SPATIAL AUTOCORRELATION AND THE ANALYSIS OF PATTERNS RESULTING FROM CRIME OCCURRENCE

by

Gary J. Ward B.A.(Hons)

Submitted to the Faculty of Arts, Rhodes University, Grahamstown for the degree of Master of Arts.

Department of Geography
Rhodes University
Grahamstown

January 1978

ACKNOWLEDGEMENTS

I am particularly indebted to my supervisor, Dr Gillian P. Cook, for her unfailing and invaluable interest, encouragement and firm guidance.

To Mr Peter Roberts, also, a special word of thanks for the many hours of consultation, advice and inspiring co-operation.

For devoting much of their valuable time to the preparation of this thesis I would like to thank Mr Oakley West and Mr John Keulder. My thanks also to Mrs Carol Price and Mrs Yvonne White for the typing involved. My appreciation is also extended to Lt. Fish of the local branch of the South African Police for allowing me access to police files.

Thanks also go to Mary, my wife, for her enthusiasm and assistance.

The financial assistance of the Human Sciences Council towards the costs of this research is hereby acknowledged. Opinions expressed or conclusions reached are those of the author and are not to be regarded as a reflection of the opinions and conclusions of the Human Sciences Research Council.

G.J. Ward
January 1978

CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

In geography during the 1950's there was a definite move away from the study of unique phenomena to the study of generalized phenomena or pattern (Mather and Openshaw, 1974). At the same time interrelationships between phenomena distributed in space and time became the topic of much interest among geographers, as well as members of other disciplines. The changing emphasis initiated acceptance of certain scientific principles (Cole, 1973), and mathematical techniques became the recognized and respected means through which objective analysis of pattern, structure, and interrelationships between areally distributed phenomena could be achieved (Ackerman, 1972; Burton, 1972; Gould, 1973).

Geographers, as do members of other disciplines, frequently borrow mathematical techniques developed for problems encountered in the pure sciences and apply these techniques to what are felt to be analogous situations in geography. The spatial nature of much geographic data presents severe difficulties "... to the application of conventional statistical techniques ..." and this "... emphasizes the difference between geographic requirements and those of non-spatial disciplines" (Mather and Openshaw, 1974, p.284). Certain distinctive characteristics of geographic data include the fact that firstly, spatial data cannot be assumed to be independent, secondly, it is often very difficult to select those variables that contribute most to the nature of the distribution, thirdly, these prime variables are not very often found to be normally distributed, and fourthly, the selection of zonal boundaries is extremely important in any research which seeks to analyse spatially distributed phenomena. Therefore, the geographer who wishes to apply a mathematical technique to spatial data is limited in his choice and is faced with the problem of selecting a particular technique which will be directly suited to the proposed analysis (Matravers, 1977). Several implications surround the choice of a particular mathematical formula, the most important being that the roles that variables play in the final formula are dependent on how the formula was derived, and as such these roles can be very different depending on the particular approach adopted by the researcher to the specific problem. Further, depending on the specific problem, the hypotheses upon which the assumptions underlying the mathematical models are built either stand or fall in the social sciences framework, and as such the onus is on the researcher to understand the assumptions underlying the chosen model prior to analysis of the problem (Matravers, 1977). A simple example of how misunderstanding of a particular technique can occur is the case of nearest neighbour analysis, where the assumptions underlying the formula are clearly non-justifiable in reality (Ebdon, 1976). Unfortunately the failure of many geographers to grasp comprehensively the formal properties of various methods of numerical analysis has led to a lack of understanding of how a particular technique functions and what its limitations and advantages are for problems that are specifically geographic and has meant that the results of much research are often misinterpreted (Harvey, 1973; Mather, 1976).

One technique commonly used by geographers to analyse relationships between sets of variables is correlation analysis, developed originally in the pure sciences (Ezekiel and Fox, 1959). In geography in particular, correlation methodology is used almost solely for the analysis of relationships between two or more differentiable sets of variables. However, the technique can also be applied to a single set of data, the process sometimes being called 'self-correlation'. The concept of self-correlation arose from the need to isolate cyclic or periodic patterns in linear arrays; where a sequence of events is compared with itself with the aim of establishing the degree of similarity or dissimilarity between corresponding segments of the array. This process of autocorrelation is defined as a "... linear correlation between a time series and the same series at a later interval of time" (Davis, 1973, p.232). Whittle (1954) extended temporal autocorrelation to the two-dimensional case where the interdependence of spatially distributed phenomena formed the focus of attention. Based on this initial research Cliff and Ord (1969-1975) developed a statistical technique, known as spatial autocorrelation, to analyse the degree of spatial interdependence of a single variable. This is particularly appropriate for geographers, as an essential component of much geographic research during recent years has been concerned with the concept of space and in particular the relationships that exist between phenomena distributed in space.

This study aims at applying Cliff and Ord's technique to a spatial pattern resulting from a particular set of phenomena of a unique form, viz., distribution of crime occurrence in an urban area. The following are the two major objectives. Firstly, to isolate and describe the problems involved in applying a particular mathematical technique to data which are unlike that for which the technique was originally developed. In this study therefore spatial autocorrelation is applied to two types of crime patterns, viz., one resulting from the occurrence of breaking and entering and one resulting from the occurrence of assault. It is reasonable to assume that the former pattern is constrained by the location of buildings, while in the latter case it may be assumed that an assault could occur anywhere within the area being considered. The second objective of the study is to evaluate the usefulness of spatial autocorrelation as a technique which can be used to measure the extent of spatial relations among areally distributed phenomena. Further, the approach has been to assume ignorance of the particular technique and attempts, by way of example, to show how precise understanding of the mathematics involved can be achieved by application to simple examples. Thus this study is not merely an exercise in statistical analysis and computing. It aims at complementing contemporary geographic inquiry in the sphere of analysis of areally distributed phenomena, by evaluating spatial autocorrelation as a tool whereby the importance of space in patterns resulting from crime occurrence can be gauged. By establishing the degree of spatial autocorrelation the significance of one variable in crime occurrence, viz., space, can be considered. An evaluation of this variable can aid crime research in three respects. Firstly, if space were shown to be important, researchers into crime patterns could focus attention on particular areas in order to determine causative factors. Secondly, other variables which have a similar spatial distribution could be studied and thirdly, an evaluation of areal correlation could lead to the simulation of crime patterns. Predictions could be made on the basis of correlation results even without understanding of the complex causal mechanisms of crime occurrence, thus aiding both preventative and control measures.

In order to fulfil the above objectives this study requires that an autocorrelation program be written. Because most computers have a fortran compiler the program is written in standard Fortran IV so that it can be universally applied. In this regard the various assumptions, limitations, and problems relating to spatial autocorrelation, particularly the problem of delimiting subareas for purposes of comparison and computation, are highlighted so that geographers with a limited knowledge of mathematics and statistics, and little experience of computing should be able to follow the research technique easily and apply it to relevant problems in other fields of geography. The perspective is provided by considering the influence of space in distributional patterns particularly of crime occurrence which forms the focus of this research. Furthermore the problem of deterministic thinking and interpreting data patterns where no geographic relationship exists is considered in detail and is associated with the concepts of directional bias and weighting both of which form an integral part of spatial autocorrelation.

The thesis is therefore organized into the following chapters. In Chapter II some of the major problems concerning the incorporation of mathematical language in geography are considered, using correlation methodology as an example of a quantitative technique used by geographers. Further, because geographers are concerned with the analysis of the spatial component in the distribution of real world phenomena this chapter outlines the conceptual problem of regarding space as an independent variable. Following this is Chapter III, which outlines the problems involved in analysing phenomena distributed in space, specifically point patterns. Various techniques available to geographers for the analysis of pattern with respect to itself are outlined in detail, and are shown to be inadequate to research which seeks to isolate the importance of space in any distribution of points. Chapter IV outlines the technique known as spatial autocorrelation, which offers a solution to the question of spatial dependence. Various limitations regarding the applied use of the technique are outlined in detail and important factors regarding interpretation of the autocorrelation coefficient are discussed. In Chapter V the study area, data base and matrix constraints are defined, which is followed by a section outlining results of the analysis of spatial dependence in crime patterns in Grahamstown. Following this is Chapter VI which evaluates the usefulness of spatial autocorrelation as an analytical tool that can be used by geographers and suggests lines for future research.

CHAPTER II

GEOGRAPHY, MATHEMATICS AND SPACE: RELEVANCE AND INTERPRETATION

The study of relationships between phenomena has always been one of the major concerns of the geographer. Prior to the quantitative revolution geographers generally stressed that variability of phenomena was directly related to certain causative factors, environmental or otherwise (Harvey, 1973). However, contemporary geographers believe most cause and effect relationships to be complex two-way interrelationships (Harvey, 1973), and further, acknowledge the fact that studies should be extended to encompass spatial relationships, without necessarily implying the incorporation of causal factors. In this regard, the extent to which the variation of a particular phenomenon is related to its spatial location as opposed to other factors is one of the concerns of contemporary geographic inquiry. In general therefore, recent geographic analyses have tended to conform to methodological and theoretical changes that have occurred since the 1950's, particularly the incorporation of mathematical language into geographic methodology and a concern with the concept of space.

Mathematical Models and Geographic Explanation

One of the possible reasons why quantitative geography and the use of mathematics is heavily criticised may be because many geographers labour under a lack of understanding of the technique they wish to use; how it should be applied, how it functions, what is achieved in the process of analysis, and what the results mean in the context of the particular problem (Marchand, 1974).

Quantification in geography, which has involved mainly the incorporation of statistical techniques to "... analyse interaction, to establish comparative relationships, and to verify hypotheses". (Beujeu-Garnier, 1976, p.39), has also involved use of the language of mathematics which facilitates the abstract formulation of geographic concepts (Nystuen, 1963). The abstraction of basic concepts means that geographic problems can be reformulated through a universal language and a basic system of logical reasoning. The product of such a reformulation of a real world problem can be seen as a model. A model is taken to mean simply a representation of reality, where representation comprises an expression of certain relevant characteristics of the observed real world situation (Echenique, 1972). As such a model cannot ever be conceived as being equal to reality, but must necessarily be seen as a likeness thereof (March, 1974, 1975).

However, in order to gain a representation of reality which is not dissimilar to the real world certain assumptions regarding the real world situation need to be made, for example, in many models an isotropic surface is assumed. The assumptions regarding the real world situation chosen by the researcher are crucial to the way in which the results of the modelling process are interpreted. If the basic assumptions are invalid any interpretation of the results will be pointless. Further, the results should be interpreted within the context of the mathematical framework as well as within the context of the particular problem, i.e., the results might have different implications for a mathematician and a geographer. It is therefore not possible for the results of any form of analysis to be wrong, assuming that there are neither computational errors nor errors in the formula. Very often the concept of incorrectness lies with either the formulation of assumptions upon which the model is based or with the interpretation of the results (Moss, 1970).

Finally, therefore, even though it is generally accepted that measuring and quantifying variables in human geography and the consequent application of mathematical techniques is a complex task, it is necessary that geographical analyses as well as the interpretation of those analyses be precise and objective; aims which cannot be achieved without the use of mathematical models (Beaujeu-Garnier, 1976).

The geographer's concern with the study of relationships has led to the incorporation of many techniques into geographic methodology in an effort to arrive at precise statements concerning the relationship between two variables. A technique which facilitates a mathematical evaluation of the linear relationship between any two variables is correlation analysis.

Correlation analysis

In geographic research it is often necessary to know to what extent a set of two variables co-vary. It may be important to know simply how strong the linear relationship between two variables is, or alternatively, how closely values of one variable can be estimated from values of another. If variation in one variable is reflected by a similar variation in another variable, and the relationship cannot be said to be one of chance or circumstance, then it is assumed that one variable will in some respect be related to the other. When all the variation in a single variable can be accounted for by association differences in the value of the accompanying variable, perfect correlation exists. Correlation analysis, therefore, involves the calculation of a correlation coefficient which is an index that reflects the "... degree to which changes in one direction and magnitude in one set of data are associated with comparable changes in the other set" (Gregory, 1963, p.189). Two very important characteristics of a correlation coefficient are firstly that it must be interpreted solely as a measure of the strength of the linear relationship between two variables. Secondly, this interpretation is purely mathematical and performs no other function than the measurement of the co-variation between two variables. Because of this no cause and effect relationship can be implied, even though the method was originally developed for cause and effect series in the physical sciences (Ezekiel and Fox, 1959). In other words, the fact that two variables co-vary does not imply that one is dependent on the other. It could be that both variables are influenced by a third unknown variable in such a manner as to give rise to a strong mathematical relationship. Further, it is possible that even the indirect link may not be present, and that the correlation coefficient is partially or perhaps even wholly spurious (for a discussion of spurious or bogus correlations see Benson, 1965).

One measure of the relationship between two random variables is the degree to which they co-vary (Harnett and Murphy, 1975), which is defined as

$$S_{xy} = {}^{1}/n \sum_{i=1}^{n} (x_i-\bar{x})\ (y_i-\bar{y}) \qquad \text{Equation 1}$$

Comparison of the two variables x and y is facilitated by calculating the amount and direction to which each pair of x and y values differ from their respective mean values $\bar{x}$ and $\bar{y}$. In order to find a value which for any single x and y pair will express the combined variation about the mean, the two separate deviations $(x_i-\bar{x})$ and $(y_i-\bar{y})$ are multiplied, which will result in either a positive or a negative value. A positive result indicates that both variable deviate in the same direction while a negative result indicates that the deviations of x_i and y_i about their mean values are in opposing directions.

Clearly an evaluation of the relationship between two variables will be detrimentally affected by the likelyhood of each variable fluctuating at a different rate. It is therefore unwise to attempt a comparison between two variables by imposing a fixed point of origin and a fixed scale of measurement. For example, to regard two variables such as average family income and distance of residence from the centre of town as having equal points of origin is unrealistic. Also, a small variation in distance could be linearly related to a large variation in income and therefore it is frequently necessary to standardize the scale of variation. In other words, if one, or both, variables are doubled or even squared the basic relationship, i.e. how the two variables vary together, should remain unaltered. As defined in equation 1, co-variance is not a good indicator of the relative strength of the relationship between two variables because its magnitude depends greatly on the units used to measure the variables (Harnett and Murphy, 1975). Consider again the variables family income and distance of residence from the centre of town. The degree of co-variance between these two variables will depend upon the scale at which the variables are measured, for example, if distance were measured in miles the co-variance will be one figure while if distance were measured in kilometers the co-variance will be another.

In order to standardize the co-variance, i.e., render differences in the scale of measurement to zero, S_{xy} is divided by the product of the standard deviations of x and y, the resulting measure r being called the correlation coefficient, where

$$r = \frac{\text{co-variance of } x \text{ and } y}{(\text{standard deviation of } x)\ (\text{standard deviation of } y)} \qquad \text{Equation 2}$$

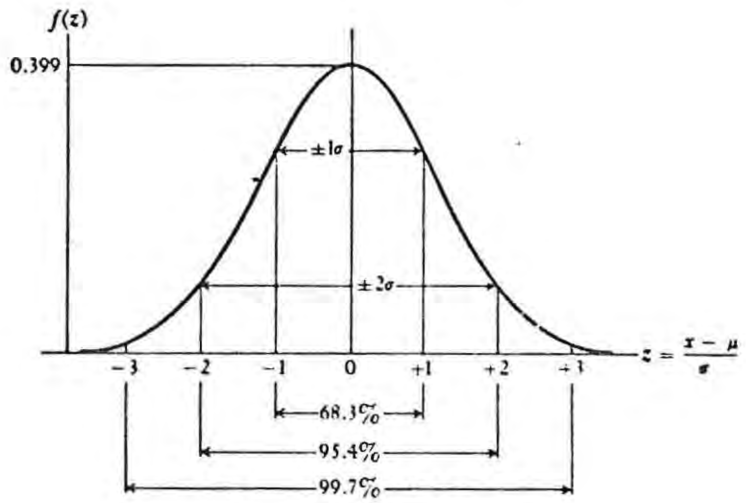
The possible values of r lie between +1 and −1. The former case indicates that both variables co-vary at the same magnitude in the same direction on all occasions while the latter case indicates that both variables co-vary at the same magnitude in opposing directions on all occasions. If the two variables do not co-vary in either a positive or negative fashion the value of r will tend towards zero.

In most practical situations encountered in research it is impossible to test the relationship between any two variables using the entire set of measurements possibly available to one, i.e. the population, simply because most populations are so vast they often have to be conceived as hypothetical entities. Therefore, since populations can seldom be studied exhaustively researchers must content themselves with samples taken from the population as a basis for arriving at conclusions about the population. A sampling distribution, as defined by Haber and Runyan (1973, p.203) is "... a theoretical probability distribution of the possible values of some sample statistic which would occur if we were to draw all possible samples of a fixed size from a given population". The universal set of all possible sampling distributions is therefore infinitely large and as such the findings of many research projects are often based on merely one sample. A major question in all research projects where sample data was employed is whether the sample characteristics adequately reflect those inherent in the population from which the sample was drawn. This is commonly achieved by a process known as significance testing.

The Significance of r

Using correlation analysis as an example, it is always necessary to know whether the coefficient calculated from analysing the relationship between two sets of sample values is an adequate reflection of the relationship that exists between the two populations. In order to test whether the coefficient occurred by chance it is necessary to conduct a test of significance. Supposing it were possible to draw, from the two populations in question, all possible samples of a certain size and calculate a correlation coefficient for each pair, the resulting coefficients, when plotted on a set of axes, would yield a distribution resembling a normal curve. Therefore, in order to test the significance of any particular correlation, researchers often choose to use the normal distribution curve on the basis of its mathematical convenience, and the fact that research has often shown distributions in reality to approximate the normal (De Groot, 1975, p.217). In order to compare an observed distribution to a normal curve some form of standardization is necessary. For this reason the standard normal curve is defined as $N(\mu, \sigma^2)$ where μ and σ^2 are the mean and the variance of the curve and indicate the shape of the distribution. The mean is usually zero and the variance one, i.e. N (0,1).

FIGURE 1. Standardized normal distribution



(from Harnett and Murphy, 1975, p.180.)

For purposes of comparison the distribution of the set of possible sample coefficients, of which the observed coefficient is one possible representative value, has to be transformed to a standardized form as reflected in Figure 1. Transformation of the sample distribution is effected by the following method

$$z = \frac{\text{observed coefficient - hypothetical mean}}{\text{standard deviation of the statistic}} \quad \frac{r - \mu}{\sigma} \qquad \text{Equation 3}$$

(Mosteller, Rourke and Thomas, 1970; Harnett and Murphy, 1975)

This new variable z is commonly known as the standardized random variable and its distribution is the same as that of the original variable, but has a specified mean and variance.

Clearly this method involves calculating the hypothetical mean of the observed population, μ , as well as the standard deviation of the statistic. The variance of the observed statistic can be expressed as $\sigma^2 = E(r^2) - (E(r))^2$ (Mosteller, Rourke and Thomas, 1970, p.209), where $E(r^2)$ is the second moment of the observed statistic and $E(r)$ is the first moment μ of the observed statistic.

Having evaluated the standardized random variable it is possible to see whether the observed coefficient could have occurred by chance alone. The test is conducted as follows. A statistical hypothesis is formulated which states that no relationship exists between the two populations being considered even though the observed sample correlation coefficient may suggest that a relationship does exist. This statistical hypothesis is referred to as the null hypothesis. The truth of the null hypothesis is tested by comparing the standardized random variable against the normal curve as shown in Figure 1. The critical limits under the normal curve are shown as percentages in Figure 1, viz. 68.3%, 95.4% and 99.7%, which refer respectively to the various acceptance or rejection zones commonly required for the test of significance. The researcher is free to choose an appropriate level at which the null hypothesis will either be accepted or rejected. The chosen level is usually a reflection of the type of problem being researched. In many geographical problems it is often the 95% level that is used as the cut off point, although such a high degree of precision is not necessarily required in geographic studies.

The acceptance level could be as low as 68.3%. One assumes the null hypothesis to be true and, on the basis of this assumption, the probability is calculated (Haber and Runyan, 1973), by reading off the tables at what probability level a standard normal deviate value would have occured by chance alone. On the basis of this reading, if the null hypothesis may be rejected then the significance of the coefficient has been established and it is possible to regard the relationship isolated, using sample data, as being representative of the respective populations.

In general, two major problems still exist which tend to hamper research that seeks to analyse the relationships between certain variables, viz., the danger of wrongly inferring dependence between variables and the problem of data collection. As far as the former problem is concerned geographers have frequently assumed that co-variability of phenomena was directly related to certain causative factors, and this has led to attempted explanation of pattern according to purely deterministic principles (Harvey, 1973). Attempting to establish degree of interdependence in the social sciences is acceptable but correlation analysis should not be used as a means towards attaining this goal, simply because the interpretation of the coefficient will depend on what assumptions were made concerning the expected nature of the relationship prior to analysis. In other words, cause and effect relationships in a social science such as geography are seldom as simple as those encountered in the physical sciences. Furthermore, it is clear that if a geographical pattern is thought to result from certain causative factors some form of evolution or temporal component is implied. The aspect of transformation of a pattern over a certain period of time is very difficult to analyse and as such inhibits any form of research which seeks to determine linkages between pattern and causative factors (Harvey, 1968). As a result geographers have realized that where causal thinking is necessary it should be used solely as a means to establishing what Harvey, (1973, p.403) calls 'working assumptions'. Clearly it is useful to think deterministically but such thinking should be restricted to a theoretical level and should be combined with recognition and acceptance of the concept of uncertainty by the inclusion of probability in geographic methodology. Geographers therefore, can use correlation methodology for purposes of prediction whenever a high coefficient is obtained, as long as this is done without inferring any direct cause and effect relationship and it is remembered that prediction will be useful solely in relation to a certain degree of probability which reflects the significance level established.

Because misinterpretation of the results gained from correlation analysis, and other techniques for that matter, is possible the true value of research findings are often undermined which has tended to inhibit a strong body of theory developing in human geography.

The various problems outlined in analysing either the relationship between two similar patterns or the relationship between a particular pattern and certain assumed causative factors, have necessitated a turn to a feasible alternative, which is the analysis of pattern with respect to itself; the form, the structure, and the nature and strength of the point distribution and any linkages. Clearly the strategy involved in this type of analysis is inductivist and the approach which is followed concerns largely the study of form (Moss, 1970). The use of statistical techniques in inductivist strategy provides a good background for the rigorous testing of hypotheses but the results gained must be regarded as tentative, simply because the interpretation is from the unique observable facts towards generalizations about the real world. Clearly, for a body of theory to be evolved generalizations concerning the real world are required. On the basis of widely accepted theory it is possible to develop experimental modelling of real world situations, this approach being clearly deductivist in strategy in that the direction is from the general to the specific. However, following the inductivist strategy many contemporary geographers are concerned with the analysis of the spatial component in the distribution of real world phenomena. In this regard it is not so much the location of phenomena and the distance separating them that is of interest but rather the way in which space itself is reflected in the distribution of phenomena, i.e. the interaction, the communication, the flow between phenomena.

The Concept of Space

The concept of space as a factor in the distribution of some phenomenon has given rise to many philosophical problems simply because it is difficult to conceive of space as an independent variable, and also, to quantify the variable space. The major problem in this respect is one of definition, i.e. the distinction between space and pattern is not always clearly understood.

Pattern can be defined simply as any non-bounded collection of points. In other words, pattern can be expressed via the relative spacing of the individual points in any distribution (Hudson and Fowler, 1976). Therefore pattern has some form of spatial quality in that a two-dimensional surface is involved in any planer distribution of points. The relationship between points can be expressed in three different ways, viz. in terms of the distance between them, in terms of the orientation of each point in relation to every other point, or finally, in terms of the influence of each point on the area surrounding that point. The concept in this latter case is known as spatial dependence, which implies that the presence of a point (or some quality) at one location on a two-dimensional surface is likely to make the presence of a similar point (or quality) in an adjacent location either more or less likely (Cliff and Ord, 1973). When assessing the degree of spatial dependence the emphasis is upon the occurrence of the point or the occurrence of the value (or quality) being measured. The question posed is: What importance does space have in the occurrence of any point or value? As far as spatial dependence is concerned two important aspects emerge. Firstly, the distance and orientation between points are taken as given and secondly, the concept of probability is incorporated.

Coupled with the problem of how space can be regarded as an independent variable is the problem that geographers often fail to establish in what manner they should analyse the specific geographic problem with which they are faced (Harvey, 1973). The question arising at this stage is what techniques are available to geographers for the analysis of the nature and extent of ordered point patterns with particular reference to the form, the structure, and the nature and strength of the point distribution and any linkages, i.e. without making assumptions regarding cause and effect relationships between pattern and causal factors?

In particular, the point pattern resulting from the distribution of crime is of concern in this study. Criminologists, sociologists, psychologists, and more recently geographers, have attempted to isolate a generally accepted theory of crime causation with little success (for example; Sutherland, 1939; Mays, 1956, 1968; McCord, McCord and Zola, 1959; Harries, 1971, 1973, 1974 respectively). The importance of space in the occurrence of crime is also unknown. Although it is possible that analyses that relate pattern to assumed causal factors such as economic, social, familial, psychological, ethnic and religious variables among others, would yield interesting results, these would have to be regarded conservatively due to the known complexity of the assumed cause and effect relationship between the motivation for crime and its relative position in space. An analysis of the pattern resulting from crime occurrence with respect to itself can provide an insight into the importance of the variable space in such a distribution, and as such aid future research which seeks to develop a generally accepted theory of crime occurrence.

CHAPTER III

ANALYSIS OF POINT PATTERNS

Geographical phenomena can be distributed in space in simple geometric forms, viz., points, lines, areas and volumes. Their distribution can be classified according to intensities, associations between them, or flows between phenomena (Harvey, 1973). None of these categorizations deal with space but deal rather with the distributional characteristics of the phenomena, i.e. distance and orientation.

A point distribution, for example, could comprise a map on which each observation of a certain variable is represented by a point. Similarly, associations between points could be represented by lines, and combination of points could be represented by areas or volumes. By conceptualizing about various distributions of phenomena in one of these forms it is possible to generalize about pattern. This can be achieved intuitively or by translating the information into mathematical language, thus facilitating an objective assessment of the nature and strength of the pattern. Since the adoption of mathematical techniques geographers, notably Chisholm (1975), Chorley (1965), Dacey (1968) and Harvey (1966, 1968, 1973), have been able to offer objective statements about the form and strength of various patterns, yet the problems discussed above still exist; namely that geographers very often lack sound knowledge of mathematical language and consequently apply techniques to data which are often unsuited to the assumptions inherent in the technique, and also that results gained from using various techniques are misinterpreted because of a lack of understanding of the assumptions underlying the technique (Harvey, 1973).

Analysis of point distributions

Three recognized means of analysing a point pattern are available to geographers. Firstly, the pattern one wishes to analyse can be compared to similar orders in the distribution of local environmental qualities, i.e. the observed pattern can be compared to a similar environmental pattern such as relief, rainfall, soil type, etc. This is known as establishing the degree of areal association (Herbert, 1977). Secondly, attempts can be made to relate the pattern to underlying mechanisms which are thought to have caused the distribution and thirdly, the pattern can be analysed with respect to itself (Herbert, 1977). In each case the emphasis is on measuring the relationship between the observed pattern and either another pattern, or a causally related pattern, or the pattern itself. According to Berry (1968) geographers have tended to direct their attention towards the processes and integrating concepts that result directly in spatial patterns, and it is possible that the focus of attention towards assumed causal relationships between pattern and independent causative factors is due to the availability of techniques developed in the physical sciences. However, in order to analyse pattern with respect to itself, a different set of techniques is required.

Analysis of Pattern with respect to itself

As the spatial pattern resulting from the distribution of crime occurrence forms the focus of this study the following techniques which are available to the geographer for the analysis of pattern with respect to itself will be evaluated with specific regard to their relevance in analysis of crime patterns. In the case of analysing pattern with respect to itself there are two general categories of techniques available to geographers, viz., generalized mathematical measures and specific mathematical representations (Harvey, 1968). The former category includes techniques which have a bias towards analysis of process while the latter group of techniques tend to focus on the analysis of actual pattern. In both cases the mathematics is essentially similar but in the former case a temporal component is involved, while in the latter case attention is focussed on pattern at one time. This separation, which is necessary to simplify matters, involves conceptually dividing activity into process on the one hand, and pattern on the other, although these two concepts are, for all practical purposes, inseparable. Geographers, however, seem to prefer to analyse spatial patterns with an emphasis on either process or pattern.

Process oriented techniques

Analysis of spatial processes in geography inevitably implies a temporal dimension simply because the term process is used to refer to any sequence of events over time. Analysis of spatial processes must therefore recognize and incorporate a temporal dimension. However, the mechanisms generating spatial patterns are regarded as extremely complex because of limited knowledge concerning temporal processes and the problem of scale of analysis (Harvey, 1968). Therefore, analytical techniques at present in use for analysis of process, viz. generalized mathematical measures "... attempt to store as much information about pattern as possible by way of some mathematical expression (which need not be linear) or to generalize objectively about that pattern" (Harvey, 1968, p.73). These are classificatory devices commonly known as smoothing or interpolation techniques which facilitate reasoned, logical inferences regarding spatial process. In this regard techniques such as trend surface analysis, fourier analysis, and spectral analysis, may aid understanding of process but in no way do they offer specific interpretation with respect to process (Harvey, 1968). In addition to the above, investigators concerned with the problem of crime feel that the complexity of the interrelations of crime variables defy synthesis into any form of general theory and as a result an accepted theory of crime causation does not exist (Harries, 1974). Clearly therefore, it is beyond the scope of a geographical study to provide answers to the question of why crime occurs and as such analysis of the complex processes underlying crime patterns with the aid of the abovementioned process oriented techniques is perhaps best suited to multidisciplinary research. However, geographers can aid research into crime occurrence by analysing the spatial aspects in particular and, in so doing, add an element of understanding that might otherwise be overlooked.

Pattern oriented techniques

Specific mathematical representations measure the deviation of a pattern from a theoretical pattern generated under certain assumptions. Among these techniques nearest neighbour analysis and quadrat analysis share the basic assumption that the pattern against which the observed one is being compared, is generated by random processes (Getis, 1964). They differ in the methodology used to compare the observed with the expected pattern. A third technique in this category, cluster analysis, relies upon the assumption that the observed pattern is clustered, because the technique provides a set of cluster boundaries for any given distribution of points. For this reason cluster analysis is usually preceeded by either nearest neighbour analysis or quadrat analysis which will provide some indication of the nature of the distribution. In all three techniques distance between points is the important factor, in that it is the spacing of a distribution of points with respect to one another which forms the basis of the technique. In this way the techniques provide a means of classification of pattern. The inter-relationships or the spatial interaction between points are not considered and as such an evaluation of the overall spatial component in the distribution is lacking. However, classificatory techniques could be usefully applied in conjunction with a technique which facilitates an isolation and evaluation of the importance of space in any distribution.

