Model Building and the Analysis of Spatial Pattern in Human Geography
Author(s): A. D. Cliff and J. K. Ord
Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 37, No. 3 (1975), pp. 297-348
Published by: Wiley for the Royal Statistical Society
Stable URL: http://www.jstor.org/stable/2984781
Accessed: 02-11-2015 04:54 UTC

# Model Building and the Analysis of Spatial Pattern in Human Geography

By A. D. CLIFF and J. K. ORD

*University of Cambridge* *University of Warwick*

[Read at a joint meeting of the ROYAL STATISTICAL SOCIETY and the INSTITUTE of BRITISH GEOGRAPHERS on Wednesday, February 12th, 1975, at a meeting organized by the RESEARCH SECTION, Professor R. L. PLACKETT in the Chair]

## 1. INTRODUCTION AND SUMMARY

It is the purpose of this paper to determine how far various statistical models and methods of statistical inference have enabled the aims of geographical research to be met in the problem areas to which they have been applied. In so doing, we hope we can indicate to the statistician questions of geographical interest which cannot readily be answered by existing statistical methods; and to the geographer, some of the insights into geographical processes which may be gained from a statistical and model building approach. We would stress that we have not tried to be all inclusive in our coverage. Instead, we have tried to select some topics which best seem to convey the flavour of the kinds of things human geographers have been doing, and which will, at the same time, be of interest to statisticians on either theoretical or empirical grounds. In addition, our own interests mean that we have concentrated upon examples in human (economic and urban), rather than physical, geography, although similar approaches have been used there to a lesser degree. More general reviews are provided by Gould (1969), Berry (1971) and Wilson (1972).

*Keywords*: AGGREGATION; DEPENDENT OBSERVATIONS; DIFFUSION; GRAVITY MODELS; HUMAN GEOGRAPHY; INTERACTION MODELS; ISOTROPY; KRIGING; MAXIMUM LIKELIHOOD; MEASLES EPIDEMICS; MISSING DATA; NEAREST-NEIGHBOUR ANALYSIS; JOINT PATTERNS; POISSON PROCESSES; QUADRAT COUNTS; RANK SIZE RULE; SETTLEMENT SIZES AND SPACING; SPACE-TIME MODELLING; SPATIAL AUTOCORRELATION; SPATIAL INTERACTION MODELS; SPATIAL SPECTRAL ANALYSIS; STATIONARITY; TREND SURFACES

## 2. THE NATURE OF GEOGRAPHY

### 2.1. *Aims*

SUBSTANTIAL treatises on the nature of geography and its methodology exist. The standard reference works are those of Hartshorne (1939, 1959) and Harvey (1969). Hartshorne's books are based upon the premise that geography is what geographers have done, and his contribution represents an empirically derived statement of the purpose of geography. Some quotations taken from Hartshorne's (1959) book help to summarize his findings.

> With astronomy and geophysics, it [geography] is one of the chorological sciences, as history, prehistory, paleontology, etc., form the chronological sciences. These two groups are both in contrast to the systematic sciences in that they study segments of space or time in terms of whatever may be their contents, whereas the systematic sciences concentrate upon particular categories of objects or phenomena wherever they may be in space or time (p. 178).

> The goal of the chorological point of view is to know the character of regions and places through a comprehension of the existence together and interrelations among the different realms of reality and their varied manifestations, and to comprehend the earth's surface as a whole in its actual arrangements in continents, larger and smaller regions, and places (p. 13).
>
> Geography is that discipline that seeks to describe and interpret the variable character from place to place of the earth as the world of man (p. 47).
>
> Geography seeks (1) on the basis of empirical observation as independent as possible of the person of the observer, to describe phenomena with the maximum degree of accuracy and certainty; (2) on this basis, to classify the phenomena, as far as reality permits, in terms of generic concepts or universals; (3) through rational consideration of the facts thus secured and classified and by logical processes of analysis and synthesis, including the construction and use wherever possible of general principles or laws of generic relationships, to attain the maximum comprehension of the specific interrelationships between phenomena; and (4) to arrange these findings in orderly systems so that what is known leads directly to the margin of the unknown (p. 169.)

In addition, Hartshorne stresses two other points. First, that while geographers are interested in the development of laws, the description of individual cases has always occupied a central place in the discipline. Second, that the analysis of areal variations through time is an integral part of the discipline.

Thus a picture emerges of geography as a subject which seeks to describe and account for spatial patterns on the earth's surface insofar as they affect man. The analysis may be purely static—that is, at a single point of time—or concerned with the temporal evolution of spatial patterns. The latter, of course, forms the basis of process studies.

The present writers accept the view of geography as a subject characterized by its spatial approach, but with one addition. We would include in the list of objectives the desire to make predictions, whether they relate to the future development of a spatial–temporal process, or whether they are concerned with purely spatial phenomena as in the mapping of geological structure. It should be noted, however, that not all geographers agree with this emphasis. For example, Eyre (1973) regards geography's spatial viewpoint as something of an "encumbrance".

Given that we believe the spatial viewpoint to be central to geography, it is of interest to look through early copies of the Society's *Journal* to determine the extent to which this approach has been adopted by statisticians. Some of the first papers in the *Journal* contained sections which were concerned with descriptive physical geography, such as that on Antigua by Tulloch (1838) in Volume 1. This tradition was continued with papers like that of Hind (1864). Turning to human geography, the main contributor to the *Journal* was E. G. Ravenstein. During a visit to the Paris Geographical Congress (described in Volume 38, 1875), he was considerably impressed by the value of maps as visual displays of statistical data. Later (1879), he produced "linguistic maps" for the Celtic languages in the British Isles which provided a graphic picture of language usage. However, his most enduring contribution was a study of spatial laws of migration in 1885—the beginnings of the gravity model still widely used today (see Section 5.1).

Early in the twentieth century, Student was the first statistician to show real concern about spatial problems. In 1907, in his analysis of cell counts using a haemacytometer, he was worried by the assumption that counts in contiguous cells of a rectangular grid could be regarded as independent variates. Thus was born the first "spatial autocorrelation coefficient" (Section 6.2). Later (Student, 1914), he discussed the elimination of spurious correlation due to the position of the observation in time *or space*, by the use of polynomials. The spatial model is not spelled out in detail, but it is quite clear from the paper. It is the technique now widely used by geographers and others under the name of "trend surface analysis".

We conclude our brief and incomplete historical survey of the Society's *Journal* with a mention of an article by Kendall (1939) which represents, possibly, the first use of truly multivariate methods in geography. Kendall carried out a principal components analysis on the yields per acre for ten main crops in the English and Welsh counties. The county scores on the first component were used as an index of productivity which formed the basis of a procedure for classifying counties into agricultural regions.†

### 2.2. *Use of Statistics and Models in Geography*

The extensive use of statistical methods by geographers to further the aims of the discipline is rather recent. Most writers [see, for example, Burton (1963) and Wilson (1972)] suggest that it dates from about 1953, and we can identify the following three main approaches.

(1) The testing of hypotheses about multivariate spatial data sets using classical aspatial statistical techniques. The ultimate purpose of these studies has been gradually to identify a series of working statements about spatial behaviour which might yield inductive theories for spatial processes. The best summary accounts of the substantive content of these studies are Haggett (1965) and Gould (1969).

(2) The use of special-purpose techniques, for example nearest-neighbour methods, to describe precisely the form of spatial patterns. It was believed that such precise description might be suggestive of the processes generating the forms. There appeared, therefore, a large number of studies which tried to impute process from form.

(3) The development of formal models of spatial processes. In many cases, the results of approaches (1) and (2) provided the empirical base from which the models were built.

In the remainder of the paper, we try to examine critically the methods and contribution of each of these three approaches. In Section 3, we discuss some properties of spatial data which make the results of analyses from any of the approaches difficult to interpret, although specific examples are given with reference to approach (1). Approach (2) is considered in Section 4, and approach (3) in Sections 5 and 6. The paper is concluded in Section 7.

## 3. THE USE OF CLASSICAL STATISTICAL METHODS IN GEOGRAPHY

There are several basic properties of spatial data which make their analysis using classical statistical models difficult. We comment upon each of these in turn in this section.

### 3.1. *Spatial Stationarity*

Let us first say what we mean by spatial stationarity. We assume that the data collected correspond to a finite set of "locations". The locations may be either points or areas ("spatial aggregates"). Let $J$ denote the set of locations and

† It is worth reading Stamp's comments in the discussion following Kendall's paper as a reflection of the attitude of geographers to quantitative methods at the time.

$j=1,\ldots,n$ index the $n$ members of the set. In addition to the observed variate value, $y_j$ on $Y_j$, there will also be some information concerning location $j$ (co-ordinates for example). We suppose that the variate $Y_j$ can be decomposed into a stochastic component $X_j$ and a deterministic component $m_j$, such that $E(X_j)=0$ or $E(Y_j)=m_j$.

*Definition* 3.1. $X_j$ is said to describe a spatially stationary process in the wide sense, or to be weakly spatially stationary, if the quantities

$$E(X_j X_{j'}) = \sigma(j, j') \tag{3.1}$$

depend only upon the *relative* position of locations $j$ and $j'$. For example, if the locations refer to equally spaced points on a line, then $\sigma(j,j')$ becomes a function of $|j-j'|$.

*Definition* 3.2. Suppose $X_j$ satisfies equation (3.1). If, in addition, the correlation between $X_j$ and $X_{j'}$ depends only upon the distance between their locations, and not upon the orientation of the chord between $j$ and $j'$, then the process is said to be (weakly) *isotropic*.

*Definition* 3.3. If the joint distribution of the $X_j$ ($j \in J$) depends only upon the relative positions of the locations, then the spatial process is stationary in the strict sense. If the joint distribution is the multivariate normal, and the $X_j$ are stationary in the wide sense, then the process is also strictly stationary.

*Definition* 3.4. The process is strictly isotropic if it is both strictly stationary and direction-invariant.

In econometrics and time series analysis, Definitions 3.1 and 3.3 suffice. However, in spatial analysis, Definitions 3.2 and 3.4 will often be required, and represent a further specialization of the model.

As Granger (1969) has argued, the "assumption of stationarity on the plane is completely unrealistic for economic variables", and he illustrates his case with an analysis of regional unemployment rates. Granger's results suggest that London and the South-East lead slightly the other regions of the United Kingdom. The time lag for the other regions varies, approximately, with their distance from London. More critically, the other regional series relate more strongly to London and the South-East than to each other.

Results of this nature, combined with geographical theory and observation, strongly suggest the existence of one or more leading regions, or growth poles, rather than spatial stationarity. The growth pole approach provides a more realistic approach to spatial economic modelling, but the identification of such poles may not be straightforward (Cliff and Ord, 1975).

In general, while the assumption of spatial stationarity is difficult to sustain, it may be possible to induce stationarity by spatial differencing, which we discuss in the next section.

### 3.2. *Spatial Dependence*

In econometrics, temporal autocorrelation among the observations on economic variables is the norm rather than the exception. Similarly, spatial dependence among geogaphical data is usual. Thus Tobler (1970a) has referred to "the first law of geography: everything is related to everything else, but near things are more related than distant things". Again in 1970b, Tobler wrote, "the central dogma in geography asserts that what happens at one place is not independent of what happens at another", while Gould (1970) stated:

> Why we should expect independence in spatial observations that are of the slightest intellectual interest or conceptual importance in geographic research I cannot imagine. All our efforts to understand spatial pattern, structure, and process have indicated that it is precisely the *lack* of independence—the *inter*dependence—of spatial phenomena that allows us to substitute pattern, and therefore predictability and order, for chaos and apparent lack of interdependence of things in time and space.

What is the effect of spatially autocorrelated observations upon tests of inference? The consequences in some cases are well known. Thus, standard applications of the $t$ and $F$ statistics for the comparison of means or the construction of confidence intervals require spatial independence (a special case of Definition 3.4). The same assumption is necessary for the error terms in regression analysis if the ordinary least squares estimators are to be BLU (best linear unbiased).

For example, if the sample mean $\bar{y}$ is used to estimate $E(Y_j) = m$ (assumed fixed for all $j$), then the usual variance estimator

$$\hat{\sigma}^2 = \sum_{j=1}^{n} (y_j - \bar{y})^2/(n-1) \tag{3.2}$$

will be biased downwards when the observations are positively autocorrelated in space. The temporal analogue of this is, of course, well known to econometricians (cf. Johnston, 1972, Section 8.2), leading to overstatement of the significance of the results. However, in the case of other statistical models which assume independence, little is known about their robustness to departures from the assumption of independence. Cliff *et al.* (1975b) have modified the $\chi^2$ goodness-of-fit test to allow for the presence of spatial dependence. The sampling distribution of $\chi^2$ under the null hypothesis can change substantially. Much more theoretical work is required in this area.

What avenues are open to workers interested in this problem? The first is explicitly to allow for the spatial dependence among the observations. This was the approach adopted by Cliff *et al.* (1975b). The second possibility is to apply a spatial variate differencing procedure to the observations in an attempt to remove the spatial dependence so that conventional models may be applied. This is clearly analogous to the variate differencing used by time series analysts to remove temporal autocorrelation. However, in the spatial case, the dependence is multilateral. That is, it can extend in all compass directions, and not just into the past, as with a time series. One possibility might be to define a first spatial difference for county $i$ as

$$\Delta y_i = y_i - \bar{y}_{j \in I}, \tag{3.3}$$

where $I$ denotes the set of counties contiguous to (first nearest neighbours of) $i$. Very little work has been done on spatial differencing filters (Curry, 1971). Martin (1974) has looked at their use in the regression case, and a further interesting study, which we now consider, is that of Lebart (1969) on the effect of spatial dependence in factor analysis.

