W_{ij}}{\sum_j W_{ij}}$$

causing $\sum_j W_{ij}^* = 1$ for all $i$ and $\sum_i \sum_j W_{ij}^* = n$. The latter is a convenient normalizing transformation which assures that the sum of the weights associated with each observation is identical for all observations. The error term, $e_i$, is simply a randomly generated standard normal deviate, with mean zero and variance one.

Cliff & Ord (1973,1975,1981), Haggett *et al.* (1977), Whitley & Clark (1985), and others, show that spatially autocorrelated $x_i$ may be generated (in matrix notation) by

[7] $$\mathbf{x} = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{e}$$

where **I** is the identity matrix, **W** is the weighting matrix, and **e** is the vector of error terms. When $\rho$ is greater than zero the vector **x** will contain $n$ positively spatially autocorrelated observations.

In the computer simulation spatially autocorrelated lattices were generated by Eq. 7 for a variety of sample sizes and mapped onto a torus to remove undesirable edge effects. For each simulation run the sample means and variances were computed for each lattice (with $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$, meeting the condition of the null hypothesis) and Student's $t$–statistic given by Eqs. 2–4 was obtained, assuming independent observations. This process was repeated 1200 times for specific sample sizes $(n_1, n_2)$ and levels of spatial autocorrelation $(\rho_1, \rho_2)$. The results of one such simulation that utilized $7 \times 7$ lattices ($n_1 = n_2 = 49$; df=96) is given in Table 10.1.

It is generally the case that under positive spatial autocorrelation conventional statistical tests greatly overstate significance and increase the probability of a Type I error (Haggett *et al.* 1977). This clearly is the case here. The distribution of the statistic is so different from the $t$–distribution (Table 10.1) that without corrections serious inferential errors can arise.

### 10.4.2 A modified spatial $t$–test

When there is first–order spatial autocorrelation in the variates $x_1$ and $x_2$, Cliff & Ord (1975, 1981) suggest that the estimated standard error of Eq. 2 should be given by

[8] $$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{S_p^2 \left( \frac{1}{m_1} + \frac{1}{m_2} \right)}$$

where (in matrix notation)

[9]
$$S_p^2 = \frac{(\mathbf{x}_1 - \bar{x}_1 \mathbf{1})' \mathbf{V}_1^{-1} (\mathbf{x}_1 - \bar{x}_1 \mathbf{1}) + (\mathbf{x}_2 - \bar{x}_2 \mathbf{1})' \mathbf{V}_2^{-1} (\mathbf{x}_2 - \bar{x}_2 \mathbf{1})}{n_1 + n_2 - 2}$$

and

[10] $$m_k = n_k (1 - \hat{\rho}_k)^2, \; k = 1,2$$

when the spatial weighting scheme is standardized as in Eq. 6. The $m_k$ may thus be interpreted as "an equivalent number of independent observations" (Cliff & Ord 1975:729). It should be clear from Eq. 10 that when there is a high level of spatial autocorrelation, $\rho_k$, the effective sample size, $m_k$, is greatly reduced from the apparent sample size $n_k$. This aspect of the modified $t$–test therefore addresses the second spatial concern of this paper.

The $\mathbf{V}_k^{-1}$ is given by

$$\mathbf{V}_k^{-1} = (\mathbf{I} - \hat{\rho}_k \mathbf{W})' (\mathbf{I} - \hat{\rho}_k \mathbf{W}), \; k = 1,2$$

which is the inverse of the covariance matrix of the first–order spatial autoregression (see Eqs. 5, 7). It denotes the spatial covariance structure among the variates which, through Eqs. 8–9, is fed directly to the $t$–test formula (Eq. 2). Finally, we note that when $\hat{\rho}_1 = \hat{\rho}_2 = 0$, $m_k = n_k$ (Eq. 10), as it should, and the standard error given by Eq. 8 reduces to the form of the conventional $t$–test (Eq. 3), which assumes independent observations.

Equations 8–10, then, provide a correction to the $t$–test formula that in theory will bring its distribution nearer to the Student form when it is applied to spatially autocorrelated or dependent data.