(i) Nearest Neighbour Analysis

This method "... gauges the departure from randomness of an observed spatial point pattern ..." (Rogers, 1974, p.8). The distance between each point and its nearest neighbour forms the basis of the method (Clark and Evans, 1954, 1955; De Vos, 1973). The mean of the distance between the various points is used to calculate a density function which, in turn, is used to indicate the degree to which the observed pattern deviates from an expected pattern generated under a stochastic process. The observed pattern is classified according to the value of the density function R and is then classified under one of three possible categories, viz., clustered, random or regular (Pinder and Witherick, 1972). However, pattern cannot be measured along a continuous scale between the three categories and therefore it is not possible to infer the relative position of any one pattern in relation to another or the same pattern at different times (Sibley, 1976). Further limitations are that the function R is density dependent which means that the location and shape of the boundary affects the results that can be gained from applying nearest neighbour analysis to any pattern. Finally, recent research has suggested that because the function is density dependent it does not provide a description of pattern characteristics but merely information concerning the distribution of nearest neighbour distances (Sibley, 1976; Vincent, 1976). Sibley (1976) has discussed the ways in which geographers have misinterpreted the functioning of nearest neighbour analysis and this provides good examples of how the mathematical capabilities of a technique are not understood, thus resulting in inappropriate applications and questionable research findings. Notwithstanding the above limitations of nearest neighbour analysis the technique is inappropriate to crime research because it is already accepted (Harries, 1971, 1973, 1974; Pyle, 1974; Corsi and Harvey, 1975) that most, if not all, patterns of crime occurrence tend to be clustered. As the technique does not provide a means by which the relative degree of clustering may be gauged, its application to the analysis of crime patterns is not warranted.

(ii) Quadrat Analysis

In this method, a study region is "... divided into a grid with cells of equal size, called quadrats, and the number of points in each cell, or in sample cells, is noted "... (Rogers, 1974, p.5). As with nearest neighbour analysis, if the observed frequency of points does not conform with one expected from a random process then the hypothesis of randomness is rejected in favour of a suitable alternative, usually a regular or a clustered pattern. The underlying procedure for quadrat analysis is simple. If the point pattern is such that every quadrat contains one point or less (or an even number of points) then the distribution is said to be regular. Where some quadrats hold most of the points a clustered distribution is said to exist, while a random distribution would fall somewhere between these two extremes. In order to ensure a statistically acceptable classification of a point distribution into one of these three categories two tests may be employed, viz., the variance - mean ratio and the chi-squared goodness of fit test (Rogers, 1974). A major problem associated with this technique is the implication of choice of quadrat size which unless carefully selected may result in the analysis hiding more than it reveals. Further, the technique tends to be more useful in analysing dispersion than pattern and can therefore be used purely as a classificatory device. Quadrat analysis is therefore inappropriate as the only means of establishing patterns in crime research because it provides little more than classification of the configuration of the distribution.

A problem which applies to both nearest neighbour analysis and quadrat analysis is that many different probability models can give rise to similar frequency distributions (Rogers, 1974). In order to generate a poisson distribution various combinations of an infinite number of possible assumptions can be employed. The resulting probability functions will be mathematically similar and will produce similar, but not identical, frequency distributions (Getis, 1977). Clearly this mathematical limitation can lead to serious problems when observed frequencies are compared with expected frequencies, and further, it is not possible to understand a spatial pattern from a knowledge of frequency distributions alone (Vincent, 1976). Notwithstanding this mathematical limitation the study of poisson processes provides one of the most appropriate formats for discussing geographic problems because of the relative ease with which empirical geographic problems can be mapped into probability theory that rests on random probability (Harvey, 1973).

(iii) Cluster Analysis

By this method a surface comprising a point distribution is scanned in an effort to establish or produce a mathematically optimum set of clusters (Welch, 1976). The technique establishes clusters of points on the basis of relating every point on the surface to a number of seeding points which are chosen by the researcher prior to analysis (Welch, 1976). Seeding points represent those points or areas which are expected to be optimum centres of the various clusters and may be chosen on the basis of logical reasoning. All the remaining points are identified by means of co-ordinates and, on the basis of their location, are related to the various seeding points. A cluster boundary around each seeding point is then assigned on the basis of the optimum total sum of the squared deviation of distance between the seeding point and the surrounding points. The process is complete when the arrangement of boundaries around each seeding point cannot mathematically be improved (Welch, 1976). Computation is frequently based on non-hierarchical clustering procedures, i.e.

clustering procedure is based on straight-line contiguity, but the technique can be adapted to suit hierarchical procedures. Because the technique will provide a set of cluster boundaries for any distribution of points it is usual to test for the existence of clusters prior to analysis. Further, although the resulting boundaries around the seeding points are mathematically optimum they are not necessarily geographically meaningful. Two limitations exist; firstly, the selection of seeding points is subject to personal bias and secondly, the researcher has to select an appropriate boundary for each cluster which could result in non-objective statements of research findings unless directly related to complimentary analysis.

The above three techniques provide the means by which pattern can be analysed with reference to the location of points on a two-dimensional surface. Thus the relationship between points is evaluated purely in terms of the relative spacing between each point and its neighbours. The spacing between points is usually measured as straight-line distance although distance is measurable in a variety of ways, i.e. among others, distance can be measured in terms of time, cost and energy expended.

A technique for the analysis of point distributions where the physical location of points is taken as given and the relationship between points is regarded as primary is spatial autocorrelation. Cliff and Ord (1969, 1971, 1972, 1973, 1975 a, b,), developed the technique from three basic sources, viz. Moran (1950 a, b), Whittle (1954) and Geary (1968). The equation, which measures the degree of co-variance between two sets of data, follows the standard correlation format. However, in the case of auto-correlation only one data set is involved and thus the relationship between values of the same data set can be investigated. This facilitates inferences regarding the degree of spatial dependence among the given data (Cliff and Ord, 1973). Statements indicating the degree of spatial dependence are particularly useful when considering data that exhibits specific distributional characteristics. For example, a pattern resulting from the occurrence of crime will reflect the following characteristics, viz. the points will have no affixed measurable value; the points will represent discrete occurrences; data will be available for every point on the map (therefore rendering the application of interpolation techniques inappropriate); knowledge of the processes giving rise to the pattern is weak, i.e. little is known about the link between the motivation for the act and where the act is perpetrated. Studies conducted overseas, notably Pyle (1974), Harries (1971, 1973, 1974) and Herbert (1977) and a pilot study conducted in Grahamstown in 1976 show evidence that spatial patterns which result from crime occurrence exhibit distinctive qualities and variations. With regard to possible locations of criminal actions, which should be considered in the context of motivation, it is clear that crime can occur at any point in space although some crimes are certainly space constrained, for example, crime against property. On the other hand, it is likely that crime against person, although relatively unconstrained, may have a probability distribution which varies with particular spatial locations.

By evaluating the spatial component of crime patterns the extent to which crime occurrences are associated in spatial terms will become evident. Clearly research of this nature will aid attempts to predict crime occurrence, i.e. the probability with which crimes occurring in one area can be related to crime occurrence in an adjacent area will become evident. Further, an evalutation of space will focus attention on particular areas, which could result in a detailed analysis of other factors that may co-vary with space and as such be related to crime occurrence. In this regard it is clear that the only technique presently available for an objective assessment of the importance of space in crime occurrence (and other patterns) is autocorrelation. It is the one technique which offers a solution to the question of spatial dependence.

CHAPTER IV

SPATIAL AUTOCORRELATION

Geographical interest in spatial autocorrelation lies in the fact that it facilitates an expression of the spatial characteristic of any distribution of data (Olson, 1975). Spatial autocorrelation involves the manipulation of simple correlation procedures for assessing the relationship between values of a single set of data in a spatial context. If neighbouring values on a two-dimensional surface co-vary then positive spatial autocorrelation exists. However, the study of spatial autocorrelation is not important simply because an observation may be seen to be a function of neighbouring values but also because every value on a two-dimensional surface is affected by the flow or interaction between itself and neighbouring values (Cliff and Ord, 1975). Measuring autocorrelation in geographic data is different from measuring autocorrelation in a time series because dependence is seen to extend in all directions. In this regard three problems arise when attempting to evaluate spatial dependence, viz. the nature of the statistic to be used, the form of the variable which is to be measured and finally, the form of the interaction between observations (Cliff and Ord, 1975). This chapter considers the various problems regarding the evaluation of spatial dependence by outlining Cliff and Ord's method with reference to a practical situation. Theoretic values are used to represent a spatial distribution of a variate X, and dummy grids provide the means of demarcating sub-areas on a map. Thus, working with hypothetical data it is hoped that understanding of the reaction of the statistic to various situations will be gained.

The Statistic

Consider a two dimensional surface comprising non-overlapping regular or irregular areas. The variate X is distributed over this surface so that each sub-area contains a value x_i which is the value of X in sub-area i . Taking the location of each x_i value as given the spatial relationship between the x_i values can be determined by correlating each x_i value with all the neighbouring x_i values where neighbour is defined according to the form of interaction one wishes to evaluate.

Cliff and Ord (1969) put forward an autocorrelation statistic developed initially by Moran (1950 a) and Geary (1968). Both Moran and Geary's statistics take the standard form of any correlation coefficient, viz. "... the numerator term ... is a measure of covariance among the x_i (individual observations) and the denominator term is a measure of variance" (Cliff and Ord, 1973, p. 8). Further, both statistics examine the relationship between contiguous x_i only, and the concept of neighbouring values is limited to spatially adjacent x_i values in all directions. For the purpose of a test analysis it is adequate to view the interaction between individual sub-areas that are adjacent and it is not necessary to make generalizations concerning interaction between sub-areas that are spatially separate, but spatial autocorrelation could be extended to take second order relationships into account in situations where non-proximate co-variance is expected to occur.

Following the standard formula which facilitates computation of the degree of linear association between two separate variables x and y

$$r = \frac{\sum_{i=1}^{n} (x-\bar{x})\ (y-\bar{y})}{n\ \sigma x\ \sigma y}$$

Equation 4
(Gregory, 1963. p.189)

it can be seen that Moran's statistic (Cliff and Ord, 1973), p. 8) is merely an extension of the same procedure.

$$r = \frac{n \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} \delta_{ij}\ (x_i-\bar{x})\ (x_j-\bar{x})}{W \sum_{i=1}^{n} (x_i-\bar{x})^2}$$

Equation 5

when n $=$ the number of x_i values

$\sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n}$ $\equiv$ summation of all pairs of x_i and x_j throughout the matrix

δ_{ij} $\equiv$ the indication of interaction between x_i and x_j

x_i $=$ the value of the current subarea

x_j $=$ the value of the neighbouring subarea

W $\equiv$ the total number of joins in the system.

In the case of a surface comprising regular, non-overlapping sub-areas δ_{ij} could be simply 1 if the sub-areas are adjacent neighbours and 0 if the sub-areas are non-adjacent or nth order neighbours. This form of δ_{ij} is a simple example of a binary weighting system.

Cliff and Ord (1969, 1973, 1975ab) have extended equation 5 to developing a specific form of δ_{ij} which if wished can allow for incorporation of areal size differences between sub-areas as well as variation in straight-line distances between the central points of these sub-areas.

$$r = \frac{n \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} w_{ij} \, z_i \, z_j}{W \sum_{i=1}^{n} z_i^2}$$

Equation 6
(Cliff and Ord, 1973, p. 12)

where w_{ij} $\equiv$ matrix of values indicating the nature of the interaction between x_i and x_j

z_i $=$ $(x_i - \bar{x})$

z_j $=$ $(x_j - \bar{x})$

By using a generalized interaction function w_{ij} Cliff and Ord's statistic has two major advantages over both the Moran and the Geary statistics. Firstly, the concept of a neighbouring value need not necessarily be restricted to first order interaction. Secondly, if the surface comprises irregular sub-areas (as is the case with many geographic surfaces) the relative sizes of these sub-areas can be standardized. Nevertheless the Cliff and Ord statistic can be applied to the regular sub-area case using adjacent cells only.

Cliff and Ord (1971, p.53) have shown that the distribution of their coefficient is affected by four variables, viz., the distribution of the variate X, the shape of the system of sub-areas and the nature of the joins (i.e. interaction), the size of the sub-area system (denoted by n), and the form of w_{ij} .

Specifically, how do each of these four variables affect the correlation coefficient derived from the statistic? We know that in a classical correlation situation we can expect a coefficient of 1.00 if the two sets of data co-vary perfectly regardless of the number of observations. However, it is clear, by reference to the above statement that even if a theoretically perfect pattern of observed values did exist, the nature of Cliff and Ord's coefficient could be affected by one or a combination of factors, and as such interpretation of the coefficient is more complex than is the case with the classical correlation coefficient. In fact Cliff and Ord (1973) imply that it is impractical to interpret the correlation coefficient yielded by the statistic mainly due to the fact that the coefficient does not vary between a fixed bounded range as does the coefficient resulting from Pearson's method. Cliff and Ord (1971) prove that as the number of observations increases so does the distribution of their coefficient tend towards normality, and as such their coefficient can be tested for significance as a standard normal deviate. Following this, Cliff and Ord (1969-1975) go to lengths to perfect the means by which the results of their statistic may be tested for significance. However, what they fail to emphasize, and what is important to the average geographer, is how their statistic reacts in a practical situation and further, how the results may best be interpreted. As geographers, familiar with classical correlation procedures, we know that it is possible for a weak correlation to be significant, and as such we are used to interpreting the strength of the coefficient in relation to its position along the continuum of -1 to $+1$.

The aim of this part of the study therefore, is to construct a computer program whereby Cliff and Ord's spatial autocorrelation statistic can be applied to establish the degree to which variation in one or more of the factors listed above affect the resultant correlation coefficient values.

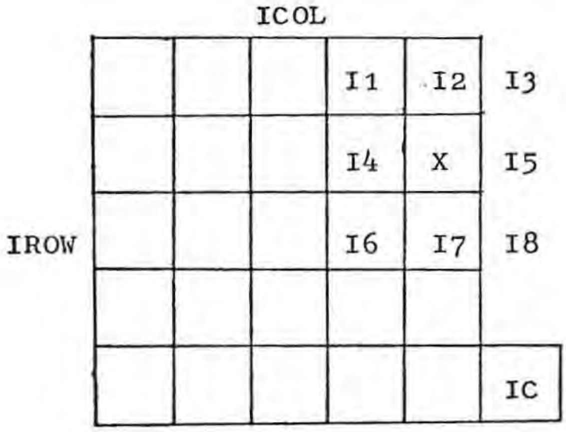
In order to fulfil the above aim the program is written in standard Fortran IV and wherever possible, has been designed to facilitate adaptability to different situations. Specifically, the program applies to the ICL 1900 T computer which has a maximum of 42K storage. Because of this a conscious effort was made to keep the storage capacity of the program as low as possible, for example, a matrix comprising 250 grid squares will require approximately 20 K storage. Finally, the approach throughout is designed with what is hoped to be logical simplicity so as to enable geographers with a minimal knowledge of computing to be able to follow the procedure.

Application of Cliff and Ord's formula to a date set, which has been subdivided according to a regular matrix, can be achieved very simply. Demarcation of sub-areas according to a regular grid effectively eliminated the problems of relative sizes of irregular sub-areas and relative distances between the central points of these sub-areas thus holding at least two of the variables mentioned by Cliff and Ord constant. In addition by considering the case where interaction occurs between neighbouring cells only the simple case can be dealt with for demonstration purposes. The following procedure is clearly only one of many alternatives:

```
C     THIS SEGMENT CREATES A GRID SQUARE OUTSIDE THE LIMITS OFTHE MATRIX
C     NEND.  REFER COMMENT STATEMENT BELOW
C     THE TOTAL NUMBER OF GRID SQUARES IN THE MATRIX IS INDICATED BY
C     NEND, IE. IROW*ICOL.
C     -----------------------------------------------------------------
      NEND=IROW*ICOL
      IC=NEND+1
      ZDIF(IC)=0.0
C     -----------------------------------------------------------------
C     THIS SEGMENT ISOLATES ALL THE ADJACENT GRID SQUARES TO EVERY SINGLE
C     GRID SQUARE IN THE MATRIX. EVERY GRID SQUARE HAS 8 ADJACENT SQUARES
C     WHICH ARE CODED I1 TO I8. ZDIF SCORES IN ADJACENT GRID SQUARES ARE
C     MULTIPLIED.
C     ALL RESULTING VALUES ARE SUMMED.
C     WAIT IS THE TOTAL NUMBER OF "JOINS" IN THE SYSTEM
C     DED IS A COUNTER FOR D,  USED IN SIGNIFICANCE.
C     BOTH ARE SET TO ZERO.  FOR EXPLANATION REFER BELOW.
C     TOTAL IS SET TO ZERO. FOR EXPLANATION OF TOTAL REFER BELOW.
C     -----------------------------------------------------------------
      WAIT=0.0
      DED=0.0
      TOTAL=0.0
      DO 50 J=1,NEND
      I1=(J-ICOL)-1
      I8=(J+ICOL)+1
      I3=(J-ICOL)+1
      I6=(J+ICOL)-1
      I2=J-ICOL
      I7=J+ICOL
      I4=J-1
      I5=J+1
C     -----------------------------------------------------------------
C     IF I1 TO I8 FALL OUTSIDE THE LIMITS OF MATRIX NEND THEY ARE ROUTED
C     TO IC , WHICH READS 0.0
C     -----------------------------------------------------------------
      IF(J.GT.ICOL) GO TO 60
      I1=IC
      I2=IC
      I3=IC
   60 NB=NEND-ICOL
      IF(J.LE.NB) GO TO 61
      I6=IC
      I7=IC
      I8=IC
   61 IF(J-ICOL*(J/ICOL).NE.0) GO TO 62
      I3=IC
      I5=IC
      I8=IC
   62 NL=(J/ICOL)*ICOL+1
      IF(J.NE.NL) GO TO 63
      I1=IC
      I4=IC
      I6=IC
```

In order to prevent hypothetical cells outside the limits of the matrix being included in the computation, it is necessary in this particular program first to exclude all possibilities in this regard. For this purpose one cell outside the limits of the matrix is given a value of zero. In this example the cell chosen is the one immediately right of the last cell in the matrix (refer Figure 2).

FIGURE 2. Coding for adjacent and off-limit cells



In the above matrix there are 25 cells (IROW X ICOL) and the 26th cell is the one which is set to zero, i.e. IROW X ICOL + 1 = 0. Further, all 25 cells have 8 immediate adjacent neighbouring cells. These are coded I1, I2 ... I8. If any of these cells falls outside the limits of the matrix it is routed to IC which has a preset value of zero. Thus any interaction with cells outside the limits of the matrix will yield an answer of zero and will not be counted as part of the total number of joins in the system. Once neighbouring cells have been isolated two problems arise, viz. the nature of the variable values which have to be compared and what form the interaction between these values should take.

The Variable

Very often data collected and measured in geography are not immediately available in a form suitable for analysis. For example, variables measured in geography are often nominally scaled or rank ordered and consequently tend to be unsuited to rigorous analysis, mainly due to the fact that many of the analytical procedures available to geographers are designed specifically to analyse data which are measured by interval or ratio scales (Wrigley, 1977; Yeates, 1974). Further, the social nature of many geographic variables determines that the underlying spatial distribution can seldom be considered continuous but rather discrete categorizations (Wrigley, 1977).

Although the statistic put forward by Cliff and Ord can be applied to data of any form the focus of interest in this study is on the analysis of a point distribution. Points on a map could represent either a single event or a categorization of events. In the latter case each point represents the number of occurrences within a certain area and the data would be classified as content data. Data consisting of points where the points represent the locations of single events would be regarded as place data (Boots and Getis, 1977).

Assuming a distribution comprising a set of points which represent random discrete occurrences of some event, for example, crime; then it is clear that each point is likely to represent one crime, as it is very seldom that the same type of crime will be committed more than once at precisely the same location. Because it is the variable quantity x_i which is of importance in the calculation of spatial autocorrelation and not the location of the x_i value itself within the sub-area the points should be indicators of content rather than indicators of place in order to enable one to view the occurrence as a variate. When considering crime for example, the number of crimes per city region may be viewed as a variable value of the variate X, which is the distribution of the total number of crimes throughout the urban area. In certain cases therefore, discrete objects or events can be grouped in sub-areas and the resultant values can be used as a reflection of how that variable varies over a two-dimensional surface.

It is now necessary to discuss the second functional prerequisite for spatial autocorrelation, which concerns the form of the interaction between the x_i values of the principal data set.

Interaction between sub-areas

The problems involved in the choice of the way in which interaction between sub-areas can be defined can be discussed under two headings, viz., the problem of demarcation and the problem of interaction. Both of the above problems directly concern the shape of the system of sub-areas and the nature of the joins (Cliff and Ord, 1971, p.53). Further, the problem of demarcation also concerns the size of the sub-area system.

The problem of demarcation

Assuming that most of the data which geographers handle will be in the form of discrete categorizations, the problem arises of how categorization can be effected in the case of spatially distributed data. There are two opinions available, viz. regular or irregular sub-areas. In the case where data are available by recognized sub-areas, for example enumerator's districts, then it is acceptable to use the demarcated counties as the sub-areas used in the analysis. However, data are not always available by sub-area (particularly in smaller South African towns) and as a result researchers are often faced with the problem of supporting their demarcations into sub-areas on the basis of some criteria. A simple solution to this problem is to categorize spatially distributed data according to a regular grid matrix. The only decisions involved in this case are those associated with the choice of a suitable grid size. MacDougal (1976) outlines several considerations which are important in the decision regarding the choice of an appropriate grid size. He states that the grid must be small or fine enough to capture the degree of detail basic to the specific type of problem being considered. However, MacDougall does not mention the possibility of a maximum grid size constraint, and it is equally important to note that if the grid square size is too large the possible variable nature of the data will be lost due to gross generalization. In general satisfactory criteria are that firstly the axes of a grid square should be less than the shortest distance across the smallest area of interest on the map. Secondly that the size of the grid square must be "... small enough to ensure that the smallest region of interest on the map is the modal type ..." in at least one grid square (MacDougall, 1976, p. 60).

Clearly therefore, the choice of a grid size involves both the size of each grid square as well as the number of cells which can be incorporated in the matrix. What Cliff and Ord fail to make explicit, and what the following example clearly shows, is that the coefficient varies according to the number of grid squares (and consequently the number of joins) in the system, even though all other influencing factors are held constant. The results of computation of spatial autocorrelation on 8 matrices of varying size (Figure 3) are shown in Table 1.

TABLE 1. Autocorrelation coefficient: matrix size

Matrix	Frequency of occurrence	Number of joins	Correlation coefficient
ATHR	27	40	0,300 **
BTHR	18	34	0,192 *
ASEV	441	312	0,756 **
BSEV	294	306	0,758 **
ATWL	2376	1012	0,866 **
BTWL	1584	1006	0,869 **
ATEE	11400	2964	0,922 **
BTEE	7600	2958	0,924 **

* Significant at the 95% level

** Significant at the 99.7% level

Although all the correlations are significant at the 95% level it is clear that the strength of the correlation increases with an increase in the number of observations. Clearly, therefore, the larger the size of the matrix the higher the correlation value is likely to be for the same data set.

FIGURE 3. Dummy Grids

SUBFILE ATHR IN CARD MODE

3 3 5
1 2 3 4 5

4	3	2
4	3	2
4	3	2

3 x 3 Matrix - A

SUBFILE BTHR IN CARD MODE

3 3 5
1 2 3 4 5

4	3	2
3	2	1
2	1	0

3 x 3 Matrix - B

SUBFILE ASEV IN CARD MODE

7 7 5
1 2 3 4 5

12	11	10	9	8	7	6
12	11	10	9	8	7	6
12	11	10	9	8	7	6
12	11	10	9	8	7	6
12	11	10	9	8	7	6
12	11	10	9	8	7	6
12	11	10	9	8	7	6

7 x 7 Matrix - A

SUBFILE BSEV IN CARD MODE

7 7 5
1 2 3 4 5

12	11	10	9	8	7	6
11	10	9	8	7	6	5
10	9	8	7	6	5	4
9	8	7	6	5	4	3
8	7	6	5	4	3	2
7	6	5	4	3	2	1
6	5	4	3	2	1	0

7 x 7 Matrix - B

SUBFILE ATWL IN CARD MODE

12 12 5
1 2 3 4 5

22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11
22	21	20	19	18	17	16	15	14	13	12	11

12 x 12 Matrix - A

SUBFILE BTWL IN CARD MODE

12 12 5
1 2 3 4 5

22	21	20	19	18	17	16	15	14	13	12	11
21	20	19	18	17	16	15	14	13	12	11	10
20	19	18	17	16	15	14	13	12	11	10	9
19	18	17	16	15	14	13	12	11	10	9	8
18	17	16	15	14	13	12	11	10	9	8	7
17	16	15	14	13	12	11	10	9	8	7	6
16	15	14	13	12	11	10	9	8	7	6	5
15	14	13	12	11	10	9	8	7	6	5	4
14	13	12	11	10	9	8	7	6	5	4	3
13	12	11	10	9	8	7	6	5	4	3	2
12	11	10	9	8	7	6	5	4	3	2	1
11	10	9	8	7	6	5	4	3	2	1	0

12 x 12 Matrix - B

SUBFILE ATEE IN CARD MODE

20 20 5
1 2 3 4 5

38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19

20 x 20 Matrix - A

SUBFILE BTEE IN CARD MODE

20 20 5
1 2 3 4 5

38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19
37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18
36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17
35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16
34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15
33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14
32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13
31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12
30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11
29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10
28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9
27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8
26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7
25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6
24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5
23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4
22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3
21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2
20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0

20 x 20 Matrix - B

In the matrices shown in Figure 3 all the A-matrices display a perfect left-right trend of theoretical values while the B-matrices display a similar trend but the range in values is double that of the A-matrices. Because of this the theoretical frequency of occurrence of the event is greater in the A-matrices than the B-matrices (Figure 3). The stronger correlation values yielded by the A matrices suggest therefore that the frequency of occurrence also affects the nature of Cliff and Ord's coefficient. Further, in the matrices shown in Figure 3 the frequency of occurrence of the event being measured clearly increases with increasing matrix size, thus sample size increases correspondingly. As sample size increases, the population is more accurately represented, therefore increasing the probability that the sample coefficient will be statistically significant. Therefore, both the number of joins in the system and the frequency of occurrence of the event in the sample are factors which are important in the assessment of autocorrelation values if comparative studies are to be carried out.

Notwithstanding the possibility that the correlation could be spurious, the results of spatial autocorrelation analysis on two matrices with random variate values (Figure 4) indicate that a coefficient of 0.09 need not occur by chance alone (Table 2). This result highlights the fact that it is necessary to attach some form of interpretation to the value of the correlation coefficient itself. In order to be useful, it is necessary to compare results gained from analysis of autocorrelation for different point patterns on the basis of something other than merely whether they reflect a significant level of autocorrelation or not. Naturally the lack of a fixed bounded range makes interpretation of Cliff and Ord's statistic difficult, yet the fact that given the same significance level the coefficient varies according to the size of the sub-area system and that the coefficient varies according to the nature of the interaction

FIGURE 4. Patterns: random values

SUBFILE RANA IN CARD MODE

11 16 5

1 2 3 4 5

20	17	42	28	23	17	59	66	38	61	2	10	86	10	51	55
74	49	4	49	3	4	10	35	53	70	11	54	48	63	94	60
94	70	49	31	38	67	23	42	29	65	40	88	78	71	37	18
22	15	78	15	69	84	32	52	32	54	15	12	54	2	1	37
93	29	12	18	27	30	50	55	91	87	50	57	58	51	49	36
45	4	77	97	36	14	99	45	52	95	69	85	3	83	51	87
44	91	99	49	89	39	94	60	48	49	6	77	64	72	59	26
16	23	91	2	19	96	47	59	89	65	27	84	30	92	63	37
4	50	65	4	65	65	82	42	70	51	55	4	61	47	88	83
32	70	17	72	3	61	66	20	24	71	22	77	88	33	17	78
3	64	59	7	42	95	81	39	6	41	20	81	92	34	51	90

11 16 5

1 2 3 4 5

62	49	1	90	67	86	93	48	31	83	19	7	67	68	49	5
61	2	95	86	98	36	14	5	48	88	51	7	33	40	6	86
89	3	90	49	28	74	21	4	9	96	60	45	22	3	52	80
1	72	33	85	52	40	60	7	6	71	89	27	14	29	55	24
27	56	49	79	34	34	32	22	60	53	91	17	35	26	44	70
49	5	74	48	10	55	35	25	24	28	20	22	35	66	66	34
49	74	37	25	97	26	33	94	42	23	1	28	59	58	92	69
20	26	22	43	83	8	19	85	8	12	47	65	65	63	56	7
48	87	77	96	43	39	76	93	8	79	22	18	54	55	93	75
8	72	87	46	75	73	1	11	27	7	5	20	30	85	22	21
95	97	98	62	17	27	31	42	64	71	46	22	32	75	19	32

(Values obtained from Lindley, D. and Miller, J. 1953. <u>Cambridge Elementry Statistical Tables</u>, Cambridge. Table 8).

between sub-areas is surely important. It is possible therefore that with further work understanding of how the coefficient varies according to these factors may result, and consequently preliminary interpretation of the coefficient may be possible which will increase the analytical value of Cliff and Ord's autocorrelation coefficient.

TABLE 2. Autocorrelation coefficient: random values

Matrix	Run	Number of joins	Correlation coefficient
RANA N = 8597	N/S	320	0,104
	W/E	330	0,023
	NW/SE	300	-0,036
	NE/SW	300	-0,000
	general	1250	0.024
RANB N = 8114	N/S	320	0,147
	W/E	330	0,096
	NW/SE	300	0,058
	NE/SW	300	0,184 *
	general	1250	0,093 *

* Significant at the 95% level

** Significant at the 99.7% level.

The problem of interaction

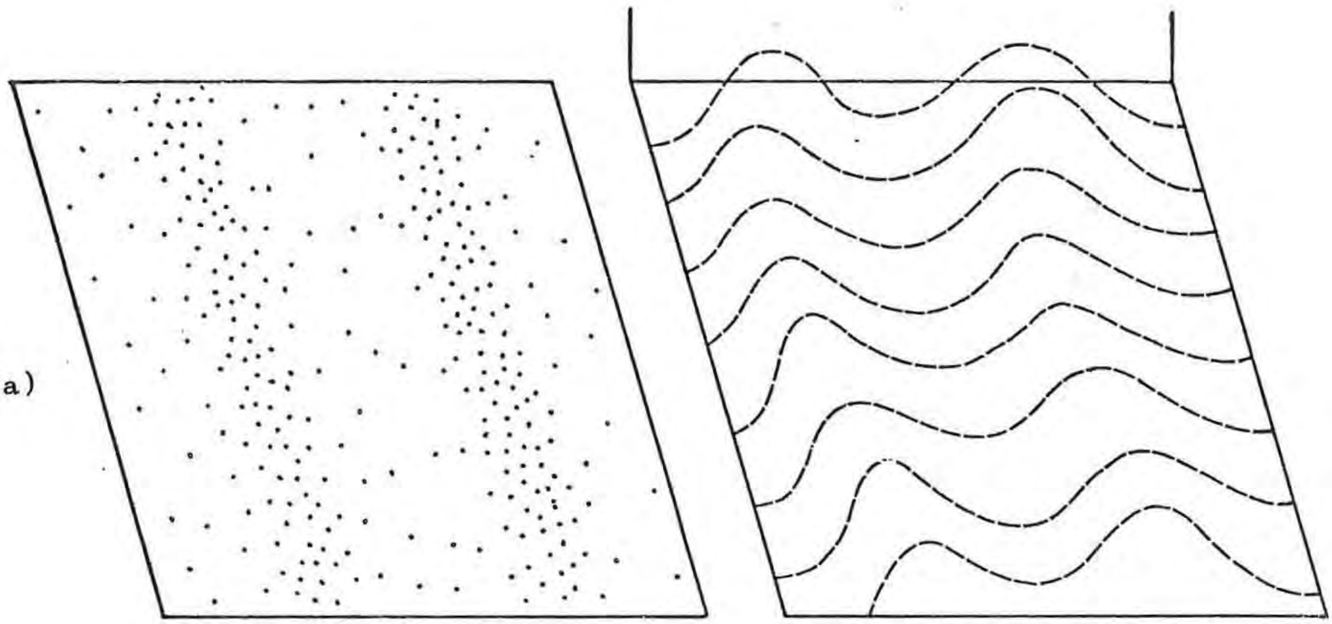
Under this heading four major problems are immediately evident, viz. whether interaction should be confined to any particular direction, how cells which fall within the limits of the matrix but which cover 'dead' or 'vacant' ground can be rendered invalid for purposes of computation, whether interaction should take place between adjacent as opposed to non-adjacent cells, and finally, whether interaction between regular sized cells is the same as interaction between irregular sized cells.