### 3.2.1. *Spatial dependence in factor analysis*

Suppose that data are collected on $p$ variates for each of $n$ counties, yielding the $(n \times p)$ data matrix $\mathbf{X}$. In addition, we construct a binary connection matrix $\mathbf{M}$, whose elements are defined as

$$\begin{aligned} m_{ij} &= 1, \quad \text{if the } i\text{th and } j\text{th counties are contiguous} \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

Then $\mathbf{M}_\alpha = \mathbf{M}^\alpha$ defines the number of paths, $\alpha$ links in length, between each pair of counties. Note that this definition includes redundant paths such as $i \to j \to i$. For each variate, Lebart computed differences between county values for all pairs of counties, $i$ and $j$, $\alpha$ links apart. He defined the sample $\alpha$-lag covariance for the variates $k$ and $l$ as

$$C_\alpha(k,l) = \frac{1}{2n_\alpha} \sum (x_{ki} - x_{kj})(x_{li} - x_{lj}), \tag{3.4}$$

where the summation is over all paths of length $\alpha$, and $n_\alpha$ is the number of such paths. In matrix terms, (3.4) can be written as

$$\mathbf{C}_\alpha = (1/n_\alpha)\mathbf{X}^{\mathrm{T}}(\mathbf{N}_\alpha - \mathbf{M}_\alpha)\mathbf{X} \quad (\alpha = 0, 1, \ldots), \tag{3.5}$$

where $\mathbf{N}_\alpha = \operatorname{diag}(\mathbf{M}_\alpha^2)$. $\mathbf{C}_0$ is the usual sample covariance matrix. Lebart argues that the use of some $\mathbf{C}_\alpha$ other than $\mathbf{C}_0$ will reduce the effect of spatial dependence upon the analysis. However, suppose that the spatial dependence in the $i$th variate may be represented by the model

$$\mathbf{N}_1 \mathbf{X}_i = \rho_i \mathbf{M}_1 \mathbf{X}_i + \mathbf{Z}_i \quad (i = 1, \ldots, p), \tag{3.6}$$

where $E(\mathbf{Z}_i \mathbf{Z}_j^{\mathrm{T}}) = \sigma_{ij} \mathbf{I}_n$ and $E(\mathbf{X}_i) = \mathbf{0}$ for simplicity. This implies that

$$E\{C_\alpha(i,j)\} = \sigma_{ij} \operatorname{tr}\{(\mathbf{N}_1 - \rho_i \mathbf{M}_1)^{-1}(\mathbf{N}_\alpha - \mathbf{M}_\alpha)(\mathbf{N}_1 - \rho_j \mathbf{M}_1)^{-1}\}. \tag{3.7}$$

When $\rho_i = \rho$ $(i = 1, \ldots, p)$, it follows that $E(\mathbf{C}_1)$ is proportional to $E(\mathbf{C}_0)$, so that a factor analysis or a components analysis of the two population covariance matrices would produce the same results, scale factors apart. In his study, Lebart applied a factor analysis to both $\mathbf{C}_0$ and $\mathbf{C}_1$, where the matrices were constructed from socio-economic data collected for the 88 départements of France. The differences between the two analyses would appear to be the result of sampling fluctuations, and there is no reason to suppose that the analysis based on $\mathbf{C}_1$ is to be preferred to that based on $\mathbf{C}_0$.

One possible approach would be to postulate a model such as (3.6) for each variate, and then to estimate the $\{\rho_i\}$ in the course of the complete analysis. However, we have no idea as to the merits of such a scheme, and one may question whether the computational effort would be justified. In this respect, Lebart's work yields the positive result that spatial dependence does not materially affect factor or principal components analyses when the $\{\rho_i\}$ in (3.6) are approximately equal. The practical value of such a conclusion remains unexplored.

Whether we incorporate spatial dependence into the model explicitly, or remove it before commencing the main analysis is partly a question of emphasis. However, from the viewpoint of geographical modelling one may argue, as does Gould (1970), that "such corrections [as differencing] will represent a throwing out of the baby and keeping the bathwater". This again represents an area where statisticians and geographers could fruitfully collaborate.

### 3.3. *Aggregation Problems*

The results of almost any analysis will depend upon the way in which the geographical study area is partitioned for data collection purposes. To illustrate this point, we may consider the correlation coefficient. Yule and Kendall (1957, pp. 310–13) computed the correlation between the yields per acre of wheat and potatoes in the

48 counties of England and Wales in 1936, and obtained $r = 0{\cdot}22$. The 48 counties were grouped into 24. The yield of each pair of combined counties was defined as the unweighted mean of the yields of the two counties grouped together. The revised $r$ value was 0·30. Repeating the process for 12, 6 and 3 counties, they obtained, successively, $r = 0{\cdot}58$, $r = 0{\cdot}76$ and $r = 0{\cdot}99$. The discussion of this problem among geographers has focused chiefly upon the idea of weighting each "county" by its size in the definition of the correlation coefficient (Robinson *et al.*, 1961; Thomas and Anderson, 1965). However, in some situations, the size of the enumeration unit (area, population etc.) may enter into a model directly as in the studies of Gould (1960) and Taaffe *et al* (1963) on the sizes of road networks in developing countries. As we have noted, one of the geographer's research aims is to make interregional comparisons. If the sizes of the enumeration units are very different (say the states of the U.S. compared with the counties of the U.K.) how can such comparisons be made?

Yule and Kendall (1957, pp. 312–13) argue that there is no complete answer to this problem. However, work on the aggregation of economic data should be useful in this context (Theil, 1971). One approach used by geographers is to employ a nested analysis of variance (AV) to determine the most important scale for study. The initial impetus for this idea came from the ecological literature (Greig-Smith, 1957, 1964). Chorley *et al.* (1966) and Moellering and Tobler (1972) have been the chief exponents of the method among geographers.

Let $X_{ijk}$ be the value of the $k$th "district" of the $j$th "county" in the $i$th "region" of the study area. Then put

$$X_{ijk} = \mu + \alpha_i + \beta_{ij} + \gamma_{ijk}, \tag{3.8}$$

where $\mu$ is the overall mean and the $\{\alpha_i\}, \{\beta_{ij}\}, \{\gamma_{ijk}\}$ represent regional, county and district effects respectively.

The Model II Analysis of Variance seems most appropriate here. We may therefore hypothesize that the different effects are zero mean, uncorrelated random variables with equal variances within each set; that is, $\sigma_\alpha^2$, $\sigma_\beta^2$ or $\sigma_\gamma^2$. The unknown variances can then be estimated (Kendall and Stuart, 1966, pp. 59–60) so that the total variation may be partitioned into scale effects attributable to each level of the hierarchy. Clearly, the number of levels is restricted only by the data available.

### 3.4. *Data Availability*

As Granger (1969) has noted, economic and urban data are usually not available on a regular grid basis. Instead, either they are values collected at some arbitrary set of spatial locations or they are values for a region. In the latter case, they are frequently transformed into point data by assigning the variate values to the geographical centroids of the regions concerned. In general, then, spatial data are irregularly spaced and discontinuous. Granger (1969) has argued that this makes spatial modelling "almost impossible". We suggest that these data limitations make it necessary to develop models conditionally upon the structure of the data collecting units. This restricts the inferences which can be made from the analysis, but means that progress can be made. The spatial autocorrelation models considered in Section 6.2 provide an example of this approach.

### 3.5. *Conclusions*

We have discussed the above difficulties in connection with the use of classical statistical methods in geography because that is the context in which the problems were first noted. It will be evident, however, that they carry over into the material considered in the other sections. Meanwhile, until more basic statistical research has been done in the areas noted, it is difficult to decide what degree of reliability can be placed in the substantive findings of geographers using these techniques.

## 4. THE SPACING AND SIZE OF SETTLEMENTS

As we noted in Section 2.2, one important research area in geography has been the development of special purpose techniques for spatial (point) pattern description. This interest has been motivated by the belief that the accurate description of spatial forms might yield insights into the processes generating these forms. In this section, we look at the principles and difficulties involved in this kind of analysis. We do this with special reference to settlement distributions, since, at the macro-scale, cities, towns, villages and hamlets have frequently been regarded as point-like objects. In addition, the analysis of the spacing and size characteristics of settlements has traditionally been a central concern of the geographer.

### 4.1. *Spatial Point Patterns*

A basic theme in the study of point patterns is the question of randomness. If points are located in the plane such that in any small area, $\delta A$,

$$\left.\begin{aligned} P\text{ (exactly one individual)} &= \lambda\delta A + o(\delta A),\\ P\text{ (two or more individuals)} &\sim o(\delta A),\\ P\text{ (no individuals)} &= 1-\lambda\delta A + o(\delta A),\end{aligned}\right\} \tag{4.1}$$

where $o$ means "of smaller order than", then it follows that the number of individuals in any area $A$ is Poisson distributed with mean $\lambda A$. If the pattern is generated by a Poisson process, we say that it is *random*. Evidently the process is strictly stationary; further, it is strictly isotropic.

Accepting this definition, how can we establish whether or not a particular process is random, given a realization of that process? Two main alternative approaches exist, based upon distance and areal methods. We shall discuss each in turn, but we first consider why the question should be asked at all.

A principal reason for the interest of geographers in randomness is the belief that some aggregate human behaviour patterns can be adequately described by random process models [the "random spatial economy" of Curry (1964, 1967)]. This belief is motivated by the hope that the underpinnings of the Second Law of Thermodynamics may apply to human activity. This law states, roughly, that molecular systems tend in the long run towards a state of maximum entropy, or randomness, in which the energy for change among the particles in the system has been dissipated. Measuring the degree of non-randomness in a realization of a process may thus provide some guide as to the state of organization in the system.

A major difficulty arises, however, when we try to infer the nature of a process from the form of a realization of that process because many processes are equifinal (Harvey, 1968); that is, more than one process may produce the same end pattern. The following example serves to illustrate this point. One way of looking at the colonization of an area by settlers is to think in terms of *clustering*, aggregation or true contagion. The first settlers choose their locations (possibly at random) and later settlers choose locations near to the early settlements. A point map of such a process would exhibit clustering of settlements, albeit with a minimum "social distance" between points.

An example of a model for such a process is the negative binomial distribution. This distribution assumes an initial Poisson pattern of points (centres). Points added later are assumed to cluster around the original nuclei according to a logarithmic series growth law. Thus, in the notation of Gurland (1957),

$$\text{negative binomial} \sim \text{Poisson} \vee \text{logarithmic}.$$

For a fuller discussion of this and similar models, see Ord (1972, Chapter 6).

A second way of viewing the colonization process is to assume that each settler chooses his site independently, but that some sites are more attractive than others. That is, $\lambda$ in the set of equations (4.1) is not constant, but is a function of location. This heterogeneity, or apparent contagion, model can lead to the same departures from randomness as the true contagion model, but it is based upon an entirely different chance mechanism. Indeed, an initial Poisson pattern with $\lambda$ specified by a gamma distribution again yields a negative binomial. That is,

$$\text{negative binomial} \sim \text{Poisson} \wedge \text{gamma}.$$

The negative binomial has been extensively used in geography [see, for example, Dacey (1968, 1969) and Cliff and Ord (1973, Chapter 3)] to examine the settlement of a landscape. Yet the basic difficulty exists that the model is over-specified, and a good fit does not enable the researcher to infer the particular generating process with certainty. Although some progress has been made by Ord (1972) and Cliff and Ord (1973, Chapter 3) in distinguishing between the spatial processes generating the negative binomial, the same problem exists with other members of the Poisson family (such as the Neyman type A), and relatively little work has been done in this area. Bearing in mind this difficulty with "form to process" studies, we can now examine some of the methods used to identify the degree of randomness or otherwise in spatial forms.

4.1.1. *Nearest-neighbour methods*

Let the distance between individual $i$ and individual $j$ be $d_{ij}$. If

$$d_i^{\min} = \min_{j \neq i} d_{ij} \quad \text{and} \quad d_i^{\min} = d_{ik},$$

say, then individual $k$ is the nearest neighbour of $i$. The relationship is not reflexive, since it does not follow that $d_k^{\min} = d_{ik}$. A remarkable and oft-rediscovered result (Hertz, 1909; Skellam, 1952; and many more recent sources) is the following.

When the spatial pattern is generated by a Poisson process with parameter $\lambda$, the distribution of the distance, $X_1$, between an individual and its nearest neighbour has the density function,

$$f(x) = 2\rho x \exp(-\rho x^2), \tag{4.2}$$

where $\rho = \pi\lambda$. Further if $Y_1 = X_1^2$, then $Y_1$ has the density function,

$$g(y) = \rho \exp(-\rho y), \tag{4.3}$$

that is, $Y_1$ is exponentially distributed. More generally, the square of the distance to the $r$th nearest neighbour, $Y_r$ say, follows a gamma distribution with parameters $r$ and $\rho$. These results also hold for the distance from *any* randomly selected point to the ($r$th) nearest individual.

A variety of tests for randomness based upon these findings has appeared in the literature, mainly for use by ecologists (see Holgate, 1972).

When $\lambda$ is unknown, the performance of these test statistics is inversely related to their robustness as estimators of $\lambda$ (see Persson, 1971). However, in geographical applications, $\lambda$ can usually be taken as given, so that direct tests are available using nearest-neighbour distances and the results

$$E(Y_r) = r\rho^{-1}, \tag{4.4}$$

$$E(X_r) = \rho^{-\frac{1}{2}}\,\Gamma(r+\tfrac{1}{2})/\Gamma(r) \tag{4.5}$$

$$= \tfrac{1}{2}\lambda^{-\frac{1}{2}}, \quad \text{for } r = 1.$$

For example, Clark and Evans (1954) suggested the statistic

$$R_1 = 2\lambda^{\frac{1}{2}} \sum_{i=1}^{n} d_i^{\min}/n, \tag{4.6}$$

while Moore (1954) and Skellam (1952) independently suggested

$$R_2 = \rho \sum_{i=1}^{n} (d_i^{\min})^2/n. \tag{4.7}$$

In a Poisson process, both statistics have expected value equal to unity. For $i = 1, 2$, $R_i$ tends to zero under conditions of maximum aggregation, while for maximum spacing, when the points fall at the vertices of a network of regular hexagons,

$$E(R_1) = 2{\cdot}15 \quad \text{and} \quad E(R_2) = 3{\cdot}63. \tag{4.8}$$

Since $Y_1$ is an exponential variate, $R_2$ will be gamma distributed with index $n$ and variance $n^{-1}$, *provided that all the observations are independent*. This proviso is not to be taken lightly, since if $k$ is the nearest neighbour for $i$, we know that $d_k^{\min} \leqslant d_{ik}$. That is, we cannot include all the individuals in the study area in the sample. Indeed, the requirement is only likely to be met if the study area is stratified so that the distance between selected individuals is much greater than $\lambda^{-1}$. This is rarely feasible in geographical applications. Further, sampling must not be carried out close to the edge of the study region since this will alter the distribution of the test statistics. One solution to this difficulty might be to map the study area onto a torus. Given that the tests assume spatially stationary processes, this may not be that unreasonable. However, if the stationarity assumption is not met, the substantive interpretation of the results of any analysis may not be easy. When it is desired to examine all points in the study area, a Monte Carlo method could be used. The procedure involves the generation of dummy realizations of the hypothesized process and the comparison of the observed and dummy results (for further details, see Section 5.2.1). The statistics (4.6) and (4.7) could be used to make the comparisons, but various improvements have been suggested in the literature. For example, Dacey (1963) and Holgate (1966) considered the use of higher order distances, while Dacey and Tung (1962) examined a sectoral nearest-neighbour method. Other possibilities are the wandering quarter of Catana (1963), and the T-square of Besag and Gleaves (1974). None of these measures is free from the operational difficulties mentioned above.

4.1.2. *Quadrat count analysis*

As implied in the discussion in Section 4.1, a different approach to distance methods for the analysis of point pattern data is to sample the area using a system of quadrats. In geographical work, this is usually done by exhaustively partitioning the whole study area (cf. Greig-Smith, 1964; Kershaw, 1964). The resulting observed distribution of cell counts may then be compared to a model derived from some theoretical spatial process. As Harvey (1968) has noted, the Poisson family has been extensively used by geographers in this context, again, as discussed in Section 4.1, to establish the degree of randomness in the point pattern.

Quadrat count analysis suffers from certain defects. First, the results are not scale free and may alter with quadrat size (Pielou, 1957). Second, as noted in Section 4.1, several models are equifinal, leading to problems of process identification. Third, a basic assumption of the Poisson models is that the counts for various quadrats are independent of each other—that is, the counts lack spatial autocorrelation (see Section 6.2). This can only hold for "point" clusters, but will be approximately true when the quadrat size greatly exceeds cluster size (Ord, 1970). Difficulties arise when clusters are allowed to spread across sampling boundaries (see Pielou, 1957; Harvey, 1968). To overcome the problem, Neyman and Scott (1952, 1958, 1972 and other papers referenced therein) and Warren (1962) have developed the centre-satellite model, which explicitly incorporates the spatial spread of clusters.