#### 10.4.2.1 *Estimation of the* $\rho_k$

An obvious candidate for the estimation of the $\rho_k$ is maximum likelihood, but multiple difficulties arise with this method (see Griffith 1987; Cliff & Ord 1975,1981). A far simpler estimator exists,

| Tabulated | | Empirical t-values | | | | | |
|---|---|---|---|---|---|---|---|
| $\alpha$ | t | .0,.0 | .0,.5 | .0,.9 | .5,.5 | .5,.9 | .9,.9 |
| .10 | 1.29 | 1.30 | 1.88 | 7.30 | 2.39 | 7.21 | 8.81 |
| .05 | 1.66 | 1.61 | 2.53 | 9.20 | 3.01 | 9.08 | 10.98 |
| .025 | 1.98 | 1.97 | 2.94 | 10.90 | 3.63 | 11.05 | 12.88 |
| .01 | 2.36 | 2.23 | 3.40 | 13.22 | 4.05 | 13.38 | 14.99 |
| .005 | 2.62 | 2.42 | 3.71 | 14.69 | 4.47 | 16.63 | 16.35 |

*Table 10.1: Simulation results based on 1200 repetitions showing the performance of the conventional Student's t–test under varying levels of positive spatial autocorrelation. These results are from 7x7 lattices ($n_1 = n_2 = 49$; df=96). Values of $\rho_1$ and $\rho_2$ are given at the head of each column. (Source: Cliff & Ord 1975:730.)*

however, that is based on Moran's $I$ statistic (Eq. 1).

It is well-known that Moran's $I$ does not range over the usual [-1,+1] interval, unlike most correlation coefficients (Cliff & Ord 1973). When standardized weights are employed (Eq. 6), however, it is possible to compute its maximum attainable value for any given spatial lattice:

$$\max|I| = \sqrt{\frac{\mathrm{var}\left[\sum_j W_{ij}^*\left(x_j - \bar{x}\right)\right]}{\mathrm{var}\left(x_j - \bar{x}\right)}}$$

where "var" is the sample variance over all $i$ observations. An estimate of $\rho$ on the usual [-1,+1] scale then is given by

[11] $$\hat{\rho} = \frac{I}{\max|I|}$$

Cliff & Ord (1975, 1981) show that this estimator is not consistent and is generally biased downward. Computer simulations indicate, however, that despite the bias the estimator given in Eq. 11 is reasonable provided that $n$ is large ($n$>25) and $\rho$ is not near unity (Cliff & Ord 1975:730). This use of Moran's $I$ as a basis for estimating $\rho$ satisfies the first spatial concern of this paper because it explicitly incorporates the locational component of the data.

*10.4.2.2 Performance of the modified test*

The Monte Carlo methods documented earlier that were employed to evaluate the conventional $t$-test when applied to spatial data also were used by Cliff & Ord (1975) to assess the performance of the modified $t$-test. That is, two samples autocorrelated by a known amount were generated (with $\mu_1 = \mu_2$ agreeing with the null hypothesis), and the performance of the modified test was evaluated under different sample sizes and levels of spatial autocorrelation. A number of important modifications to the simulation were inserted, however. First, in addition to the several sizes of square lattices mapped onto a torus that were applied in the earlier simulation, an irregular, real-world lattice representing the county system of Ireland (excluding Dublin for $n$=25 counties) was employed, also with binary, standardized weights. Additionally, the pooled sample variance, $S_p^2$, was estimated first by the theoretically-derived form of Eq.9, and then by its simpler form, Eq. 4 of the conventional $t$-test, allowing a comparison of the suitability of the two estimators. The results for the simulation dealing with 7 × 7 lattices (comparable with the conventional $t$-test outcome of Table 10.1) are given in Table 10.2.

It is clear in Table 10.2 that the test statistic with the pooled variance, $S_p^2$, estimated by the simpler form of Eq. 4, gives much better results than the more complex estimator given by Eq. 9. The latter is highly unstable and yields completely untrustworthy results even for moderate levels of $\rho$. The modified $t$-statistic with $S_p^2$ in its simpler form, however, performs remarkably well, particularly when the $\rho_k$ are not too close to unity (Table 10.2).