(i) Directional restriction of autocorrelation

Although Olson (1975) has indicated that a restriction of autocorrelation to certain directions would affect the value of the coefficient and as such be undesirable when attempting to measure the degree of organization on the map it can be shown that autocorrelation of the surface as a whole can limit research findings due to the inherent variable nature of most geographic distributions. It is quite possible that spatial dependence may not exist equally in all directions and as such the onus is on the researcher to isolate possible directional biases which may exist due to barriers, communication linkages or other factors. An example of a typical directional bias which could occur in reality is outlined in Figure 5, where it is likely that spatial dependence will exist in the columns as opposed to the rows. In the case of the surface depicted in Figure 5(a) it is probable that a significant autocorrelation coefficient for the whole surface, i.e. autocorrelation in all directions, would be low. Yet intuitively it seems likely that the actual relationship between the depicted values is strong. By calculating only the generalized autocorrelation coefficient the researcher might remain unaware of the fact that spatial dependence in certain directions only is far stronger than that for all directions.

FIGURE 5. Patterns: idealised crest and trough

(a)



(b)

SUBFILE CRTR IN CARD MODE

11 16 5
1 2 3 4 5

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

SUBFILE CVER IN CARD MODE

11 16 5
1 2 3 4 5

16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10
10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20
20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20
10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

The data matrices as shown in Figure 5(b) both illustrate idealized crest and trough surfaces. Autocorrelation was computed for four cardinal directions across both surfaces and finally for each surface as a whole, i.e. autocorrelation in all directions. The following segment from the program illustrates the procedure followed in isolating various directions for analysis. The results are shown in Table 3 and indicate the need to check a distribution for spatial dependence in different directions rather than merely calculating the generalized coefficient. With regard to matrix CRTR the negative correlation values appear because of the range in values across the matrix from top to bottom. On the other hand the probability of finding similar values to the left or right of any value on the matrix is great, hence the high positive correlation coefficient for the second run.

```
C     ----------------------------------------------------------------------
C     THE DIRECTIONAL LIMITS ARE NOW SET
C     FOR THE FIRST RUN N/S, FOR THE SECOND RUN E/W,FOR THE THIRD AND
C     FOURTH RUNS BOTH DIAGONALS, AND FOR THE FIFTH RUN THE WHOLE SURFACE.
C     I1 REFERS TO THE GRID SQUARE NORTHWEST OF THE CURRENT GRID
C     SQUARE.  FURTHER, I2=N, I3=NE, I4=W, I5=E, I6=SW, I7=S, I8=SE,
C     OF THE CURRENT GRID SQUARE.
C     ----------------------------------------------------------------------
   63 IF(ISEL(MM).GT.1) GO TO 64
      I1=IC
      I4=IC
      I6=IC
      I3=IC
      I5=IC
      I8=IC
      GO TO 67
   64 IF(ISEL(MM).GT.2) GO TO 65
      I1=IC
      I2=IC
      I3=IC
      I6=IC
      I7=IC
      I8=IC
      GO TO 67
```

```
65 IF(ISEL(MM).GT.3) GO TO 66
   I2=IC
   I3=IC
   I4=IC
   I5=IC
   I6=IC
   I7=IC
   GO TO 67
66 IF(ISEL(MM).GT.4) GO TO 67
   I1=IC
   I2=IC
   I4=IC
   I5=IC
   I7=IC
   I8=IC
```

TABLE 3. Autocorrelation Coefficient: generalized versus directional runs

Matrix	Run	Number of joins	Correlation coefficient
CRTR N = 1992	N/S	320	-0,774 **
	W/E	330	0,989 **
	NW/SE	300	-0,774 **
	NE/SW	300	-0,774 **
	general	1250	-0,308 **
CVER N = 2984	N/S	320	0,200 *
	E/W	330	0,964 **
	NW/SE	200	0,237 *
	NE/SW	300	0,237 *
	general	1250	0,419 **

* Significant at the 95% level

** Significant at the 99.7% level

In order to isolate a trend in the values using Cliff and Ord's statistic it is therefore necessary to isolate autocorrelation for certain directions. In the example of the A-matrices (Figure 3) the coefficient for the top to bottom direction viewed in conjunction with the coefficient for the left to right direction would indicate where the bias in the pattern exists. It is likely that situations such as these often exist in many distributions encountered in reality and as such it is impractical to assume that a given distribution will possess the same amount of spatial organization regardless of direction.

(ii) Vacant cells

In addition to the problem of directional bias is that concerning the isolation of those grid squares within the matrix which cover 'dead' ground or 'non-content' cells. A typical example of this is the surface area of a lake which may lie in an urban centre. If the urban centre is of interest to the researcher it is likely that some grid squares will cover the lake surface once a matrix has been superimposed over the area. These squares, hereafter called blank squares, require isolation for the purposes of computation but at the same time need to be distinguishable from cells which fall outside the limits of the matrix. As shown above, when using Fortran IV isolation of grid squares falling outside the limits of the matrix can be effected by way of a code. However, assuming that blank squares can be situated anywhere within the limits of the matrix, the problem becomes a little more complex. The following segments indicate one way in which the problem of blank cells can be dealt with.

```
C     -------------------------------------------------------------------------
C     THIS SEGMENT REFERS TO THE CLIFF AND ORD METHOD ONLY,
C     AND CALCULATES THE DIFFERENCE BETWEEN ALL THE
C     OBSERVED VALUES IX AND THE MEAN, THE ANSWER IN EACH CASE
C     BEING LABELLED ZDIF.  IF THE OBSERVED VALUE IX IS ZERO,
C     THEN THE ZDIF VALUE FOR THAT GRID SQUARE IS SET TO -999.0 FOR
C     IDENTIFICATION PURPOSES.  IF THE ZDIF VALUE IS CALCULATED
C     AND THE RESULT IS ZERO, IE. ZDIF IS NATURALLY ZERO BECAUSE
C     IX=MEAN, THEN ZDIF IS SET TO +999.0, ALSO FOR IDENTIFICATION
C     PURPOSES.  A COUNTER K IS INTRODUCED FOR THE CALCULATION OF ZDIF.
C     -------------------------------------------------------------------------
      K=0
      SUMZ=0.0
      SUMB=0.0
      DO 30 I=1,IROW
      DO 30 J=1,ICOL
      IF(IX(I,J).GT.0) GO TO 33
      K=K+1
      ZDIF(K)=-999.0
      GO TO 34
   33 K=K+1
      ZDIF(K)=FLOAT(IX(I,J))-AVER
   34 IF(ZDIF(K).NE.-999.0) SUMZ=SUMZ+ZDIF(K)*ZDIF(K)
      IF(ZDIF(K).NE.-999.0) SUMB=SUMB+ZDIF(K)*ZDIF(K)*ZDIF(K)*ZDIF(K)
      IF(ZDIF(K).EQ.0.0) ZDIF(K)=999.0
   30 CONTINUE

C     -------------------------------------------------------------------------
C     T1 IS THE PRODUCT OF THE CURRENT GRID SQUARE, ZDIF(J), MULTIPLIED
C     BY ADJACENT ZDIF(I1), ZDIF(I2), ZDIF(I3),... ZDIF(I8).
C     TOT(J) IS THE SUM OF THESE 8 PRODUCTS. THERE WILL BE
C     NNNN   TOT(J) VALUES.
C     TOTAL IS THE SUM OF ALL THE TOT(J) VALUES
C     THE RESULTS ARE PRINTED WITH TITLES  TOT VALUES, AND  TOTAL.
C     BETWEEN STATEMENT NUMBERS 67 AND 626 GRID SQUARES HAVING A
C     NATURAL VALUE OF ZERO, AND THOSE WHICH WERE SET TO ZERO ARE
C     DIFFERENTIATED. JOINS BETWEEN ADJACENT GRID SQUARES ARE ALSO NOTED.
C     -------------------------------------------------------------------------
   67 ZZ=ZDIF(J)
      IF(ZDIF(J).EQ.-999.0) ZDIF(J)=0.0
      T1=0.0
      S1=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I1).NE.999.0) GO TO 611
      IF(ZDIF(I1).NE.0.0) S1=1.0
      GO TO 612
  611 IF(ZDIF(I1).NE.-999.0) T1=ZDIF(J)*ZDIF(I1)
  612 T2=0.0
      S2=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I2).NE.999.0) GO TO 613
      IF(ZDIF(I2).NE.0.0) S2=1.0
      GO TO 614
  613 IF(ZDIF(I2).NE.-999.0) T2=ZDIF(J)*ZDIF(I2)
  614 T3=0.0
      S3=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I3).NE.999.0) GO TO 615
      IF(ZDIF(I3).NE.0.0) S3=1.0
      GO TO 616
  615 IF(ZDIF(I3).NE.-999.0) T3=ZDIF(J)*ZDIF(I3)
  616 T4=0.0
      S4=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I4).NE.999.0) GO TO 617
      IF(ZDIF(I4).NE.0.0) S4=1.0
      GO TO 618
  617 IF(ZDIF(I4).NE.-999.0) T4=ZDIF(J)*ZDIF(I4)
  618 T5=0.0
      S5=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I5).NE.999.0) GO TO 619
      IF(ZDIF(I5).NE.0.0) S5=1.0
      GO TO 620
  619 IF(ZDIF(I5).NE.-999.0) T5=ZDIF(J)*ZDIF(I5)
  620 T6=0.0
      S6=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I6).NE.999.0) GO TO 621
      IF(ZDIF(I6).NE.0.0) S6=1.0
      GO TO 622
  621 IF(ZDIF(I6).NE.-999.0) T6=ZDIF(J)*ZDIF(I6)
  622 T7=0.0
      S7=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I7).NE.999.0) GO TO 623
      IF(ZDIF(I7).NE.0.0) S7=1.0
      GO TO 624
```

```
623 IF(ZDIF(I7).NE.-999.0) T7=ZDIF(J)*ZDIF(I7)
624 T8=0.0
    S8=0.0
    IF(ZDIF(J).NE.999.0.AND.ZDIF(I8).NE.999.0) GO TO 625
    IF(ZDIF(I8).NE.0.0) S8=1.0
    GO TO 626
625 IF(ZDIF(I8).NE.-999.0) T8=ZDIF(J)*ZDIF(I8)
626 IF(ZDIF(J).EQ.0.0) ZDIF(J)=ZZ
    TOT(J)=T1+T2+T3+T4+T5+T6+T7+T8
    TOTAL=TOTAL+TOT(J)
C   ------------------------------------------------------------------------
C   FOR PURPOSES OF CALCULATION INSIDE LOOP NUMBER 50 WAIT=WW
C   CLEARLY WW WILL CHANGE ACCORDING TO DIRECTION
C   WW=WAIT ONCE AGAIN OUTSIDE OF THE LOOP
C   WW(J) IS A COUNTER WHICH IS SET TO ZERO.
C   IF EITHER THE CURRENT GRID SQUARE OR AN ADJACENT GRID SQUARE
C   REFLECTS A PRESET  VALUE OF ZERO THEN THERE IS NO "JOIN" AND THAT
C   PARTICULAR X AND Y PAIR IS NOT INCLUDED IN CALCULATION OF THE
C   COEFFICIENT.  DEE(J) IS A COUNTER FOR D.
C   ------------------------------------------------------------------------
    WW(J)=S1+S2+S3+S4+S5+S6+S7+S8
    IF(T1.EQ.0.0) GO TO 601
    WW(J)=WW(J)+1.0
601 IF(T2.EQ.0.0) GO TO 602
    WW(J)=WW(J)+1.0
602 IF(T3.EQ.0.0) GO TO 603
    WW(J)=WW(J)+1.0
603 IF(T4.EQ.0.0) GO TO 604
    WW(J)=WW(J)+1.0
604 IF(T5.EQ.0.0) GO TO 605
    WW(J)=WW(J)+1.0
605 IF(T6.EQ.0.0) GO TO 606
    WW(J)=WW(J)+1.0
606 IF(T7.EQ.0.0) GO TO 607
    WW(J)=WW(J)+1.0
607 IF(T8.EQ.0.0) GO TO 608
    WW(J)=WW(J)+1.0
608 WAIT=WAIT+WW(J)
    DEE(J)=WW(J)*(WW(J)-1.0)
    DED=DED+DEE(J)
    D=DED/2.0
 50 CONTINUE
```

In this segment any grid square within the limits of the matrix that reflects a value of zero is set to the unlikely score of -990.0. (In this case ZDIF = $x_i - \bar{x}$). However, when ZDIF is calculated it could happen that ZDIF = 0.0, i.e. when the observed value x_i is equal to the mean value $\bar{x}$. If this situation occurs then the ZDIF value is set to +990.0. This is important because we still wish to distinguish between a zero value within the matrix which is an indicator of a blank square, a zero value which occurs naturally, and a zero value which is the result of directional selection (refer to the process of directional selection, illustrated above).

Because of the above directional settings it is again possible to find a ZDIF value of zero inside the matrix. There is however, a subtle difference between the three zero values mentioned. Firstly, the blank cells will never be considered in computation. They will therefore always be zero and will not be counted as part of N. Secondly, the naturally occurring zero value is one which needs to be counted as part of N, yet it too will yield zero results. Thirdly, those squares set to IC (0) for directional purposes will not always be zero. As the direction setting changes so the ZDIF values, which were set to zero for the previous direction, will revert to their original values.

It is extremely important to note that blank squares are given the value zero on the original data matrix, i.e. before the ZDIF values are calculated, because we do not wish to include them in the analysis and neither do we want to recognize a join between a vacant square and a non-vacant square. They could be given any unlikely value, say for example -1. However, because the blank squares are denoted by zero the lowest value which can be recorded in those squares which represent non-blank areas is 1. The data matrices shown in Figure 6 both illustrate a hypothetical surface which, when overlaid by a regular grid, produces a number of grid squares that fall on vacant ground. MAPA illustrates the surface where the vacant squares are not included in the computation while MAPB reflects the situation where all the squares in the matrix are included in the analysis. Figure 7 shows the matrix of ZDIF scores for the matrix MAPA shown in Figure 6, i.e. all vacant squares reflect the value -990.0. The lack of a value of +999.0 on the surface indicates that nowhere did a ZDIF value of zero occur naturally.

Table 4 shows the results of autocorrelation for MAPA and MAPB which illustrates the differences in results gained from Cliff and Ord's statistic when vacant squares are included in the analysis and when they are not.

FIGURE 6. Patterns: blank cells

SUBFILE MAPA IN CARD MODE

```
  11    16    5
   1     2    3    4    5
```

0	0	1	2	3	4	5	0	0	0	0	0	0	0	0	0
0	1	2	3	4	5	6	0	0	0	0	0	0	0	0	0
1	2	3	4	5	0	6	0	7	0	0	0	0	0	0	0
4	5	6	7	8	9	10	11	12	13	0	0	0	0	0	0
11	12	13	14	15	16	17	18	19	20	0	0	0	23	24	0
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
0	0	0	13	14	15	16	17	18	19	20	21	22	23	24	25
0	0	0	8	9	10	11	12	13	14	0	0	0	0	0	0
0	0	0	2	3	4	5	6	7	8	9	0	0	0	0	0
0	0	0	0	1	2	3	4	5	6	7	0	0	0	0	0
0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0

SUBFILE MAPB IN CARD MODE

```
  11    16    5
   1     2    3    4    5
```

1	1	1	2	3	4	5	1	1	1	1	1	1	1	1	1
1	1	2	3	4	5	6	1	1	1	1	1	1	1	1	1
1	2	3	4	5	1	6	1	7	1	1	1	1	1	1	1
4	5	6	7	8	9	10	11	12	13	1	1	1	1	1	1
11	12	13	14	15	16	17	18	19	20	1	1	1	23	24	1
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	1	1	13	14	15	16	17	18	19	20	21	22	23	24	25
1	1	1	8	9	10	11	12	13	14	1	1	1	1	1	1
1	1	1	2	3	4	5	6	7	8	9	1	1	1	1	1
1	1	1	1	1	2	3	4	5	6	7	1	1	1	1	1
1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1

FIGURE 7. Matrix of ZDIF scores for MAPA

HE ZDIF SCORES ARE

999.0	-999.0	-10.2	-9.2	-8.2	-7.2	-6.2	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0
999.0	-10.2	-9.2	-8.2	-7.2	-6.2	-5.2	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0
-10.2	-9.2	-8.2	-7.2	-6.2	-999.0	-5.2	-999.0	-4.2	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0
-7.2	-6.2	-5.2	-4.2	-3.2	-2.2	-1.2	-0.2	0.8	1.8	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0
-0.2	0.8	1.8	2.8	3.8	4.8	5.8	6.8	7.8	8.8	-999.0	-999.0	-999.0	11.8	12.8	-999.0
-0.2	0.8	1.8	2.8	3.8	4.8	5.8	6.8	7.8	8.8	9.8	10.8	11.8	12.8	13.8	14.8
999.0	-999.0	-999.0	1.8	2.8	3.8	4.8	5.8	6.8	7.8	8.8	9.8	10.8	11.8	12.8	13.8
999.0	-999.0	-999.0	-3.2	-2.2	-1.2	-0.2	0.8	1.8	2.8	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0
999.0	-999.0	-999.0	-9.2	-8.2	-7.2	-6.2	-5.2	-4.2	-3.2	-2.2	-999.0	-999.0	-999.0	-999.0	-999.0
999.0	-999.0	-999.0	-999.0	-10.2	-9.2	-8.2	-7.2	-6.2	-5.2	-4.2	-999.0	-999.0	-999.0	-999.0	-999.0
999.0	-999.0	-999.0	-999.0	-999.0	-10.2	-9.2	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0

TABLE 4. Autocorrelation coefficient: inclusion and exclusion of vacant cells

Matrix	Run	Number of Joins	Correlation coefficient
MAPA N = 1046	N/S	150	0.753
	W/E	158	0.954
	NW/SE	142	0.744
	NE/SW	138	0,758
	general	588	0.806
MAPB N = 1129	N/S	320	0,606
	W/E	330	0,869
	NW/SE	300	0,620
	NE/SW	300	0,554
	general	1250	0,666

All values significant at the 99.7% level.

There is a distinct drop in the correlation values for each direction, and consequently for the surface as a whole, when vacant cells are included in the analysis. The reason for this is probably that by including the vacant cells an artificial pattern is actually being added to the existing point pattern which, because of the difference in values, is causing existing features to lose their impact. Clearly therefore, if accurate results regarding spatial dependence are to be obtained vacant squares should not be included in the analysis.

(iii) Interaction between non-adjacent cells

Cliff and Ord's extension of Moran and Geary's interaction function δ_{ij} enables the researcher to apply the statistic to a situation where interaction, not only between adjacent cells but also between non-adjacent cells, can be taken into account. In this way Cliff and Ord have been able to incorporate the effect of friction of distance between non-adjacent sub-areas (Yeates, 1974). This extension could be extremely useful in a practical situation where non-proximate co-variance (or hierarchical co-variance) is expected to occur. An example of a possible hierarchical situation could be the measurement of autocorrelation of population size where it is expected that population migrates from the peripheral areas of a country to certain core areas. In this case an autocorrelation with a lag to the n^{th} neighbour could reveal a strong pattern of population change. Research in this field has been attempted by Bannister (1976). However, both Bannister (1975) and Olson (1975) maintain that first lag relationships (covariance between adjacent neighbours) tend to be most meaningful in terms of autocorrelation. Further it is very likely that many research problems encountered in geography as well as other disciplines will require evaluation of proximate spatial dependence as opposed to non-proximate dependence.

Second, third or n^{th} order interaction may be calculated for a regular matrix simply by effecting a change in the code system. The following segment shows how the standard form of the program, i.e. that useful for first order interaction can be changed to achieve both first and second order autocorrelation.

```
      WAIT=0.0
      DED=0.0
      TOTAL=0.0
      DO 50 J=1,NEND
      I1=(J-LAG*ICOL)-LAG
      I8=(J+LAG*ICOL)+LAG
      I3=(J-LAG*ICOL)+LAG
      I6=(J+LAG*ICOL)-LAG
      I2=J-LAG*ICOL
      I7=J+LAG*ICOL
      I4=J-LAG
      I5=J+LAG
C     ------------------------------------------------------------------
C     IF I1 TO I8 FALL OUTSIDE THE LIMITS OF MATRIX NEND THEY ARE ROUTED
C     TO IC , WHICH READS 0.0
C     ------------------------------------------------------------------
```

```
    KR=((J-1)/ICOL)+1
    IF(J.GT.ICOL*LAG) GO TO 60
    I1=IC
    I2=IC
    I3=IC
60  NB=NEND-ICOL*LAG
    IF(J.LE.NB) GO TO 61
    I6=IC
    I7=IC
    I8=IC
61  IF(((J+1)-(ICOL*KR)+((J+1)/(ICOL*KR))).LT.1) GO TO 62
    I3=IC
    I5=IC
    I8=IC
62  IF(J.GT.((KR-1)*ICOL+LAG)) GO TO 63
    I1=IC
    I4=IC
    I6=IC
```

The value for lag which is read in from the data matrix can be the number 1 or 2. If a lag-two autocorrelation is required only those cells which lie along the four cardinal directions are included in the analysis. Referring to Figure 8 directional interaction can best be referred to in chess terminology - rook's case and the bishop's case are included while second order joins for the knights case are excluded from the analysis because they are not two joins removed in any linear direction from the current grid square X .

FIGURE 8. Coding for second order interaction

I1		I2		I3
I4		X		I5
I6		I7		I8

(The arrow indicates the knight's case which is not included in the analysis.)

The results for first and second order autocorrelation when applying the data matrix, MAPA, are shown in Table 5.

TABLE 5. Autocorrelation coefficient: first and second order interaction

Matrix	Run	Number of joins	Correlation coefficient
MAPA first order N = 1046	N/S	150	0,753 **
	W/E	158	0,954 **
	NW/SE	142	0,744 **
	NE/SW	138	0,758 **
	general	588	0,805 **
MAPA second order N = 1046	N/S	118	0,238
	W/E	138	0,869 **
	NW/SE	112	0,190
	NE/SW	102	0,242
	general	470	0,413 **

* Significant at the 95% level

** Significant at the 99.7% level.

The only significant correlations yielded by computation of second order lag relationships are for the W/E directional and the general runs, which strongly suggests that second order relationships do not exist for the data shown in MAPA. It is therefore clear that if Bannister (1975) and Olson (1975) are correct in suggesting that most geographic problems require evaluation of first order lag relationships, then evaluation of second order relationships will not be of much practical value to the geographer unless he has some prior knowledge of the distribution with which he is concerned and expects second order relationships to exist.

(iv) Interaction between irregular cells

Because it is the degree of linear association between x_i values we wish to measure by means of the autocorrelation technique it is necessary to standardize any factor which could lend weight to particular x_i values within the set. For example, the situation where sub-areas are irregular in size may require that some form of general function be employed to negate the effect of size differences when computing the autocorrelation coefficient.

Besides facilitating interaction between non-adjacent cells Cliff and Ord's extension to Moran and Geary's interaction function δ_{ij} also facilitates the standardization of irregular areas, where interaction occurs between sub-areas of different sizes. It is in the context of the interaction function w_{ij} that Cliff and Ord introduce the concept of weighting. However, they do point out that interaction between sub-areas can also be weighted to take into account the effects of natural barriers, communication linkages, and so on (Cliff and Ord, 1969, p.31).

Using MAPB, data were grouped in such a way as to form a 6 x 9 matrix composed of irregular areas such that the x_i value for the irregular area was a combination of x_i values originally reflected in MAPB (for further information on this procedure see Robinson, 1956). For this new matrix comprised of irregular squares autocorrelation was computed twice, viz., firstly for the irregular squares without adjusting the weights and secondly for the irregular squares where the values contained in the squares had been multiplied by a fraction which generalized the size differences (see matrices IRRN and IRRW respectively - Figure 9). The values were therefore weighted according to areal size differences between the various squares.

Results for autocorrelation of MAPB, IRRN and IRRW (Table 6) show that the grouping of values reflected in MAPB caused a marked decrease

FIGURE 9. Patterns: areal size differences

SUBFILE MAPB IN CARD MODE

11 16 5
1 2 3 4 5

1	1	1	2	3	4	5	1	1	1	1	1	1	1	1	1
1	1	2	3	4	5	6	1	1	1	1	1	1	1	1	1
1	2	3	4	5	1	6	1	7	1	1	1	1	1	1	1
4	5	6	7	8	9	10	11	12	13	1	1	1	1	1	1
11	12	13	14	15	16	17	18	19	20	1	1	1	23	24	1
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	1	1	13	14	15	16	17	18	19	20	21	22	23	24	25
1	1	1	8	9	10	11	12	13	14	1	1	1	1	1	1
1	1	1	2	3	4	5	6	7	8	9	1	1	1	1	1
1	1	1	1	1	2	3	4	5	6	7	1	1	1	1	1
1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1

SUBFILE IRRN IN CARD MODE

6 9 5
1 2 3 4 5

7	12	9	15	2	4	2	4	2
3	3	16	1	7	5	1	1	2
55	105	41	91	50	76	49	48	78
6	44	25	87	33	43	23	49	26
1	2	9	11	7	17	3	1	1
4	4	5	5	5	13	10	4	4

SUBFILE IRRW IN CARD MODE

6 9 5
1 2 3 4 5

11	27	41	23	9	9	9	9	9
14	27	36	9	63	9	9	9	9
83	105	123	137	150	114	74	144	117
9	99	113	131	149	97	104	110	117
9	9	27	50	63	76	9	9	9
9	9	11	23	23	29	23	9	9

in autocorrelation values as expected. Theoretically, the coefficients reflecting autocorrelation levels, after weighting had been carried out, should have been similar to those reflected for MAPB. The simple weighting function used to standardize areal size differences was not adequate and presumably needs further adjustment for distance from centres and differences in boundary length. Both aspects are mentioned by Cliff and Ord when discussing the function for w_{ij} used to standardize areal differences between cells. From the above it appears that much research is required in this field before it can be shown with confidence which forms of w_{ij} do not distort the natural pattern inherent in any distribution.

TABLE 6. Autocorrelation coefficient: irregular cells

Matrix	Run	Number of joins	Correlation coifficient
MAPB N = 1129	N/S	320	0,606 **
	W/E	330	0,869 **
	NW/SE	300	0,620 **
	NE/SW	300	0,554 **
	general	1250	0,666 **
IRRN	N/S	90	0.132
	W/E	96	0.654 **
	NW/SE	80	0.036
	NE/SW	80	0,022
	general	346	0,229 **
IRRW	N/S	90	0,304 *
	W/E	96	0,854 **
	NW/SE	80	0.338 *
	NE/SW	80	0.210
	general	346	0,443 **

* Significant at the 95% level
** Significant at the 99.7% level

Tests of significance

Once the degree of autocorrelation has been evaluated it is necessary to assess whether the given coefficient could have occurred by chance alone. In order for the significance of Cliff and Ord's statistic to be tested, and as such render the method useful as an analytical tool, it is necessary for there to be a constant expected value under the null hypothesis of no correlation (Cliff and Ord, 1969). Further, it is necessary that a distribution theory be developed for the statistic, which will serve as a context in which the actual recorded coefficients may be tested for significance.

Cliff and Ord (1969) prove that their statistic can be tested for significance as a standard normal deviate. Standardization is effected by calculating the first two moments of the coefficient, which serve to specify the mean and the variance of the curve.

Distribution theory

Cliff and Ord (1969, 1973) formulate a distribution theory for their statistic under the following basic assumptions.

(i) Assumption of Normality. In this case it is assumed that the x_i values are the result of a number of independent drawings from a normal population.

(ii) Assumption of Randomness. In this case no assumptions are made regarding the distribution of the population, and the coefficient is considered relative to "... the set of all possible values ..." which it could take were the x_i values repeatedly permitted around the surface in a random fashion (Cliff and Ord, 1973, p.8).

In order to effect a transformation of the observed statistic to a form where it can be compared to the normal curve the statistic is standardized by means of the following procedure (Refer Equation 3).

$$z = \frac{\text{observed coefficient} - E(r)}{\sqrt{E(r^2) - [E(r)]^2}} \qquad \text{Equation 7}$$

(Cliff and Ord, 1969. p.28).

The first and second moments of the statistic ($E(r)$ and $E(r^2)$ respectively) are evaluated under the null hypothesis of no autocorrelation. Further, evaluation of the moments is based on an assumption regarding the underlying distribution of the population (refer above - and also Cliff and Ord, 1969, 1973). Once evaluated z can be compared to the standard normal curve, since Cliff and Ord (1969, p.36) have shown that their coefficient, like Pearson's r, is asymptotically normally distributed, and will increaseingly approximate the normal curve as n becomes greater. Because the expected values are derived under the null hypothesis of no correlation the mean of the observed statistic will approximate the hypothetical mean of the normal curve if the observed data exhibits no spatial autocorrelation. Referring to Figure 1 it is clear that if we wish to reject the null hypothesis at the 95,4% level of confidence the absolute value of z must exceed +2. The figure shows clearly the limits of z for various levels of confidence. For example, if the absolute value of z is greater than 1.96 the null hypothesis of no spatial autocorrelation can be rejected at the 95% level of confidence.

Calculation of Moment

Cliff and Ord (1969, 1973) derive formulas for the calculation of the first and second moments of their statistic under both the assumptions of normality and randomness ("For any random variable x and any positive integer k, the expectation $E(x^k)$ is called the k^{th} moment of x." De Groot, 1975, p.161). The first crude moment of their statistic r for both assumptions is

$$E_N(r) = E_R(r) = -(n-1)^{-1}$$ Equation 8

(Cliff and Ord, 1969. p.35).

The second crude moments of r are as follows

(i) Assumption of randomness

$$E_R(r^2) = \frac{\left[n \left\{(n^2-3n+3)\, S_1 - nS_2 + 3W^2\right\} - b_2 \left\{(n^2-n)\, S_1 - 2nS_2 + 6W^2\right\}\right]}{(n-1)^{(3)}\, W^2}$$

Equation 9

(Cliff and Ord,1969. p.36)

(ii) Assumption of normality

$$E_N(r^2) = \frac{(n^2 S_1 - nS_2 + 3W^2)}{(n^2-1)\, W^2}$$ Equation 10

(Cliff and Ord, 1969. p.36)

Where n = the number of variate values in the sample or population

$W \equiv \sum(2)\, w_{ij} \equiv$ Summation of all the weights

$S_1 = \frac{1}{2} \sum(2)\, (w_{ij} + W_{ji})^2$ where $\sum(2) \equiv \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n}$

$S_2 = \sum(1)\, (w_{i\cdot} + w_{\cdot j})^2$

b_2 = the sample coefficient of kurtosis, $^{m_4}/_{m_2^2}$ (refer below)

$(n-1)^{(3)} = (n-1)\,(n-2)\,(n-3)$

If a binary weighting system is used (as has been done in this study) a simple method for evaluating S_1 and S_2 is available to the researcher, viz. $S_1 = 4A$ and $S_2 = 8(A+D)$ (Cliff and Ord, 1973, p.16). In this regard $A = \frac{1}{2} \sum_{i=1}^{n} L_i$ where L_i denotes the number of grid squares joined to the i^{th} grid square. Further, $D = \frac{1}{2} \sum_{i=1}^{n} L_i\,(L_i-1)$. The total number of joins in the matrix is denoted by W, and therefore $A = \frac{1}{2}W$. In the case of D the situation is more complicated in that it is the product of L_i and (L_i-1) that is summed. For example, referring to Figure 2 there are eight squares adjacent to X, but only 5 of these lie within the limits of the matrix. In this situation D would therefore be equal to 10 ($D = {}^{1}/_{2} \sum 5(4)$).