The centre-satellite process is made up of three components in the following way:

(i) an initial (Poisson) pattern of cluster centres;

(ii) a cluster size distribution, such as the logarithmic: independence between different clusters is usually assumed;

(iii) a location mechanism which describes the position of each member within a cluster with respect to the cluster centre. Usually, individuals are assumed to locate independently of each other. The isotropic bivariate normal distribution is often used in this context.

The greater complexity of this model means that analytical solutions are available only for special cases, although it provides a suitable framework for simulation studies. Even here, care is needed in interpretation, since the doubly stochastic Poisson models described in Section 4.2 can produce equivalent results.

### 4.2. *The Spacing of Settlements*

The idea of the random spatial economy and the tests of spatial randomness discussed above provide a backcloth against which we can consider work by Dacey (1960, 1964, 1965, 1966a, b, c, 1968, 1969) and Curry (1967) on the spacing between settlements in an area. In his classic (1960) paper, Dacey took data from Brush (1953) on the spacing of towns in South-west Wisconsin. He analysed the data by nearest-neighbour methods and concluded that the spacing of towns in the area was random, rather than the hexagonal (uniform) arrangement postulated by economists in the body of literature known as *central place theory*.† In the remaining papers cited, Dacey developed a whole series of models based upon Poisson processes for the spacing of towns in the American Mid-West, Puerto Rico and Japan. In his later papers, Dacey (1968, 1969) obtained particular success by partitioning up maps with a system of regular quadrats, and then fitting the negative binomial to the frequency distribution of the number of quadrats in the map with $0, 1, 2, \ldots, k$ settlements in them. As noted in Section 4.1.2, this kind of analysis is dependent upon the size of the quadrats used. Cliff and Ord (1973, Chapter 3) have exploited this fact to overcome the equifinality problem. They have shown that the parameters of the negative binomial vary in systematic and different fashion with changes in quadrat size,

† An extensive review and bibliography of the central place literature is given in Berry and Pred (1961).

depending on whether the negative binomial is generated by a true contagion process (the generalized Poisson) or an apparent contagion process (the compound Poisson). Application of this finding to the Japanese settlement data used by Dacey suggests that the compound process is more plausible than the generalized process. This implies that the spacing of settlements in the region of Japan considered is essentially Poissonian, but that the density of the process varies from area to area, possibly because of differential soil fertility.

Medvedkov (1967) has used entropy measures to analyse Brush's (1953) data, as well as maps of settlements in the plains of France, Northern Italy and Czechoslovakia. He concluded that of the 99 central places in South-west Wisconsin, 46 fitted a random lattice and 53 a uniform grid. The corresponding figures for France were 23 and 99, for Italy 78 and 115, and for Czechoslovakia 39 and 95.

It is apparent from these results that neither the simple random model nor the uniform model is wholly satisfactory. This result is confirmed by Rayner and Golledge (1972) who detected, using two-dimensional spectral analysis, an important random component in the settlement pattern of Oregon but not in that of North Dakota or Pennsylvania. The random spatial economy assumes an isotropic surface, and many of the study areas selected by Dacey, for example, were chosen deliberately to approximate this condition as closely as possible. Departures from the random spatial economy formulation may only reflect departures in the physical and economic characteristics of a study area from homogeneity, rather than implying different aggregate behaviour patterns. Rayner and Golledge are currently working on the idea of constructing filters which represent the transport pattern of an area, and elements of the physical and social landscape, in order to estimate the degree to which they act as distorting influences on the underlying settlement patterns. Looked at from another point of view, however, one must expect some regularity simply because settlements occupy a finite area. In this context, the centre-satellite model of Section 4.1.2 or the doubly stochastic Poisson processes of Matérn (1960, 1971), Bartlett (1963) and Grandell (1972) may offer more realistic bases for theoretical developments. In the doubly stochastic models, the Poisson process is described by $\lambda$, which is itself generated by a stochastic mechanism. In most practical geographical examples, the model would have to be used with spatially aggregated data. In that case, the Poisson process might give the number of settlements in each area, while the underlying mechanism generating $\lambda$ described the proportion of land suitable for settlement.

The clear implication, however, of the empirical work described in this section is that there is a significant random component in the spacing between settlements. Curry (1964) has argued that settlements will not only have a random component in their spacings, but also in their rank-size characteristics, and we now examine the extent to which this contention is borne out substantively.

### 4.3. *The Size of Settlements*

Several writers (Berry and Garrison, 1958) have noted that if the $n$ cities in a nation are ranked from largest (rank $n$) to smallest (rank 1) in terms of population, and the ranks are plotted against city population size, then the relationship

$$(n-r_i+1)\,p_i^q = k \tag{4.9}$$

appears to hold. Here $r_i$ = the rank of the $i$th city, $p_i$ is its population, and $q$ and $k$ are constants. Various explanations of Zipf's (1949) rank-size rule for cities given in equation (4.9) have been offered (Berry and Garrison, 1958), while rank-size relationships of this form have attracted a good deal of attention from workers in many other fields, for example biology (ranking genera by number of species) and linguistics (ranking words by frequency of occurrence).

The rank-size rule has been reformulated in a slightly different fashion by Cliff *et al.* (1975b). Suppose that we calculate the quantity

$$g_{(i)} = p_i \Big/ \sum_{i=1}^{n} p_i \tag{4.10}$$

and arrange the data so that $g_{(i)}$ is the proportion of the total urban population of a nation in the $i$th smallest city, $i = 1, \ldots, n$. Then Zipf's rank size rule, with $q = 1$, is suggestive of the expressions

$$\begin{aligned} E\{g_{(n)}\} &= k, & E\{g_{(i)}\} &= \frac{k}{n-i+1}, \\ E\{g_{(n-1)}\} &= \frac{k}{2}, & &\vdots \\ E\{g_{(n-i)}\} &= \frac{k}{i+1}, & E\{g_{(1)}\} &= \frac{k}{n}, \end{aligned} \tag{4.11}$$

where $k = 1/\sum_{i=1}^{n} i^{-1}$, and $g_{(n)}$ is the share of the largest city. From equation (4.11)

$$E\{g_{(i)} - g_{(i-1)}\} = \frac{k}{(n-i+1)(n-i+2)}, \quad i \geqslant 2. \tag{4.12}$$

If, however, Curry's suggestion of a random population share size distribution holds, a suitable model may be obtained from Whitworth's (1934) work; see also Kendall and Moran (1963, pp. 28–31). Whitworth took a line of unit length cut at $(n-1)$ points located at random along it. That is, a line of unit length was broken at random into $n$ segments. If the segments are ranked from smallest (1) to largest ($n$), and the $i$th segment is taken as $g_{(i)}$ given by equation (4.10), then the difference between the population shares of the $i$th and $(i-1)$th smallest cities, according to the Whitworth model, is

$$E\{g_{(i)} - g_{(i-1)}\} = \frac{1}{n(n-i+1)}, \quad i = 2, 3, \ldots, n. \tag{4.13}$$

Alternatively, Cohen (1966) argued that there is a threshold minimum share size, $\Delta$ say, for the smallest city; that is, $g_{(1)} \geqslant \Delta$. If there is a threshold, then under Whitworth's random splitting process,

$$E\{g_{(i)}\} = \left\{\left(\frac{1}{n} - \Delta\right) \sum_{r=1}^{i} \frac{1}{n+1-r}\right\} + \Delta, \quad i = 1, 2, \ldots, n. \tag{4.14}$$

The $g_{(i)}$ defined above are referred to as "spacings" in the statistics literature (Pyke, 1965). Cliff *et al.* (1975), using Pyke's paper and Durbin's comments thereon, have developed test procedures which enable the goodness-of-fit of each of the models (Zipf, Whitworth and Cohen) to real data to be evaluated. The three models were applied to nine sets of urban and county population data taken from the (1967) *Redcliffe-Maud Report*. In seven out of the nine cases, the best fit was obtained with the Cohen model, and the observed population shares did not depart significantly from the expected values under the Cohen model. The only exceptions were provided by data sets containing a mixture of different units (such as cities and rural counties). This result again confirms the importance of Curry's idea that random component models may be very valuable in examining aggregate human behaviour.

Cliff and Robson (1975) have tried to relate the above work to the long run maximum entropy notion discussed in Section 4.1. They have taken the rank-size distribution for all urban centres in England and Wales with a population in excess of 2000 for each census from 1801–1911, and have fitted the Whitworth–Cohen models to these data. Some interesting points emerged. First, the observed population share-size distribution was markedly non-random throughout the nineteenth century, with an increasing concentration of the population into the larger cities. The fit of the models became consistently worse with the passage of each census throughout the century. The 1901 census represented a turning point in that the fit of the models to the data improved for the first time, and 1911 registered yet another improvement. We know from the results reported above that by the time of Redcliffe-Maud (1967), the rank-size distribution for administrative regions was largely random. These findings together suggest that, although the Industrial Revolution administered a severe shock to the system, there may be a general tendency towards a dynamic equilibrium, maximum entropy, distribution. Certainly Robson (1973) has argued that city sizes throughout the nineteenth century were in disequilibrium. The Industrial Revolution made entirely new cities, while others failed. Robson goes on to argue that the forces of the Industrial Revolution in this context were largely worked out by about 1900. Conversely, the twentieth century has been dominated by shifts between existing urban centres and centre/suburb relationships, rather than by the birth and death of towns. That is, the system has been in dynamic equilibrium. It will be intriguing to see what the censuses of 1921–71 ultimately reveal in this connection.

### 4.4. *Conclusions*

In this section, we have shown how the analysis of spatial forms using nearest neighbour, quadrat and rank-size methods can yield valuable insights into the processes generating the forms [approach (2) of Section 2.2]. The approach has been illustrated with reference to settlement studies. Despite the limitations of the methods, the general conclusion that there is an important random component in the spacing between, and sizes of, settlements seems reasonable. The reader is asked to bear this conclusion in mind when reading Section 6.5. There, spectral methods are considered as one of a series of spatial modelling techniques, and we discuss some substantive work on settlement patterns using these methods which support the present findings.

## 5. SPATIAL INTERACTION AND DIFFUSION MODELS

### 5.1. *Spatial Interaction*

Given that the spatial dimension is so central to geographical studies, it is both natural and essential that an important part of the model building effort in the subject should have been devoted to the development of models to describe flows between locations. As noted in Section 2.1, the study of interactions between different areas started with Ravenstein's (1885) analysis of migration flows [also considered by Stouffer (1940, 1960) and by Morrill and Pitts (1967)]. Similar models have been developed for journey-to-work trips between zones of a city (Chicago Area Transportation Study, 1960) and shopping trips (Reilly, 1931; Huff, 1963; Lakshmanan and Hansen, 1965). A general review and bibliography is provided by Olsson (1965).

Most of these papers have used the so-called *gravity* (or *potential* or *interaction*) model in some form or another. The standard form of this model is as follows:

$$I_{ij} = A_i B_j P_i^m Q_j^n / d_{ij}^{\alpha}. \tag{5.1}$$

Here, $m$, $n$ and $\alpha$ (usually taken to be $>0$), the $A_i$ and the $B_j$ are parameters. $I_{ij}$ is the interaction or flow between the $i$th and $j$th locations, $P_i$ and $Q_i$ represent, respectively, the total mass to be "transferred" from and to the $i$th location (outflows and inflows of migrants, for example) and $d_{ij}$ represents the distance between the $i$th and $j$th locations.

As an example of this kind of study, we consider the work of Mackay (1958), who wished to measure the effect of political boundaries on spatial interaction. He fitted (5.1) to data on "interactions" (frequency of long-distance telephone calls and marriages) between pairs of English-speaking cities within Quebec province, as opposed to interactions between pairs of English-speaking cities the same distance apart but separated by the Quebec/Ontario border. For this study, he took $A_i B_j = k$ for all $i$ and $j$, $Q_j = P_j$ and $m = n$. The remaining parameters were estimated by ordinary least squares after a logarithmic transform to produce a linear model. The assumption of a multiplicative error term seems more acceptable in this context. Among other things, Mackay found that the "friction of distance" parameter, $\alpha$, was approximately five times greater for interactions between cities across the border than for interactions between cities in Quebec alone.

The "principle of least effort" formulated by Zipf (1949) provided a heuristic basis for gravity-type models, but a proper theoretical framework was lacking until the work of Wilson (1970). Wilson argues that the flows should be arranged so that the information content of the system is minimized, or the entropy, $E$, is maximized. Thus, we have the formulation

$$\text{maximize}\, E = -\sum I_{ij} \log_e I_{ij}, \tag{5.2}$$

subject to various constraints. Here the summation is taken over the $n$ locations. When $P_i$ and $Q_i$ represent total mass, as suggested above, the constraints are

$$\sum_i I_{ij} = Q_j, \quad j = 1, \ldots, n, \tag{5.3}$$

$$\sum_j I_{ij} = P_i, \quad i = 1, \ldots, n, \tag{5.4}$$

and

$$\sum I_{ij} d_{ij} = C, \tag{5.5}$$

where (5.5) represents a budget constraint for "transport costs" (mass $\times$ distance). Maximization of (5.2) subject to (5.3)–(5.5) yields

$$I_{ij} = A_i B_j P_i Q_j \exp(-\alpha d_{ij}). \tag{5.6}$$

The quantities $A_i$ and $B_j$ are functions of the $d_{ij}$ and the Lagrangean multipliers; $\alpha$ is the multiplier corresponding to (5.5). If transport costs are assumed to be a declining function of distance such as

$$\sum I_{ij} \log(d_{ij} + a) = C, \tag{5.7}$$

then model (5.1) is obtained (when $a \to 0$). The reader is referred to Wilson's monograph and the papers cited therein for alternative derivations of the model and a fuller discussion. The formulation of model (5.6) is also in the spirit of Curry's (1964) notion of the random spatial economy, subjected to certain constraints.

### 5.2. *The Diffusion of Innovations*

A second research area in which a substantial amount of model building has been done is the diffusion of innovations. Geographical work in this area stems very much from the pioneering paper of Hägerstrand (1953). Hägerstrand studied the diffusion of some agricultural innovations among farmers in the Asby district of Sweden. For example, he collected data on the spatial pattern of acceptance of a subsidy granted by the Swedish government from 1928 onwards to all farmers of small units (less than 8 hectares of tilled land). The subsidy was granted if they enclosed woodland on their farms and converted it to pasture. Hägerstrand developed a Monte Carlo model to simulate the space–time pattern of acceptance of the innovations by the farmers. The model was based on two main assumptions: (i) the chief mechanism by which information about an innovation spreads through the population is by oral communication at pairwise meetings of adopters and potential adopters of the innovation; (ii) the probability of such person-to-person contacts has a strong inverse relationship with the geographical distance between the teller and the receiver. Hägerstrand also considered briefly the problem of evaluating the spatial and aspatial goodness-of-fit of the simulated to the observed patterns (maps).