A. $S_p^2$ estimated by Eq. 4.

| Tabulated | | Empirical t-values | | | | | |
|---|---|---|---|---|---|---|---|
| α | t | .0,.0 | .0,.5 | .0,.9 | .5,.5 | .5,.9 | .9,.9 |
| .10 | 1.29 | 1.34 | 1.28 | 1.27 | 1.34 | 1.26 | 1.01 |
| .05 | 1.66 | 1.69 | 1.86 | 1.59 | 1.74 | 1.65 | 1.25 |
| .025 | 1.98 | 1.94 | 2.17 | 2.12 | 2.09 | 2.14 | 1.45 |
| .01 | 2.36 | 2.35 | 2.61 | 3.28 | 2.49 | 2.60 | 1.84 |
| .005 | 2.62 | 2.54 | 2.70 | 3.55 | 2.67 | 2.87 | 2.71 |

B. $S_p^2$ estimated by Eq. 9.

| Tabulated | | Empirical t-values | | | | | |
|---|---|---|---|---|---|---|---|
| α | t | .0,.0 | .0,.5 | .0,.9 | .5,.5 | .5,.9 | .9,.9 |
| .10 | 1.29 | 1.36 | 1.41 | 2.35 | 1.51 | 2.33 | 2.50 |
| .05 | 1.66 | 1.72 | 1.96 | 4.04 | 1.93 | 3.97 | 4.09 |
| .025 | 1.98 | 1.96 | 2.34 | 6.08 | 2.34 | 6.34 | 6.34 |
| .01 | 2.36 | 2.34 | 2.76 | 10.09 | 2.71 | 10.92 | 9.88 |
| .005 | 2.62 | 2.56 | 2.85 | 20.34 | 2.98 | 11.00 | 12.17 |

*Table 10.2: Simulation results based on 1200 repetitions showing the performance of the modified spatial t-test under two estimators of $S_p^2$ and varying levels of positive spatial autocorrelation. These results are from 7 × 7 lattices ($n_1 = n_2 = 49$; df=96). Values $\rho_1$ and $\rho_2$ are given at the head of each column. (Source: Cliff & Ord 1975:732)*

Further insight into the performance of the spatially modified test is gained in Table 10.3 where three things are apparent: as expected the modified test works better as

1) $n_1$ and $n_2$ increase;
2) when $\rho_1$ and $\rho_2$ are not close to unity;
3) when both sample sizes are equal.

With regard to the first point, Cliff & Ord (1981:189) suggest that the modified test should

| Data Set | Percent Observations Exceeding Tabled 5% Level | | | | | |
|---|---|---|---|---|---|---|
| (Sample 1,2) | .0,.0 | .0,.5 | .0,.9 | .5,.5 | .5,.9 | .9,.9 |
| 3x3, 3x3 | 7.8 | 12.4 | 29.4 | 13.8 | 28.4 | 34.0 |
| 5x5, 5x5 | 5.3 | 7.3 | 11.2 | 8.2 | 8.5 | 7.7 |
| Ireland,Ireland | 6.2 | 6.8 | 8.7 | 7.2 | 6.7 | 2.5 |
| 7x7, 5x5 | 5.0 | 7.5 | 13.8 | 6.7 | 12.8 | 6.0 |
| 7x7, 7x7 | 5.5 | 6.3 | 4.5 | 5.7 | 5.0 | 1.7 |
| 10x10, 10x10* | 3.0 | 5.5 | 4.5 | 3.5 | 2.5 | 1.5 |

*Based on 400 repetitions

*Tabel 10.3: Simulation results based on 1200 repetitions showing the performance of the modified spatial t–test under varying levels of positive spatial autocorrelation, sample sizes, and spatial configurations. Values of $\rho_1$ and $\rho_2$ are given ath the head of each column. (Source: Cliff & Ord 1975:733)*

only be applied when $n_1$ and $n_2$ both exceed 25. If the $\rho_k$ are near unity the test becomes conservative causing results to be more significant than they appear to be. Finally, and of some importance, the performance of the test on the Irish data suggests that irregular spatial configurations do not present a problem to the test.