However, if a generalized weighting system is used (refer Cliff and Ord, 1969, 1973) then it is necessary to evaluate S_1 and S_2 in the manner outlined following equation 10. Finally if a generalized weighting system is used in conjunction with a regular grid matrix then $w_{ij} = w_{ji}$, i.e. the weights will be symmetric.

The sample coefficient of kurtosis, b_2, where $b_2 = {}^{m_4}/{m_2}^2$, is simply the fourth sample moment of the observed data divided by the square of the second sample moment (Cliff and Ord, 1969, p.36; p.33), where any sample moment M_p about the mean is given by the expression

$$M_p = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^P}{n}$$

Equation 11

(King, 1969. p.25)

The sample coefficient of kurtosis can therefore be expressed as

$$b_2 = \frac{1/n \sum_{i=1}^{n} (x_i - \bar{x})^4}{\left[1/n \sum_{i=1}^{n} (x_i - \bar{x})^2\right]^2}$$

Equation 12

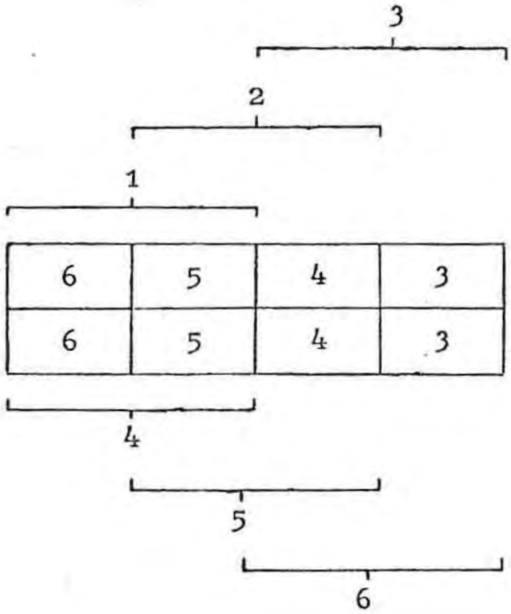
For an outline of the segment of the program which evaluates the significance of the coefficient refer to Appendix I (statement number 77-94).

Interpretation of the Coefficient

Geographers often attach some sort of tentative interpretation to correlation coefficients gained from classical autocorrelation procedures. If we wish to treat coefficients resulting from Cliff and Ord's method in a similar fashion it is necessary to compare the reactions of both Cliff and Ord's method and one of the classical methods to the same situation. Pearson's product moment correlation method is introduced for this purpose. Further, it is likely that Pearson's method would be useful in confirming trends isolated by Cliff and Ord's method outlined above.

The recognition of x_i and x_j pairs differs from the method outlined above for Cliff and Ord's statistic. For Pearson's method dependence is seen to extend in one direction only. If for example, data were organized in a fashion that displayed a perfect left to right trend, as shown in the A-matrices in Figure 3, then Pearson's coefficient would reflect a correlation of 1.00. This occurs because x_i and x_j pairs are isolated according to the rationale of a single directional dependence, the procedure being depicted in Figure 10. In the example the x_i value is always one unit greater than the x_j value. Bearing in mind that correlation in the classical sense is the "... degree to which changes in one direction and magnitude in one set of data are associated with comparable changes in the other set" (Gregory, 1963, p.189), this will yield a coefficient of 1.00 when a standard product moment correlation formula is applied to the data (refer to equation 4). In this respect Pearson's method, when computed in addition to Cliff and Ord's method, will show in which direction and the magnitude of any trends in the distribution of the variable being measured. For the final run of the program, which computes autocorrelation for all directions, the various arrays of x_i and x_j pairs which have been isolated for the four directional runs are merely combined to form a single array which represents all directions. Where Cliff and Ord's coefficient shows a similar pattern of distribution to Pearson's coefficient it is suggested that tentative interpretation of the relative strength of Cliff and Ord's coefficient can be attempted in the same way that

FIGURE 10. Pearson's method: isolation of x_i and x_j pairs.



Pair number	x_i	x_j
1	6	5
2	5	4
3	4	3
4	6	5
5	5	4
6	4	3

Pearson's coefficient is often interpreted by geographers, i.e. a correlation of 0,3 is seen as being different from a correlation of 0,9 although both are significant at the 95% level (refer for example Table 1 and 7). The results of Pearson's method for computation of correlation for the matrices shown in Figure 3 are tabulated below in Table 7.

TABLE 7. Autocorrelation coefficients: Cliff and Ord's and Pearson's

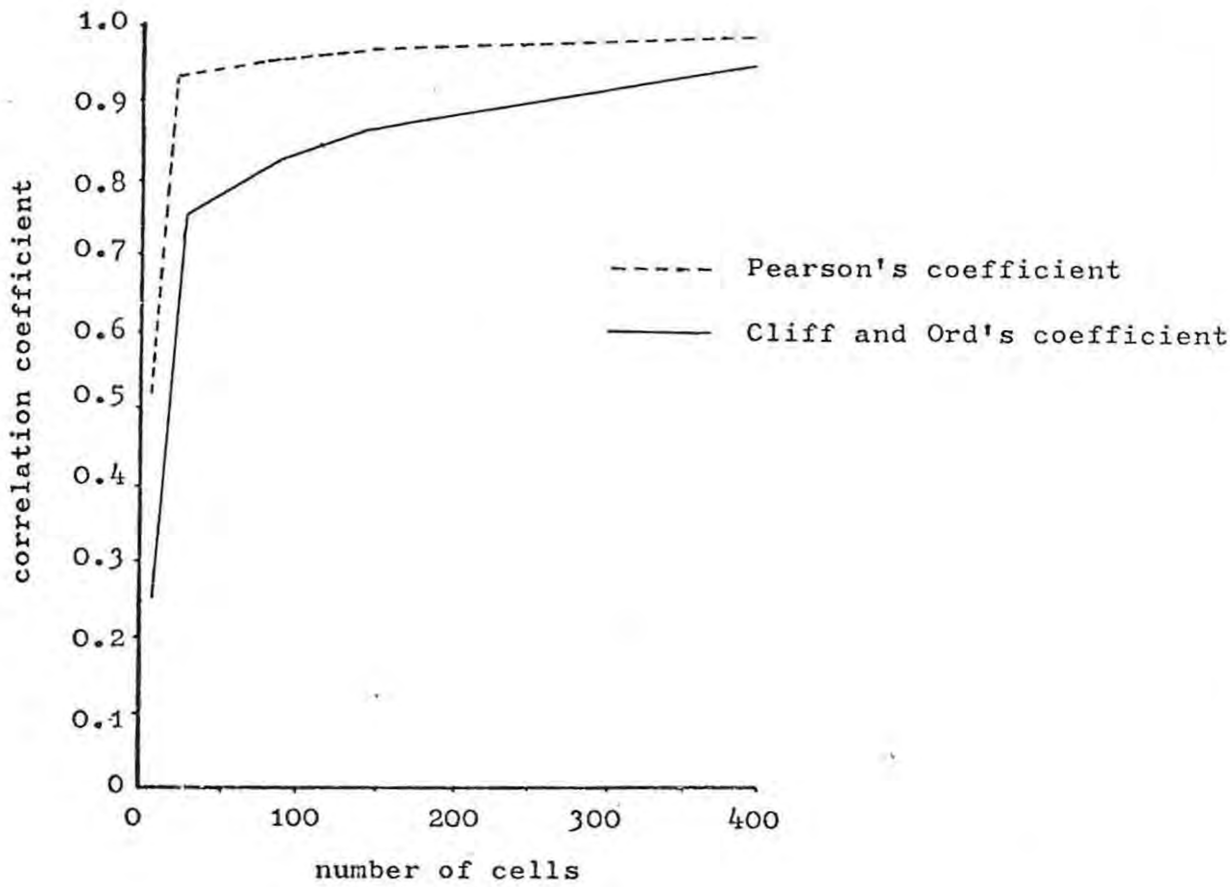
	Cliff and Ord		Pearson		
Matrix	Number of joins	Correlation coefficient	Number of pairs	Correlation coefficient	t-value
ATHR	40	0,300 **	20	0,421	1.97
BTHR	34	0,192 *	17	0,667 **	3.47
ASEV	312	0,756 **	156	0,902 **	25.99
BSEV	306	0,758 **	153	0,962 **	43.23
ATWL	1012	0,866 **	506	0,969 **	87.38
BTWL	1006	0,869 **	503	0,988 **	145.50
ATEE	2964	0,922 **	1482	0,989 **	258.03
BTEE	2958	0,924 **	1479	0,966 **	429.42

* Significant at the 95% level

** Significant at the 99.7% level

The above results were plotted and the resulting trend is shown in Figure 11 which illustrates that the calculated values tend to converge as n increases. The large discrepancies in results when n is small is possibly due to the fact that the relative significance of the one-unit difference between the x_i values is greater than when n is large.

FIGURE 11. Matrix size: Cliff and Ord's statistic and Pearson's statistic.



When comparing coefficients calculated using Ord's and Pearson's method respectively the results from the dummy matrices, MAPA and MAPB (Figure 6) it is clear that the results from Pearson's method are very similar to those recorded for Cliff and Ord's method (Table 8). Therefore, the manner in which geographers have traditionally interpreted coefficients resulting from Pearson's method can possible be applicable to the Cliff and Ord method. However, from the results reflected in Table 8 it is evident that Pearson's statistic is more sensitive to directional dependence

TABLE 8. Autocorrelation coefficients: Cliff and Ord's and Pearson's, directional bias

		Cliff and Ord		Pearson		
Matrix	Run	Number of joins	Correlation coefficient	Number of pairs	Correlation coefficient	t-values
MAPA	N/S	150	0,753	75	0,849	13.76
	W/E	158	0,954	79	1,000	OO
	NW/SE	142	0,744	71	0,851	13.47
	NE/SW	138	0,758	69	0,853	13.40
	general	588	0,806	294	0,887	32.84
MAPB	N/S	320	0,606	160	0,575	8.84
	W/E	330	0,869	165	0,874	23.04
	NW/SE	300	0,620	150	0,590	8.89
	NE/SW	300	0,554	150	0,536	7.54
	general	1250	0,666	625	0,643	20,98

All values significant at the 99.7% level.

than Cliff and Ord's method. For example, the coefficient for the MAPA pattern are all slightly stronger, with the perfect left to right trend yielding a correlation coefficient of 1.00. This could be expected from the manner in which the pairs are coupled. However, in the case of MAPB, where the pattern consists of both the non-vacant and vacant grid squares Pearson's method, being directionally biased, yields weaker coefficients for all the runs except the left to right direction, thus indicating greater sensitivity to the reduced frequencies imposed by the addition of black squares in the computation. Pearson's method adequately confirms that directional dependence exists and is incorporated in the spatial dependence isolated by Cliff and Ord's method. It is suggested that Pearson's coefficient can be calculated when strong directional trends are suspected after calculating coefficients according to Cliff and Ord's method. However, it must be remembered that the two techniques are conceptually different and are not interchangeable, i.e. Pearson's method merely compares each cell with its immediate neighbour on the one side only whereas the Cliff and Ord technique compares each cell with its neighbours on both or all sides and thus measures a truly spatial rather than a directional relationship.

This chapter has provided a means whereby an objective impression of the spatial character of any distribution can be gained. The program has been constructed simply and in a manner which can be applied to a variety of geographical situations. By using a regular matrix as a means of demarcation the variable with which Cliff and Ord (1969, 1973) were most preoccupied in their practical example, viz. subarea size, was held constant. Consequently it has been possible to highlight pitfalls that may be encountered in any practical application of Cliff and Ord's statistic, viz. the effect of the size and number of subareas included in the analysis and the level of interaction between these subareas on the nature of the coefficient. It is also evident that it is necessary to look for directional trends in a point distribution rather than to be concerned with a generalized value and finally, that it is essential to attach some form of interpretation to the coefficient because if the technique is to be a useful tool of analysis it is not sufficient merely to know whether a surface displays autocorrelation or not.

CHAPTER V

A PRACTICAL APPLICATION OF SPATIAL AUTOCORRELATION

In order to extend the application of the method while bearing in mind the limitations outlined in the previous chapter, it is necessary to apply it to a point pattern resulting from a real world situation. Within the sphere of human geography there are many different point patterns which feasibly could be suited to an analysis of spatial dependence, such as the distribution of certain diseases, people belonging to particular religious denominations, people having certain political affiliations, and crimes. In this study the distribution of crime in the urban area of Grahamstown is the pattern which will be analysed by means of the spatial autocorrelation technique.

The subject of crime occurrence in urban areas has received much attention in Britain and the U.S.A. in recent years. The spiralling crime rates in large American Centres lead researchers to seek answers to questions of why crime occurs, and how to control it. More recently there has been research conducted into the question of crime and its relationship to individual behaviour and space, but it appears that the internalized behaviour characteristics of criminal offenders have received more attention than the spatial distributions of crime. This is due to the fact that traditionally, disciplines such as sociology and psychology (among others) have tended to dominate research in the sphere of crime and consequently the spatial aspects of crime have been neglected in favour of the behavioural aspects. It is only recently that geographers have begun to investigate the spatial dimension of crime occurrence (see for example Harries, 1971, 1973, 1974; Pyle, 1974).

In South Africa very little research has been conducted into either spatial or behavioural aspects of crime. Midgley, Steyn and Graser (1975) point out that academic interests regarding crime in South Africa has followed developments in other countries, and consequently the work which has been conducted also has tended to neglect the spatial aspects of crime occurrence.

By investigating the degree of spatial autocorrelation in the distribution of crime occurrence in Grahamstown not only will the technique be tested but if the results prove useful, as suggested by Cliff and Ord, then a means of comparison between a South African pattern of crime occurrence and specific patterns isolated overseas will be possible. Further, by isolating, describing and establishing the extent to which space is one particular variable in the distribution of crime, using the spatial autocorrelation technique, "... elements of understanding that might otherwise be overlooked or underestimated" (Harries, 1974, p.116) may be added to crime theory.

Crime occurrence and the spatial component

As far as the spatial distribution of crime is concerned two broad classes of interest may be identified, viz. the distributional characteristics and the ecological characteristics (Herbert, 1977).

The former category refers to the mapping of incidence rates while the latter refers to the analysis of the degree to which distributions of crime can be cross-correlated with similar distributions of environmental characteristics. Analysis concerning the latter category are restricted to techniques defined in Chapter 3 under the heading of areas association, while for the former category analysis of pattern with respect to itself can be considered. With regard to the analysis of pattern with respect to itself the use of the spatial autocorrelation technique will not facilitate an evaluation of the distributional characteristics of crime but will rather establish the degree to which crime in one location is associated with crime in adjacent locations. Therefore, Cliff and Ord's technique is useful as an analytical tool which facilitates an evaluation of the relative importance of space in the distribution of crime occurrence.

The attempted isolation of reasons for the occurrence of crime in any community has not met with much success in the past and as such there are as yet no broadly accepted theories of crime causation (Harries, 1974). However, on the basis of empirical research conducted overseas it is thought that the following factors affect crime occurrence:

(a) Economic factors, evident in features such as

   (i) The physical deterioration of property (Mays, 1956; Pyle, 1974; Corsi and Harvey, 1975; Lotter, 1975).

   (ii) Conditions of low rental (Chilton, 1964; Boggs, 1965; Klein, 1967).

   (iii) Land-use change (Lee and Egan, 1972; Brantingham, 1975).

(b) Social status factors (Mays, 1956; Mays, 1968; Boggs, 1965; Pyle, 1974; Corsi and Harvey, 1975; Lotter, 1975).

(c) Familial factors (McCord, McCord and Zola, 1959; Douglas, Ross Hammond and Mulligan, 1966; Jeffrey, 1971).

(d) Ethnic factors (Moses, 1947; Klein, 1967; Carter, 1972).

(e) Mobility factors (Mays, 1956; Schmid, 1960; Harries, 1973; Harries, 1974).

(f) The concept of "anomie" (or "not belonging") (Merton, 1938; Jonassen, 1949; Cohen, 1965; Gibbs, 1966; Retief, 1975; Midgley, 1975).

Although only a few studies (for example Harries, 1971, 1973, 1974; Pyle, 1974; Brantingham, 1974; Corsi and Harvey, 1975) have integrated the spatial aspects of the above variables in their analyses, they all possess, to a greater or lesser extent, some form of spatial component. Further, some factors are clearly more constrained by space than others.

For example, in South Africa the ethnicity factor has distinct spatial connotations due to legislation regarding place of residence of racial groups as opposed to something like familial status, which is perhaps less obviously constrained by space. The familial environment need not necessarily be directly related to the spatial location of the criminal's home, although its situation could be related to economic and social factors, which could in turn influence the familial environment (refer for example McCord, McCord and Zola, 1959).

Alternatively it is possible to group all the variables listed above under the general heading of environmental factors, i.e. these variables encompass a wide range of stimuli external to the mind of the potential criminal. Most studies therefore focus attention solely on investigating the link between crime and those factors thought to cause crime. However, the possible combinations of the variables listed above are innumerable and further, it is likely that different combinations of these variables will be related to different types of crime. It is therefore very difficult to isolate a distinct set of causative factors and in that way work towards a generally accepted theory of crime causation. Yet by isolating the importance of space in the distribution of crime occurrence a further element of understanding will be added to the existing body of knowledge regarding crime causation.

Of course, certain types of crime are clearly more space constrained than others. Using the broad dichotemy between crime against property and crime against person as far as the spatial component is concerned a major difference between these two types of crime is that the distance separating residential and other units, which are regarded as 'targets' for crime against property in an urban area, must inevitably be greater than that separating two individuals involved in a particular crime against person. Thus, analysis of the spatial component is likely to provide interesting results when applied to each of the two types of crime.

Analytical framework

In order for crime against property or crime against person to occur it is necessary for there to be interaction between people and property, and between people respectively. Clearly therefore, for crime to occur a movement of people is required. Further, studies conducted for urban centres in America have shown that definite patterns of crime exist, which have been linked to various socio-economic, ethnic, religious and familial factors. In Grahamstown, as in most South African towns, much inhomogeneity exists regarding the various factors listed above and, in addition, there is a daily movement of people to and from the central area of town where most of the employment opportunities in Grahamstown exist, therefore it is likely that spatial patterns of crime can be isolated in Grahamstown.

Although it is not an industrial centre the urban area of Grahamstown, which lies between Port Elizabeth and East London, is a fourth order town in the South African urban hierarchy (Davies and Cook, 1968), and in 1970 ranked thirty third in South Africa with a total population of 41 086 persons comprised of 10 447 whites, 4 875 coloureds, 242 Asians and 25 522 Bantu (Government printer, 1970. This figure for Bantu should be seen in relation to the estimate for 1977 of 37 000 Bantu, supplied by the Bantu Affairs Administration Board in Grahamstown). Grahamstown plays an important educational and cultural role in the Eastern Cape (Daniel, 1974) and has been classified on the Nelson Scheme as primarily a service centre although commercial activities are next in importance (Davies and Young, 1969).

The central area (Figure 12) has a well developed CBD where most of the retail and professional activity is located. There is evidence of ribbon development extending along Beaufort Street, especially towards the black areas in the east. There is some white housing which predominates away from the CBD towards the east and this area also includes the homes of a few Asiatics, located adjacent to their shops. Commuting patterns of workers centre on the CBD area, which is the focus of interaction during working hours in Grahamstown and as such the potential for crimes such as shoplifting, robbery and assault is high.

Although racial groups are residentially separate, almost all employment opportunities are to be found in the white sector of Grahamstown, i.e. the western areas. The major sources of employment are located in and around the CBD and on the university campus. As a result there is a movement of non-white working population into the white areas in the morning, and a movement outwards in the evening. Because of the relative size of Grahamstown many of the black workers walk to and from work. The majority of this commuting occurs along the major thoroughfares through Grahamstown which all run in approximately a west to east direction, viz., Fitzroy, African, New, High, Beaufort and Market Streets (Figure 12). The bus service for blacks in Grahamstown operates only in the mornings and evenings from 5.00 a.m. to 8.30 a.m. and from 4.30 to 8.00 p.m. respectively with peaks from 6.00 a.m. to 7.00 a.m. and 5.00 p.m. to 7.00 p.m. respectively. These buses make ten trips, viz., one trip which runs as far as the market square and then returns to Makana's Kop and one which makes a round trip through Grahamstown.

The boundaries between each racial group have been officially demarcated (Figure 13) and there is little evidence of racial mixing in residential areas except among Asiatics who, as already mentioned, have not yet been moved into the zoned area. Further exceptions occur in the 'frozen zone' (Figure 12), which is predominantly white, and the Fingo Village (Figures 12 and 13), which is predominantly black. Both the 'frozen zone' and the Fingo village areas are affected by zoned changes and there is evidence of population change and physical deterioration. In addition, the former area is experiencing land-use change from intensive low-quality housing mixed with light industry to high density economic

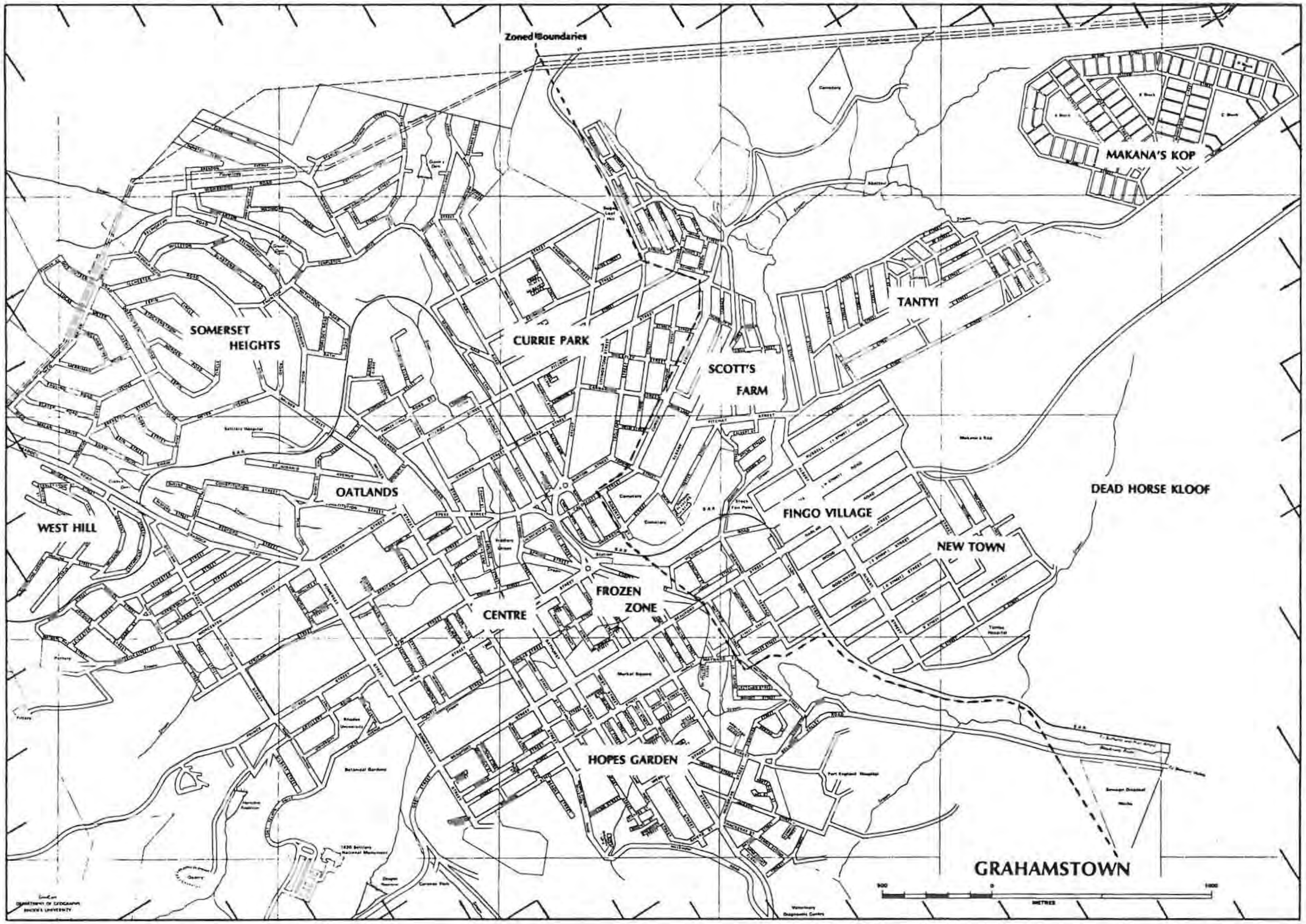
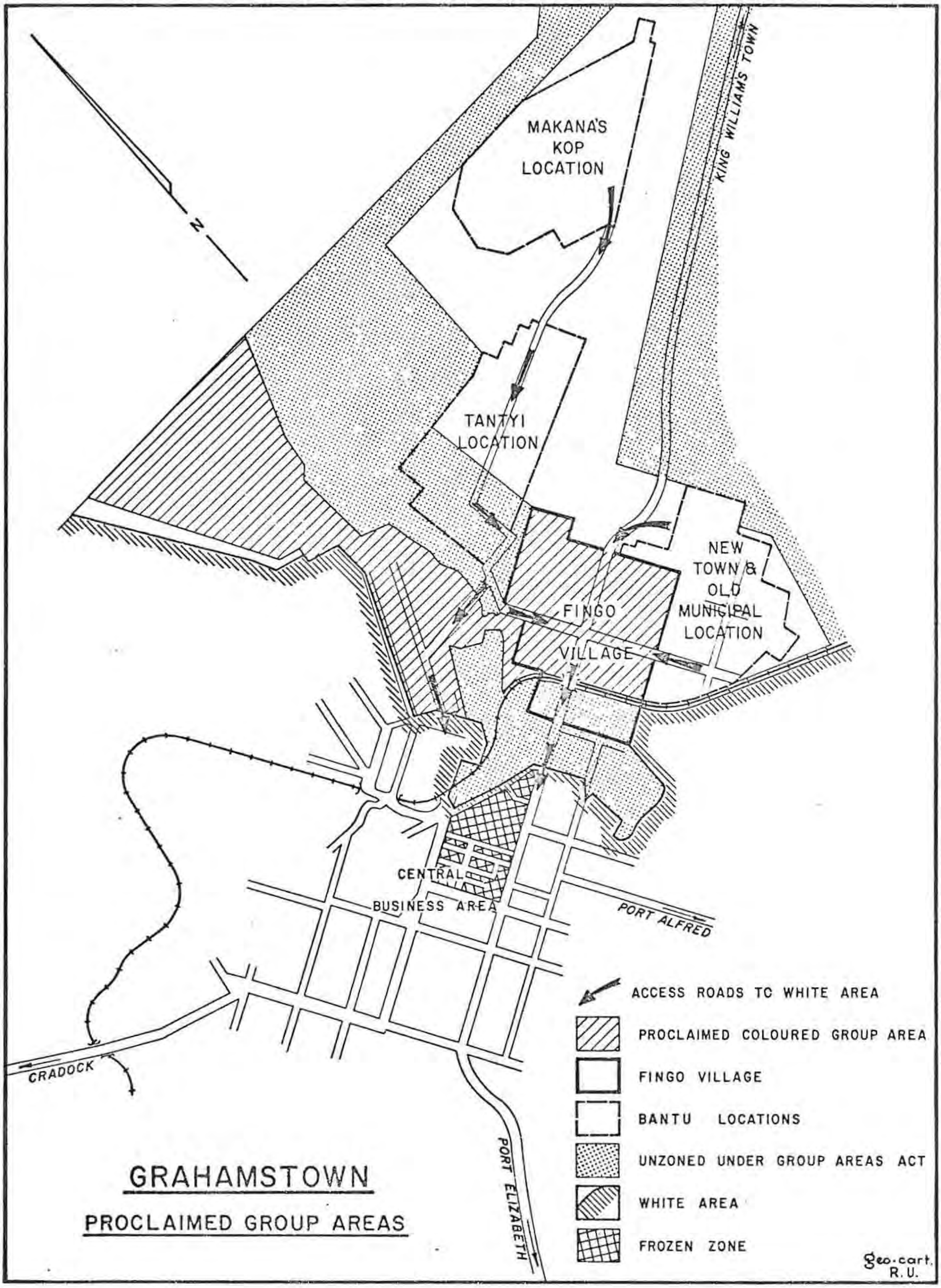


FIGURE 12: Grahamstown: residential areas

FIGURE 13: Grahamstown: proclaimed non-white areas



housing with some open spaces in the form of recreational and car parks. For the rest residential differentiation in white areas is not very marked when municipal valuation is considered, but distinctly separate areas of socio-economic status have been recognised on the basis of sample surveys conducted by Rhodes University students. These areas are shown in Figure 12. The relative housing densities for each of the white sub-areas (Table 9) tend to support the abovementioned distinction of separate white residential sub-areas according to socio-economic status.

TABLE 9. Grahamstown: White housing, relative densities

Area	Number of homes per 250 metres x 250 metres
Frozen zone	21.9
Currie park	18.8
Somerset Heights	5.0
Oatlands	15.2
Centre	34.8
Hopes Garden	37.5
West Hill	10.3

In the black townships, either all the housing is provided by the municipality, as at Makana's Kop (Figures 12 and 13) or either municipal housing or sites, on which people are free to erect their own houses, are provided. This has resulted in approximately equal proportions of the two types of housing in both Tantyi and Newtown locations (Figures 12 and 13). The system whereby sites are provided upon which private construction may occur is known as 'site and service', the service consisting of one tap approximately every 500 yards along the streets, at least one toilet per site, street lights and refuse removal. This scheme has resulted in unplanned and congested living conditions where, in some cases, as many as 10 families share one toilet. Such living conditions must clearly exert some influence on the social and familial aspects of the lives of the people living under these conditions and may be associated with spatial patterns of crime occurrence. The remaining location, Fingo Village, comprises freehold land which was initially divided into 364 plots. Landlords have allowed the erection of backyard shacks which yield between R2-00 and R12-00 per month for a room (Eastern Province Herald, August 22nd, 1975). As a result of these unrestricted and unplanned extensions to housing units, many plots house numerous black families According to research conducted in 1975 an average of 19.96 people lived on one plot in Fingo Village while in some cases this figure was as high as 96 people per plot (Eastern Province Herald, August 22nd, 1975). Once again, from the conditions described above, it would be realistic to expect crime rates to be high in these areas if the empirical research cited is correct in postulating conditions of low rental, physical deterioration, economic, familial and social status being causative factors in crime occurrence.

The main coloured area, Scott's farm (Figure 12), comprises approximately 25% private housing situated predominantly along Albany, Middle Terrace and Currie Streets, and 75% low-rental municipal housing situated predominantly in the northern area of Scott's farm and in the southeastern areas along the western bank of the river.