#### 5.2.1. *Later developments*

An extensive review of the work done in geography and in other disciplines since Hägerstrand's paper is provided by Brown and Moore (1969). The research effort has focused largely upon the identification of diffusion processes. Attempts at modelling these processes have been mainly modifications of the basic Hägerstrand model rather than new departures. A major reason for this is that while it is often possible to write down the equations for quite sophisticated spatial models, general solutions are rarely possible. This fact has been noted particularly in the epidemiological literature (Bailey, 1957, 1967; Bartlett, 1960; Bartholomew, 1973).

The structure of many spatial diffusion processes for innovations has been summarized in Casetti (1969a), Casetti and Semple (1969) and Hudson (1969). First, the process of spread of information about an innovation in the plane postulated by Hägerstrand, that of person-to-person communication, seems to be substantially correct. However, the two-step hypothesis of communication flow suggested by the sociologists (Katz and Lazarsfeld, 1955), while recognizing that personal communication is the principal medium, states that mass media may be responsible for causing initial awareness. The importance of person-to-person contact in the spread of information means, as Hägerstrand (1953) has noted, that the spatial pattern of acceptance of an innovation is likely to be highly contagious, producing Hägerstrand's so-called "neighbourhood effect" of clustered growth. Work by Cliff (1968) and McClellan (1973) suggests, however, that the neighbourhood effect may be a function of scale. Viewed at the macro-level, personal communication fields may decay regularly with distance as postulated by Hägerstrand. At the micro-scale, though, the spatial pattern of individual kinship and acquaintance circles can be much more irregular. The spatial pattern of innovation acceptance may also be modified away from strict contagious growth by the tendency of innovations to diffuse downward through the central place hierarchy. Thus Hägerstrand (1967; cited in Hudson, 1969) states:

> A closer analysis shows that the spread along the initial "frontier" is led through the urban hierarchy. The point of introduction in a new country is its primate city, sometimes some other metropolis. Then centres next in rank follow. Soon, however, this order is broken up and replaced by one where the neighbourhood effect dominates over the pure size succession.

This idea is explored further by Pedersen (1970).

Over time, the proportion of adopters in the population at time $t$, $p_t$, often starts slowly, then increases rapidly as the innovation "takes off", and finally levels out as saturation is approached. This empirical result has led to models for the rate of change in $p_t$ with respect to time. For example, a model which incorporates the variations in growth rate, $r_t$, just mentioned is

$$r_t = \frac{d}{dt} p_t = -bp_t(1-p_t), \tag{5.8}$$

leading to the familiar logistic model

$$p_t = \{1+\exp(a-bt)\}^{-1}. \tag{5.9}$$

Equation (5.9) has been widely used as a simple growth model in several disciplines—see Casetti and Semple (1969) for example.

A major problem, discussed by Hudson (1969), is the implicit assumption underlying (5.9) of homogeneous mixing of adopters and potential adopters. The quantity, $p_t(1-p_t)$, then represents the probability that a random meeting between two individuals is between an adopter and a potential adopter, while the parameter $b$ represents the rate at which meetings take place. However, this assumption is in conflict with the neighbourhood effect discussed earlier. A possible extension of the model, which allows homogeneous mixing within $n$ regions ($j = 1, \ldots, n$) but less mixing between regions, is

$$r_{jt} = -(1-p_{jt})\left\{b_j\, p_{jt} + \sum_{i\neq j} c_{ij}\, p_{it}\right\}. \tag{5.10}$$

Equation (5.10) reduces to (5.8) for each $j$ when all $c_{ij} = 0$.

A continuing difficulty, noted by Hägerstrand (1967), Brown and Moore (1969) and Casetti (1969b), is how to evaluate the spatial goodness-of-fit between real world patterns and realizations of any diffusion model. The difficulty arises because the nature of the diffusion process implies that neighbouring regions will produce counts of numbers of adopters that are positively correlated. Cliff and Ord (1973, Chapter 4) have suggested a testing procedure based on the approach of Hope (1968). The method involves the following steps.

*Step* 1. Generate $m$ independent realizations of the diffusion model, and from these *and the real world map* compute an "average expected" map by averaging over the $(m+1)$ realizations.

*Step* 2. For each model map and for the real world map, compute a goodness-of-fit measure between that map and the average expected map. For example, one of the spatial autocorrelation measures discussed in Section 6.2, Pearson's product moment correlation coefficient or sum of squared differences could be used. Under the null hypothesis, these $(m+1)$ measures will be identically distributed and equicorrelated.

*Step* 3. Rank the $(m+1)$ measures and reject $H_0$ at the $100\{(j+1)/(m+1)\}$ per cent level if the measure between the real world and average expected map has rank $(m-j+1)$ or worse (one-tailed test). Rules for two-tailed tests may be formulated in a similar manner. In ranking the measures, if a correlation coefficient is used, call the highest positive value rank 1; if the sum of squared differences is used, call the smallest sum rank 1.

## 6. SPATIAL MODELLING AND FORECASTING

The natural framework for modelling in geography is the spatial–temporal process, since we need to capture both the components of evolution through time and spatial dependence. The purpose of this section is to outline possible approaches to spatial modelling and to indicate some of the statistical difficulties involved. Not surprisingly, our methods draw heavily on the time series literature, although we also consider purely spatial processes, since either

(a) the process may have converged to an apparent equilibrium, and observations are available only for this equilibrium state (a situation common in geology); or
(b) data are available for only a single time period (as tends to occur for data obtained from special purpose or infrequent surveys, such as the Population Census).

We consider the specification of purely spatial models in Section 6.1, and explore the inferential problems related to such models in Sections 6.2 and 6.3. In Section 6.4, we discuss spatial–temporal models, and conclude with a review of spectral methods in Section 6.5.

### 6.1. *Purely Spatial Data*

A basis for spatial model building was spelled out in Section 3.1. Two principal approaches are possible: either we could specify the mean and covariance structures, which will in turn yield a model, or we could specify the model and then generate the mean and covariance structures. Implicitly, our discussion is restricted to consideration of the first two moments ("wide sense"), unless we add assumptions of normality. The two approaches may be illustrated with an example from time series analysis. We might specify a discrete time zero mean process with the covariance structure

$$\operatorname{cov}(Y_t, Y_{t+\tau}) = \sigma_y^2 \rho^{|\tau|} \quad (\tau = 0, \pm 1, \pm 2, \ldots). \tag{6.1}$$

Equation (6.1) gives rise to the spectrum

$$g(z) = \sigma^2\{1+\rho^2-\rho(z+z^{-1})\}^{-1}, \tag{6.2}$$

where $\sigma^2 = \sigma_y^2(1-\rho^2)$ and $z = \exp(i\omega)$. Now $g(z)$ factorizes into $\sigma^2[(1-\rho z)(1-\rho z^{-1})]$, so that it corresponds to the model

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \tag{6.3}$$

where $\varepsilon_t$ is a random disturbance term with zero mean and $E(\varepsilon_t, \varepsilon_{t+\tau}) = \sigma^2$ when $\tau = 0$ (but $=0$ otherwise). That is, $\varepsilon_t$ is "white noise". Equally, the first-order Markov scheme given in (6.3) could be specified, and it would lead to the covariance structure given by (6.1). The choice of model building approach is irrelevant in this case. Unfortunately, as Whittle (1954) has demonstrated, things are not so easy when we attempt symmetric formulations in the plane. For example, consider variates located at the vertices of a square grid, and postulate the model, when $E(Y_{rc}) = 0$,

$$Y_{rc} = \rho(Y_{r-1,c} + Y_{r+1,c} + Y_{r,c-1} + Y_{r,c+1}) + \varepsilon_{rc}. \tag{6.4}$$

The subscripts $r$ and $c$ in equation (6.4) identify the vertex located in the $r$th row and $c$th column of the grid. Then, ignoring boundary problems, we find that the spectrum is of the form

$$g(z) = \sigma^2\{h(z)\}^2 = \sigma^2\{1-\rho(z_1+z_2+z_1^{-1}+z_2^{-1})\}^{-2}, \tag{6.5}$$

where $z_j = \exp(i\omega_j)$, $j = 1, 2$. The bilateral dependence in model (6.4) is reflected in the lack of independence between the $Y$ variables on the right-hand side and $\varepsilon_{rc}$. A natural alternative approach would be to consider the spectrum of the form $g(z) = \sigma^2 h(z)$. This spectrum gives rise to *conditional expectation* schemes of the form

$$E\{Y_{rc}|\text{values of } Y_{ij} \text{ at all other locations}\} = \rho(y_{r-1,c}+y_{r+1,c}+y_{r,c-1}+y_{r,c+1}). \tag{6.6}$$

This approach has several attractive features and is fully described by Besag (1974); Besag's paper also contains a useful bibliography on this area.

Given the similar but conceptually quite different structure of the models, it is of interest to ask how each might arise as a cross-sectional representation of a spatial–temporal process. Ord (1974), in the discussion on Besag's paper, has noted that when $Y_{rc}(t)$, the value at time $t$, depends only upon *past* values of the other $Y$ variates, then a conditional expectations model results (cf. Bartlett, 1971). However, when *simultaneous* dependence, possibly upon other variables, is allowed, joint models such as (6.4) result. The joint models have found greater favour in the econometric literature, but the conditional specification deserves careful consideration.

We now turn to questions of inference for spatial models; the statistical procedures are similar in form for both the joint and the conditional versions.

### 6.2. *Tests for Spatial Dependence*

If a pattern of spatial dependence is postulated, as in equations (6.4) and (6.6), we need a procedure for testing whether or not such dependence is present. We refer to these methods as tests for spatial autocorrelation. Formally, we wish to test the null hypothesis $H_0$: $\rho = 0$ against alternatives such as $H_1$: $\rho \neq 0$, where $\rho$ is defined by equations such as (6.4) or (6.6).

When data are available only in binary form, we may use the join count statistics originally proposed by Moran (1948) and Krishna Iyer (1949). For example, if the $i$th county is coded black, $B\,(Y_i = 1)$, or white, $W\,(Y_i = 0)$, then we might use the statistics

$$BB = \tfrac{1}{2}\textstyle\sum_{(2)} w_{ij}y_i y_j \quad \text{and} \quad BW = \tfrac{1}{2}\textstyle\sum_{(2)} w_{ij}(y_i - y_j)^2, \tag{6.7}$$

where $\sum_{(2)} \equiv \sum_{i=1}^n \sum_{j=1}^n \,(i \neq j)$ and $\{w_{ij}\}$ indicate the *weights* associated with pairs of counties. The weights derive from the form of the alternative hypothesis, such as

$$Y_i = \rho \sum_j w_{ij} Y_j + \varepsilon_i \quad (i = 1, \ldots, n) \tag{6.8}$$

or the appropriate conditional version. The statistics $BB$ and $BW$ are the (weighted) numbers of $B$–$B$ and $B$–$W$ links in the study area. Intuitively, it can be seen that a "lot" of $BB$ and a "few" $BW$ links imply spatial clustering, while the reverse implies more or less uniform spacing. The test statistics have been extended to $k$, rather than two, classes by Krishna Iyer (1949) and Cliff (1969).

In general, there is no uniformly most powerful test statistic for this situation, but when normality is assumed, we can obtain a locally efficient statistic as $\rho \to 0$ for either the joint or the conditional specification. This statistic is of the form

$$I = (n/W) \sum_{(2)} w_{ij} z_i z_j \Big/ \sum_{i=1}^{n} z_i^2. \tag{6.9}$$

where $z_i = y_i - \bar{y}$, $n\bar{y} = \sum y_i$ and $W = \sum_{(2)} w_{ij}$. It was suggested originally by Moran (1950), for the situation where $\{w_{ij} = 0 \text{ or } 1\}$, and is akin to the serial correlation coefficient in time series analysis. Geary (1954) proposed a similar statistic, but based upon the weighted sums of squared differences, $\sum_{(2)} w_{ij}(y_i - y_j)^2$. Following the approach of Durbin and Watson (1950, 1951, 1971), Cliff and Ord (1973, Chapters 2 and 5) have shown that $I$ is asymptotically normally distributed when $\rho = 0$, under mild conditions. They also give the moments for various assumptions, and extend the analysis (Cliff and Ord, 1972a) to cover the situation where the $\{y_i\}$ are least squares regression residuals, rather than original observations. Cliff and Ord (1975a) give both theoretical and Monte Carlo comparisons of the power of the various test statistics which generally confirm the superiority of $I$ over its competitors, including the Geary form.

The choice of weights is much more subjective than for the time series case. Regular lattices suggest natural steps for one, two,... units from the reference cell, but there is no clear guide for irregular lattices. When spatial dependence decays rapidly with the distance, $d_{ij}$, between locations $i$ and $j$, weights of the form $w_{ij} \propto d_{ij}^{-\alpha}$ or $w_{ij} \propto \exp(-\alpha d_{ij})$ seem reasonable (cf. Whittle, 1956); or, if we are dealing with areal aggregates, $w_{ij} \propto$ length of common boundary between $i$ and $j$. The notion of a decline in interaction with distance is commonly accepted (cf. Zipf, 1949), so that such a formulation is reasonable, but is by no means necessary. At this stage, all that can be said is that the $\{w_{ij}\}$ must be specified in accordance with the pattern of autocorrelation the researcher wishes, *a priori*, to examine. It is only through the specification of these weights that the map structure enters explicitly into the analysis.

### 6.2.1. *Spatial dependence in measles outbreaks*

The $I$ statistic can be used to construct spatial correlograms; that is, to determine how the dependence between variate values decays over space. Cliff *et al.* (1975) examined a 222-week time series of the number of measles cases reported per week from 1966 (week 40) to 1970 (week 52) for each of the 27 General Register Office (GRO) districts of Cornwall. The writers scaled the variable to give the number of measles cases per 1,000 children under 15 years of age (the best measure in the available data of the size of the susceptible population). They then determined, for each of the 222 weeks, the level of spatial autocorrelation on the variable between GROs which were first, second,..., eighth nearest neighbours in a graph theoretic sense. The coefficient, $I$, given in equation (6.9) was used with $w_{ij}(r) = 1$ if the GROs $i$ and $j$ were $r$th nearest neighbours and $w_{ij}(r) = 0$ otherwise. The average correlograms for weeks 1–50 and 186–204 are given in Fig. 1. These sets of weeks coincide broadly with the two major measles epidemics which occurred in the South-west in the period studied, while weeks 51–185 represent the "fade-out" period between the epidemic peaks. $I$ is plotted as a standard deviate on the ordinate of each graph. The spatial lag is plotted on the abscissa. In Table 1 we give the number of positive and negative standard deviates for $I$ at each spatial lag. The basic pattern is fairly clear. For weeks 1–50, it is evident that, although many of the individual $I$ values were not statistically significant, positive levels of autocorrelation predominate at spatial lags 1, 6 and 8 and negative levels at lags 2, 3 and 4. The positive spatial autocorrelation at lag 1 and negative autocorrelation at lags 2–4 implies that the measles outbreaks are, as suggested by Haggett (1972), clustered spatially. That is, if a GRO has an outbreak/no outbreak, contiguous GROs are likely to behave similarly. The puzzling feature of the average correlogram is the positive spatial autocorrelation at lags 6 and 8.