*10.4.2.3 Issues and concerns*

In addition to the foregoing limitations of the spatial *t*–test, a concern that may be more serious has been emphasized by Haining (1980). Cliff & Ord's (1975) modifications of the *t*–test were initially established assuming a first–order spatial autoregressive process (Eq. 5). Haining (1980:23) argues that under some other spatial process, and the first–order autoregressive is one of the simplest compared to most real–world processes, different modifications or corrections to the test would be required. This point deserves closer attention.

Cliff & Ord (1975) did, indeed, establish their modifications to the *t*–test based on corrections theoretically derived from an assumed first–order spatial autoregressive model. The corrections resulted in the complex form $S_p^2$ given by Eq. 9. The performance of the modified test based on this statistic, however, was shown by simulation to be very poor under the condition of positive spatial autocorrelation (Table 10.2). Cliff & Ord (1975: 733, 1981:188–9) therefore recommend against its use. They advocate, rather, that the simple aspatial form of $S_p^2$ (Eq. 4) be employed which yields superior results (Tables 10.2–3). Consequently, the test as it stands (Eq. 2,4,8,10) does not employ a modification based on a first–order spatial autoregressive assumption.

The single remaining modification to the test, the correction of the apparent sample sizes to effective sample sizes (Eq. 10), is based on an empirical estimate of the $\rho_k$ (Eq. 11) that is derived entirely from Moran's *I* statistic (Eq. 1). Moran's *I*, of course, is a general measure that allows assessment of the strength and nature of empirical spatial autocorrelation generated by any spatial process. Consequently, proper application of Cliff & Ord's (1975) modified *t*–test may not be as limited as Haining (1980) suggests.

In any case, there is poor theoretical understanding of real–world spatial processes at present, even by geographers who have been confronting the issue for several decades (Cliff & Ord 1973,1981; Haining 1980; Griffith 1987). This is particularly true in fields like archaeology where little work has been caried out in this area. In other words, we usually do not know, and can barely guess, the true nature of the spatial processes that generated realized patterns in the archaeological spaces we deal with. In such situations a usual, and acceptable, procedure is to assume the simplest form; in terms of spatial processes this turns out to be the first–order spatial autoregressive model (Griffith 1987:28). Cliff & Ord's (1975) modified *t*–test, then, might be applied to real–world situations when there is a possibility that the data are spatially autocorrelated. At worst, such applications will give more correct results than use of the test in its conventional aspatial form.

## 10.5 ARCHAEOLOGICAL APPLICATION

An archaeological data set is employed that illustrates application of the modified *t*–test and many of the issues discussed in this paper. The application also shows the benefits that can be derived when conducting spatial investigations of this nature in a GIS setting. With regard to the latter, a program called SPTTEST has been written by the author that performs the modified spatial *t*–test and which interfaces with several commonly available microcomputer GIS programs, including the IDRISI GIS (Eastman 1990). The data for the application were obtained from an archaeological mapping project conducted on Seiber Ridge, located in a remote wilderness area of western Colorado, USA (Kvamme 1990b).

Seiber Ridge, lying at an altitude of nearly 2000 m, is a geologically deflated ridge, caused in part by a century of overgrazing by the cattle industry, but also by a forest fire that occurred in the area more than 30 years ago. These circumstances, together with the general high visibility of the ground surface caused by the region's aridity and consequent lack of vegetation, have enabled the detailed surface mapping of the abundant artifacts that occur there. Over two field seasons approximately 24,000 artifacts have been mapped in an area of approximately 6.5 hectares. The archaeology, spanning nearly 7000 years and consisting entirely of lithic scatters, represents deposits left primarily by hunter–gatherers and possibly by roving bands of part–time horticulturalists that frequented the area during a brief 300 year period.

Much of the project data has been encoded within a comprehensive GIS database to facilitate data retrieval, analysis, and display. Most of the database exists in a raster data structure with a grid resolution of four meters. The nature of the artifact density over the region and the ridge upon which the artifacts occur is illustrated in Figure 10.4 through simple GIS graphics.