Studies conducted in American urban centres (for example Klein, 1967) have shown that crime very often occurs across zones of transition from one residential area to the next. Assuming this to be true Grahamstown displays potential firstly for crime against property to occur across boundaries between the various residential zones, based on the aspect of differing socio-economic status between the various areas, and secondly, for crime against person to occur across transition zones, based on the concept of territoriality. Further, it is likely that the 'frozen zone' and Fingo Village will experience higher crime rates than other areas due to mainly economic factors such as physical deterioration of property (Corsi and Harvey, 1975).

Thus Grahamstown, having areas which reflect characteristics related to crime occurrence in overseas urban areas, as well as a movement of people from the various residential areas to the centre of town, has potential for crime in any area within the urban limits. However, given that crime occurs in one area, what is the probability that it will occur in adjacent areas, i.e. does crime occurrence exhibit any form of spatial dependence? Further, if some degree of spatial dependence does exist, does it exist equally for all types of crime, for example, is there any difference in spatial dependence for crime against property as opposed to crime against person? If either of these crime types exhibit some degree of spatial dependence it is clearly necessary to test this finding for different time periods. If the pattern is persistent it is necessary to see whether the spatial component varies during the day as opposed to the hours after dark and also, whether the pattern of spatial dependence in the white areas is similar to that isolated in the non-white areas, and finally, whether there is any marked change in the degree of spatial dependence in the transition zones between white and non-white areas. With this in mind the technique of spatial autocorrelation should provide a useful tool in the analysis of patterns of crime occurrence, and by application to a variety of situations as discussed above the practical usefulness of the technique may be illustrated.

Data Base

In order to arrive at patterns of crime occurrence in Grahamstown offences reported to the central police station in New Street, Grahamstown, over the two-year period 1 January, 1975 to 31 December 1976, were recorded for analysis. From the value of data available from official files, two categories, viz. crime against property and crime against person, out of a possible six categories, were selected for analysis, the former category of crime being more space constrained than the latter. From these two categories, four types of crime against property, (breaking and entering; theft; shoplifting; and auto-theft) and four types of crime against person, (assault; robbery; rape; and murder) were considered because it was felt that each group of four crimes was representative of the respective categories.

Within the first group of crimes against property, breaking and entering was selected for detailed analysis of spatial patterns of crime in the study area because it is the most common crime against property and one for which a precise location can be pinpointed. Theft, shoplifting and auto-theft were not considered for the detailed analysis of spatial patterns for the following reasons. Theft comprises a wide range of sub-categories for each of which complex spatial patterns could be expected to exist and consequently is not suitable for testing the technique of analysis. Also, the location of theft was unrecorded in certain cases or occurred outside the limits of the urban area of Grahamstown. On the other hand, the spatial distribution of shoplifting is limited to retail establishments which in Grahamstown are restricted to the CBD and Beaufort Street extension and do not have a sufficiently wide spatial distribution for analysis. Finally, auto thefts are infrequent and need not necessarily have occurred at the owners' address and are therefore also inappropriate for the technique.

Within the second group of crimes, i.e. those against person, assault was selected for detailed investigation of spatial patterns because it has a high frequency of occurrence and is of major concern to many communities (Harries, 1973). Reported robberies, rapes and murders occurred relatively infrequently and therefore analysis over a one, or even two-year period could not be expected to give enough data to test. In addition to the reasons given above for breaking and entering and assault being representative of their respective categories, the results gained from the pilot study (Ward, 1976) further supported the premise that spatial patterns exist, mainly for the following reason. When mapped, reported breaking and entering and assault for 1975 yielded usually different spatial patterns in the major concentrations of the respective offences, viz. breaking and entering predominated in the white sub-areas of Grahamstown, where 76.0% of all reported breaking and entering occurred, while assault predominated in the non-white sub-areas, where 80.9% of all reported assaults occurred, thus supporting hypotheses postulated for these two crime types in American and British urban centres (refer for example: Harries, 1971, 1973, 1974; Pyle, 1974; Corsi and Harvey, 1975). This difference is strengthened when it is born in mind that of the estimated total number of residential units in Grahamstown, only 30.1% are situated in the white area and further, of the total number of people in Grahamstown only 25.6% are white (Government printer, 1970).

Although breaking and entering has specific spatial constraints, i.e. the location of buildings, empirical research conducted in America has shown that distance travelled for crimes against property is generally more than that for crimes against person (Phillips, 1972; Pyle, 1974). Coupled with the fact that for the two-year period 1975/76, 82.5% of the offenders in cases of reported assault in Grahamstown were black, it is not unlikely that patterns of crime occurrence will be evident in Grahamstown.

However, the question arising from this is to what extent the visual patterns of crime occurrence reflect various degrees of spatial dependence, i.e. will measurement of the degree of first order spatial autocorrelation support the visual differences noted in the patterns for assault and breaking and entering?

When the data for breaking and entering and assault for 1975 and 1976 were plotted the distributions shown in Figures 14-20 were revealed. These figures show very little apart from the fact that breaking and entering occurs predominantly in the white urban areas while assault occurs predominantly in the non-white residential areas. Therefore, gross patterns appear to exist for these two crime types, but on a smaller scale it is not possible to distinguish any pattern. For example, data in the pilot study indicated that breaking and entering was primarily restricted to major west to east thoroughfares in the white area, while assaults were chiefly reported along major west to east thoroughfares as well, but in the black area (Figure 21). However, when adjusting the figures to take into account the length of the street it is clear that inferences cannot be made on a gross level (Table 10). When seen in relation to street distance units, one of the shorter streets in the white area, viz. New Street, experiences as many occurrences of reported breaking and entering as both Beaufort and Fitzroy Streets, which are both more than double its length. Similarly two of the streets in the non-white area of Grahamstown where the occurrence of reported assaults does not appear to be particularly great, viz. E street and Aiken Street, both reflect as many cases of reported assault per 500 metres as both Raglan Road and M Street, superficially both appearing in Figure 21 to experience a predominance of assault. Visual patterns therefore may be misleading.

FIGURE 14: Grahamstown: crime patterns

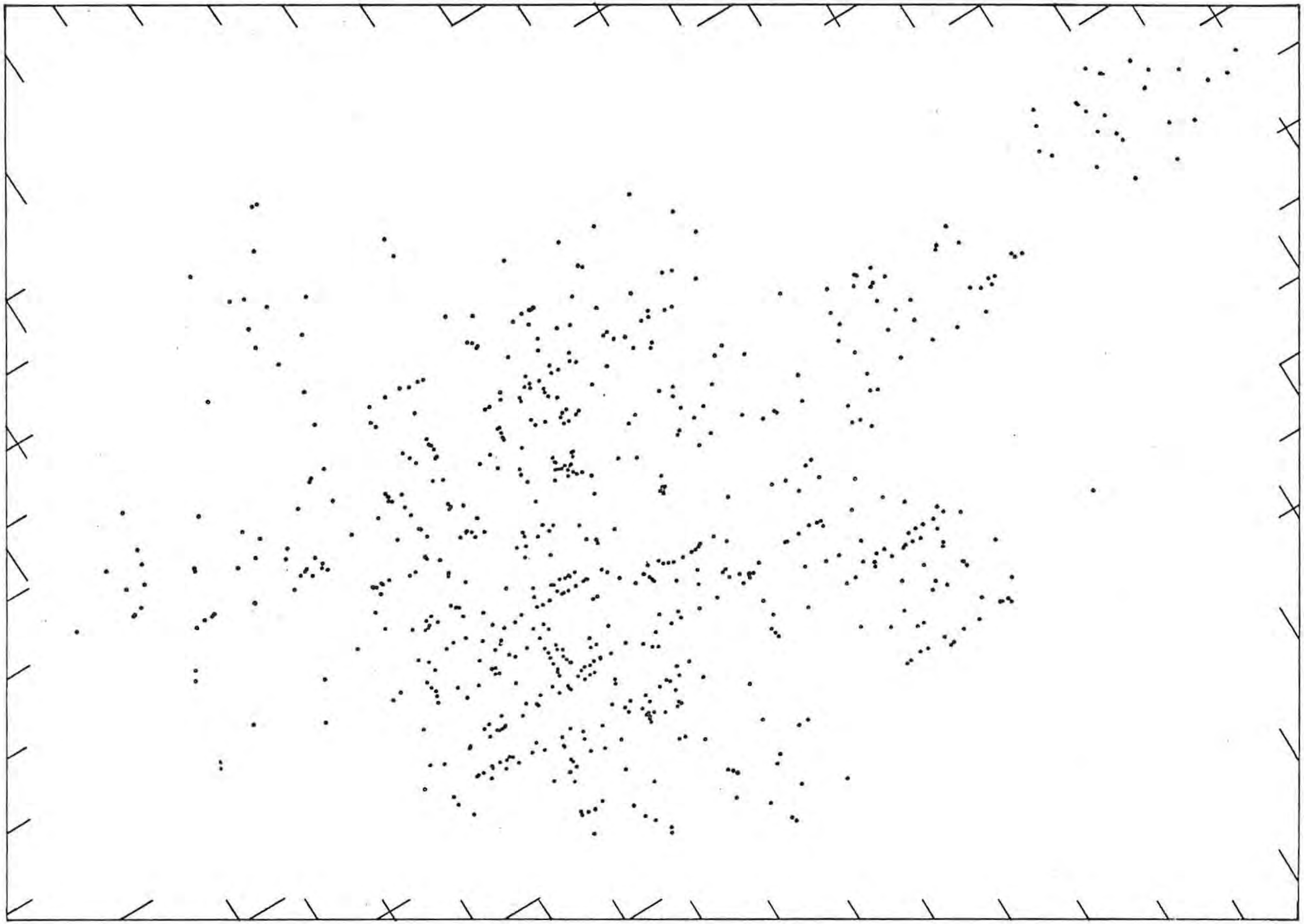
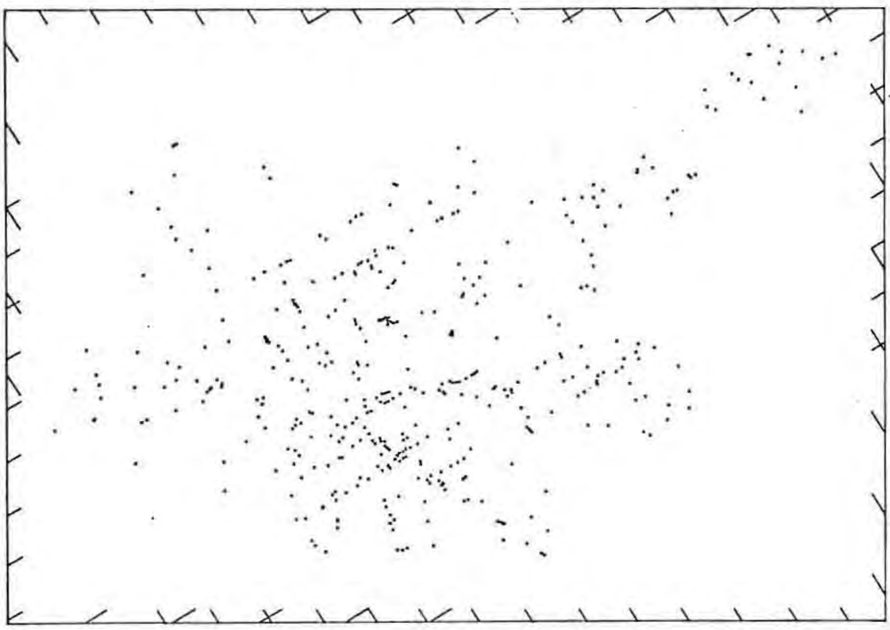


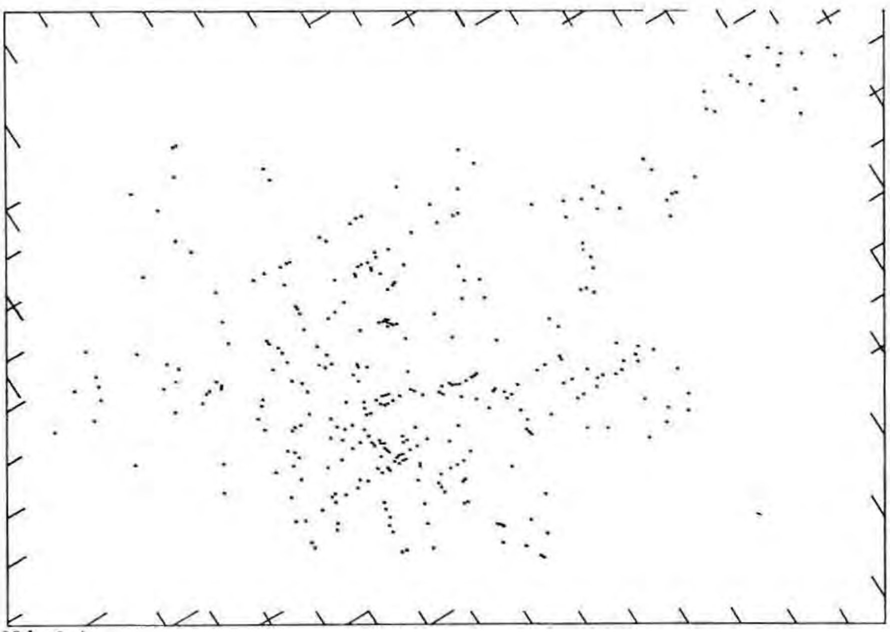
FIGURE 15: Grahamstown: point pattern, total breaking and entering

FIGURE 16: Grahamstown: point pattern, total assault

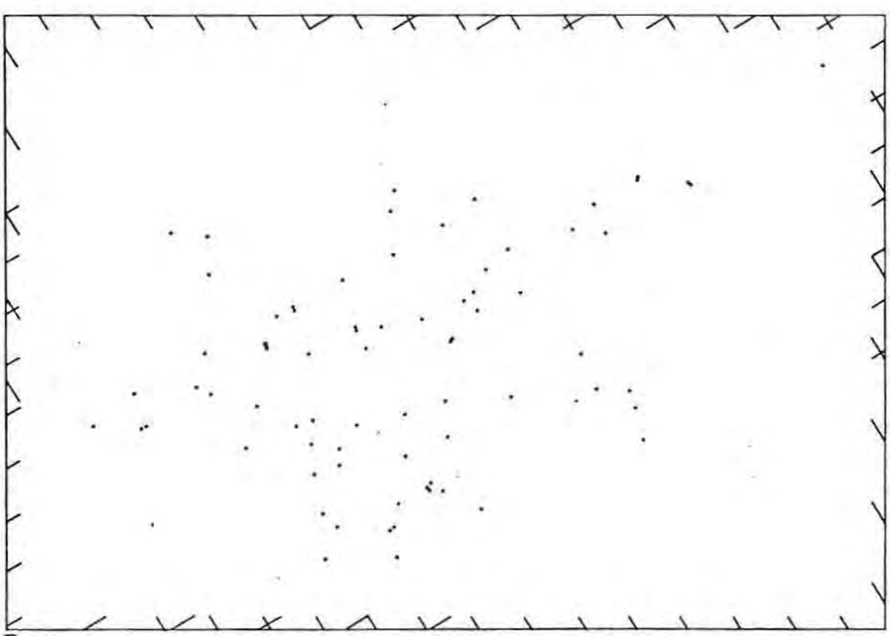
FIGURE 17: Grahamstown: point pattern, breaking and entering 1975



Total

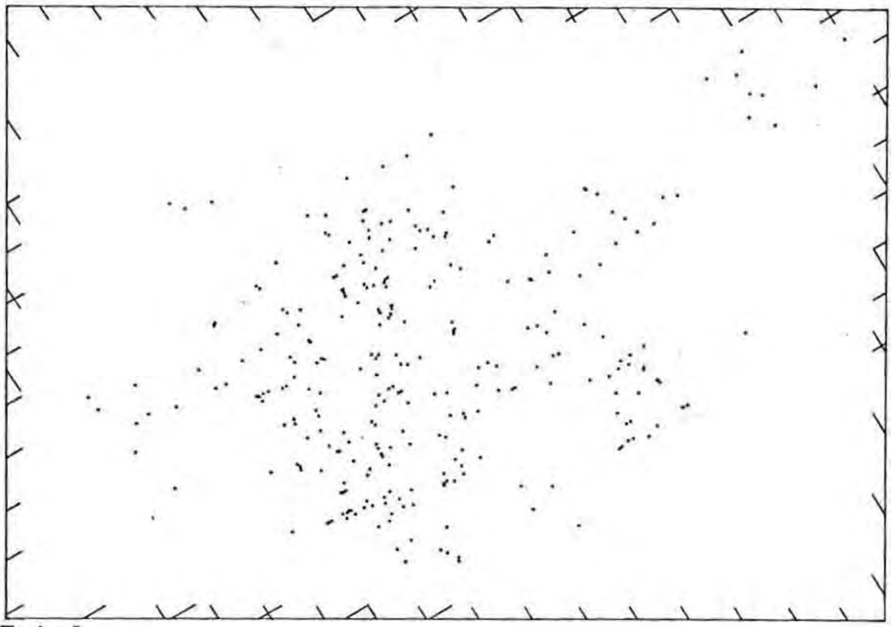


Night

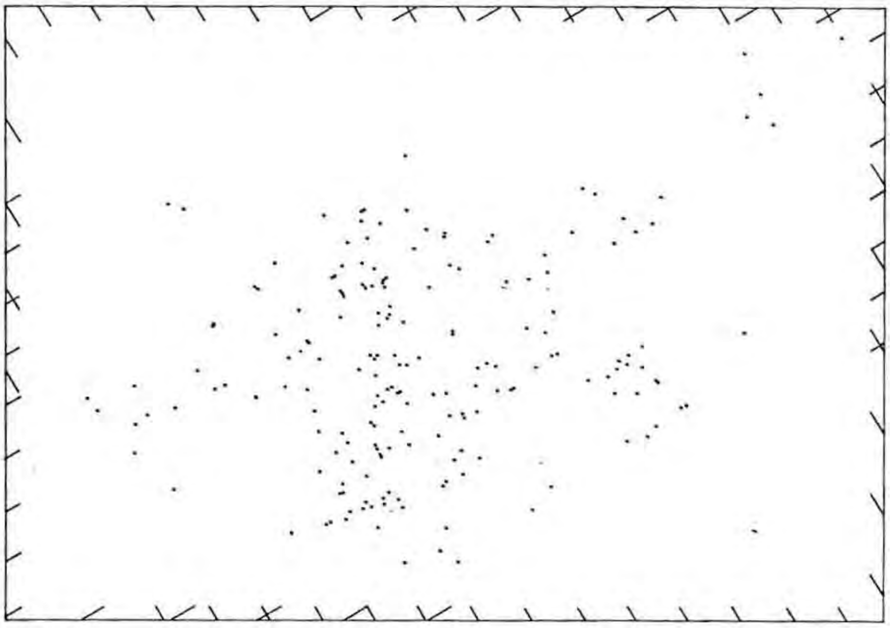


Day

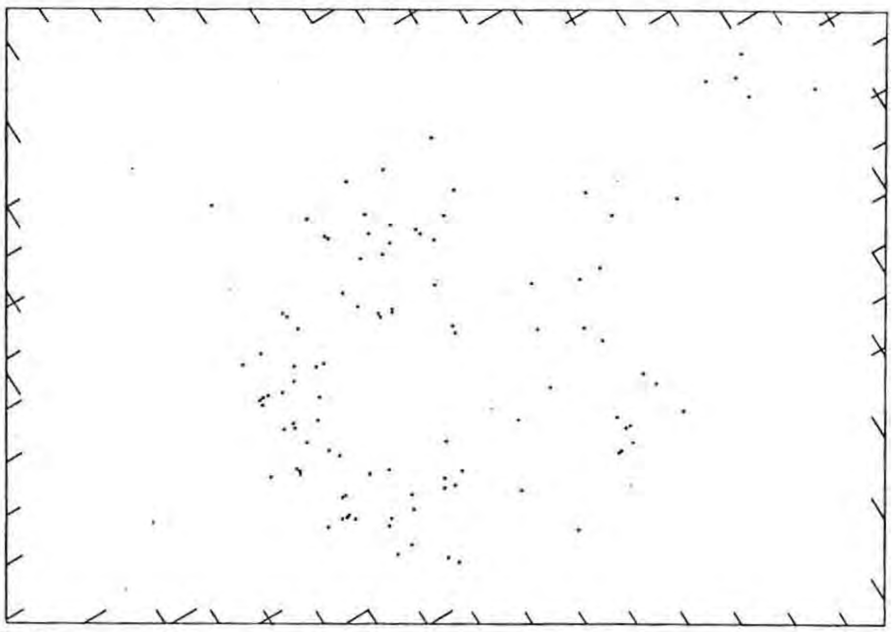
FIGURE 18: Grahamstown: point pattern, breaking and entering 1976



Total

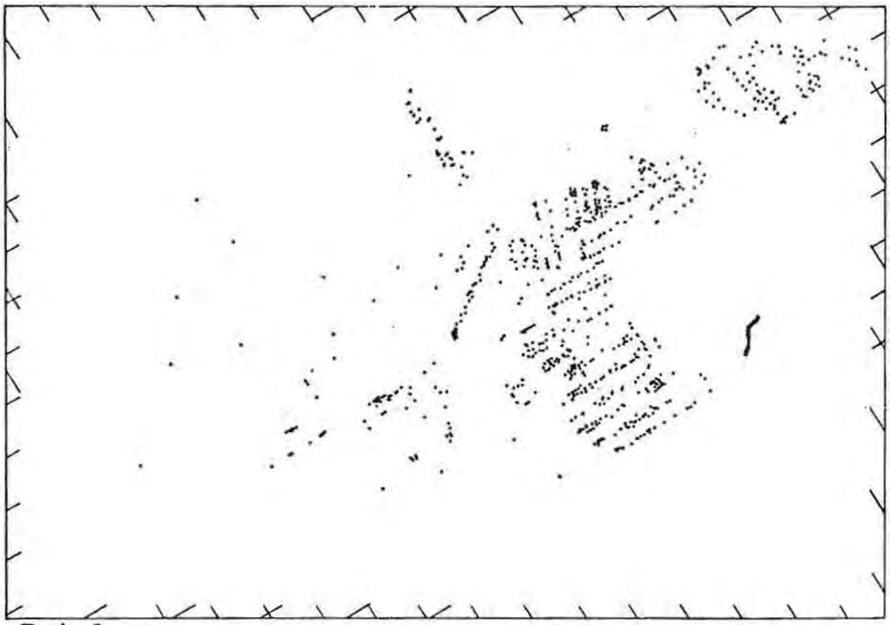


Night

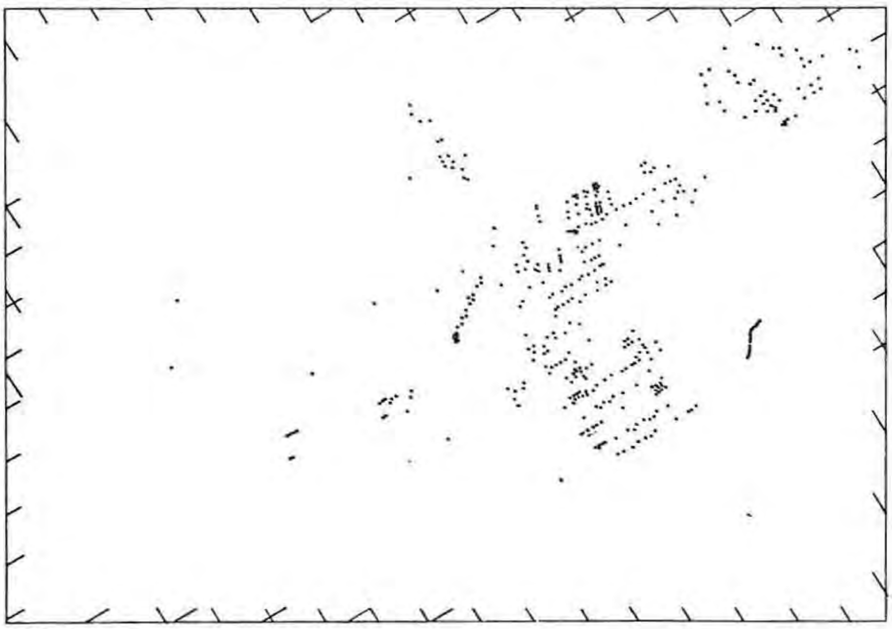


Day

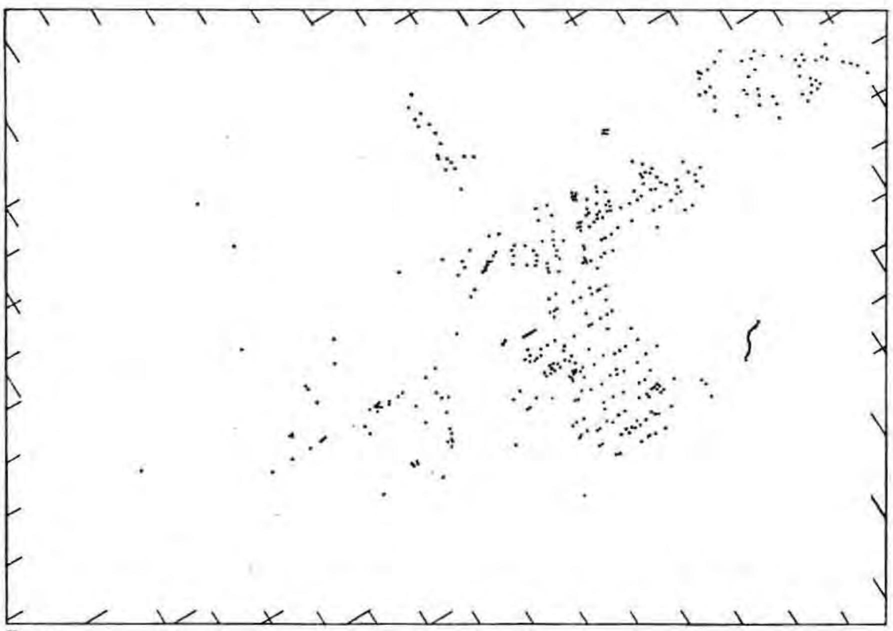
FIGURE 19: Grahamstown: point pattern, assault 1975



Total

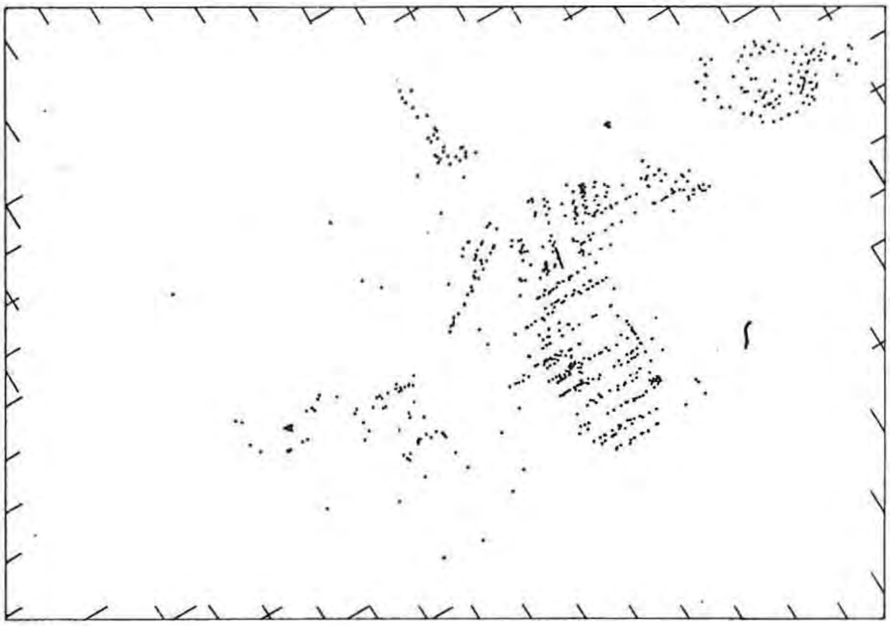


Night



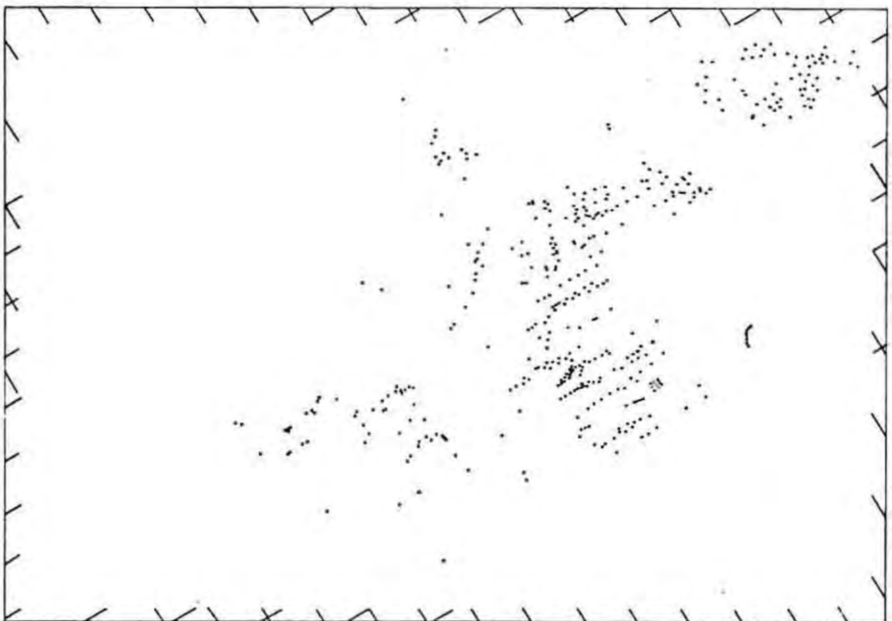
Day

FIGURE 20: Grahamstown: point pattern, assault 1976



Total

Night



Day

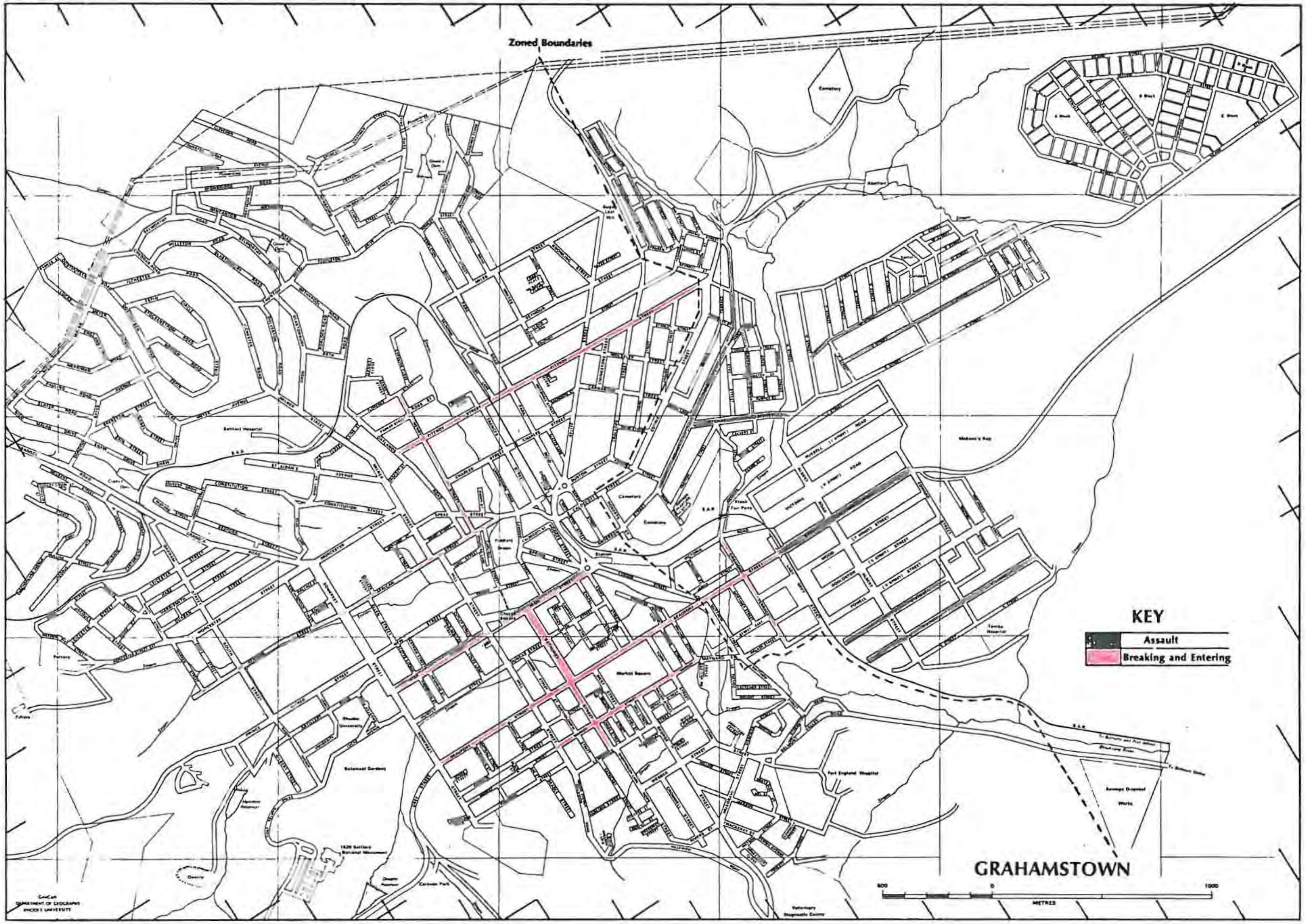


FIGURE 21: Grahamstown: reported offences, by road

TABLE 10. Grahamstown: number of reported offences relative to street length

Street	No. of houses	Breaking and entering per 500 metres	Assault per 500 metres
Beaufort	86	9	2
Fitzroy	60	5	0
Raglan	68	4	17
M	57	1	20
New	63	7	7
Milner	38	6	0
E	62	9	26
Aiken	30	0	17

Even where either breaking and entering or assault is characterized by an apparent pattern in a west to east direction along particular thoroughfares, there is no means of establishing whether given a crime anywhere along one of these streets what the probability is of there being a reported incident of the same crime in an adjacent area, either up or down the street. The one technique which can be used to indicate the degree of spatial dependence, and consequently the importance of spatial proximity in a particular direction, e.g. along predominant streets, is spatial autocorrelation.