To help interpret this, the authors called all GROs which were Rural Districts, "rural", and all GROs which were Urban Districts or Metropolitan Boroughs,

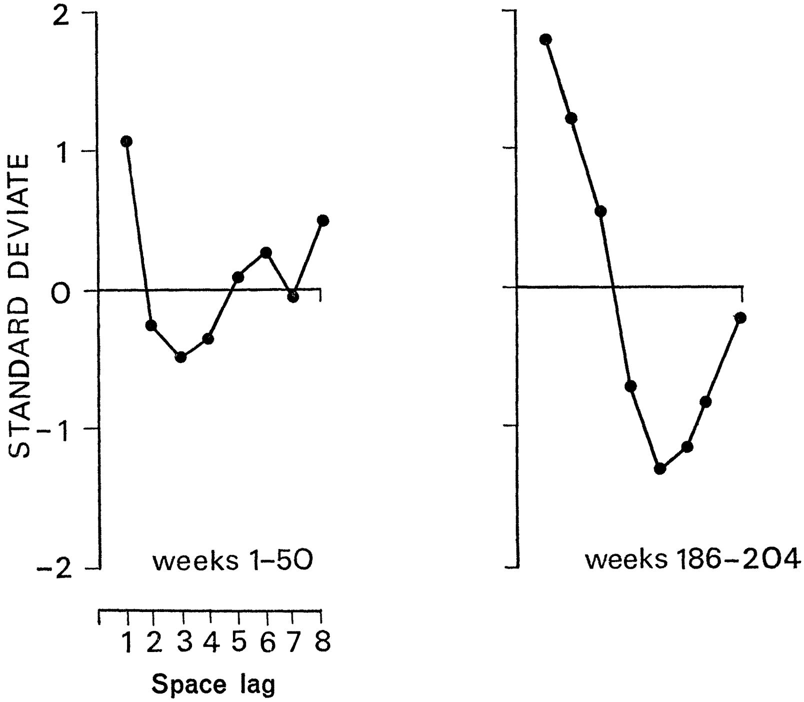


FIG. 1. Average spatial correlograms for measles outbreaks in Cornwall.

TABLE 1

*Number of positive and negative standard deviates for I at each spatial lag*

| *Item* | *Spatial lag* 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| *Weeks* 1–50 | | | | | | | | |
| Positive standard deviates | 43 | 20 | 14 | 19 | 24 | 31 | 25 | 47 |
| Negative standard deviates | 7 | 30 | 36 | 31 | 26 | 19 | 25 | 3 |
| *Weeks* 186–204 | | | | | | | | |
| Positive standard deviates | 19 | 15 | 15 | 4 | 2 | 3 | 4 | 8 |
| Negative standard deviates | 0 | 4 | 4 | 15 | 17 | 16 | 15 | 11 |

"urban". The numbers of urban–urban, rural–rural and urban–rural links at each spatial lag, 1–8, were then determined. These counts are given in Table 2, along with the expected numbers in brackets under the assumption of independence between link type and spatial lag.

TABLE 2

*Expected and actual numbers of links at each spatial log for Cornish GROs*

| | *Link type* | | | |
|---|---|---|---|---|
| *Spatial lag* | *Urban–urban* | *Rural–rural* | *Urban–rural* | *Totals* |
| 1 | 1 (13·6) | 13 (4·5) | 21 (16·9) | 35 |
| 2 | 20 (28·4) | 14 (9·3) | 39 (35·3) | 73 |
| 3 | 33 (32·3) | 7 (10·6) | 43 (40·1) | 83 |
| 4 | 31 (24·5) | 6 (8·1) | 26 (30·4) | 63 |
| 5 | 20 (17·9) | 4 (5·9) | 22 (22·2) | 46 |
| 6 | 18 (13·2) | 1 (4·3) | 15 (16·4) | 34 |
| 7 | 11 (5·8) | 0 (1·9) | 4 (7·2) | 15 |
| 8 | 3 (1·2) | 0 (0·4) | 0 (1·4) | 3 |
| Totals | 137 | 45 | 170 | 352 |

It is evident from Table 2 that spatial lags 1 and 2 are dominated by urban–rural and rural–rural links, whereas lags 4–8 are dominated by urban–urban links. In addition, lags 6–8 include predominantly those GROs which are linked in an east–west direction by the main transport arteries. A picture therefore emerges of similar levels of measles cases in (1) non-contiguous urban areas and (2) contiguous rural–urban and rural–rural districts. A possible interpretation of this pattern is to postulate initial outbreaks of measles in an epidemic in urban areas (hence the positive spatial autocorrelation at lags 6 and 8). This could be called a central place effect. This is supported by a spread of the disease from the towns into surrounding rural areas by a spatial diffusion process. This would account for the positive spatial autocorrelation at lag one where urban–rural and rural–rural links predominate.

Turning to weeks 186–204, the spatial clustering of measles outbreaks is again confirmed by the positive autocorrelation at lags 1–3, but there is negative autocorrelation at lags 4–8. This suggests that the central place effect was less important in the second epidemic than in the first.

### 6.3. *Estimation for Spatial Models*

The analysis of spatial data may have any of three (overlapping) objectives: smoothing, interpolation or modelling. Typically, for smoothing or trend elimination, we assume that the variate is non-stationary in the mean, but that the covariance structure is the simplest possible (uncorrelated, homoscedastic disturbance terms). Frequently, the analysis is handled by standard regression methods, such as the trend surface techniques already mentioned in Section 2.1. An alternative approach is the use of spatial moving, or cascaded, averages (Curry, 1970, 1971), which is particularly useful for regularly spaced data. This technique has been used with considerable success in time series analysis (Kendall and Stuart, 1966, Chapter 47) and is equivalent to the local fitting of low-order polynomials.

Interpolation methods have two principal uses, as follows:

(i) the estimation of missing values;

(ii) the estimation of the mean value for a given volume or area.

The second question assumes considerable practical importance in mining, when it is desired to estimate the ore content of a volume of rock which must be mined *en bloc* or not at all. For a definitive discussion, see Matheron (1970, 1971). The problem may be formulated in the following way. Given $n$ observations $\{y_j$ at location $l_j\}$, find an estimator of the quantity $U$ of the form

$$\hat{U} = \sum_{j=1}^{n} \lambda_j y_j = \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{y}, \tag{6.10}$$

where the weights $\{\lambda_j\}$ are at choice. The standard approach (known as *Kriging* in the geological literature) is to assume that $Y$ is a mean stationary spatial process, and to derive the best linear unbiased estimator (BLUE), "best" meaning minimum variance in this context. That is, we use

$$\boldsymbol{\lambda} = \lambda \boldsymbol{\Sigma}^{-1} \mathbf{1}, \tag{6.11}$$

where $\boldsymbol{\Sigma}$ has elements $\Sigma_{jk} = \operatorname{cov}(Y_j, Y_k)$, $\lambda = (\mathbf{1}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{1})^{-1}$ and $\mathbf{1}^{\mathrm{T}} = (1, 1, \ldots, 1)$. It follows that $\operatorname{var}(\hat{U}) = \lambda$. Unfortunately, $\boldsymbol{\Sigma}$ is usually unknown and must be estimated from the sample data. The usual Kriging approach assumes that $Y$ is a weakly isotropic process, and takes the covariance function to be a simple function in $h$, such as a low-order polynomial, which can be estimated from the data. For fuller details, including extensions to anisotropic models and processes with "drift" (non-stationarity in the mean), see Matheron (1969) and Huijbreghts (1975).

### 6.3.1. *Linear spatial models*

For modelling purposes, the natural starting point is the specification of a model such as (6.4) or (6.6); see Whittle (1954) and Besag (1974). Following Box and Jenkins (1970, Chapter 3), we can readily incorporate moving average components into the joint model, but this does not appear to be meaningful for the conditional form. The general joint model may be written as

$$\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} + \lambda \mathbf{A} \boldsymbol{\varepsilon}, \tag{6.12}$$

where $\mathbf{W}$ and $\mathbf{A}$ represent matrices of known weights and $\mathbf{X}$ is a matrix of known regressor variables; $\rho$, $\lambda$ and $\boldsymbol{\beta}$ are unknown parameters.

$\lambda = 0$ (*the autoregressive model*)

If $\boldsymbol{\varepsilon}$ is $N(\mathbf{0}, \sigma^2 \mathbf{I})$, the maximum likelihood estimators are (Mead, 1967, 1971; Ord, 1975)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} \hat{\mathbf{z}}, \tag{6.13}$$

$$\hat{\sigma}^2 = \hat{\mathbf{z}}^{\mathrm{T}} \mathbf{M} \hat{\mathbf{z}} / n \tag{6.14}$$

and that value of $\rho$, $\hat{\rho}$ say, which minimizes

$$\hat{\sigma}^n |\mathbf{I} - \rho \mathbf{W}|^{-1}, \tag{6.15}$$

where $\hat{\mathbf{z}} = (\mathbf{I} - \rho \mathbf{W}) \mathbf{y}$ and $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}}$. The best computational form for (6.15) is obtained by rewriting the determinant as $\prod_{i=1}^{n} (1 - \rho \omega_i)$, where $\{\omega_i\}$ are the eigenvalues of $\mathbf{W}$. See Ord (1975) for further details. If the model is extended to higher order "spatial lags", which give rise to determinants such as $|\mathbf{I} - \rho_1 \mathbf{W} - \rho_2 \mathbf{W}^2|$, the same approach can be used since the determinant can be rewritten as $\prod_{i=1}^{n}(1-\rho_1\omega_i-\rho_2\omega_i^2)$. We note that $\mathbf{W}^2$ may include some circular routes $(i \to j \to i)$. These circular routes do not affect the validity of the maximum likelihood estimators, but if $\boldsymbol{\Delta}$ is the diagonal matrix containing the leading diagonal of $\mathbf{W}^2$, the alternate version

$$(\mathbf{I}-\rho_2\boldsymbol{\Delta})\mathbf{Y} = \rho_1\mathbf{W}\mathbf{y}+\rho_2(\mathbf{W}^2-\boldsymbol{\Delta})\mathbf{Y}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \tag{6.16}$$

seems preferable as a basis for interpretation. Similar corrections can be made for higher order lags (see Cliff and Ord, 1973, Chapter 8, for a discussion of this topic in a different context). Further computational procedures are described in the Appendix. Examples are given by Besag (1974) and Ord (1975).

### 6.4. *Spatial–Temporal Models*

When data are available for several time periods, models which recognize spatial dependence as lagged in time can be formulated as mixed autoregressive, moving average, regression models (cf. Box and Jenkins, 1970, Chapter 4). Such models are very flexible, and have been used to good effect by Tobler (1967, 1969b, 1970a) to estimate the linear spatial transfer function which best transforms a map at time $t$ into that at $t+1$. A variant of the usual regression approach is that of Tinline (1971) who used minimum mean absolute deviation (MAD) estimators in a similar analysis of Hägerstrand's (1953) data (cf. Section 5.2).

If simultaneous dependence is allowed, we have a model such as

$$\mathbf{Y}_t = \mathbf{G}\mathbf{Y}_t+\mathbf{B}\mathbf{X}_t+\boldsymbol{\varepsilon}_t \quad (t=1,\ldots,T), \tag{6.17}$$

where the general notation of equation (6.12) has been followed, except that $\mathbf{B} = (\boldsymbol{\beta}_1^{\mathrm{T}},\ldots,\boldsymbol{\beta}_n^{\mathrm{T}})$ is a matrix of parameters, $\mathbf{G}$ is a linear function of one or more parameters and $\mathbf{X}_t$ may include temporally lagged values of the dependent variables. Two special cases will be briefly mentioned:

(i) $\mathbf{G} = \rho\mathbf{W}$: the general approach of Section 6.3.1 may be followed, although the determinant in the likelihood function will now appear $T$ times;

(ii) a more general $\mathbf{G}$, but with sufficient *a priori* structure to ensure that all the parameters are estimable, or that all equations are identified (Johnston, 1972, p. 352; Theil, 1971, p. 443). Standard simultaneous equation econometric methods might then be applied.

*Example*

Suppose that $\mathbf{G} = (\rho_1\mathbf{w}_1,\ldots,\rho_n\mathbf{w}_n)$, where the $\{\mathbf{w}_i\}$ are known column vectors. Thus the equation for the $j$th area becomes

$$\mathbf{Y}_j = \rho_j\mathbf{Y}\mathbf{w}_j+\mathbf{X}_j\boldsymbol{\beta}_j+\boldsymbol{\varepsilon}_j, \tag{6.18}$$

where $\mathbf{Y}_{(T\times n)} = (\mathbf{Y}_1,\ldots,\mathbf{Y}_n)$ and the $\mathbf{X}_j, \boldsymbol{\varepsilon}_j$ are grouped by area rather than by time period. $\mathbf{X}_j$ includes only those predetermined variables for which the corresponding elements in $\boldsymbol{\beta}_j$ are assumed to be non-zero. $\mathbf{X}$ denotes the complete matrix of predetermined variables. The two-stage least squares estimators are given by

$$\begin{pmatrix} \tilde{\rho}_j \\ \tilde{\boldsymbol{\beta}}_j \end{pmatrix} = \begin{pmatrix} \mathbf{z}_j^{\mathrm{T}}\mathbf{z}_j & \mathbf{w}_j^{\mathrm{T}}\mathbf{y}_j^{\mathrm{T}}\mathbf{X}_j \\ \mathbf{X}_j^{\mathrm{T}}\mathbf{y}\mathbf{w}_j & \mathbf{X}_j^{\mathrm{T}}\mathbf{X}_j \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{z}_j^{\mathrm{T}}\mathbf{y}_j \\ \mathbf{X}_j^{\mathrm{T}}\mathbf{y}_j \end{pmatrix}, \tag{6.19}$$

where $\mathbf{z}_j = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}\mathbf{w}_j$ and $\mathbf{y}$ denotes the observations on $\mathbf{Y}$. Since $w_{jj}=0$, equation (6.19) is in accordance with standard notation (Johnston, 1972, pp. 380–2).

A practical difficulty in the application of (6.19) is the potential size of $\mathbf{X}^T\mathbf{X}$. In econometric work this has often been resolved by the use of principal components. For a critique and an alternative approach, see Theil (1971, pp. 532–6).

Another line of development which is likely to attract increasing attention is the use of variable parameter models (Harrison and Stevens, 1971; Mendel, 1973). Cliff and Ord (1971) employed a simple regression procedure of this type which gave encouraging results. Also, in a later paper (Cliff and Ord, 1972b), the authors used an analysis of covariance approach, breaking the data down into temporal, spatial and interaction components. Such methods often allow a mass of data to be simplified, with little or no loss of information.

### 6.5. *Spectral Methods*

Spatial spectral methods have been used by geographers on only a few occasions (Rayner, 1971). However, such methods represent a powerful tool when the assumption of spatial stationarity (Section 3.1) can be sustained. Spectral methods may be used for either point patterns (Bartlett 1963, 1964, 1972) or for data recorded on a variable at regular intervals.