The chief means of dating surface collections in this region lies in projectile point styles. One of the major artifact clusters, shown on the far right of Figure 10.4, contains numerous early projectile points, dating from the Archaic Period (ca. 7000–3000 BP). A second major concentration, located along the lower left edge of the study area (Figure 10.4), contains many late projectile points, dating primarily from the last 1000 years. Visual inspection of Figure 10.4 alone suggests that these artifact clusters represent highly autocorrelated data sets. A number of differences between these clusters also are suggested by the data, which may be examined by the spatial *t*–test.

The first analysis question pertains to artifact density, computed as artifacts per square meter in the GIS database. In Figure 10.4a it appears that there may be some differences in artifact density between the early and late clusters (indicated by gray–level shading). To implement the spatial *t*–test to assess the null hypothesis of no mean difference between the two concentrations the following GIS steps might be caried out:

1) A density cut–point might be employed to help isolate the clusters by indicating, through map reclassification, those localities at or above a specified artifact density level. Alternatively, simple on–screen digitizing can be used to trace the perimeters of the clusters.
2) When the clusters are defined they are converted to simple binary masks.
3) The primary image containing the variate to be analyzed, here artifact density, together with the two masks that indicate the artifact sample regions provide input to the spatial *t*–test. In the raster data structure the masks indicate the grid cell elements that constitute each sample. Other software that exists in the GIS package might be employed to offer basic descriptive statistics, histograms, and other data about the clusters.

The results of the spatial autocorrelation analysis and basic descriptive statistics for each artifact cluster, output by SPTTEST, are given in Table 10.4. These data indicate that the sample means are indeed different between the clusters, with $\bar{x}_1 = 1.81/\text{m}^2$ and $\bar{x}_2 = 2.38/\text{m}^2$, for the Archaic and Late Periods, respectively. The question is, are these differences significant?

Assuming first (incorrectly) independent observations, and using the conventional (aspatial) t–test of Eqs. 2–4 with the data in Table 10.4, a *t*–statistic of $t$=-2.02 is achieved which, when compared against Student's *t*–distribution with $n_1 + n_2 - 2 = 98 + 117 - 2 = 213$ degrees of freedom, indicates a significant mean difference ($p$=.044). This result is instructive, because when Cliff & Ord's (1975) spatial *t*–test is employed that recognizes the interdependencies between these matrices of related grid cells, a different conclusion is reached. From Table 10.4 we note that the estimated level of spatial autocorrelation, $\hat{\rho}$, for each sample is approximately 0.5, yielding effective

| | 1. ARCHAIC | | 2. LATE PERIOD |
|---|---|---|---|
| **A. Artifact Density** | | | |
| $\bar{x}_1$ | = 1.8112 | $\bar{x}_2$ | = 2.3830 |
| $S_1^2$ | = 2.3887 | $S_2^2$ | = 5.8125 |
| $n_1$ | = 98 | $n_2$ | = 117 |
| $I_1$ | = .3014 ($p$ <.001) | $I_2$ | = .3099 ($p$ <.001) |
| $\hat{\rho}_1$ | = .4796 | $\hat{\rho}_2$ | = .4982 |
| $m_1$ | = 26.54 | $m_2$ | = 29.46 |
| **B. Slope** | | | |
| $\bar{x}_1$ | = 5.5185 | $\bar{x}_2$ | = 7.9401 |
| $S_1^2$ | = 9.1707 | $S_2^2$ | = 6.2355 |
| $n_1$ | = 98 | $n_2$ | = 117 |
| $I_1$ | = .6688 ($p$ <.001) | $I_2$ | = .5990 ($p$ <.001) |
| $\hat{\rho}_1$ | = .8809 | $\hat{\rho}_2$ | = .8101 |
| $m_1$ | = 1.39 | $m_2$ | = 4.22 |

*Table 10.4: Descriptive statistics and spatial autocorrelation test results for the Seiber Ridge archaeological data.*