In addition to the visual patterns noted above distinct daily and temporal variations in breaking and entering are strongly evident in Table 11 and it is possible that the associated spatial pattern could differ between the hours of 6.00 a.m. to 5.59 p.m. and the hours of 6.00 p.m. to 5.59 a.m. Here again the technique of spatial autocorrelation will be a useful tool to establish the extent of any variation that may exist. On the other hand, reported assaults reflect a fairly even distribution between the two time periods but because of the movement to and from work made every day by the black working population the similar temporal patterns may hide different spatial characteristics. It is therefore not unreasonable to test whether spatial patterns of assaults differ for those occurring between the hours of 6.00 a.m. to 5.59 p.m. and those recorded between 6.00 p.m. to 5.59 a.m.

TABLE 11. Grahamstown: temporal distribution of reported offences

Time	Breaking and entering %		Assault %	
	1975	1976	1975	1976
6.00a.m. - 5.59 p.m.	19.0	35.0	50.4	54.1
6.00p.m. - 5.59 a.m.	81.0	65.0	49.6	45.9
	100.0	100.0	100.0	100.0

In the light of the above characteristics of Grahamstown and crime occurrence in Grahamstown, it is reasonable to assume that certain parts of Grahamstown will experience greater frequencies of crime than others, and also that different types of crime will produce different patterns of crime occurrence. The concern at this stage is to isolate what proportion of this pattern is related to spatial proximity as opposed to other factors and in this regard it is appropriate to make use of the technique of spatial autocorrelation.

Matrix

In order to establish the extent to which the various patterns of crime in Grahamstown are related to spatial proximity it is necessary to organize the data such that it corresponds with the input requirements of the technique. Individual occurrences of crime are of no immediate concern because the exact location of the occurrence is taken as given and also that the incidenc, represented by a single point, has no measurable value. However, when grouped into sub-areas, the number of points within each sub-area may be regarded as a variate value representing crime in that sub-area. Grouping of point based data can be effected by means of a regular matrix without interfering with the requirements of the spatial autocorrelation method, i.e. the location of the area is taken as given and further, the relative sizes of the different sub-areas is held constant which greatly simplifies computation. It has been shown that the nature of the autocorrelation coefficient is influenced by irregular sub-areas and therefore, if data were available by irregular areas, e.g. census tracts or enumerator's sub-districts would have to be standardized before computation could proceed. Therefore, in order to establish whether spatial autocorrelation does exist for the data collected regarding breaking and entering and assault in Grahamstown the urban area was divided into sub-areas according to a regular grid matrix. In fact, because data are not available for small South African towns by any recognized sub-system of delimitation, the use of a regular matrix is the one means by which data can be grouped objectively, the only decision necessary being the choice of grid square size.

In this case the rule of thumb that was applied was that the length of the axis of every grid square should be shorter than the shortest distance across the smallest area of interest on the map. Although it has been argued that different crime types require a different grid size due to the fact that the relative size of areas of interest will differ according to the type of crime being studied, a residential block has been taken to be the smallest area of interest. This decision is based on the fact that every residential block will have at least one house, which represents a target for burglary, and further that is is probable to find people where there are houses, people being necessary if interaction and consequently assault, is to occur. If the shortest distance across the smallest block in the urban area of Grahamstown were taken as a suitable grid square base the squares would be unrealistically small and in many cases would not contain residential units. As a practical alternative therefore it was decided to use the average distance across the larger residential blocks in Grahamstown as a basis for grid size. The selected length for the sides of the grid squares was one centimeter (refer block sizes on Figure 12).

Orientation of the grid matrix also posed problems. Finally it was decided to orientate the axes according to dominant street patterns in Grahamstown which tend to be approximately WSW to ENE and NNW to SSE (refer Figure 12). This decision was based on two factors. Firstly, many of the streets in Grahamstown form a regular grid pattern and as such the superimposition of the matrix would capture more complete residential blocks per grid square than would otherwise be the case. Secondly, orientation was arranged so that the rows and columns of the matrix would incorporate complete streets, for example, if a street fell along a row it would be sub-divided only by the column axis and not by the row axis of the matrix, therefore facilitating the analysis of linear spatial dependence along cardinal directions. With the matrix oriented in this fashion it is likely that certain directional biases, if they exist according to major/ street direction in Grahamstown, will appear. These will be referred to as column (N/S), row W/E), left diagonal (NW/SE), and right diagonal (NE/SW) directions respectively and correspond directly to the first four runs of the program.

It has been shown that when testing for spatial autocorrelation using different grid sizes differences in the value of the coefficient could result (refer Chapter 3). It is therefore necessary to establish whether the decision regarding grid size for analysis of crime patterns in Grahamstown was appropriate or not. By superimposing a different grid size on the crime data this can be achieved. In addition the fact that the matrix was orientated in an arbitarily chosen direction also necessitates that the effect of this decision on the value of the coefficient should be noted. Although Olson (1975) states that the pattern inherent in the distribution should be reflected in the autocorrelation coefficient regardless of the orientation of the map this may not be the case and it is possible that the coefficient will vary according to grid matrix orientation simply because the data are being grouped in a different manner. In order to take these factors into account, firstly, a slightly larger grid square size was used, viz. 1.25 cms x 1.25 cms, on the data base. The decision to increase the size of grid was based on the fact that during the pilot study, larger irregular sub-areas of similar socio-economic characteristics were found to be associated with concentrations in crime patterns and the shortest distance across the smallest of these was 1.25 cms. Secondly the matrix was superimposed according to the cardinal or "true north" direction.

Results

The results of the application of Cliff and Ord's spatial autocorrelation statistic to crime data in Grahamstown are discussed firstly with regard to the nature of the computed statistic, after which attempts are made to interpret them in a geographic context. Although the primary matrix was oriented according to strict direction the various runs of the program will be referred to as north to south (N/S), west to east (W/E), northwest to southeast (NW/SE), and northeast to southwest (NE/SW) respectively and relate to town, and consequently grid layout rather than cardinal directions. The size of the matrix (13 x 20), coupled with the high frequencies and widespread nature of crime occurrence in Grahamstown meant that nearly all of the results were significant at the 99.7% level. As a result it seems essential to attempt interpretation of the coefficients where possible on the basis of knowledge concerning the study area.

The importance of space in crime patterns

Analysis of breaking and entering and assault, when combined over the two year period 1975/1976 (Figure 14) showed that autocorrelation in all directions yielded a coefficient of 0,56, which means that 31% of the variance in the distribution of crime may be attributed to the spatial component (Table 12). It is evident therefore that the very fact that crime occurs at a particular location means that there is an increased probability of another crime occuring nearby in Grahamstown. When the different directions are considered separately the highest coefficient for this distribution is in a N/S direction where the coefficient of determination rises to 39%. In geographic terms this means that although the frequency of crime may be distinctly different for the west and eastern parts of town at least 39% of the probability of the occurrance or non-occurrance of crime is associated with location in areas immediately north or south of any cell within the matrix. Spatial location therefore plays a distinct role in crime patterns in Grahamstown.

(i) Breaking and entering and assault: total patterns

When crimes against person are separated from those against property in order to establish if the overall pattern already discussed reflected a particularly strong spatial component in only one of the two types of crime it is evident that this is not the case (Figure 15 and 16; Table 12). Although the pattern for assault yielded coefficients that were slightly stronger than those yielded by patterns for breaking and entering spatial dependence in both patterns of crime occurrence is clearly very similar. In both cases the N/S directions yield the strongest autocorrelation coefficients while the NW/SE directions yield the

TABLE 12: Autocorrelation coefficient: total crime patterns

Run	Number of joins	1975/1976 Breaking and Entering, and Assault N = 2215	1975/1976 Breaking and Entering N = 690	1975/1976 Assault N = 1525
N/S	214	0,625	0,617	0.683
W/E	228	0,612	0,569	0,640
NW/SE	210	0,477	0,516	0,503
NE/SW	196	0,519	0,555	0,577
general	848	0,560	0,565	0,602

All values significant at the 99,7% level.

weakest coefficients (Table 12). These directional trends were confirmed by calculation of Pearson's correlation coefficient which shows a slight decrease in the strength of the coefficients for both the above directions for breaking and entering and a slight increase in the strength of the coefficients for the same directions for the pattern of assault. These results indicate that there is a greater probability of the values for assault exhibiting an even trend in any direction than for the values for breaking and entering, which are possibly more variable in nature (for an example of an even trend of values refer Figure 3). Nevertheless, the coefficients for both patterns are sufficiently strong to ensure that given either breaking and entering or assault at any one location in Grahamstown the probability that either will occur in an adjacent area will be highest immediately north and south of that location. This pattern of spatial dependence may be interpreted as reflecting the socio-economic conditions in Grahamstown as the one direction in which there is least change across the surface in the housing and living characteristics of Grahamstown residents is from north to south. On the other hand along the NW/SE axis great changes in the housing and residential characteristics are evidnet, i.e. expensive white housing is situated in the northwestern area of Grahamstown while at the other end of the axis Grahamstown's squatter settlement is found. In view of this fact and in line with theories developed overseas about crime occurrence it can be expected that there will be a lesser degree of spatial dependence along this axis with regard to crime occurrence.

(ii) Breaking and entering: yearly and daily patterns

When each of the two years are regarded separately the differences noted in the autocorrelation coefficients are minimal (Table 13), indicating that the relative importance of the spatial component has not changed from 1975 to 1976 (Figures 17 and 18 respectively). This finding makes it clear that the spatial component in the pattern of breaking and entering is not ephemeral especially as for each year separately and for the two combined, autocorrelation along the N/S axis is strongest. When seen in relation to the degree of general dependence exhibited by the combined 1975/76 pattern for breaking and entering, viz. 31%, the degree of variance attributable to the spatial component is lower for the individual patterns for 1975 and 1976, viz. 19.9% and 22.7% respectively, a drop which can be accounted for by the fact that the number of occurrences is reduced by 41.8% and 58.2% respectively. Therefore, by keeping the coefficients for the 1975 and 1976 patterns significant at the 99.7% level there are insufficient cases to attribute more than 20% of the variance in the pattern to a spatial component with such a high degree of confidence. A factor which does differ between the 1975 and 1976 pattern is the direction in which the weakest coefficient appears, viz., the NE/SE axis for the 1975 pattern as opposed to the NW/SE axis for the 1976 pattern. This suggests that the probability that breaking and entering will occur in any direction other than north to south is variable. These directional correlations were all supported by the Pearson's correlation coefficient which, as for the total pattern of breaking and entering, yielded coefficients which were slightly lower than those yielded by Cliff and Ord's method.

In order to establish whether the patterns of spatial dependence was evident at all times patterns for each year were subdivided into two categories, viz., the day time period from 6-00 a.m. to 5.59 p.m. and the period after dark from 6.00 p.m. to 5.59 a.m. The degree of spatial dependence reflected by the night patterns for both years is very similar to the total patterns for the two years, whereas daily patterns for both years exhibit a far lower level of spatial autocorrelation. In fact three of the directional runs for the 1975 pattern, viz. N/S, NW/SE and

TABLE 13. Autocorrelation coefficient: breaking and entering

Pattern	Run	Number of joins	1975 Breaking and Entering N = 402	1976 Breaking and Entering N = 288
Total	N/S=	214	0,517 **	0,550 **
	W/E	228	0,442 **	0,450 **
	NW/SE	210	0,438 **	0,415 **
	NE/SW	196	0,382 **	0,494 **
	general	848	0,446 **	0,477 **
Night			n = 333	n = 187
	N/E	214	0,517 **	0,448 **
	W/E	228	0,399 **	0,409 **
	NW/SE	210	0,412 **	0,378 **
	NE/SW	196	0,327 **	0,564 **
	general	848	0,416 **	0,447 **
Day			n = 69	n = 101
	N/S	214	-0,005	0,449 **
	W/E	228	0,354 **	0,327 **
	NW/SE	210	0,166	0,172
	NE/SW	196	0,082	0,290 **
	general	848	0,154 **	0,311 **

* Significant at the 95% level

** Significant at the 99.7% level

NE/SW, and one run for the 1976 pattern, viz., NW/SE, yielded non-significant correlation coefficients. The degree of variance attributable to the spatial component for the daily pattern for 1975 is a mere 2.3% while for the corresponding 1976 pattern the value only rises to 9.6%. The apparent decrease in spatial dependence in the daily patterns could be a reflection of the number of occurrences reported during that period. During 1975, only 69 incidents were reported during the day, which represents 17% of the total number of breaking and entries during that year, while for 1976 the daytime frequency was 101, or 35% of the total number of occurrences. On the other hand these results could indicate that breaking and entering occurring during the day is conducted on a far more random basis than is the case for the hours after dark. However, other variables would have to be considered if this idea were to be comprehensively tested. One explanation could be simply that a different range of possible targets is exposed during the hours of daylight as opposed to the hours after dark. For example, shops and warehouses are likely to be targets during the night rather than during the hours of daylight while residences could be burgled both during the day and at night, although certain locations may be more conducive to burglary under the cover of dark.

(iii) Assault: yearly and daily patterns

When both yearly and daily patterns of assault were analysed consistent support for the relative importance of space in the patterns was obtained (refer Figures 19 and 20). The difference in autocorrelation levels between the 1975 and the 1976 pattern are minimal (Table 14) and reflect the total pattern that 36% of the variance is attributable to the spatial component inherent in the distribution. Although the dominant direction of spatial dependence in the 1975 pattern is W/E ($r = 0{,}688$), that for the 1976 pattern is a NS/ direction ($r = 0{,}667$) as is the case in

TABLE 14. Autocorrelation coefficient; assault

Pattern	Run	Number of Joins	1975 Assault N = 791	1976 Assault N = 734
Total	N/S	214	0,659	0,667
	W/E	228	0,688	0,557
	NW/SE	210	0,504	0,470
	NE/SW	196	0,564	0,570
	general	848	0,607	0,567
Night			n = 392	n = 328
	N/S	214	0,647	0,653
	W/E	228	0,648	0,469
	NW/SE	210	0,473	0,450
	NE/SW	196	0,587	0,550
	general	848	0,590	0,529
Day			n = 399	n = 406
	N/S	214	0,546	0,621
	W/E	228	0,652	0,575
	NW/SE	210	0,444	0,474
	NE/SW	196	0,398	0,533
	general	848	0,515	0,552

All values significant at the 99,7% level.

the total pattern (r = 0,683). The 1975 results may be an anomaly associated with assaults concentration in W-E streets for it appears that when assaults occur they are more likely to be repeated in areas north or south of the first occurrence than in any other direction. This finding may again be related to the fact that the white parts of Grahamstown are dissimilar to the black areas and as such there are more differences than similarities between the western to eastern parts of Grahamstown in the pattern for assault. As the differences between spatial autocorrelation coefficients for daily patterns of assault, as opposed to assaults which occurred after dark are minimal, the degree of spatial dependence in the occurrence of assault is apparently not related to time of occurrence. Neither does the frequency of assault vary between day and night as the ratio is approximately one to one for both years. The fact that space accounts consistently for at least 30% of the occurrence or non-occurrence of assault, as opposed to other factors such as temporal, environmental, psychological and many other motivational factors, indicates that space is positively associated with the distributional characteristics of the occurrence of assault.

(iv) Assault: Racial patterns

In the light of the higher correlations for the N/S direction, viz., 0.683 (Cliff and Ord's method) and 0.707 (Pearson's method) evident in the overall assault patterns and its relatively greater importance in the 1976 breakdown (0.667 (N/S) versus 0.470 (NW/SE)) it was decided to consider the black and white parts of town separately. By separating the two residential areas the frequency of assaults per cell dropped particularly in the white area, where only 193 were reported in contrast to the 1 332 incidents which occurred in the black area. When coupled with the fact that the total number of joins in the black area amounted to 234 as against 554 in the white area it is not surprising that the values of the calculated coefficients for both the white and the black areas of Grahamstown dropped (Table 15). As a result the coefficients of 0,33 and 0,38 indicate that at the 97,7% level of significance approximately 12.25% of the variance can be attributed to the spatial component and that it is equally important in both distributions. When directional runs are considered the white area clearly reflects a stronger W/E dependence while for the black area the N/S direction is dominant, both trends being supported by Pearson's method. This is perhaps due to the fact that in the white area residential and associated socio-economic characteristics change from the northern to the southern side of town and as a result there is more similarity in an east-west direction. On the other hand in the black area the differences in housing characteristics are more marked between the west and east parts. In addition, there are many more open spaces in the eastern parts of the black area and thus north south similarities are to be expected.

When the transition zone between the white and non-white residential areas is considered separately a very different picture emerges for the N/S correlation in particular. In this case four columns from the original matrix were used in the analysis - two from either side of the white black transition zone. The coefficient of 0,64 (0,64 Pearson's method) for the N/S direction, when seen in relation to the coefficient of 0,33 (0,36 Pearson's method) for the W/E direction, and dropping to 0,26 (0,32 Pearson's method) for NW/SE, all reflect a distinct difference in spatial dependence between the two directions. The obvious lack of dependence across the transition zone may be related to the postulate that crime will be higher in the lower income, disorganized, deteriorating areas (Mays, 1956; Boggs, 1965; Pyle, 1974; Corsi and Harvey, 1975). On the one side of the transition zone is the socially and economically stable white area whereas on the other side from north to south there is

TABLE 15. Autocorrelation coefficient: assault, residential patterns

Pattern	Run	Number of joins	1975/1976 Assault N = 1525
Total	N/S	214	0,683 **
	W/E	228	0,640 **
	NW/SE	210	0,503 **
	NE/SW	196	0,577 **
	general	848	0,602 **
White area			n = 193
	N/S	146	0,334 **
	W/E	142	0,445 **
	NW/SE	138	0,317 **
	NE/SW	128	0,223 *
	general	554	0,332 **
Black area			n = 1332
	N/S	64	0,467 **
	W/E	66	0,377 *
	NW/SE	54	0,258
	NE/SW	50	0,425 *
	general	234	0,425
Transition zone			n = 703
	N/S	72	0,641 **
	W/E	58	0,335 *
	NW/SE	54	0,268
	NE/SW	54	0,368 *
	general	238	0,420 **

* Significant at the 95% level

** Significant at the 99,7% level

firstly Scott's farm which is an elongated area of coloured housing oriented on a N/S direction adjacent to the transition zone. At the southern end of this zone is the coloured beer hall and part of Grahamstown's industrial area. South of this is Fingo Village with its beer hall and congested living conditions.

Effects of grid size and matrix orientation

(i) Grid size

Patterns for breaking and entering and for assault for 1975 were used to test the effect of applying a larger grid square size, viz. 1.25 cms x 1.25 cms, to the study area. One of the reasons for selecting both assaults and breaking and entering was because the differences between the coefficients for autocorrelation of these two patterns using a 1 cm x 1 cm grid were most marked during 1975. It was therefore considered necessary to establish whether this difference was due to the choice of grid size and its relationship to frequency of occurrence. In addition, breaking and entry being a crime against property and assault a crime against person suggests that the level of spatial interaction and hence the spatial component inherent in each distribution may be distinctly different.

When the larger matrix was used, the coefficients for the pattern of breaking and entering increased and reflected a greater degree of spatial dependence, while for the pattern of assault the coefficients dropped and suggested a lesser degree of autocorrelation (Table 16). Specifically, using the larger grid size the amount of the variance attributable to the spatial component for the pattern of breaking and entering rose from 19.9% to 30.5%, while for the pattern of assault this factor dropped from 36.8% to 21.5%. As the crime frequency remained the same and the results were of the same level of significance it is clear that the choice

TABLE 16. Autocorrelation coefficient: effects of grid size

Pattern	Run	Number of Joins	1975 Breaking and Entering N = 402	1975 Assault N = 791
1 cm x 1 cm grid	N/S	214	0,517	0,659
	W/E	228	0,442	0,688
	NW/SE	210	0,438	0,504
	NE/SW	196	0,382	0,564
	general	848	0,446	0,607
1,25 cm x 1,25 cm grid	N/S	140	0,598	0,534
	W/E	152	0,621	0,492
	NW/SE	140	0,508	0,338
	NE/SW	126	0,466	0,492
	general	558	0,552	0,464

All values significant at the 99,7% level.

of grid size does affect the nature of the results that can be gained from using Cliff and Ord's statistic. The complete reversal in terms of the importance of space for the two sets of da5a can be attributed to the level of interaction for the two types of crime. It seems evident that it is necessary to increase the grid size for breaking and entering so as to incorporate a larger number of targets in each cell if the tehcnique is to comprehensively evaluate the importance of space in any distribution of crime against property. For assault on the other hand, where closer interaction si important grid size could possible be reduced to more adequately represent the level of interaction in any crime against person, especially assault and get a true evaluation of the relative importance of space when calculating the correlation coefficient.

(ii) Matrix orientation

The effect of superimposing a 1 cm x 1 cm grid of square size matrix according to a position other than that aligned along the predominant street direction was tested on the pattern of assault for 1975. The matrix was rotated until paralled to the cardinal north to south and west to east direction. Not unexpectedly the differences in the various coefficients were not very great (Table 17). The degree of variance in the distribution attributable to the spatial component dropped from 36.8% (street orientation) to 33.3% (cardinal orientation). However, the existing N/S pattern emerged more clearly as the percentage of the variance increased from 43.4% to 43.8%. (The Pearson's run for this direction gave values of 50.4% versus 53.3%). The variance in the distribution of assault along the W/E axis according to the intial analysis, i.e. 47.3%, dropped to 37.6% when the grid was orientated in the cardinal direction (Table 6).

TABLE 17. Autocorrelation coefficient: effects of grid orientation

Pattern	Run	Number of joins	1975 Assault N = 791
Street orientation	N/S	214	0,659
	W/E	228	0,688
	NW/SE	210	0,504
	NE/SW	196	0,564
	general	848	0,607
Cardinal orientation	N/S	206	0,662
	W/E	218	0,613
	NW/SE	192	0,492
	NE/SW	196	0,529
	general	812	0,577

All values significant at the 99,7% level.

This change can be accounted for by the fact that although a "... given pattern of values possesses the same amount of organization regardless of the orientation of the map." This law refers to any distribution of data which is not grouped or categorized in any manner (Olson, 1975, p.203). In the case of crime patterns the point data are grouped to become areas area base data. Rotation of the matrix therefore provides an alternative grouping of the data, with resultant slight changes in the autocorrelation levels noted in Table 6, and illustrates the point that some knowledge of the expected distribution is required before orientating the matrix or alternatively when using Cliff and Ord's statistic the matrix should be rotated according to various directions in order that no directional biases are being overlooked.

In conclusion therefore, Cliff and Ord's statistic, as applied in this study, is clearly a technique that can be used by geographers in practical situations. However, it is necessary to bear the following aspects of the technique in mind when interpreting the results yielded by the method. The choice of grid square size, and grid matrix orientation are clearly crucial to the analysis of any distribution of data. Further, the frequency of occurrence of the event will also affect the strength of the coefficient that can be gained from using Cliff and Ord's statistic given that a particular level of significance is maintained. Notwithstanding the above limitations regarding the application of the statistic and interpretation of the results; the statistic functions adequately in the role for which it si primarily intended, which is to indicate whether or not spatial dependence exists. Further, by applying the technique as outlined in this study it is possible to focus attention on the various directions in which dependence is particularly strong. Finally, it is possible that the nature of the results which can be gained from applying Cliff and Ord's statistic can indicate directions for future research which aims at developing generally accepted theories of crime occurrence.

CHAPTER VI

EVALUATION

Spatial autocorrelation, as has been outlined in this study, can be applied to any geographic situation to achieve an objective evaluation of the degree to which individual observations belonging to a single set of data co-vary spatially. The application of the technique has been outlined in such a manner as to facilitate its use by geographers in the analysis of the spatial component inherent in point patterns.

In order to establish the extent to which Cliff and Ord's technique is sensitive to variations in the parameters a regular matrix was used as a means by which demarcation of any two-dimensional surface into comparable sub-areas is possible. Demarcation according to a regular matrix facilitates isolation of the spatial component in the simplest possible manner and thus paves the way for a more critical examination of spatial interaction between sub-areas. When delimitation is effected according to regular matrix three factors which influence the nature of the interaction between sub-areas, viz., sub-area size, relative distances between the central points of sub-areas, as well as relative boundary lengths are all held constant. This method of demarcation is therefore suited to most practical situations where analysis is aimed at isolating the importance of the spatial component in a point distribution and especially where data are not available by recognized sub-areas.

There are three major advantages in demarcating a two-dimensional surface according to a regular matrix when applying Cliff and Ord's statistic, viz. concise and objective demarcation of a distribution of data is achieved, the possibility of isolating directional biases in spatial dependence along any cardinal direction is introduced, and a simple and standardized form of w_{ij} can be employed and thus the researcher can be assured that the original pattern is in no way distorted by a generalized interaction formula. Demarcation according to a regular matrix provides an objective means of transforming point base data (i.e. discrete variates each representing the frequency of the phenomenon in a certain area). Finally, when using a regular matrix it must be remembered that unless the frequency of the dot distribution is sufficiently high to give values in a large proportion of the cells, the real strength of the spatial component may be underestimated given that a certain level of significance is regarded as acceptable. Although this problem became apparent when computing autocorrelation for the initial dummy matrices (Figure 3), the extent of the problem with regard to practical application of Cliff and Ord's technique as outlined in this study was clarified when blank cells were included in the analysis (refer to the difference in coefficients for MAPA and MAPB, Table 4). The effect was also noticeable when the frequency of crime occurrence differed markedly between two patterns (refer Table 13. Breaking and entering 1975, daytime versus night-time pattern), which resulted in much lower coefficients for the pattern with the lower frequency of occurrence. This led directly to interpretation problems because the indication that a lower degree of spatial dependence existed could have been co-incidental rather than real. The obvious conclusion to be drawn from this is that despite Cliff and Ord's comment it is not sufficient merely to know whether or not a surface exhibits autocorrelation at a significant level, but rather that it is necessary that researchers be able to interpret the coefficient itself in the light of the factors known to influence the statistic.

Closely associated with the above factor is the influence of the number of grid squares, and consequently the number of joins present in the analysis on the strength of the coefficient, given a particular level of significance. Although it is only possible to attain a coefficient of 1.00 using Cliff and Ord's statistic if all the values are equal, coefficients as high as 0,924 using hypothetical values exhibiting a strong pattern result if the number of joins present in the system is high. Analysing the same pattern but using an increasing number of grid squares the coefficient yielded by Cliff and Ord's statistic values ranged from 0.3 for a 3 x 3 matrix to 0,924 for a 20 x 20 matrix and clearly illustrate how the precision with which the co-variance can be measured increases with an increase in the data input.

Further, it was shown that grid square size affects the results gained from application of Cliff and Ord's statistic. As grid square size increases so does the probability that at least one occurrence of the phenomena being measured will exist per grid square. It is necessary to arrive at an optimum choice for the grid square size as it has been shown that both the number of joins in the system as well as the number of vacant cells influences the strength of the autocorrelation coefficient. In this regard it is important to remember that when the grid squares approach a size where many of the occurrences appear in one grid square any real analysis of spatial variation in the data will become impossible because when demarcating patterns of breaking and entering and assault according to two different grid square sizes the former pattern yielded a stronger coefficient while the latter pattern yielded a weaker coefficient for the larger grid square size. These results indicate the need to know about the nature and the level of interaction of the data being measured prior to analysis or alternatively, indicate the need to experiment with different grid square sizes before finally deciding on a suitable matrix.

Where grouping of the data is necessary for the purposes of analysis the orientation of the matrix provides various possible combinations of data. In the Grahamstown case it was assumed that if the matrix were oriented according to the predominant street pattern so that as many grid squares as possible contained complete residential blocks (which were considered as important to the distribution of crime occurrence) the level of autocorrelation would be high. For the alternative matrix orientation no presupposition regarding the distribution of the data was made and the matrix was oriented according to the cardinal N/S direction. Not unexpectedly in the former case the pattern for assault 1975 yielded a coefficient for the general run which was higher than the corresponding coefficient for analysis according to the latter matrix. However the N/S run for the cardinal oriented matrix showed up a stronger coefficient. The orientation should be isolated prior to actual analysis in order to achieve the optimum results in the light of the aims of the particular survey.

Finally therefore, all of the above factors should be carefully considered before any analysis proceeds in order to be certain that the manner in which demarcation is affected is appropriate for the particular data case to which it is applied.