A nice illustration of the use of such methods and of some of the problems involved is given by Tobler (1969a). He took U.S. Highway 40 from Baltimore to San Francisco and recorded smoothed population densities (suitably defined) at 1 mile intervals. The resulting spectral density functions showed high power at low frequencies, but a rapid tailing off. However, Rayner and Golledge (1973), in a re-examination of Tobler's data, suggest that when unsmoothed population data are used, the observed spectrum is consistent with a purely random process. In addition, Rayner and Golledge examined the point pattern of settlement centres and found evidence of regular spacing, with greater power at higher frequencies. This evidence suggests that distance is an important factor in locating centres (as indicated by standard settlement theory), but that it is relatively unimportant in determining the subsequent size of such settlements. These findings are not inconsistent with the discussion of Sections 4.2 and 4.3. Tobler's original results, with high power at low frequencies, may have been partly due to spatial non-stationarity in the data, since there is a greater degree of settlement on the seaboards than elsewhere. However, the later results seem free from such criticisms.

Even when spatial stationarity is implausible, progress may still be possible, as shown by Granger (1969). Starting with a time series for each of several regions, Granger computed cross and partial spectra for each (pair of) series. The degree of interdependence between different regions can be explored, in particular with regard to the distance between regions. The results of Granger's analysis are outlined in Section 3.1.

While for economic data the length of series and assumptions of stationarity often rule out spectral methods as viable alternatives, in other areas, the possibility of large regularly spaced data sets makes the approach very attractive, as in the analysis of a $120 \times 120$ grid of settlement data for Iowa reported by Rayner and Golledge (1973). These data showed a rather flat spectrum, suggesting a random process.

## 7. CONCLUSION

In this paper, we have supported the view that a central concern of the geographer is to describe spatial patterns and to identify the processes producing those patterns.

Such identification should then permit us to forecast. We have argued that statistical methods have been used in three main ways to further these aims: (1) the use of classical (aspatial) methods to develop inductive theories about spatial processes; (2) the development of special-purpose techniques for spatial pattern description, in the hope that the results obtained from the application of such techniques will yield insights into the process producing the patterns (form to process studies); and (3) formal modelling of spatial processes.

We discussed in Section 3 several properties of geographical data which became evident in connection with approach (1) that hinder the application of statistical methods to spatial data. Non-stationarity and irregularly located datum points were particularly stressed. In Section 4, we considered nearest-neighbour, quadrat count and rank-size methods as examples of the special-purpose techniques for pattern description, and showed how, despite their limitations, they were highly suggestive of a strong random component in settlement spacing and sizes (form to process studies). Finally, we examined formal model building in geography in Sections 5 and 6. The specific cases of spatial interaction and diffusion models were considered in Section 5, while more general spatial model building strategies were discussed in Section 6. If the material we have presented encourages closer collaboration between the statistician and the geographer in an attempt to solve some of the problems we have presented, it will have served its purpose.

## ACKNOWLEDGEMENTS

The authors are very grateful both to the referees and to several colleagues, notably Dr B. T. Robson of Cambridge University, for their comments on an earlier version of this paper. Remaining errors and omissions are, of course, our responsibility alone.

Part of this work was carried out under a Social Science Research Council grant, and we should like to thank the Council for their support.

## REFERENCES

BAILEY, N. T. J. (1957). *The Mathematical Theory of Epidemics*. London: Griffin.
—— (1967). The simulation of stochastic epidemics in two dimensions, In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4 (J. Neyman, ed.), pp. 237–257. Berkeley: University of California Press.
BARTHOLOMEW, D. J. (1973). *Stochastic Models for Social Processes*, 2nd ed. New York: Wiley.
BARTLETT, M. S. (1960). *Stochastic Population Models in Ecology and Epidemiology*. London: Methuen.
—— (1963). The spectral analysis of point processes. *J. R. Statist. Soc.* B, **25**, 264–296.
—— (1964). The spectral analysis of two-dimensional point processes. *Biometrika*, **51**, 299–311.
—— (1971). Physical nearest neighbour models and non-linear time series. *J. Appl. Prob.*, **8**, 222–232.
—— (1972). Physical nearest neighbour models and non-linear time series II: Further discussion of approximate solutions and exact equations. *J. Appl. Prob.*, **9**, 76–86.
BERRY, B. J. L. (1971). Problems of data organisation and analytical methods in geography. *J. Amer. Statist. Ass.*, **66**, 510–523.
BERRY, B. J. L. and GARRISON, W. L. (1958). Alternative explanations of urban rank size relationships. *Ann. Ass. Amer. Geogr.*, **48**, 83–91.
BERRY, B. J. L. and PRED, A. (1961). *Central Place Studies: A Bibliography of Theory and Applications*. Philadelphia: Regional Science Research Institute. (Bibliography Series No. 1.)
BESAG, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc.* B, **36**, 192–236.
BESAG, J. E. and GLEAVES, J. T. (1974). On the detection of spatial pattern in plant communities. *Bull. Int. Statist. Inst.*, **45**, Book 1, 153–158.

BOX, G. E. P. and JENKINS, G. M. (1970). *Times Series Analysis, Forecasting and Control.* San Francisco: Holden-Day.

BROWN, L. A. and MOORE, E. G. (1969). Diffusion research in geography: A perspective. In *Progress in Geography*, Vol. 1 (C. Board *et al.*, eds), pp. 119–157.

BRUSH, J. E. (1953). The hierarchy of central places in Southwestern Wisconsin. *Geogr. Rev.*, **43**, 380–402.

BURTON, I. (1963). The quantitative revolution and theoretical geography. *Canad. Geogr.*, **7**, 151–162.

CASETTI, E. (1969a). Why do diffusion processes conform to logistic trends? *Geogr. Anal.*, **1**, 101–105.

—— (1969b). *Innovation Diffusion as a Spatial Process*, by Torsten Hägerstrand, translated by A. Pred. A Review. *Geogr. Anal.*, **1**, 318–230.

CASETTI, E. and SEMPLE, R. K. (1969). Concerning the testing of spatial diffusion hypotheses. *Geogr. Anal.*, **1**, 254–259.

CATANA, A. J. Jr. (1963). The wandering quarter method of estimating population density. *Ecology*, **44**, 349–360.

CHICAGO AREA TRANSPORTATION STUDY (1960). *Final Report* (in 3 volumes). Printed by authority of the State of Illinois.

CHORLEY, R. J., STODDART, D. R., HAGGETT, P. and SLAYMAKER, H. O. (1966). Regional and local components in the areal distribution of surface sand facies in the Breckland, Eastern England. *J. Sediment. Petrol.*, **36**, 209–220.

CLARK, P. J. and EVANS, F. C. (1954). Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology*, **35** 23–30.

CLIFF, A. D. (1968). The neighbourhood effect in the diffusion of innovations. *Transactions and Papers*, Institute of British Geographers, **44**, 75–84.

—— (1969). Some measures of spatial association in areal data. Ph.D. Thesis, University of Bristol.

CLIFF, A. D., HAGGETT, P., ORD, J. K., BASSETT, K. A. and DAVIES, R. B. (1975b). *Elements of Spatial Structure.* London: Cambridge University Press.

CLIFF, A. D., MARTIN, R. L. and ORD, J. K. (1975b). A test for spatial autocorrelation based upon a modified $\chi^2$ statistic. *Transactions and Papers*, Institute of British Geographers (in press).

CLIFF, A. D. and ORD, J. K. (1971). A regression approach to univariate spatial forecasting. In *Regional Forecasting* (M. D. I. Chisholm *et al.*, eds), pp. 47–70. London: Butterworth.

—— (1972a). Testing for spatial autocorrelation among regression residuals. *Geogr. Anal.*, **4**, 267–284.

—— (1972b). Regional forecasting with an application to school leaver patterns in the United Kingdom. In *International Geography* (W. P. Adams and F. M. Hilleiner, eds), pp. 956–958. Montreal: Proceedings of 22nd I.G.U. Conference.

—— (1973). *Spatial Autocorrelation.* London: Pion.

—— (1975a). The choice of a test for spatial autocorrelation. In *Proceedings of the NATO Institute on the Display and Analysis of Spatial Data* (J. Davis and M. McCullagh, eds), pp. 54–77. New York: Wiley.

—— (1975b). Space-time modelling with an application to regional forecasting. *Transactions and Papers*. Institute of British Geographers (in press).

CLIFF, A. D. and ROBSON, B. T. (1975). The random spatial economy and the rank-size rule (to appear).

COHEN, J. E. (1966). *A Model of Simple Competition.* Cambridge, Mass.: Harvard University Press.

COTTAM, G. and CURTIS, J. T. (1949). A method for making rapid surveys of woodlands by means of pairs of randomly selected trees. *Ecology*, **30**, 101–104.

CURRY, L. (1964). The random spatial economy: An exploration in settlement theory. *Ann. Ass. Amer. Geogr.*, **54**, 138–146.

—— (1967). Central places in the random spatial economy. *J. Regional Sci.*, **7** (supplement), 217–238.

—— (1970). Univariate spatial forecasting. *Econ. Geogr.*, **46**, (supplement), 241–258.

—— (1971). Applicability of space–time moving average forecasting. In *Regional Forecasting* (M. D. I. Chisholm *et al.*, eds), pp. 11–24. London: Butterworth.

DACEY, M. F. (1960). Analysis of central place and point patterns by a nearest neighbour method. *Lund Studies in Geography, B*, No. 23, 55–75.

—— (1963). Order neighbor statistics for a class of random patterns in multidimensional space. *Ann. Ass. Amer. Geogr.*, **53**, 505–515.

DACEY, M. F. (1964). Modified Poisson probability law for point pattern more regular than random. *Ann. Ass. Amer. Geogr.*, **54**, 559–565.
—— (1965). The geometry of central place theory. *Geografiska Annaler*, B, **47**, 111–124.
—— (1966a). A compound probability law for a pattern more dispersed than random and with areal inhomogeneity. *Econ. Geogr.*, **42**, 172–179.
—— (1966b). A probability model for central place locations. *Ann. Ass. Amer. Geogr.*, **56**, 550–568.
—— (1966c). A county seat model for the areal pattern of an urban system. *Geogr. Rev.*, **56**, 527–542.
—— (1968). An empirical study of the areal distribution of houses in Puerto Rico. *Transactions and Papers*, Institute of British Geographers, **45**, 51–70.
—— (1969). Similarities in the areal distributions of houses in Japan and Puerto Rico. *Area*, **3**, 35–37.
DACEY, M. F. and TUNG, T. (1962). The identification of randomness in point patterns. *J. Regional Sci.*, **4**, 83–96.
DURBIN, J. and WATSON, G. S. (1950, 1951, 1971). Testing for serial correlation in least squares regression, I, II, III. *Biometrika*, **37**, 409–428; **38**, 159–78; **58**, 1–19.
EYRE, S. R. (1973). The spatial encumbrance. *Area*, **5**, 320–324.
GEARY, R. C. (1954). The contiguity ratio and statistical mapping. *Incorp. Statist.*, **5**, 115–141.
GOULD, P. R. (1960). *The Development of the Transportation Patterns in Ghana*. Evanston: Northwestern University. (Studies in Geography, No. 5).
—— (1969). Methodological developments since the fifties. In *Progress in Geography*, Vol. 1 (C. Board *et al.*, eds), pp. 1–49. London: Arnold.
—— (1970). Is *Statistix Inferens* the geographical name for a wild goose? *Econ. Geogr.*, **46** (supplement), 439–484.
GRANDELL, J. (1972). Statistical inference for doubly stochastic Poisson processes. In *Stochastic Point Processes* (P. A. W. Lewis, ed.), pp. 90–121. New York: Wiley.
GRANGER, C. W. J. (1969). Spatial data and time series analysis. In *Studies in Regional Science* (A. J. Scott, ed.), pp. 1–24. London: Pion.
GREIG-SMITH, P. (1957–64). *Quantitative Plant Ecology*, 1st and 2nd eds. London: Butterworth.
GURLAND, J. (1957). Some interrelations among compound and generalised distributions. *Biometrika*, **44**, 265–268.
HÄGERSTRAND, T. (1953). *Innovationsförloppet ur korologisk synpunkt*. Lund: Gleerup. English translation and postscript by Allan Pred appeared in 1967 as *Innovation Diffusion as a Spatial Process*. Chicago: University of Chicago Press.
—— (1967). On Monte Carlo simulation of diffusion. In *Quantitative Geography* (W. L. Garrison and D. F. Marble, eds), pp. 1–32. Evanston: Northwestern University (Studies in Geography, No. 13.)
HAGGETT, P. (1965). *Locational Analysis in Human Geography*. London: Arnold.
—— (1972). Contagious processes in a planar graph: an epidemiological application. In *Medical Geography* (N. D. McGlashan, ed), pp. 307–324. London: Methuen.
HARRISON, P. J. and STEVENS, C. F. (1971). A Bayesian approach to short term forecasting. *Oper. Res. Quart.*, **22**, 341–362.
HARTSHORNE, R. (1939). *The Nature of Geography*. Lancaster, Pa: Association of American Geographers.
—— (1959). *Perspective on the Nature of Geography*. Chicago: Rand McNally.
HARVEY, D. W. (1968). Some methodological problems in the use of the Neyman type A and negative binomial probability distributions for the analysis of spatial point patterns. *Transactions and Papers*, Institute of British Geographers, **44**, 85–95.
—— (1969). *Explanation in Geography*. London: Arnold.
HERTZ, P. (1909). Uber die gegenseitigen durchshnittlichen Abstand von Punkten, die mit bekannter mittlerer Dichte im Raum angeordnet sind. *Math. Ann.*, **67**, 387–398.
HIND, H. Y. (1864). On the commercial progress and resources of Central British America; the Lake Winnipeg and Saskatchewan districts. *J. R. Statist. Soc.*, **27** 82–105.
HOLGATE, P. (1966). Some new tests of randomness. *J. Ecol.*, **53**, 261–266.
—— (1972). The use of distance methods for the analysis of spatial distributions of points. In *Stochastic Point Processes*, (P. A. W. Lewis, ed.), pp. 122–135. New York: Wiley.
HOPE, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *J. R. Statist. Soc.* B, **30**, 582–598.
HUDSON, J. C. (1969). Diffusion in a central place system. *Geogr. Anal.*, **1**, 45–58.