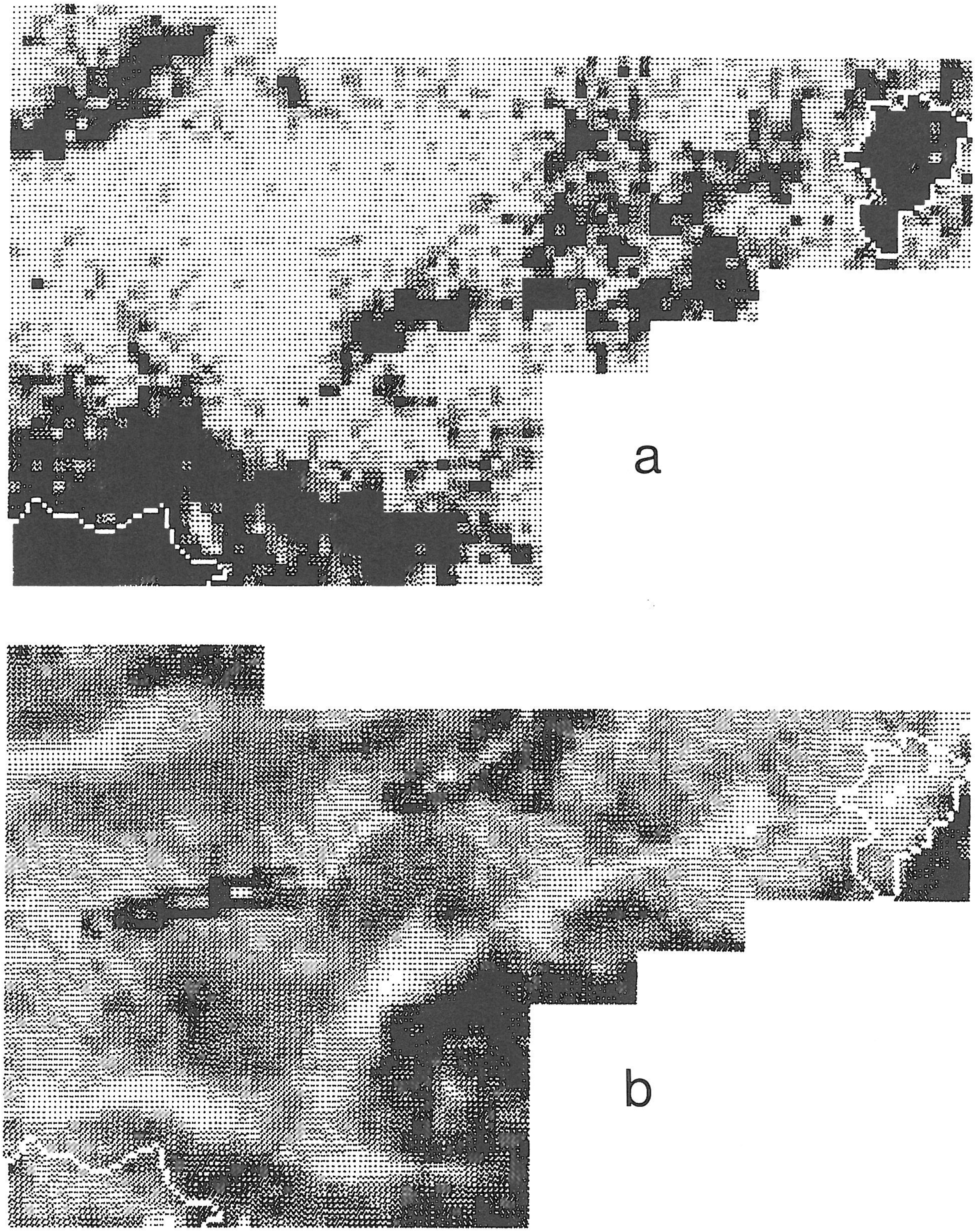



*Figure 10.4: Seiber Ridge, Colorado, USA, showing a) artifact density and b) gradient (slope) data. Darkest gray tones represent high values in both images. The site clusters of interest are highlighted.*

sample sizes of only $m_1 = 26.54$ and $m_2 = 29.46$ for the Archaic and Late Periods, respectively (down from their apparent sample sizes of $n_1 = 98$ and $n_2 = 117$, Table 10.4). These moderate levels of spatial autocorrelation have reduced the effective sample sizes to one–quarter of their initial level. A modified *t*–statistic of $t$=-1.036 results that with 213 df clearly is not significant ($p$=.301). Thus, the modified spatial *t*–test, by considering the spatial relationships and removing the redundancies in the data, has allowed a more realistic and correct inference by indicating no significant mean difference.