This study has shown that to limit analysis of spatial autocorrelation to the surface as a whole is not necessarily reasonable in that it is quite possible that spatial dependence may not exist equally in all directions. Further, if some form of directional dependence does exist, care must be taken in the manner in which the results for the various directional runs are interpreted. For example, the strongest pattern or trend immediately evident from the autocorrelation coefficient is the case where all the values in one direction are equal. A lower coefficient for one of the remaining directions may, however, reflect a perfect trend of increasing or decreasing values. If such a situation does exist calculation of the Pearson's coefficient will indicate whether the trend exists for the whole of the directional run. Interpretation of a directional coefficient must also take into account the geographic meaning of the relationship (refer for example to the results for the transitional zone between the white and black residential areas - Table 15). Results gained from analysis using Cliff and Ord's method should be done carefully, taking into account both the characteristics of the study area and if necessary, extending the analysis to calculate Pearson's product moment correlation coefficient.

Cliff and Ord (1969-1975) extend the application of their statistic to include the analysis of autocorrelation between non-adjacent areas by incorporating specialized forms of the function for w_{ij}, the aim of which is to neutralize the effect of friction of distance on the nature of the interaction between non-adjacent cells. In this regard, the onus is on the researcher to choose the form of w_{ij} prior to analysis on the basis of his perception of the level of interaction between any particular pair of cells (Cliff and Ord, 1969). Decisions regarding a form of w_{ij} will necessarily be subjective and in order to reduce the arbitrary nature of this type of choice much empiric analysis is required before the effects of the chosen forms of w_{ij} on the statistic can be comprehensively gauged. In this regard, an objective means of delineating non-adjacent sub-areas is by using cluster analysis. For example, it is generally accepted that crime occurs in clusters (Pyle, 1974), and as such a technique whereby mathematically optimum cluster boundaries are drawn could provide a means of establishing where the 'breaks' between the various areas are located. It is then possible to adjust the w_{ij} to relate the areas directly although the actual choice of the form of the interaction function across the 'break' would still be a purely subjective consideration. However, first order relationships have been emphasized in this study since both Bannister (1975, 1976) and Olson (1975) maintain that first lag relationships tend to be most meaningful in terms of autocorrelation and further, that most research problems in geography will require evaluation of proximate spatial dependence.

A further extension of Cliff and Ord's (1969-1975) statistic is the manipulation of w_{ij} to facilitate interaction between irregular sub-areas. In this study it was shown that irregularly sized sub-areas affect the strength of the coefficient and following this, how standardization of differences between sub-area size by adjusting only for relative areal size differences between sub-areas affected the strength of the coefficient. The results shown in Table 6 show clearly that this is not sufficient means of standardization to negate the effects of differing areal size on the strength of the coefficient. However, Cliff and Ord (1969, 1973) put forward an example of a weighting function which standaridzes the areal differences between subareas by taking into account the relative sizes of the subareas, relative distances between the central points of sub-areas and the boundary lengths along joins between adjacent sub-areas. Therefore it is clear that much research is necessary in this regard before it can be shown with confidence which forms of w_{ij} effectively nullify the effects of irregular sub-areas on the strength of the coefficient. It is not yet certain just how sensitive the statistic is to various forms of w_{ij} and as such it would seem that for most practical situations the evaluation of the importance of the spatial component in any distribution can be most simply effected by using a regular matrix, which automatically standardizes the factors highlighted by Cliff and Ord.

This study has clearly shown that the usefulness of Cliff and Ord's technique is a direct function of the ease with which the results that are yielded by the method can be interpreted. The fact that the coefficient does vary according to certain factors makes it necessary for the researcher to interpret the coefficient in the light of characteristics prevalent for the particular situation with which he is concerned. Further, one of the means by which preliminary evaluation of the importance of space can be achieved is by calculating the coefficient of determination, which provides an indication of what extent of the variance in the data can be attributed to the spatial component. This method also provides a means by which the degree of spatial dependence in separate patterns can be compared.

Grahamstown clearly provides a good example of a practical situation, i.e. data regarding crime occurrence were discrete, data were not available by enumerator's districts or any other recognized system of sub-areas, and finally, data were distributed over an irregularly shaped surface, thus exposing many of the problems which could be encountered in a real world situation where the focus of interest is on spatial interaction. When applying Cliff and Ord's statistic to the crime data the following factors clearly influenced the strength of the coefficient, viz. the frequency of occurrence of the event, the grid square size, the number of grid squares in the matrix, grid squares which encompassed vacant ground, and finally, matrix orientation.

With regard to the above factors it was clear that significance was more easily reached when there were sufficient occurrences of the event as well as a large number of joins in the system. The choice of grid square size was an important consideration in that this determined the level of assumed interaction between the variate values, and it was clearly shown that assumptions regarding the level of interaction would have to be carefully considered on the basis of prior understanding of the nature of the data being analysed. Finally, the problem of vacant ground and matrix orientation clearly affected the extent to which directional biases and non-interaction areas on the surface could be isolated and examined.

In the light of the above problems regarding the application of Cliff and Ord's statistic two points emerge. Firstly, the need to understand the technique, viz., how it should be applied, how it functions, what is achieved in the process of analysis, and what the results mean in the context of the particular problem. Secondly, the need to understand something about the nature of the relationship that is being analysed. With regard to the technique Cliff and Ord (1969-1975) adequately explain how it functions and what is achieved in the process of analysis. What they fail to make explicit, and what is crucial to any practical application of the statistic, is how the technique can best be applied, i.e. for what situations is it most suitable as an analytical tool, and also, how one should go about interpreting the results in for example, a specifically geographic context. This study has highlighted a variety of the problems involved in the practical application and consequent interpretation of results of Cliff and Ord's technique by evaluating the degree of spatial dependence in patterns of crime occurrence in Grahamstown.

Further, this study has emphasized the need to have some understanding of the real world situation prior to analysis so that the results may be seen in the context of what actually exists and as such lessen the chances of misinterpreting the purely mathematical answers.

Finally therefore, if all the abovementioned limitations are considered, and the coefficients resulting from the technique are interpreted so as to allow a certain degree of comparison between results, Cliff and Ord's technique is extremely useful as an analytical tool which facilitates the isolation of the importance of space in any distribution of data on a two-dimensional surface. Following the method as outlined in this study, it is possible that geographer's with a limited knowledge of statistics and computing can apply the technique in any research where evaluation of the degree of spatial dependence is necessary.

REFERENCES

Ackerman, E. 1972 Where is a research frontier, in The Conceptual Revolution in Geography, W.K.D. Davies (ed.), London, Univ. of London Press, 264-281.

Bannister, G. 1975 Population Change in Southern Ontario. A.A.A.G., Vol. 65, 2, 177-188.

Bannister, G. 1976 Space-Time components of Urban Population Change, Economic Geography. Vol. 52, 228-240.

Beaujeu-Garnier, J. 1976 Methods and Perspectives in Geography. London, Longman.

Benson, M.A. 1965 Spurious Correlation in Hydraulics and Hydrology. Journal of Hydraulics Division, Proceedings of the American Society of Civil Engineers. Vol. 91. Part HY4. July 1965. Paper 4393. p. 35-42.

Berry, B.J.L. 1968 Approaches to Regional Analysis: A Synthesis, in Spatial Analysis, B.J.L. Berry & D.F. Marble (eds), Englewood Cliffs, Prentice-Hall.

Berry, B.J.L. 1971 Problems of Data Organization and Analytical Methods in Geography, Jnr. of Amer. Stat. Ass., Vol 66, 510-523.

Boggs, S.L. 1965 Urban Crime Patterns, Amer. Sociological Rev., Vol 30, 6, 899-908.

Boots, B.N. and Getis, A. 1977 Probability model approach to map pattern analysis. Progress in Human Geography Vol 1, 2, 264-286.

Brantingham, P.L. & Brantingham, P.J. 1974 Residential Burglary and Urban Form. Urban Studies, Vol 12, 3, 273-284.

Burton, I. 1972 The Quantitative Revolution and Theoretical Geography, in The Conceptual Revolution in Geography, W.K.D. Davies (ed.), London, Univ. of London Press, 140-156.

Carter, H. 1972 The Study of Urban Geography. London, Edward Arnold.

Chilton, R.J. 1964 Continuity in Delinquency area Research: A comparison of studies for Baltimore, Detroit and Indianapolis. American Sociological Review, Vol 29, 1, 71-83.

Chisholm, M. 1975 Human Geography: Evolution or Revolution. Harmondsworth, Penguin.

Chisholm, M., Frey, A.E. & Haggett, P.L. (eds) 1970 Regional Forecasting. London, Butterworth.

Chorley, R.J. & Haggett, P. 1965 Trend Surface mapping in Geographical Research, I.B.G.Trans., Vol 37, 47-67.

Clark, P.J. & Evans, F.C. 1954 Distance to nearest neighbour as a measure of spatial relationships in populations. Ecology, Vol. 35, 445-453.

Clark, P.J. & Evans, F.C. 1955 On some aspects of spatial patterns in Biological population, Science, Vol. 121, 397-398.

Cliff, A.D., Haggett, P., Ord, J.K., Bassett, K.A., Davies, R.B. 1975 Elements of Spatial Structure: A Quantitative Approach. Cambridge Univ.Press.

Cliff, A.D., Martin, R.L. & Ord, J.K. 1974 Evaluating the friction of distance Parameter in Gravity Models, Reg.Studs., Vol. 8, 281-286.

Cliff, A.D., Martin, R.L. and Ord, J.K. 1975 A test for Spatial Autocorrelation in Choropleth Maps based upon a modified χ^2 statistic. IBG Transactions No. 65, 109-129.

Cliff, A.D., Martin, R.L. and Ord, J.K. 1975 Map pattern and friction of distance parameters: Reply to comments by R.J. Johnston, and by L. Curry, D.A. Griffith and E.S. Sheppard, Reg. Studs., Vol. 9, 285-288.

Cliff, A.D., Martin, R.L. and Ord, J.K. 1976 A Reply to the Final Comment, Reg. Studs., Vol. 10, 3, 341-342.

Cliff, A.D. & Ord, J.K. 1969 The Problem of Spatial Autocorrelation, Studs in Reg.Sci., A.T. Scott (ed.), London, Pion.

Cliff, A.D. & Ord, J.K. 1971 Evaluating the percentage points of a Spatial Autocorrelation Coefficient, Geog.Analysis, Vol. 3, 1, 51-62.

Cliff, A.D. & Ord, J.K. 1972 Testing the Spatial Autocorrelation among Regression Residuals, Geog.Analysis, Vol. 4, 267-284.

Cliff, A.D. & Ord, J.K. 1973 Spatial Autocorrelation. London, Pion.

Cliff, A.D. & Ord, J.K. 1975a Model Building & the Analysis of Spatial Pattern in Human Geography, Journ.of the Royal Statistical Society, Vol. 37, 3, 297-348.

Cliff, A.D. & Ord, J.K. 1975b The choice of a test for Spatial Autocorrelation, Display & Analysis of Spatial Data, J. Davis & McCullach (eds), London, John Wiley.

Cliff, A.D. and Ord, J.K. 1977 Large sample size distribution of Statistics used in testing for spatial correlation: A comment. Geographical Analysis, Vol. 9, 3, 297-300.

Cohen, A.K. 1965 The Sociology of the Deviant Act: Anomic Theory & Beyond, Amer.Sociological Rev., Vol. 30, 1, 5-14.

Cole, J.P. 1973 Mathematics and Geography, in Perspectives in Geog. Ed., J. Bale, N. Graves, R. Walford (eds), Edinburgh, Oliver & Boyd.

Corsi, T.M. & Harvey, M.E. 1975 The Socio Economic Determinants of Crime in the City of Cleveland, T.E.S.G., Vol. 66,

Curry, L., Griffith, D.A., Sheppard, E.S. 1975 Those Gravity Parameters again, Reg. Studs., Vol. 9, 289-296.

Dacey, M.F. 1968 A Review on Measures of Contiguity for two and K-Colour maps, in Spatial Analysis, B.J.L. Berry & D.F. Marble (eds), Englewood Cliffs, Prentice-Hall.

Daniel, J.B. McI. (ed.) 1974 Grahamstown and its environment. Grahamstown, I.S.E.R. (Rhodes University)

Davies, R.J. and Cook, G.P. 1968 Reappraisal of the South African Urban hierarchy. South African Geographical Journal, Vol. 50, 2, 116-120.

Davies, R.J. and Young, B.S. 1969 The Economic Structure of South African cities. South African Geographical Journal, Vol. 51, 19-37.

Davis, J.C. 1973 Statistics and Data Analysis in Geology. New York, John Wiley.

De Groot, M. 1975 Probability and Statistics. Addison-Wesley, Reading (Mass.).

De Vos, S. 1973 The Use of Nearest Neighbour Methods. T.E.S.G.,Vol, 64, 5, 307-319.

Douglas, J.W.B., Ross, J.M., Hammond, W.A., & Mulligan, D.G. 1966 Deliquency & Social Class. British Journ. of Criminology, Vol. 6, 3, 294-302.

Ebdon, D. 1976 On the underestimation inherent in the commonly-used formulae. Area, Vol. 8, 3, 165-169.

Echenique, M. 1972 Models; a discussion, in Urban Space and Structures, L. Martin and L. March (eds), Cambridge, Cambridge Univ.Press.

Ezekiel, M. & Fox, K.A. 1959 Methods of Correlation & Regression Analysis. New York, John Wiley.

Fielding, E.V.C. 1975 The design & implementation of a computer package to perform cluster analysis. Unpublished Honours thesis, Rhodes Univ., Grahamstown.

Geary, R.C. 1968 The Contiguity Ratio & Statistical Mapping, in Spatial Analysis, B.J.L. Berry & D.F. Marble (eds), Englewood Cliffs, Prentice Hall.

Getis, A. 1964 Temporal land-use pattern analysis with the use of nearest neighbour and quadrat methods. A.A.A.G., Vol. 54, 3, 391-399.

Getis, A. 1977 On the use of the term "random" in Spatial Analysis. The Professional Geographer, Vol. 29, 1, 59-61.

Gibbs, J.P. 1966. Crime, unemployment and status integration. The British Journal of Criminology, Vol. 6, 1, 49-58.

Gould, P.R. 1973 The New Geography, in Perspectives in Geographic Education, J. Bale, N. Graves, R. Walford (eds), Edinburgh, Oliver & Boyd.

Gregory, S. 1963 Statistical Methods & the Geographer. London, Longman.

Haber, A. & Runyan, R.P. 1973 General Statistics: Second Edition. Reading, Assison-Wesley.

Harnett, D.L. and Murphy, J.L. 1975 Introductory Statistical Analysis, Reading (Mass.), Addison-Wesley.

Harries, K.D. 1971 The Geography of American Crime, 1968, Journ. of Geog., Vol. 70, 204-213.

Harries, K.D. 1973 Spatial Aspects of Violence and Metropolitan Population, The Professional Geographer, Vol. 25, 1, 1-6.

Harries, K.D. 1974 The Geography of Crime and Justice, New York, McGraw-Hill.

Harvey, D.W. 1966 Geographical processes & the analysis of point patterns, I.B.G.Trans., Vol. 40, 81-95.

Harvey, D.W. 1968 Pattern, Process & the Scale Problem, in Geog.Research, I.B.G. Trans., Vol. 45, 71-78.

Harvey, D.W. 1973 Explanation in Geography, London, Edward Arnold.

Herbert, D.T. 1977 An Areal and Ecological Analysis of Delinquency Residence, Cardiff 1966 & 1971, T.E.S.G., Vol. 68, 2, 83-99.

Hudson, J.C. & Fowler, P.M. 1976 The concept of pattern in Geography. Discussion paper No. 1. Discussion Paper Series - Dept. of Geography, University of Iowa. (Iowa City, Iowa).

Jeffery, C.R. 1971 Crime Prevention through Environmental Design, London, Sage Publications.

Jonassen, C.T. 1949 A Re-evaluation & critique of the logic and some methods of Shaw, McKay, Amer.Sociological Rev., Vol. 14, 5, 608-617.

King, L.J. 1969. A Quantitative expression of the pattern of urban settlements in selected areas of the United States, Analytical Human Geog., P.J. Ambrose (ed.), London, Longman.

King, L.J. 1969 The Analysis of Spatial Form and its relation to Geographic theory, A.A.A.G., Vol. 59, 573-595.

King, L.J. 1969 Statistical Analysis in Geog., Englewood Cliffs, Prentice-Hall.

Klein, M.W. (ed.) 1967 Juvenile Gangs in Context, Englewood Cliffs, Prentice Hall.

Lee, Y. & Egan, F.J. 1972 The Geography of Urban Crime: The Spatial Pattern of Serious crime in the City of Denver, Proceedings A.A.G., Vol. 4, 59, 64.

Lötter, J.M. 1975 The Ecology of Crime in a S.A. City, in Crime & Punishment in S.A., J. Midgley, J.H. Steyn, R. Graser (Eds), Isando, McGraw-Hill.

McDougall, E.B. 1976 Computer Programming for Spatial Problems, London, Edward Arnold.

March, L. 1974 Quantifying the environment, in Models, Evaluations and Information Systems for Planners, J. Perraton and R. Baxter (eds), LUBFS Conference Proceedings No.1. University of Cambridge, Department of Architecture.

March, L. 1975. Models, Modes and Mores, in Urban Development Models, R. Baxter, M. Echenique and J. Owen (eds), LUBFS Conference Proceedings No. 3. Univ. of Cambridge, Department of Architecture. (Land use and Built form studies.)

Marchand, B. 1974 Quantitative Geography: Revolution or Counter-Revolution. Geoforum, Vol. 17, pp.15-23.

Mather, P.M. 1976 Computational methods of multivariate Analysis in Physical Geography, London, John Wiley.

Mather, P. & Openshaw, S. 1974 Multivariate methods and Geographical Data, The Statistician, Vol. 23, 3/4, 283-308.

Matravers, D. 1977 Interaction models: Are they gravity or entropy models? Unpublished paper, Dept. of Applied Maths, Univ. of Cape Town.

Mays, J.B. 1956 Growing up in the City, Liverpool, Liverpool Univ.Press.

Mays, J.B. 1968 Crime and the Urban Pattern, Sociological Rev., Vol, 16, 2, 241-255.

McCord, W., McCord, J. & Zola, 1959 Origins of Crime, New York, Columbia Univ. Press.

Merton, R.K. 1938 Social Structure and Anomie, Amer.Sociological Rev., Vol. 3, 5, 672-682.

Midgley, J., Steyn, J.H. and Graser, R. (eds) 1975 Crime and Punishment in South Africa. Isando, McGraw-Hill.

Midgley, J. 1975 Crime and Normlessness: Anomie in an Urban South African Community, in Crime and Punishment in South Africa, J. Midgley, J.H. Steyn, R. Graser, (eds), Isando, McGraw-Hill.

Moran, P.A.P. 1950 a Notes on Continuous Stochastic Phenomena, Biometrika, Vol. 37, 17-23.

Moran, P.A.P. 1950 b A test for serial independence of residuals, Biometrika, Vol. 37, 178-181.

Moses, E.R. 1947 Differentials in Crime Rates between Negroes and Whites, based on comparisons of four socio-economically equated areas, Amer.Sociological Rev., Vol. 12, 4, 411-420.

Moss, R.P. 1970 Authority and Charisma: Criteria of Validity in Geographical Methods, S.A.G.J., Vol. 52, 13-37.

Mosteller, F., Rourke, R.E.K. & Thomas, G.B. 1970. Probability with statistical Applications, Reading (Mass.), Addison-Wesley.

Nystuen, J.D. 1963 Identification of some fundamental spatial aspects, Papers of the Michigan Academy of Science, Arts, Letters, Vol. XLVIII (48), Univ. of Michigan.

Olson, J.M. 1975 Autocorrelation and Visual Map complexity, A.A.A.G., Vol. 65, 2, 189-204.

Phillips, P.D. 1972 A prologue to the Geography of Crime, Proceedings A.A.G., Vol. 4, 86-91.

Pinder, D.A. & Witherick, M.E. 1972 The principles, practice and pitfalls of nearest neighbour analysis, Geography, Vol. 57, 277-288.

Pinder, D.A. & Witherick, M.E. 1973 Nearest Neighbour Analysis of linear Point Patterns, T.E.S.G., Vol. 64, 3, 160-163.

Pyle, G.F. et al. 1974 The Spatial Dynamics of Crime, Chicago, Univ. of Chicago.

Retief, G.M. 1975 Social disorganization, crime and the urban Bantu people of South Africa, in Crime and Punishment in South Africa, J. Midgley, J.H. Steyn, R. Graser (eds), Isando, McGraw-Hill.

Robinson, A.H. 1956 The necessity of weighting values in correlation analysis of Areal Data, A.A.A.G., Vol. 46, 233-236.

Rogers, A. 1974 Statistical Analysis of Spatial Dispersion: The Quadrat Method, London, Pion.

Schmid, C.F. 1960 Urban Crime Areas: Part 1, Amer.Soc.Rev., Vol. 25, 4, 527-542.

Schmid, C.F. 1960 Urban Crime Areas: Part II, Amer.Soc.Rev., Vol. 25, 5, 655-678.

Sibley, D. 1976 On pattern and dispersion. AREA, Vol. 8, 3, 163-165.

Sutherland, E.H. 1939 Principles of Criminology, New York, Lippincott.

Vincent, P. 1976 The general case: how not to measure spatial point patterns. AREA, Vol. 8, 3, 161-163.

Ward, G.J. 1976 Patterns of selected crimes in Grahamstown. Unpublished study, Dept. of Geography, Rhodes University, Grahamstown.

Welch, S. 1976 Spatial Patterning of South African Towns: A Preliminary Study, Unpublished paper, Dept. of Geography, Rhodes University, Grahamstown.

Whittle, P. 1954 On Stationery Processes in the Plane, Biometrika, Vol. 41, 434-449.

Wrigley, N. 1977 Probability surface mapping: a new approach to trend surface mapping. Transactions IBG, Vol. 2, 2, 129-140.

Yeates, M. 1974 An Introduction to Quantitative Analysis in Human Geography. New York, McGraw-Hill.

APPENDIX

THE COMPUTER PROGRAM

APPENDIX

The Computer Program

The program is written in standard Fortran IV for the ICL 1900 T computer and the stages of development are explained below in detail and as simply as possible to enable geographers with a limited knowledge of computing to follow each step in the design of the final program. The program itself is, by design, set out in segments which correspond to various procedural requirements for spatial autocorrelation. No attempt has been made to shorten the program as this could lead to confusion in the case of geographers with little knowledge of Fortran IV and further, compaction of the program would increase the already sizeable core requirements.

The program consists of two parts, viz. the first part being designed to evaluate spatial autocorrelation according to Cliff and Ord's method, and the second part being designed to evaluate directional autocorrelation according to Pearson's method. A few points regarding the sequential logic are outlined below, and a complete version of the program follows this text.

The equation used in the evaluation of spatial autocorrelation according to Cliff and Ord's method is equation 6 , where w_{ij} takes the form of a system of symmetric binary weights. In other words, w_{ij} refers to a simple code of weights where a join between two sub-areas is given either a value of 1 or zero. For example, referring to Figure 2 the joins between x and I_1, I_2, I_4, I_6, and I_7 would be 'one' while all other adjacent joins, viz. I_3, I_5, and I_8 would be zero. Each join is given the weight 'one' in both directions, i.e. if $x \longrightarrow I_1 = 1$, then $I_1 \longrightarrow x = 1$. The binary weight is thus symmetric. Further, in order to compute Z_i Z_j it is necessary to calculate the mean value of all the x_i values in the data matrix.

Format statements 1-11 refer to the READ statements which collect the x_i values, in matrix form, into the computer memory. DO LOOP 20 and DO LOOP 30 affect the calculation of ISUM (sum of all the x_i values) and ZDIF ($x_i - \bar{x}$). It is important to note that $Z_i = x_i - \bar{x}$, which is the ZDIF value of the current grid square. The adjacent grid squares to grid square i are each labelled j. Therefore Z_j refers to the $x_i - \bar{x}$ values in each of the adjacent grid squares respectively. Four possible values of ZDIF exist.

(a) ZDIF which is set to Ic which is given the value zero.

(b) A naturally occurring ZDIF score which can be any real number (However, if ZDIF = 0.0 then it is set to 999.0.

(c) A ZDIF score of +999.0.

(d) A ZDIF score of -999.0.

Adjacent squares are isolated depending on the direction in which autocorrelation is being evaluated. If for example, the direction is North/South then only grid squares I_2 and I_7 are considered as lying adjacent to the current grid square, and so on. Because of the fact that dependence is seen to extend Northwards and Southwards of the current grid square or, as in the case of the final run of the program where dependence extends in <u>all</u> directions, the value in any grid square on the matrix can serve as both an x_i value and an x_j value. This is the major difference between autocorrelation in a time series and autocorrelation in a plane, where dependence in the former case only extends in a forward direction. This 'inversion' of the variate values affects both the nature of the coefficient and the test for significance.

Within DO LOOP 50 and between statement numbers 67 and 608 the sum of Z_i Z_j for the entire matrix is calculated and the total number of joins in the system is evaluated. Using the section between statement numbers 67 and 611 as an example, the sequence of calculation is as follows. Firstly the values of ZDIF are stored in the ZZ memory for the next directional run. Secondly, if ZDIF is equal to -999.0, it is set to zero. This is because we do not wish to include -999.0 in the calculation of Z_i Z_j. Thirdly, if ZDIF is equal to +999.0 (in this indicating a natural occurrence of ZDIF = 0) then the join between Z_i and Z_j is counted but the calculation of Z_i Z_j remains at zero. Fourthly, if ZDIF is not zero (by natural means) and neither is it set to zero the T_i is calculated as the ZDIF score of the current grid square multiplied by the ZDIF score of the adjacent grid square, which could be any ZDIF value coded $I_1 \ldots I_8$. The number of joins (or the number of Z_i and Z_j pairs) for each loop under DO LOOP 50 is summed under ww. If $T_1 - T_8$ is a non-zero figure then a join is registered and counted. Further, a join is also counted under $S_1 - S_8$ for all those joins where ZDIF = 0.0 by natural means. These joins are added to ww, which are finally added to WAIT, which is the counter for the total number of joins in the matrix. Following DO LOOP 50 all the information necessary for the calculation of Cliff and Ord's coefficient is available (refer Statement number 56).

Following this, the expected values $E(r)$ and $E(r^2)$ are calculated under both the assumptions of normality and randomness (refer equations 8 , 9 and 10), and finally, the observed statistic is standardized according to the format shown in equation 7 . (Note; the absolute value of the standardized normal variate is calculated).

In the program FRED and FINA refer to the standardized variate following the assumptions of randomness and normality respectively. Calculation of FRED and FINA are subdivided into a series of subsections which facilitates checking of results when handling theoretical matrices.

Pearson's method.