HUFF, D. L. (1963). A probability analysis of shopping centre trading areas. *Land Econ.*, **39**, 81–90.
HUIJBREGHTS, C. (1975). Regionalized variables and quantitative analysis spatial data. In *Proceedings of the NATO Institute on the Display and Analysis of Spatial Data* (J. C. Davis and M. J. McCullagh, eds). pp. 38–53. New York: Wiley.
JOHNSTON, J. (1972). *Econometric Methods*, 2nd ed. New York: McGraw-Hill.
KATZ, E. and LAZARSFELD, P. (1955). *Personal Influence*. Glencoe: The Free Press.
KENDALL, M. G. (1939). The geographical distribution of crop productivity in England. *J. R. Statist. Soc.*, **102**, 21–48.
—— (1957). *A Course in Multivariate Analysis*. London: Griffin.
KENDALL, M. G. and MORAN, P. A. P. (1963). *Geometrical Probability*. London: Griffin.
KENDALL, M. G. and STUART, A. (1966). *The Advanced Theory of Statistics*. Vol. 3, *Design and Analysis and Time Series*. London: Griffin.
KERSHAW, K. A. (1964). *Quantitative and Dynamic Ecology*. London: Arnold.
KRISHNA IYER, P. V. A. (1949). The first and second moments of some probability distributions arising from points on a lattice and their application. *Biometrika*, **36**, 135–141.
LAKSHMANAN, T. R. and HANSEN, W. G. (1965). A retail market potential model. *J. Amer. Inst. Planners*, **3**, 134–143.
LEBART, L. (1969). Analyse statistique de la contiguité. *Publ. Inst. Statist. Univ. Paris*, **18**, 81–112.
MACKAY, J. R. (1958). The interactance hypothesis and boundaries in Canada: a preliminary study. *Canad. Geogr.*, **11**, 1–8.
MCCLELLAN, D. (1973). A re-examination and re-consideration of the neighbourhood effect in the diffusion of innovations at the micro scale. Unpublished prize-winning essay for Royal Geographical Society.
MARTIN, R. L. (1974). On autocorrelation, bias, and the use of first spatial differences in regression analysis. *Area*, **6**, 185–194.
MATERN, B. (1960). *Spatial Variation: Stochastic Models and their Application to Some Problems in Forest Surveys and Other Sampling Investigations*. Maddelanden Fran Statens Skogsforskningsinstitut, Band 49, pp. 1–144.
—— (1971). Doubly stochastic Poisson processes in the plane. In *Statistical Ecology*, Vol. 1 (G. P. Patil *et al.*, eds), pp. 195–213. University Park: Penn. State University Press.
MATHERON, G. (1969). *Le Krigeage Universal*. Les Cahiers du C.M.M., Vol. 1. Fontainebleau: Centre de Morphologie Mathématique.
—— (1970). Structures aléatoires et géologie mathématique. *Int. Statist. Rev.*, **38**, 1–11.
—— (1971). *The Theory of Regionalised Variables and its Applications*. Les Cahiers du C.M.M., Vol. 5. Fontainebleau: Centre de Morphologie Mathématique.
MEAD, R. (1967). A mathematical model for the estimation of interplant competition. *Biometrics*, **23**, 189–205.
—— (1971). Models for interplant competition in irregularly spaced populations. In *Statistical Ecology*, Vol. 2 (G. P. Patil *et al.*, eds), pp. 13–30. University Park: Penn. State University Press.
MEDVEDKOV, YU. V. (1967). The concept of entropy in settlement pattern analysis. *Regional Sci. Ass. Papers and Proc.*, **18**, 165–168.
MENDEL, J. M. (1973). *Discrete Techniques of Parameter Estimation*. New York: Dekker.
MOELLERING, H. and TOBLER, W. R. (1972). Geographical variances. *Geograph. Anal.*, **4**, 34–50.
MOORE, P. G. (1954). Spacing in plant populations. *Ecology*, **35**, 222–227.
MORAN, P. A. P. (1948). The interpretation of statistical maps. *J. R. Statist. Soc.* B, **10**, 245–251.
—— (1950). Notes on continuous stochastic phenomena. *Biometrika*, **37**, 17–23.
MORRILL, R. L. and PITTS, F. R. (1967). Marriage, migration, and the mean information field: a study in uniqueness and generality. *Ann. Ass. Amer. Geogr.*, **57**, 401–422.
NEWLING, B. E. (1965). *A Partial Theory of Urban Growth: Mathematical Structure and Planning Implications*. Rutgers University: Urban Studies Centre.
—— (1966). Urban growth and spatial structure: mathematical models and empirical evidence. *Geogr. Rev.*, **56**, 213–225.
NEYMAN, J. and SCOTT, E. L. (1952). A theory of the spatial distribution of the galaxies. *Astrophys. J.*, **116**, 144–163.
—— (1958). Statistical approach to problems of cosmology. *J. R. Statist. Soc.* B, **20**, 1–29.
—— (1972). Processes of clustering and applications. In *Stochastic Point Processes* (P. A. W. Lewis, ed.), pp. 646–681. New York: Wiley.
OLSSON, G. (1965). *Distance and Human Interaction: A Review and Bibliography*. Pennsylvania: Regional Science Research Institute. (Bibliography Series No. 2.)

ORD, J. K. (1970). The negative binomial model and quadrat sampling. In *Random Counts in Scientific Work*. Vol. 2, *Biomedical and Social Sciences* (G. P. Patil, ed.), pp. 151–163. University Park: Penn. State University Press.
—— (1972). *Families of Frequency Distributions*. London: Griffin.
—— (1974). Discussion on paper by J. E. Besag. *J. R. Statist. Soc.* B, **36**, 229.
—— (1975). Estimation methods for models of spatial interaction. *J. Amer. Statist. Ass.*, **70** (to appear).
PEDERSEN, P. O. (1970). Innovation diffusion within and between national urban systems. *Geogr. Anal.*, **2**, 203–254.
PERSSON, O. (1971). The robustness of estimating density by distance measurements. In *Statistical Ecology*, Vol. 2 (G. P. Patil *et al.*, eds), pp. 175–190. University Park: Penn. State University Press.
PIELOU, E. C. (1957). The effect of quadrat size on the estimation of the parameters of Neyman's and Thomas' distributions. *J. Ecol.*, **45**, 31–47.
PYKE, M. R. (1965). Spacings. *J. R. Statistical Soc.* B, **27**, 395–449.
RAVENSTEIN, E. G. (1875). Statistics at the Paris Geographical Congress. *J. R. Statist. Soc.*, **38**, 422–429.
—— (1879). On the Celtic languages in the British Isles: a statistical survey. *J. R. Statist. Soc.*, **42**, 579–643.
—— (1885). The laws of migration. *J. R. Statist. Soc.*, **48**, 167–235.
RAYNER, J. N. (1971). *An Introduction to Spectral Analysis*. London: Pion.
RAYNER, J. N. and GOLLEDGE, R. G. (1972). Spectral analysis of settlement patterns in diverse physical and economic environments. *Environment and Planning*, **4**, 347–371.
—— (1973). The spectrum of U.S. route 40 reexamined. *Geogr. Anal.*, **4**, 338–350.
REILLY, W. J. (1931). *The Law of Retail Gravitation*. New York: Reilly.
ROBINSON, A. H., LINDBERG, J. B. and BRINKMAN, L. W. (1961). A correlation and regression analysis applied to rural farm population densities in the Great Plains. *Ann. Ass. Amer. Geogr.*, **51**, 211–221.
ROBSON, B. T. (1973). *Urban Growth: An Approach*. London: Methuen.
SKELLAM, J. G. (1952). Studies in statistical ecology: I, Spatial pattern. *Biometrika*, **39**, 346–362.
STOUFFER, S. A. (1940). Intervening opportunities: a theory relating mobility and distance. *Amer. Sociol. Rev.*, **5**, 845–867.
—— (1960). Intervening opportunities and competing migrants. *J. Regional Sci.*, **2**, 1–26.
STUDENT (1907). On the error of counting with a haemacytometer. *Biometrika*, **5**, 351–360.
—— (1914). The elimination of spurious correlation due to position in time or space. *Biometrika*, **10**, 179–180.
TAAFFE, E. J., MORRILL, R. L. and GOULD, P. R. (1963). Transport expansion in underdeveloped countries: a comparative analysis. *Geogr. Rev.*, **58**, 503–529.
THEIL, H. (1971). *Principles of Econometrics*. New York: Wiley.
THOMAS, E. H. and ANDERSON, D. L. (1965). Additional comments on weighting values in correlation analysis of areal data. *Ann. Ass. Amer. Geogr.*, **55**, 492–505.
TINLINE, R. R. (1971). Linear operators in diffusion research. In *Regional Forecasting* (M. D. I. Chisholm *et al.*, eds), pp. 71–91. London: Butterworth.
TOBLER, W. R. (1967). Of maps and matrices. *J. Regional Sci.*, **7**, 275–280.
—— (1969a). The spectrum of U.S. route 40. *Papers and Proc.*, *Regional Sci. Ass.*, **23**, 45–52.
—— (1969b). Geographical filters and their inverses. *Georg. Anal.*, **1**, 234–253.
—— (1970a). A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.*, **46** (supplement), 234–240.
—— (1970b). *Regional Analysis or Time Series Extended to Two Dimensions*. Mathematical Social Sciences Board Conference on the Mathematics of Population, University of Chicago.
TULLOCH, A. M. (1838). On the sickness and mortality among the troops in the West Indies. *J. R. Statist. Soc.*, **1**, 129–142.
WARREN, W. G. (1962). Contributions to the study of spatial point processes. University of North Carolina, Institute of Statistics, Mimeo Series No. 337.
WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika*, **41**, 434–449.
—— (1956). On the variation of yield variance with plot size. *Biometrika*, **43**, 336–343.
WHITWORTH, W. A. (1934). *Choice and Chance*. New York: Steichert.
WILSON, A. G. (1970). *Entropy in Urban and Regional Modelling*. London: Pion.
—— (1972). Theoretical geography: some speculations. *Transactions and Papers*, Institute of British Geographers, **57**, 31–44.

YULE, G. U. and KENDALL, M. G. (1957). *An Introduction to the Theory of Statistics*. London: Griffin.
ZIPF, G. K. (1949). *Human Behaviour and the Principle of Least Effort*. Cambridge, Mass.: Addison-Wesley.

## APPENDIX

### *Computational Schemes for the Maximum Likelihood Estimators of Linear Spatial Models*

*Case* 1. Moving average model.

Consider the general moving average model of the form

$$\mathbf{Y} = \mathbf{B}\boldsymbol{\varepsilon}, \tag{A1}$$

where

$$\mathbf{B} = \mathbf{I} + \sum_{j=1}^{q} \theta_j \mathbf{W}^j. \tag{A2}$$

Initially, suppose that all the eigenvalues of $\mathbf{W}$ are distinct, so that all the eigenvectors are uniquely determined. Let $\boldsymbol{\Lambda}$ be the diagonal matrix of eigenvalues and $\mathbf{U}$ the matrix of (column) eigenvectors. Then, if $\mathbf{V}^{-1} = \mathbf{U}$,

$$\mathbf{W}^j = \mathbf{U}\boldsymbol{\Lambda}^j \mathbf{V},$$

and, by the Cayley–Hamilton theorem,

$$\begin{aligned} \mathbf{B} &= \mathbf{U}\left(\mathbf{I} + \sum_{j=1}^{q} \theta_j \boldsymbol{\Lambda}^j\right)\mathbf{V} \\ &= \mathbf{U}\boldsymbol{\Omega}\mathbf{V}, \quad \text{say.} \end{aligned} \tag{A3}$$

The log-likelihood is

$$l = \text{const} - n \ln \sigma - \ln|\mathbf{B}| - \frac{1}{2\sigma^2}\mathbf{y}^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{y}. \tag{A4}$$

This can now be simplified using the following results.

(i) $$\ln|\mathbf{B}| = \sum_{i=1}^{n} \ln \omega_i = \sum_{i=1}^{n} \ln\left(1 + \sum_{j=1}^{q} \theta_j \lambda_i^j\right), \tag{A5}$$

where $\omega_i$ and $\lambda_i$ are the $i$th diagonal elements of $\boldsymbol{\Omega}$ and $\boldsymbol{\Lambda}$ respectively.

(ii) $$\mathbf{y}^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{y} = \mathbf{x}^{\mathrm{T}}(\mathbf{I}-\mathbf{H})\mathbf{U}^{\mathrm{T}}\mathbf{U}(\mathbf{I}-\mathbf{H})\mathbf{x}, \tag{A6}$$

where $\mathbf{x} = \mathbf{V}\mathbf{y}$ and $\mathbf{H} = \mathbf{I} - \boldsymbol{\Omega}^{-1}$. $\mathbf{H}$ is a diagonal matrix with elements $h_i = 1 - (1 + \sum_{j=1}^{q} \theta_j \lambda_i^j)^{-1}$. Results (A5) and (A6) follow from the properties of determinants and (A3) respectively. When $\mathbf{W}$ is symmetric, (A6) simplifies further to

$$\sum_{i=1}^{n} (y_i - h_i x_i)^2 = \sum_{i=1}^{n} (1 - h_i)^2 x_i^2.$$

The value of results (A5) and (A6) lies in the fact that the $\{\lambda_i\}$ and $\mathbf{x}$ can be computed once and for all from $\mathbf{y}$ and $\mathbf{W}$. The maximum of $l$ can then be found iteratively (e.g. by Newton–Raphson) without requiring repeated inversion of large matrices.

When eigenvalues $\{\lambda_i, i = m+1, \ldots, n\}$ are zero, let

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

and let $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$ be the corresponding partition of $\mathbf{U}$. Likewise, set $\mathbf{V}^{\mathrm{T}} = (\mathbf{V}_1^{\mathrm{T}}, \mathbf{V}_2^{\mathrm{T}})$. Then results (A5) and (A6) can still be used, save that only the first $m$ terms in (A5) are non-zero, while $\mathbf{x}_1 = \mathbf{V}_1\mathbf{y}$ and $\mathbf{H}$, $\mathbf{U}$ are replaced by $\mathbf{H}_1$, $\mathbf{U}_1$ in (A6). For $\mathbf{W}$ symmetric, (A6) reduces to $\sum_{i=1}^{m}(1-h_i)^2 x_{ii}^2$. If two or more eigenvalues are identical but non-zero, any choice of corresponding eigenvectors will suffice, provided that the orthogonality conditions are preserved.

*Case* 2. Autoregressive moving average model.

Let the model be

$$\mathbf{AY} = \mathbf{B}\boldsymbol{\varepsilon}, \tag{A7}$$

where $\mathbf{B}$ is given by (A2) and $\mathbf{A} = (\mathbf{I} - \sum_{j=1}^{p} \phi_j \mathbf{W}^j)$. The log-likelihood function reduces, for $\mathbf{W}$ symmetric, to

$$l = \text{const} - n \ln \sigma + \sum_{i=1}^{n} \ln g_i + \sum_{i=1}^{n} g_i^2 x_i^2, \tag{A8}$$

where $\mathbf{x}$ is defined as before, and

$$g_i = \left(1 - \sum_{i=1}^{p} \phi_j \lambda_i^j\right) \bigg/ \left(1 + \sum_{j=1}^{q} \theta_j \lambda_i^j\right) \quad (i = 1, \ldots, n).$$

The form of $l$ for non-symmetric $\mathbf{W}$ is similar, but includes $\mathbf{U}^{\mathrm{T}}\mathbf{U}$, as in (A6).

*Case* 3. Autoregressive model, conditional formulation.

The analysis can be carried out in essentially the same way, except that in equation (A4), $\mathbf{BB}^{\mathrm{T}}$ should be replaced by $\mathbf{B}$ and $|\mathbf{B}|$ by $|\mathbf{B}|^{\frac{1}{2}}$ (cf. Besag, 1974).

In general, the computational benefits obtained from forcing $\mathbf{A}$ or $\mathbf{W}$ into a block diagonal structure when $n$ is large are considerable, since the corresponding eigenvalues and eigenvectors can be computed much more rapidly. Whittle's (1954) large sample procedure is a useful alternative for regular lattices.