A second analysis indicates how pronounced the difference between the two tests can be and, consequently, the magnitude of the inferential error that can result when employing conventional tests in spatial contexts. This analysis examines mean differences in ground steepness, or slope, between the two site clusters. The GIS graphic in Figure 10.4b indicates that the Archaic cluster occurs primarily on the broad and level ridge crest while the Late Period concentration is scattered somewhat down the slope of one of the ridges. Descriptive statistics, output by SPTTEST and given in Table 10.4, indicate that there are differences in mean slope, with $\bar{x}_1 = 5.52$ and $\bar{x}_2 = 7.94$ percent grade (subscripts as before).

Slope was computed using a common GIS algorithm that fits a least–squares plane to a $3 \times 3$ window centered on a grid element in the elevation layer; the maximum slope on that plane is obtained and stored in a corresponding grid element in a slope layer. The 3x3 window then is moved, and centered, on the next grid cell element and the process is repeated until an entire matrix of slope values is obtained (Kvamme 1990c). Obviously, six of the nine elevations in each $3 \times 3$ window also occur in each previous window. With this commonality we might expect a raster slope surface to exhibit high levels of spatial autocorrelation. This indeed is the case in the current analysis (Table 10.4). Both Moran's *I* and the estimates of $\rho$ are high and statistically significant. The contrast between the *t*–tests is dramatic: for the Archaic class the apparent sample size of $n_1 = 98$ drops to an effective sample size of only $m_1 = 1.39$; for the Late Period $n_2 = 117$ drops to only $m_2 = 4.22$! Although the conventional *t*–test yields a highly significant result under the independence assumption ($t$=-6.43; $p$<.001), the pronounced spatial autocorrelation yields a modified spatial *t*–statistic of only $t$=-.89, clearly a non–significant result ($p$=.37).

## 10.6 CONCLUSIONS

Conventional statistical tests generally are aspatial in nature and therefore do not consider the locational component of information when applied to geographically distributed data. Their use in spatial contexts can lead to erroneous and misleading findings. Spatial statistical tests that incorporate the locational component of data are necessary to make correct inferences in spatial contexts. The latter, however, are difficult to undertake without specialized software and efficient means to input, organize, retrieve, and display the data. GIS methods therefore greatly facilitate spatial investigations because they provide general data input, retrieval, manipulation, and display capabilities, but also a vehicle for incorporation of specialized spatial–statistical software, like the SPTTEST program. Moreover, the advanced computer graphics common to most GIS allows ready visualization of pattern in data, an essential adjunct to quantitative analysis.

The modified *t*–test of Cliff & Ord (1975) is a good example of a spatial statistical test because it illustrates two key domains of concern in spatial analysis. The first is the incorporation of the locational component of the data which was achieved through use of Moran's *I*. The second is correcting for spatial dependencies in the data, obtained through estimation of effective sample sizes. Although some researchers (e.g., Haining 1980) have claimed restrictions to the test's general applicability because it was initially established assuming a particular type of spatial process (the first–order spatial autoregressive), the test as ultimately presented relies only on modifications obtained from a general measure of spatial autocorrelation, Moran's *I*. This issue may be a moot point, however. The nature of the spatial processes that formed the archaeological record we deal with is so poorly understood at present that typically we must rely on the simplest processes as models, such as the first–order spatial autoregressive (Griffith 1987). Consequently, the spatial *t*–test might be considered a useful alternative that can provide superior results in spatial contexts when autocorrelation is present.

**Acknowledgements**
Much of this work was conducted during a sabbatical leave in the Netherlands, under the kind auspices of Roel Brandt of the RAAP Foundation, University of Amsterdam. Fieldwork at Seiber Ridge, Colorado, USA, has been supported by the

Arizona Foundation and the U.S. Bureau of Land Management, Grand Junction District, Colorado. The SPTTEST program, and other programs for evaluating spatial autocorrelation, are available from the author.

**References**

Berry, J.K.
1987 Fundamental Operations in Computer–Assisted Map Analysis. *International Journal of Geographical Information Systems* 1:119–136.