After the results for computation and significance testing of Cliff and Ord's method have been written the program isolates pairs of x_i and x_j values for the calculation of Pearson's product moment correlation coefficient. For the fifth run all the values used in the previous runs of the program are used, i.e. all the x_i and x_j values isolated thusfar are set out in two single arrays and used in the formula, which is set out in subroutine PROD. Finally, significance is evaluated according to the student's - t test.

```
---------------------------------------------------------------------------------
        JOB REAL,GGGW,WARD                               RUN BY GEORGE 2/MK9F ON
        DISPLA 'GGGW      '
        MAXTIME 500
        VOLUME 3000
        FORTRANCOMP REAL
        EXECUTE REAL
        ****

FORTRAN COMPILATION BY #XFAE MK 4D    DATE  08/12/77  TIME  11/49/51


      LIST
      PROGRAM(REAL)
      INPUT 5=CR0
      OUTPUT 6=LP0
      TRACE 2
      END

      MASTER REAL
C     ##################################################################
C     THIS PROGRAM CALCULATES THE DEGREE OF PAIRWISE SPATIAL
C     AUTOCORRELATION FOR VARIOUS CARDINAL DIRECTIONS ACROSS A SURFACE
C     THE PROGRAM CALCULATES A CORRELATION COEFFICIENT FIRSTLY ACCORDING
C     TO CLIFF AND ORD'S METHOD AND SECONDLY ACCORDING TO PEARSON'S
C     PRODUCT MOMENT METHOD. IN THE FORMER CASE SPATIAL DEPENDENCE IN
C     ALL DIRECTIONS IS CONSIDERED WHEREAS IN THE LATTER IT IS A SINGLE
C     DIRECTIONAL DEPENDENCE WHICH IS OF IMPORTANCE.
C     ##################################################################
      DIMENSION IX(11,16),ZDIF(177),TOT(176),WW(176),DEE(176),ISEL(5)
      DIMENSION KX(625),KY(625),JX(625),JY(625)
C     ******************************************************************
C     IX REFERS TO THE OBSERVED VALUES
C     ZDIF REFERS TO THE DIFFERENCE BETWEEN THE OBSERVED AND THE MEAN
C     TOT REFERS TO THE PRODUCT OF CURRENT ZDIF AND ADJACENT ZDIF
C     WW REFERS TO THE NUMBER OF "JOINS" IN THE SYSTEM
C     DEE REFERS TO THE CALCULATION OF D, USED IN THE TEST FOR SIGNIFICANCE
C     DIMENSIONS KX, KY, JX, JY, = THE TOTAL NUMBER OF POSSIBLE
C     X AND Y PAIRS IN THE SYSTEM.  CALCULATED AS FOLLOWS:
C     (IROW*(ICOL-1))+((IROW-1)*ICOL)+(((IROW-1)*ICOL-1))*2)
C     DATA IS PUNCHED BY ROW
C     THIS SEGMENT READS THE DATA
C     READ THE NUMBER OF ROWS,COLUMNS AND DIRECTIONAL POSSIBILITIES
C     ANY COMBINATION OF 5 DIRECTIONS MAY BE CHOSEN WHERE NUMM=NUMBER
C     OF DIRECTIONS REQUIRED AND ISEL DENOTES THE DIRECTIONS
C     THE NUMM VALUES OF ISEL ARE GIVEN BY:
C     ISEL(1)= A NORTH / SOUTH DIRECTION
C          (2)= A WEST / EAST DIRECTION
C          (3)= A NORTHWEST / SOUTHEAST DIRECTION
C          (4)= A NORTHEAST / SOUTHWEST DIRECTION
C          (5)= ALL DIRECTIONS, IE. THE SURFACE AS A WHOLE.
C     ******************************************************************
      READ (5,1) IROW,ICOL,NUMM
    1 FORMAT (3I5)
      READ (5,2) (ISEL(J),J=1,NUMM)
    2 FORMAT (5I5)
      DO 10 I=1,IROW
   10 READ (5,11) (IX(I,J),J=1,ICOL)
   11 FORMAT (16I3)
C     ------------------------------------------------------------------
C     THIS SEGMENT CALCULATES THE MEAN VALUE
C     USE ISUM AS AN ACCUMULATOR.  NNNN IS A COUNTER WHICH COUNTS ALL
C     THOSE OBSERVATIONS THAT HAVE A VARIATE VALUE GREATER THAN ZERO.
C     A VALUE OF ZERO IN ANY GRID SQUARE INDICATES THAT THAT PARTICULAR
C     GRID SQUARE IS NOT INCLUDED IN CALCULATION OF THE COEFFICIENT.
C     NNNN, ICON, AND ISUM ARE SET TO ZERO.
C     ICON IS A COUNTER WHICH IS USED IN THE CALCULATION OF PEARSON'S R
C     MM IS A DIRECTIONAL COUNTER. SPATIAL AUTOCORRELATION IS
C     CALCULATED 5 TIMES FOR BOTH CLIFF AND ORD'S R AND PEARSON'S R
C     ------------------------------------------------------------------
      MM=1
      ICON=0
   13 ISUM=0
      NNNN=0
C     ------------------------------------------------------------------
C     SUM ALL THE VALUES GREATER THAN ZERO.
C     ------------------------------------------------------------------
      DO 20 J=1,ICOL
      DO 20 I=1,IROW
```

```
      IF(IX(I,J).EQ.0) GO TO 20
      ISUM=ISUM+IX(I,J)
      NNNN=NNNN+1
   20 CONTINUE
      MSUM=ISUM-NNNN
C     ------------------------------------------------------------------
C     DIVIDE ISUM BY THE NUMBER OF ELEMENTS IN THE MATRIX GREATER THAN
C     ZERO, VIZ. NNNN, TO GET MEAN.
C     ------------------------------------------------------------------
      AVER=FLOAT(ISUM)/FLOAT(NNNN)
C     ------------------------------------------------------------------
C     THE RESULTS FOR COMPUTATION OF AUTOCORRELATION IN THE VARIOUS
C     DIRECTIONS FOR BOTH CLIFF AND ORD'S COEFFICIENT AND PEARSON'S
C     COEFFICIENT ARE PRINTED WITH TITLES.
C     ------------------------------------------------------------------
      IF(ISEL(MM).GT.1) GO TO 211
      WRITE (6,22)
   22 FORMAT (1H1,1X,45HTHIS IS THE FIRST RUN. DIRECTION NORTH/SOUTH ,/,
     12X,44(1H-),//)
      GO TO 27
  211 IF(ISEL(MM).GT.2) GO TO 212
      WRITE (6,23)
   23 FORMAT (1H1,1X,44HTHIS IS THE SECOND RUN. DIRECTION WEST/EAST ,/,2
     1X,43(1H-),//)
      GO TO 27
  212 IF(ISEL(MM).GT.3) GO TO 213
      WRITE (6,24)
   24 FORMAT (1H1,1X,53HTHIS IS THE THIRD RUN. DIRECTION NORTHWEST/SOUTH
     1EAST ,/,2X,52(1H-),//)
      GO TO 27
  213 IF(ISEL(MM).GT.4) GO TO 214
      WRITE (6,25)
   25 FORMAT (1H1,1X,54HTHIS IS THE FOURTH RUN. DIRECTION NORTHEAST/SOUT
     1HWEST ,/,2X,53(1H-),//)
      GO TO 27
  214 WRITE (6,26)
   26 FORMAT (1H1,1X,52HTHIS IS THE FIFTH RUN. WHOLE SURFACE,ALL DIRECTI
     1ONS ,/,2X,51(1H-),//)
   27 WRITE (6,28)
   28 FORMAT (1H0,27H(A) CLIFF AND ORD'S METHOD ,/,1X,26(1H-),//)
      WRITE (6,14)
   14 FORMAT (1H0,3X,5HISUM ,5X,5HMEAN ,5X,15HCORRECTED ISUM ,/)
      WRITE (6,15) ISUM,AVER,MSUM
   15 FORMAT (2X,I6,2X,F8.2,12X,I8,//)
C     ------------------------------------------------------------------
C     THIS SEGMENT REFERS TO THE CLIFF AND ORD METHOD ONLY,
C     AND CALCULATES THE DIFFERENCE BETWEEN ALL THE
C     OBSERVED VALUES IX AND THE MEAN, THE ANSWER IN EACH CASE
C     BEING LABELLED ZDIF.  IF THE OBSERVED VALUE IX IS ZERO,
C     THEN THE ZDIF VALUE FOR THAT GRID SQUARE IS SET TO -999.0 FOR
C     IDENTIFICATION PURPOSES.  IF THE ZDIF VALUE IS CALCULATED
C     AND THE RESULT IS ZERO, IE. ZDIF IS NATURALLY ZERO BECAUSE
C     IX=MEAN, THEN ZDIF IS SET TO +999.0, ALSO FOR IDENTIFICATION
C     PURPOSES.  A COUNTER K IS INTRODUCED FOR THE CALCULATION OF ZDIF.
C     ------------------------------------------------------------------
      K=0
      SUMZ=0.0
      SUMB=0.0
      DO 30 I=1,IROW
      DO 30 J=1,ICOL
      IF(IX(I,J).GT.0) GO TO 33
      K=K+1
      ZDIF(K)=-999.0
      GO TO 34
   33 K=K+1
      ZDIF(K)=FLOAT(IX(I,J))-AVER
   34 IF(ZDIF(K).NE.-999.0) SUMZ=SUMZ+ZDIF(K)*ZDIF(K)
      IF(ZDIF(K).NE.-999.0) SUMB=SUMB+ZDIF(K)*ZDIF(K)*ZDIF(K)*ZDIF(K)
      IF(ZDIF(K).EQ.0.0) ZDIF(K)=999.0
   30 CONTINUE
C     ------------------------------------------------------------------
C     SUMB IS THE SUM OF ZDIF TO THE POWER FOUR. USED IN CALCULATION OF
C     SKEWNESS
C     SUMZ IS THE TOTAL OF (IX-XMEAN) SQUARED.--REFER VARIANCE.
C     PRINT THE RESULTS WITH TITLE
C     ------------------------------------------------------------------
      WRITE (6,31)
   31 FORMAT (1H0,20HTHE ZDIF SCORES ARE ,/)
      WRITE (6,32)(ZDIF(N),N=1,K)
   32 FORMAT (16(1X,F6.1),//)
C     ------------------------------------------------------------------
```

```
C     THIS SEGMENT CREATES A GRID SQUARE OUTSIDE THE LIMITS OFTHE MATRIX
C     NEND.  REFER COMMENT STATEMENT BELOW
C     THE TOTAL NUMBER OF GRID SQUARES IN THE MATRIX IS INDICATED BY
C     NEND, IE. IROW*ICOL.
C     ------------------------------------------------------------------
      NEND=IROW*ICOL
      IC=NEND+1
      ZDIF(IC)=0.0
C     ------------------------------------------------------------------
C     THIS SEGMENT ISOLATES ALL THE ADJACENT GRID SQUARES TO EVERY SINGLE
C     GRID SQUARE IN THE MATRIX. EVERY GRID SQUARE HAS 8 ADJACENT SQUARES
C     WHICH ARE CODED I1 TO I8. ZDIF SCORES IN ADJACENT GRID SQUARES ARE
C     MULTIPLIED.
C     ALL RESULTING VALUES ARE SUMMED.
C     WAIT IS THE TOTAL NUMBER OF "JOINS" IN THE SYSTEM
C     DED IS A COUNTER FOR D,  USED IN SIGNIFICANCE.
C     BOTH ARE SET TO ZERO.  FOR EXPLANATION REFER BELOW.
C     TOTAL IS SET TO ZERO. FOR EXPLANATION OF TOTAL REFER BELOW.
C     ------------------------------------------------------------------
      WAIT=0.0
      DED=0.0
      TOTAL=0.0
      DO 50 J=1,NEND
      I1=(J-ICOL)-1
      I8=(J+ICOL)+1
      I3=(J-ICOL)+1
      I6=(J+ICOL)-1
      I2=J-ICOL
      I7=J+ICOL
      I4=J-1
      I5=J+1
C     ------------------------------------------------------------------
C     IF I1 TO I8 FALL OUTSIDE THE LIMITS OF MATRIX NEND THEY ARE ROUTED
C     TO IC , WHICH READS 0.0
C     ------------------------------------------------------------------
      IF(J.GT.ICOL) GO TO 60
      I1=IC
      I2=IC
      I3=IC
   60 NB=NEND-ICOL
      IF(J.LE.NB) GO TO 61
      I6=IC
      I7=IC
      I8=IC
   61 IF(J-ICOL*(J/ICOL).NE.0) GO TO 62
      I3=IC
      I5=IC
      I8=IC
   62 NL=(J/ICOL)*ICOL+1
      IF(J.NE.NL) GO TO 63
      I1=IC
      I4=IC
      I6=IC
C     ------------------------------------------------------------------
C     THE DIRECTIONAL LIMITS ARE NOW SET
C     FOR THE FIRST RUN N/S, FOR THE SECOND RUN E/W,FOR THE THIRD AND
C     FOURTH RUNS BOTH DIAGONALS, AND FOR THE FIFTH RUN THE WHOLE SURFACE.
C     I1 REFERS TO THE GRID SQUARE NORTHWEST OF THE CURRENT GRID
C     SQUARE.  FURTHER, I2=N, I3=NE, I4=W, I5=E, I6=SW, I7=S, I8=SE,
C     OF THE CURRENT GRID SQUARE.
C     ------------------------------------------------------------------
   63 IF(ISEL(MM).GT.1) GO TO 64
      I1=IC
      I4=IC
      I6=IC
      I3=IC
      I5=IC
      I8=IC
      GO TO 67
   64 IF(ISEL(MM).GT.2) GO TO 65
      I1=IC
      I2=IC
      I3=IC
      I6=IC
      I7=IC
      I8=IC
      GO TO 67
   65 IF(ISEL(MM).GT.3) GO TO 66
      I2=IC
      I3=IC
      I4=IC
```

```
      I5=IC
      I6=IC
      I7=IC
      GO TO 67
   66 IF(ISEL(MM).GT.4) GO TO 67
      I1=IC
      I2=IC
      I4=IC
      I5=IC
      I7=IC
      I8=IC
C     ----------------------------------------------------------------------
C     T1 IS THE PRODUCT OF THE CURRENT GRID SQUARE, ZDIF(J), MULTIPLIED
C     BY ADJACENT ZDIF(I1), ZDIF(I2), ZDIF(I3),... ZDIF(I8).
C     TOT(J) IS THE SUM OF THESE 8 PRODUCTS. THERE WILL BE
C     NNNN   TOT(J) VALUES.
C     TOTAL IS THE SUM OF ALL THE TOT(J) VALUES
C     THE RESULTS ARE PRINTED WITH TITLES  TOT VALUES, AND  TOTAL.
C     BETWEEN STATEMENT NUMBERS 67 AND 626 GRID SQUARES HAVING A
C     NATURAL VALUE OF ZERO, AND THOSE WHICH WERE SET TO ZERO ARE
C     DIFFERENTIATED. JOINS BETWEEN ADJACENT GRID SQUARES ARE ALSO NOTED.
C     ----------------------------------------------------------------------
   67 ZZ=ZDIF(J)
      IF(ZDIF(J).EQ.-999.0) ZDIF(J)=0.0
      T1=0.0
      S1=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I1).NE.999.0) GO TO 611
      IF(ZDIF(I1).NE.0.0) S1=1.0
      GO TO 612
  611 IF(ZDIF(I1).NE.-999.0) T1=ZDIF(J)*ZDIF(I1)
  612 T2=0.0
      S2=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I2).NE.999.0) GO TO 613
      IF(ZDIF(I2).NE.0.0) S2=1.0
      GO TO 614
  613 IF(ZDIF(I2).NE.-999.0) T2=ZDIF(J)*ZDIF(I2)
  614 T3=0.0
      S3=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I3).NE.999.0) GO TO 615
      IF(ZDIF(I3).NE.0.0) S3=1.0
      GO TO 616
  615 IF(ZDIF(I3).NE.-999.0) T3=ZDIF(J)*ZDIF(I3)
  616 T4=0.0
      S4=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I4).NE.999.0) GO TO 617
      IF(ZDIF(I4).NE.0.0) S4=1.0
      GO TO 618
  617 IF(ZDIF(I4).NE.-999.0) T4=ZDIF(J)*ZDIF(I4)
  618 T5=0.0
      S5=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I5).NE.999.0) GO TO 619
      IF(ZDIF(I5).NE.0.0) S5=1.0
      GO TO 620
  619 IF(ZDIF(I5).NE.-999.0) T5=ZDIF(J)*ZDIF(I5)
  620 T6=0.0
      S6=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I6).NE.999.0) GO TO 621
      IF(ZDIF(I6).NE.0.0) S6=1.0
      GO TO 622
  621 IF(ZDIF(I6).NE.-999.0) T6=ZDIF(J)*ZDIF(I6)
  622 T7=0.0
      S7=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I7).NE.999.0) GO TO 623
      IF(ZDIF(I7).NE.0.0) S7=1.0
      GO TO 624
  623 IF(ZDIF(I7).NE.-999.0) T7=ZDIF(J)*ZDIF(I7)
  624 T8=0.0
      S8=0.0
      IF(ZDIF(J).NE.999.0.AND.ZDIF(I8).NE.999.0) GO TO 625
      IF(ZDIF(I8).NE.0.0) S8=1.0
      GO TO 626
  625 IF(ZDIF(I8).NE.-999.0) T8=ZDIF(J)*ZDIF(I8)
  626 IF(ZDIF(J).EQ.0.0) ZDIF(J)=ZZ
      TOT(J)=T1+T2+T3+T4+T5+T6+T7+T8
      TOTAL=TOTAL+TOT(J)
C     ----------------------------------------------------------------------
C     FOR PURPOSES OF CALCULATION INSIDE LOOP NUMBER 50 WAIT=WW
C     CLEARLY WW WILL CHANGE ACCORDING TO DIRECTION
C     WW=WAIT ONCE AGAIN OUTSIDE OF THE LOOP
C     WW(J) IS A COUNTER WHICH IS SET TO ZERO.
C     IF EITHER THE CURRENT GRID SQUARE OR AN ADJACENT GRID SQUARE
```

```
C      REFLECTS A PRESET  VALUE OF ZERO THEN THERE IS NO "JOIN" AND THAT
C      PARTICULAR X AND Y PAIR IS NOT INCLUDED IN CALCULATION OF THE
C      COEFFICIENT.  DEE(J) IS A COUNTER FOR D.
C      ------------------------------------------------------------------
       WW(J)=S1+S2+S3+S4+S5+S6+S7+S8
       IF(T1.EQ.0.0) GO TO 601
       WW(J)=WW(J)+1.0
  601  IF(T2.EQ.0.0) GO TO 602
       WW(J)=WW(J)+1.0
  602  IF(T3.EQ.0.0) GO TO 603
       WW(J)=WW(J)+1.0
  603  IF(T4.EQ.0.0) GO TO 604
       WW(J)=WW(J)+1.0
  604  IF(T5.EQ.0.0) GO TO 605
       WW(J)=WW(J)+1.0
  605  IF(T6.EQ.0.0) GO TO 606
       WW(J)=WW(J)+1.0
  606  IF(T7.EQ.0.0) GO TO 607
       WW(J)=WW(J)+1.0
  607  IF(T8.EQ.0.0) GO TO 608
       WW(J)=WW(J)+1.0
  608  WAIT=WAIT+WW(J)
       DEE(J)=WW(J)*(WW(J)-1.0)
       DED=DED+DEE(J)
       D=DED/2.0
   50  CONTINUE
       WRITE (6,51)
   51  FORMAT (1H0,19HTHE TOT VALUES ARE ,/)
       WRITE (6,32)(TOT(K),K=1,NEND)
       WRITE (6,52) TOTAL
   52  FORMAT (1H0,6HTOTAL ,F12.4,//)
C      ------------------------------------------------------------------
C      THIS SEGMENT CALCULATES THE AUTOCORRELATION COEFFICIENT ACCORDING
C      TO CLIFF AND ORD'S METHOD.
C      IF THE VARIANCE, SUMZ, IS ZERO, CORRELATION IS SET TO ZERO
C      THE VALUES FOR NEND, WAIT, SUMZ, SUMB, AND CORRELATION ARE
C      PRINTED WITH TITLES
C      ------------------------------------------------------------------
       IF(SUMZ.NE.0.0) GO TO 56
       CORR=0.0
       GO TO 57
   56  CORR=(FLOAT(NNNN)*TOTAL)/(WAIT*SUMZ)
   57  WRITE (6,70)
   70  FORMAT (2X,5HNNNN ,2X,5HNEND ,9X,5HWAIT ,15X,5HSUMZ ,15X,5HSUMB ,/
      1)
       WRITE (6,71) NNNN,NEND,WAIT,SUMZ,SUMB
   71  FORMAT (1X,I5,2X,I5,2X,F12.2,2(2X,F18.2),//)
       WRITE (6,72) CORR
   72  FORMAT (1H0,31HTHE CORRELATION COEFFICIENT IS ,F12.6,/,1X,30(1H-),
      1//)
C      ------------------------------------------------------------------
C      THIS SEGMENT CALCULATES THE SIGNIFICANCE OF CLIFF AND ORD'S
C      COEFFICIENT.
C      THERE ARE NO FIXED LIMITS FOR THIS COEFFICIENT AND THEREFORE
C      SIGNIFICANCE IS EVALUATED FROM THE VALUE FOR R, PLUS THE MEAN
C      AND VARIANCE OF THE DATA.   H1 IS ACCEPTED ONLY IF HO (HYPOTHESIS
C      OF NO CORRELATION) IS REJECTED AT THE 95% LEVEL.   SIGNIFICANCE
C      IS EVALUATED ACCORDING TO BOTH THE ASSUMPTIONS OF A RANDOMLY
C      DISTRIBUTED XI AND A NORMALLY DISTRIBUTED XI.
C      WHEN JOINS BETWEEN ADJACENT GRID SQUARES ARE NOT WEIGHTED, A
C      SYSTEM OF BINARY WEIGHTS CAN BE USED WHERE  W=2A,  S1=4A,  AND
C      S2=8(A+D).  (REFER CLIFF AND ORD, 1973,PAGE 16.)
C      A=HALF THE SUM OF LI  WHERE LI=THE NUMBER OF GRID SQUARES
C      ADJOINING GRID SQUARE  I.  "A" IS THEREFORE W/2.
C      D=HALF THE SUM OF  LI(LI-1)
C      SONE=4A,   STWO=8(A+D)
C      SKEW=SAMPLE COEFFICIENT OF KURTOSIS, IE. THE FOURTH SAMPLE MOMENT
C      OF X DIVIDED BY (THE SECOND SAMPLE MOMENT OF X) SQUARED.
C      EXPS=EXPECTED VALUE OF THE COEFFICIENT UNDER ASSUMPTION OF RANDOMNESS
C      EXSP=EXPECTED VALUE OF THE COEFFICIENT UNDER ASSUMPTION OF NORMALITY
C      PONE, PTWO, AND PTHR ARE SECTIONS OF THE FORMULA FOR EXPS
C      FRED=ABSOLUTE VALUE OF (CORRELATION PLUS THE FIRST MOMENT OF X
C      DIVIDED BY THE SQUARE ROOT OF EXPS-FIRST MOMENT SQUARED)
C      TOOK AND TALK ARE SECTIONS OF THE FORMULA FOR EXSP
C      FINA=ABSOLUTE VALUE OF THE FORMULA USED IN FRED, USING EXSP INSTEAD
C      OF EXPS.
C      IF FRED OR FINA ARE GREATER THAN 1.96, WHICH IS THE VALUE OF TWO
C      STANDARD DEVIATIONS FROM THE MEAN UNDER A STANDARD NORMAL CURVE,
C      THEN THE HO IS REJECTED IN FAVOUR OF H1
C      RESULTS FOR D, A, S1, S2, B2, PONE, PTWO, PTHR, PLUS BOTH EXPECTED
C      AND ABSOLUTE VALUES ARE PRINTED WITH TITLES.
```

```
C     ------------------------------------------------------------------------
   77 XN=FLOAT(NNNN)
      XW=WAIT
      A=WAIT/2.0
      SONE=4.0*A
      STWO=(8.0*(A+D))
      SKEW=((SUMB/XN)/((SUMZ/XN)**2.0))
      WRITE (6,771) D
  771 FORMAT (1X,15HTHE D-VALUE IS ,F20.4,/)
      WRITE (6,772) A
  772 FORMAT (1X,15HTHE A-VALUE IS ,F20.4,/)
      WRITE (6,773) SONE
  773 FORMAT (1X,18HTHE SONE-VALUE IS ,F20.4,/)
      WRITE (6,774) STWO
  774 FORMAT (1X,18HTHE STWO-VALUE IS ,F20.4,/)
      WRITE (6,775) SKEW
  775 FORMAT (1X,31HTHE VALUE FOR SKEWNESS (B2) IS ,F10.4,//)
      PONE=(XN*((((XN*XN)-(3.0*XN)+3.0)*SONE)-(XN*STWO)+(3.0*(XW*XW))))
      PTWO=(SKEW*((((XN*XN)-XN)*SONE)-((2.0*XN)*STWO)+(6.0*(XW*XW))))
      PTHR=((XN-1.0)*(XN-2.0)*(XN-3.0)*(XW*XW))
      EXPS=((PONE-PTWO)/PTHR)
      WRITE (6,111) PONE
  111 FORMAT (1H0,8HPONE IS ,F28.4,/)
      WRITE (6,112) PTWO
  112 FORMAT (1X,8HPTWO IS ,F28.4,/)
      WRITE (6,113) PTHR
  113 FORMAT (1X,8HPTHR IS ,F28.4,//)
      IF(EXPS.GT.0.0) GO TO 777
      WRITE (6,776)
  776 FORMAT (1H0,46HTHE EXPECTED VALUE IS NEGATIVE.RUN TERMINATED ,//)
      GO TO 83
  777 WRITE (6,778) EXPS
  778 FORMAT (1X,37HTHE EXPECTED VALUE ASSUMING IX TO BE ,/,1X,
     138HDRAWN FROM A NON-NORMAL POPULATION IS ,F16.10,//)
      ROSE=(EXPS-((1.0/(XN-1.0))**2.0))
      IF(ROSE.GT.0.0) GO TO 800
      WRITE (6,779)
  779 FORMAT (1X,44HVALUE UNDER ROOT IS NEGATIVE.RUN TERMINATED ,//)
      GO TO 83
  800 ROOT=(ROSE**0.5)
      FRED=ABS((CORR+(1.0/(XN-1.0)))/ROOT)
      WRITE (6,90) FRED
   90 FORMAT (1X,22HTHE ABSOLUTE VALUE IS ,F16.10,/,1X,21(1H-),//)
      TOOK=(((XN*XN)*SONE)-(XN*STWO)+(3.0*(XW*XW)))
      TALK=(((XN*XN)-1.0)*(XW*XW))
      EXSP=TOOK/TALK
      IF(EXSP.GT.0.0) GO TO 91
      WRITE (6,776)
      GO TO 83
   91 WRITE (6,92) EXSP
   92 FORMAT (1X,37HTHE EXPECTED VALUE ASSUMING IX TO BE ,/,1X,
     134HDRAWN FROM A NORMAL POPULATION IS ,F16.10,//)
      SAMM=(EXSP-((1.0/(XN-1.0))**2.0))
      IF(SAMM.GT.0.0) GO TO 93
      WRITE (6,779)
      GO TO 83
   93 RUUT=(SAMM**0.5)
      FINA=ABS((CORR+(1.0/(XN-1.0)))/RUUT)
      WRITE (6,94) FINA
   94 FORMAT (1X,22HTHE ABSOLUTE VALUE IS ,F16.10,/,1X,21(1H-),//)
C     ------------------------------------------------------------------------
C     PEARSON'S METHOD
C     THIS SEGMENT CALCULATES AN AUTOCORRELATION COEFFICIENT IMPLEMENTING
C     PEARSON'S PRODUCT MOMENT METHOD.  THE DIRECTIONS REMAIN THE SAME,
C     IE. 1=N/S, 2=W/E, AND SO ON.
C     PAIRS OF ADJOINING GRID SQUARES ARE ISOLATED. THE VARIATE VALUE IN
C     GRID SQUARE J BECOMES THE X-VALUE, WHILE THE VARIATE VALUE IN
C     GRID SQUARES I7, I5, I8, AND I6 RESPECTIVELY BECOME THE Y-VALUE.
C     DEPENDENCE IS THEREFORE MEASURED IN ONE DIRECTION SPECIFICALLY,
C     IE. 1=N TO S, 2=E TO W, 3=NW TO SE, AND 4=NE TO SW.
C     X AND Y PAIRS FOR EACH OF THESE RUNS ARE ACCUMULATED UNDER COUNTER
C     ICON AND ARE ALL USED FOR THE FINAL OPTION OF ALL DIRECTIONS.
C     HOWEVER, IF EITHER X OR Y = ZERO THEN THAT PAIR IS NOT USED IN
C     THE CALCULATION OF THE COEFFICIENT.
C     SOME INDICATION OF AUTOCORRELATION OVER THE ENTIRE SURFACE IS THUS
C     GAINED.  ONCE THE APPROPRIATE X AND Y PAIRS HAVE BEEN ISOLATED
C     FOR EACH DIRECTIONAL OPTION A SUBROUTINE IS CALLED FOR THE
C     CALCULATION OF PEARSON'S R.
C     ------------------------------------------------------------------------
      IPER=0
   83 IF(ISEL(MM).GT.1) GO TO 799
```

```
      K=1
      DO 911 J=1,ICOL
      K=K-1
      L=0
      DO 911 I=1,IROW
      K=K+1
      L=I+1
      IF(L.GT.IROW) GO TO 911
      IF(IX(I,J).GT.0) GO TO 901
      K=K-1
      GO TO 911
  901 IF(IX(L,J).GT.0) GO TO 902
      K=K-1
      GO TO 911
  902 IPER=IPER+1
      KX(K)=IX(I,J)
      KY(K)=IX(L,J)
      ICON=ICON+1
      JX(ICON)=KX(K)
      JY(ICON)=KY(K)
  911 CONTINUE
      CALL PROD(KX,KY,IPER)
      GO TO 999
  799 IF(ISEL(MM).GT.2) GO TO 801
      K=1
      DO 912 I=1,IROW
      K=K-1
      L=0
      DO 912 J=1,ICOL
      K=K+1
      L=J+1
      IF(L.GT.ICOL) GO TO 912
      IF(IX(I,J).GT.0) GO TO 903
      K=K-1
      GO TO 912
  903 IF(IX(I,L).GT.0) GO TO 904
      K=K-1
      GO TO 912
  904 IPER=IPER+1
      KX(K)=IX(I,J)
      KY(K)=IX(I,L)
      ICON=ICON+1
      JX(ICON)=KX(K)
      JY(ICON)=KY(K)
  912 CONTINUE
      CALL PROD(KX,KY,IPER)
      GO TO 999
  801 IF(ISEL(MM).GT.3) GO TO 802
      K=1
      KEND=IROW-1
      DO 913 I=1,KEND
      K=K-1
      DO 913 J=1,ICOL
      K=K+1
      II=I+1
      JJ=J+1
      IF(JJ.GT.ICOL) GO TO 913
      IF(IX(I,J).GT.0) GO TO 905
      K=K-1
      GO TO 913
  905 IF(IX(II,JJ).GT.0) GO TO 906
      K=K-1
      GO TO 913
  906 IPER=IPER+1
      KX(K)=IX(I,J)
      KY(K)=IX(II,JJ)
      ICON=ICON+1
      JX(ICON)=KX(K)
      JY(ICON)=KY(K)
  913 CONTINUE
      CALL PROD(KX,KY,IPER)
      GO TO 999
  802 IF(ISEL(MM).GT.4) GO TO 803
      K=0
      LEND=IROW-1
      DO 914 I=1,LEND
      DO 914 J=2,ICOL
      K=K+1
      IN=I+1
      JN=J-1
      IF(IX(I,J).GT.0) GO TO 907
```

```
      K=K-1
      GO TO 914
  907 IF(IX(IN,JN).GT.0) GO TO 908
      K=K-1
      GO TO 914
  908 IPER=IPER+1
      KX(K)=IX(I,J)
      KY(K)=IX(IN,JN)
      ICON=ICON+1
      JX(ICON)=KX(K)
      JY(ICON)=KY(K)
  914 CONTINUE
      CALL PROD(KX,KY,IPER)
      GO TO 999
  803 CALL PROD(JX,JY,ICON)
  999 IF(MM.EQ.NUMM) GO TO 79
      MM=MM+1
      GO TO 13
   79 STOP
      END

END OF SEGMENT, LENGTH  3004, NAME  REAL

      SUBROUTINE PROD(LX,LY,IPER)
C     ##############################################################
C     SUBROUTINE CALCULATES R ACCORDING TO THE PRODUCT MOMENT METHOD
C     DIMENSIONS REFLECT THE POSSIBLE TOTAL NUMBER OF X AND Y PAIRS IN
C     THE GRID.  IPER IS THE NUMBER OF POSSIBLE X AND Y PAIRS IN ONE
C     DIRECTION WHILE ICON REFLECTS ALL JOINS.
C     TOPP, VARX, AND VARY ARE SECTIONS OF THE EQUATION.  SIGNIFICANCE
C     IS EVALUATED ACCORDING TO STUDENTS T TEST.
C     RESULTS ARE PRINTED WITH TITLES
C     ##############################################################
      DIMENSION LX(625),LY(625),X(625),Y(625)
      DO 921 I=1,IPER
      X(I)=FLOAT(LX(I))
  921 Y(I)=FLOAT(LY(I))
      SXTY=0.0
      SXSQ=0.0
      SYSQ=0.0
      XSUM=0.0
      YSUM=0.0
      DO 922 I=1,IPER
      XVAL=X(I)
      YVAL=Y(I)
      XSUM=XSUM+XVAL
      YSUM=YSUM+YVAL
      XSQ=XVAL*XVAL
      YSQ=YVAL*YVAL
      SXSQ=SXSQ+XSQ
      SYSQ=SYSQ+YSQ
      XTY=XVAL*YVAL
      SXTY=SXTY+XTY
  922 CONTINUE
      TOPP=SXTY-((XSUM*YSUM)/FLOAT(IPER))
      VARX=SXSQ-((XSUM*XSUM)/FLOAT(IPER))
      VARY=SYSQ-((YSUM*YSUM)/FLOAT(IPER))
      RUTE=(VARX*VARY)**0.5
      COCO=TOPP/RUTE
      COEF=ABS(COCO)
      NODF=IPER-2
      AWON=FLOAT(NODF)
      ONE=COEF*(AWON**0.5)
      IF(COEF.LE.1.0) GO TO 940
      TWO=0.0
      GO TO 941
  940 TWO=((1.0-(COEF*COEF))**0.5)
  941 IF(TWO.NE.0.0) GO TO 928
      TEST=-99.0
      GO TO 929
  928 TEST=ONE/TWO
  929 WRITE (6,930)
  930 FORMAT (1H0,36H(B) PEARSON'S PRODUCT MOMENT METHOD ,/,1X,35(1H-),/
     1/)
      WRITE (6,931) IPER
  931 FORMAT (1X,31HTHE NUMBER OF X AND Y PAIRS IS ,I12,/)
      WRITE (6,932) TOPP
  932 FORMAT (1X,31HPART ONE OF THE COMPUTATION IS ,F20.4,/)
      WRITE (6,933) RUTE
```

```
  933 FORMAT (1X,31HPART TWO OF THE COMPUTATION IS ,F20.4,//)
      WRITE (6,934) COCO
  934 FORMAT (1H0,31HTHE CORRELATION COEFFICIENT IS ,F18.6,/,1X,30(1H-),
     1//)
      IF(TEST.GE.0.0) GO TO 936
      WRITE (6,937)
  937 FORMAT (1X,31HTHE T-VALUE APPROACHES INFINITY ,/)
      GO TO 938
  936 WRITE (6,935) TEST
  935 FORMAT (1X,24HTHE STUDENTS T VALUE IS ,F20.4,/)
  938 RETURN
      END

END OF SEGMENT, LENGTH   346, NAME  PROD

      FINISH

END OF COMPILATION -  NO ERRORS

S/C SUBFILE:       61 BUCKETS USED
CONSOLIDATED BY XPCK 12F      DATE      08/12/77    TIME 11/51/03
PROGRAM REAL
EXTENDED DATA (22AM)
COMPACT PROGRAM (DBM)
CORE                17728
                    SEG        REAL
                    SEG        FLOAT
                    SEG        ABS
                    SEG        PROD
                    ENT        FTRAP
                    ENT        FRESET
```