Bivand, R.
1980 A Monte Carlo Study of Correlation Coefficient Estimation with Spatially Auto–Correlated Observations. *Questiones Geographicae* 6:5–10.

Burrough, P. A.
1986 *Principles of Geographical Information Systems for Land Resource Assessment*. Clarendon Press, Oxford.

Chadwick, A. J.
1978 A Computer Simulation of Mycenaean Settlement. In I. Hodder (ed.) *Simulation Studies in Archaeology*. Cambridge University Press, Cambridge. pp.47–58.

Cliff, A.D. & J.K. Ord
1973 *Spatial Autocorrelation*. Pion, London.
1975 The Comparison of Means when Samples Consist of Spatially Autocorrelated Observations. *Environment and Planning A* 7:725–734.
1981 *Spatial Processes: Models and Applications*. Pion, London.

Eastman, J.R.
1990 *IDRISI: A Grid–Based Geographic Analysis System* (software documentation, version 3.2.2). Graduate School of Geography, Clark University, Worcester, Massachusetts.

Gorenflo, L. & N. Gale
1986 Population and Productivity in the Teotihuacan Valley: Changing Patterns of Spatial Association in Prehispanic Central Mexico. *Journal of Anthropological Archaeology* 5:199–228.

Griffith, D.A.
1987 *Spatial Autocorrelation: A Primer*. Resource Publications in Geography, Association of American Geographers, Washington, D.C..

Haggett, P., A.D. Cliff & A. Frey
1977 *Locational Methods* (2nd Edition). John Wiley, New York.

Haining, R.
1980 Spatial Autocorrelation Problems. In D.T. Herbert & R.J. Johnson (eds.) *Geography and the Urban Environment: Progress in Research and Application*, Vol. 3, John Wiley, New York. pp.1–44.

Harris, T.M. & G.R. Lock
1990 The Diffusion of a New Technology: A Perspective on the Adoption of a Geographic Information System within UK Archaeology. In K.M.S. Allen, S.W. Green & E.B.W. Zubrow (eds.) *Interpreting Space: GIS and Archaeology*, Taylor & Francis, London. pp.33–53.

Hays, W.L.
1988 *Statistics* (4th Edition). Holt, Rinehart and Winston, New York.

Hodder, I. & C. Orton
1976 *Spatial Analysis in Archaeology*. Cambridge University Press, Cambridge.

Hubert, L.J. & R.G. Golledge
1982 Comparing Rectangular Data Matrices. *Environment and Planning A* 14:1087–1095.

Hubert, L.J., R.G. Golledge, C.M. Costanzo & N. Gale
1985 Measuring Association between Spatially Defined Variables: An Alternative Procedure. *Geographical Analysis* 17:36–46.

Kvamme, K.L.
1990a Spatial Autocorrelation and the Classic Maya Collapse Revisited: Refined Techniques and New Conclusions. *Journal of Archaeological Science* 17:197–207.
1990b *Final Report: High Resolution Cultural Resource Mapping Project at Seiber Canyon, Mesa County, Colorado, 1989 Season*. Unpublished technical report submitted to U.S. Bureau of Land Management, Grand Junction District, Colorado.
1990c GIS Algorithms and their Effects on Regional Archaeological Analysis. In K.M.S. Allen, S.W. Green & E.B.W. Zubrow (eds.) *Interpreting Space: GIS and Archaeology*, Taylor & Francis, London. pp.112–126.

Siegel, S.
1956 *Nonparametric Statistics for the Behavioral Sciences*. McGraw–Hill, New York.

Tjostheim, D.
1978 A Measure of Association for Spatial Variables. *Biometrika* 65:109–114.

Tobler, W.R.
1970 A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46 (Supplement):234–240.

Whitley, D.S. & W.A.V. Clark
1985 Spatial Autocorrelation Tests and the Classic Maya Collapse: Methods and Inferences. *Journal of Archaeological Science* 12:377–395.

**Author's address**

Arizona State Museum, Building 26
University of Arizona,
Tucson, AZ 85721, U.S.A.
E–mail: kvamme@ccit.arizona.edu