## Discussion of the Paper by Dr Cliff and Dr Ord

Professor R. M. Cormack (University of St Andrews, Scotland): The desire of geographers to convince others that geography is a quantitative science is clearly displayed in the authors' selected bibliography. Unnoticed by most statisticians, an imposing corpus of quantitative theories has been created about spatial pattern on the surface of the earth. As was noted by Professor Cox a year ago in proposing the vote of thanks to Mr Besag for a more restricted and more statistically theoretical paper "the topic is important and notoriously difficult". Tonight we have gained a wider view over the area, a time-lapse aerial film of a world of central places, random economies, aggregated agriculture, spectral diseases, gravity-governed gallants and entropy-conscious colonists. We are in the authors' debt for opening our eyes not only to the fascinating theories which have been propounded about social pattern and behaviour, but also to the wealth of statistical problems raised by these theories and the need to test them. Some of the literature on quantitative geography consists of ill-conceived play with partially digested mathematics or with uncertainly determined numbers. This paper shows what can be achieved when both geographical and mathematical assumptions are understood. We are in a different world from a recent book in which a geographer insisted that, in assessing by $\chi^2$ the goodness of fit of a Poisson distribution, a set of data in which only two different numbers are observed has no degrees of freedom left to test the $\chi^2$.

I must also thank the speakers for the historical sidelights with which they opened their paper. In general, statisticians are not sufficiently conscious of the body that early developments in our subject can provide for the modern surface. After the authors' introduction I retain a delightful vision of Student gaining inspiration for trend-surface analysis from the changing configurations of the head on his Guinness.

Ecology is another science with spatial problems, and methods developed for the one subject should be of value in the other, perhaps with some modifications. Pielou's models for mosaics should have relevance both in geology and in the dispersal of ethnic groups. Where ecological techniques are translated into geographical terms they may be distorted in such a way as to strain the assumptions. Greig-Smith's pattern analysis has been developed largely for data in the form of counts, the index of dispersion being comparable at different hierarchical levels of scale with the Poisson value. This permits consideration in terms of conditional binomial allocation, as has been used by Mead (1974) to provide exact randomization and rank-order tests. With homoscedastic continuous variables, a Model II Analysis of Variance would allow estimation of components of variance, as the authors suggest. In the standard statistical context of randomized allocation of treatments to plots, these can be compared hierarchically, but there is no basis of assessment for the statistical significance of the smallest scale of variability. In the geographical context there is no such randomization. Of course geographers are seeking scales of variability which have geographical importance rather than statistical significance. In their paper, Moellering and Tobler insist—as this paper does not—that a complete census is required, and that no inferential process is involved. In this type of analysis the spatial ordering of subgroups within groups is completely ignored. And it seems strange to a statistician to find in the balanced case essentially sums of squares, rather than mean squares, as the basis for any conclusions.

One general point, which has been raised in much geographical literature, emerges from the above. I feel that our speakers should not be allowed to sidestep it, as they have done in the paper. Are geographical data on populations or samples? If the latter, in what sense are they random and from what population are they taken? If the former, in what way is statistical inference being used?

This question is one which arises in much of the recorded work on the spacing of settlements. To focus the discussion let me concentrate on Medvedkov's results, reported without comment in this paper. From Brush's original map Medvedkov apparently counted the 99 towns in all the squares of a 16 mile square lattice. We are not told how either the grid size or the origin was chosen. I have been unable to reconstruct the process, since Brush gives two maps defining towns differently—by size and by function—containing respectively 93 and 107 places. By an entropy argument Medvedkov essentially fits a Poisson distribution starting at $\theta$,

$$Pr = \frac{\mu^{r-\theta}\exp(-\mu)}{(r-\theta)!}, \quad r = \theta, (\theta+1), \ldots,$$

to the numbers of towns in each square, and deduces that the proportion of towns which fit a random lattice is $\hat{\mu}/(\hat{\mu}+\hat{\theta})$. I regret that space prevented our speakers from commenting on the technical details of this and other papers. My main point, however, concerns randomization and inference, both statistical and geographical. The analysis could be repeated with a grid of random size, origin and orientation. Would this yield more insight into, and perhaps valid inferences about, the form of the data? At a yet more general level how do the isolated analyses on isolated data sets combine to extend geographical theory? My own geographical colleagues dispute the authors' claim that they do.

The authors insist, rightly in my view, that spatial considerations are paramount. And yet, in one instance at least, the authors ignore their own injunction. In inferring process from form workers are stymied by the equifinality of, for example, the negative binomial distribution. The authors' suggested escape compares the estimates of the parameters of a sequence of such distributions fitted to data from quadrats combined together. The neighbouring quadrats so combined will be spatially correlated: the derivation of the compound distribution implicitly assumes independent observations on the compounding $\Gamma$-distribution. By the authors' argument the same distribution would be obtained by combining non-neighbouring quadrats.

I have done the authors less than justice in treating this paper wholly as a review. Many of the ideas discussed, notably the important one of spatial autocorrelation, have been formulated and developed by the authors themselves. Their models for spatial-temporal processes, briefly outlined in Section 6 should arouse interest in time series analysts and econometricians, and point to a new era ahead—even if one must entertain grave doubts about the number of geographical areas in which stationarity and isotropy can safely be assumed. Oceanic studies of plankton have been found to be depressingly non-stationary, in a medium that might be expected to be more homogeneous than the land. In their closing lines they plead for closer collaboration between the statistician and the geographer. The example of their own joint work is one to be followed. I am sure that many valuable developments will be stimulated by this paper.

I have much pleasure in proposing the vote of thanks.

Professor S. Gregory (Institute of British Geographers): May I begin this vote of thanks by addressing myself not so much to the authors as to the Royal Statistical Society, in my present capacity as President of the Institute of British Geographers and in one of my earlier capacities as an initiator of the quantitative study group. As geographers we greatly appreciate the invitation which has been made to us, the co-operation which we have received and for this opportunity to meet together.

Secondly, however, I must make some sort of disclaimer. I had assumed that seconding a vote of thanks would take place at the end of a long discussion, by which time everyone would have become somnolent and looking for food, so that I could get away with a few words without saying anything profound. The latter will still be true, but I suppose that I must speak for a little longer. My disclaimer must also indicate that my own statistical competence is at the introductory level rather than at the more advanced levels about which we have been hearing tonight. A further disclaimer is that, except in a superficial way, I am not a human geographer.

The presentation of geography, as given tonight, is one with which all geographers would not necessarily agree. My geographer friends here could think of many of their colleagues who would not only disagree with the definition of geography given but would also not appreciate the approach and method used. I stress this to warn statisticians who may be thinking of approaching geographers in their own universities, not to pick these at random. A biased selection is essential, and that bias should be in the direction of the younger geographer. Certainly, no one as long in the tooth as myself should really be consulted.

It is the younger geographers who, over the past 5 or 6 years, have been pushing towards a critical review of statistical and quantitative methods in geography rather than continuing the somewhat uncritical use which was made earlier in the mid-1950s to mid-1960s. Within these younger geographers, trying to make the quantitative approach to geography more valid, sound and acceptable, our two speakers tonight clearly have played a major role and will no doubt continue to do so in the future.

In terms of raising critical issues which ought to be discussed I am at a slight disadvantage in that having been brought up within the geographical fraternity the ideas which have been raised are at least known to me. In that sense I suppose there is an element of autocorrelation between myself and the speakers, even if it is a relatively weak one.

As Professor Cormack pointed out in proposing the vote of thanks the major value of tonight's gathering is the bringing together of geographers and statisticians, so that some of the current approaches and ideas in geography are exposed to a critical review by the "professional" statistician. Most of us have had in the past, or have at the present, valuable contacts with our own tame statistician in our own university. When Professor Plackett and I both lived west of the Pennines, for example, he pointed out to me some of the defects of my earlier thinking, and probably all of us have had this sort of relationship at some time. It is important at this stage, however, that the whole approach of the geographer to the use of statistical methodology in his geographical problems should be exposed to full critical consideration, discussion and perhaps attack by the statistician in a formal way. It is for this reason that, on behalf of all geographers here tonight, I express our great appreciation to Dr Cliff and Dr Ord for their presentation which has brought before this audience the type of problems we are facing, the approaches being made to try to solve some of these problems and the invitation to statisticians to suggest how improvements could be effected. May I conclude, therefore, by expressing my sincere appreciation to our speakers tonight and heartily seconding the vote of thanks.

The vote of thanks was passed by acclamation.

Professor M. S. Bartlett (Oxford University): I welcome the opportunity this paper gives to learn something of what statistical geography is about, though I suppose the environment for this meeting may favour the statistical as against the geographical. I hope the final mixture for statistical geography will prove to be a balanced one; statistical psychology is perhaps a warning of the possible dangers of lack of balance. The geographical references provide of course further information for the statistician, but I cannot claim to have followed these up—I must admit, however, to being intrigued by one of P. R. Gould's titles: 'Is *Statistix Inferens* the geographical name for a wild goose?"

My further comments are confined to some technical points. Firstly, as the authors refer in Section 6.5 to the spectral analysis of point processes, it may be worth noting (a) an ingenious device by French and Holden (1971) for simultaneous filtering and sampling of a one-dimensional point process in order to use Fast Fourier Transform computing methods, viz. replacing $dN(\tau)$ by, say,

$$X(t) = \int \frac{\sin\{(t-\tau)\pi\}}{(t-\tau)\pi}\, dN(\tau);$$

(b) the obvious extension of this device where appropriate to two-dimensional processes, viz.

$$X(x,y) = \iint \frac{\sin\{(x-u)\pi\}}{(x-u)\pi}\frac{\sin\{(y-v)\pi\}}{(y-v)\pi}\, dN(u,v).$$

Secondly, the use of spatial–temporal models to generate autonormal spatial lattice models may be extended to multivariate models. The theoretical correlational or equivalent spectral properties of the resulting *vector* Markov field may readily be investigated, e.g. by the method used by Besag (1974, Section 4). As the authors refer to diffusion models (Section 5), let me note the bivariate analogue of Whittle's spatial–temporal model which I discussed in Bartlett (1974). This model was

$$(\partial/\partial t + \kappa^2 - \nabla^2)\, X(t, \mathbf{r}) = Y(t, \mathbf{r}),$$

with spatial spectrum $C(\kappa^2+\omega^2)^{-1}$. If

$$(\partial/\partial t + \mathbf{A} - \mathbf{D}\nabla^2)\, \mathbf{X}(t, \mathbf{r}) = \mathbf{Y}(t, \mathbf{r}),$$

where $\mathbf{D}$ is a diagonal matrix, or equivalently

$$[\partial/\partial t + \mathbf{A} + \mathbf{D}\omega^2]\, d\boldsymbol{\chi}(t, \boldsymbol{\omega}) = (d\eta(t, \boldsymbol{\omega}),$$

where

$$X(t,\mathbf{r}) = \int \exp(i\boldsymbol{\omega}^{\mathrm{T}}\mathbf{r})\, d\chi(t,\boldsymbol{\omega})$$

and similarly for $Y$ and $\eta$, we obtain for the marginal spatial spectrum (under conditions of stationarity)

$$(\mathbf{A}+\mathbf{A}^{\mathrm{T}}+2\mathbf{D}\omega^2)^{-1}.$$

One final remark. So-called spatial "patchiness", which is of interest to ecologists as well as geographers, needs some clarification. The conditions for stationarity imply that $\mathbf{A}+\mathbf{A}^{\mathrm{T}}$ must be positive-definite, and neighbouring correlations will not be different in kind from the univariate case (which is not oscillatory). The bivariate case can be much more interesting under some other conditions, and is then best studied from the direct solution for $d\boldsymbol{\chi}(t, \boldsymbol{\omega})$. In ecological (or morphogenetic) contexts, the linear equations are also usually approximations to non-linear ones (see, for example, Maynard Smith, 1968; Steele, 1974). While I was interested in the shape of the spatial correlogram for measles in the authors' paper, which exhibits a space periodicity effect, the appropriate model here would seem complicated by further spatial heterogeneities, either the rural–urban one referred to by the authors, or stochastic local extinction effects which one would expect to be relevant in this case.

Professor P. HAGGETT (Bristol University): I would like to add my thanks to the authors for this wide-ranging review paper, and to the Society for initiating this joint meeting with the I.B.G. I hope this evening's successful experiment will lead to future exchanges for geography is rich in unsolved problems which may prove to be of statistical interest. Some of these are theoretical, others mundane and practical. Let me illustrate the latter from one referred to in Section 6.3 of the Cliff and Ord paper. In mapping, we frequently encounter problems in developing suitable models for producing smoothed or interpolated contour maps from sets of observations (control points) that are arrayed in a regular or, more commonly, irregular fashion in space. Ordinarily our contours are drawn so that they extend out to the control points that lie on the far edges of the area to be mapped. But as we approach this edge, it is evident that our contours are based on either fewer points (where our smoothing model involves a local operator) or are drawn from an increasingly remote and asymmetric set of control points (as in a polynomial trend surface model). Some experimental work has been done in geography on the confidence bands that we can associate with contour maps, but formal work by statisticians would be welcome. We can of course overcome the worst of these "edge-effects" in practice, by using control points *outside* the area to be mapped, thereby creating a buffer zone (or *cordon sanitaire*!) around the edge of the map. But since this solution is not always feasible and it involves information loss, we are interested in estimating how wide such a buffer zone would need to be to achieve a desired reliability.

In a few circumstances we may wish to extrapolate (i.e. spatially forecast) our contours slightly beyond the perimeter control points. Work by Krumbein (1966) has shown how some of the smoothing models used by geographers work in this situation. For example, polynomial trend surfaces may show wildly plunging or escalating values, while double Fourier surfaces show a predictably greater stability. More formal work on the performance of smoothing models around the map edge would be useful, particularly in view of the increased interest in time–space forecasting described in Section 6.4.

Tonight's carefully presented paper is welcome in showing areas where problems lie and where some progress is being made. Some problems are of course likely to remain intractable for a very long time. For example, Jan Tinbergen (1968, p. 65) was forced to conclude on the topic of settlement size distribution (see Section 4.3) that "... no scientific explanation worthy of the name has been advanced so far". Inevitably some of the equations in tonight's paper describe phenomena like human settlement patterns which are so varied and complex that we can only tackle them by proposing models which appear to be either grossly oversimplified or intractably complex. In this situation one can either curse the darkness and turn back, or try to light a candle or two and move forward. Clearly, this paper has shown that Dr Cliff and Dr Ord are in business together to light candles, and geographers present tonight wish them well on their joint research.

Mr. R. B. DAVIES (University of Wales Institute of Science and Technology): I should like to add my appreciation of Dr Cliff and Dr Ord's valuable paper to that already expressed, and to develop a few points about the use of spectral analysis in geography.

Spatial spectral methods involve the transformation of a regular data array from the two-dimensional space domain into the two-dimensional frequency domain. Orientation is preserved by the transformation; directional invariance is not assumed. This partitioning of variance by both frequency and orientation clearly has considerable potential in spatial analysis, for example in detecting significant orientations in specified frequency ranges. But, as Drs Cliff and Ord note, there have been few applications within geography.

If directional invariance may be assumed, or be inferred from the two-dimensional spectrum, the variance may be reduced to a function of frequency only by summing the variance in constant frequency bands about the origin of the two-dimensional spectrum. Rayner and Golledge (1972) use such a frequency domain formulation to aid in the interpretation of two-dimensional spectra. They also stress the value of this average spectrum for pattern comparisons.

The spatial correlogram provides another means of transforming spatial data into the one-dimensional frequency domain if directional invariance may be assumed. For example, we may suggest

$$g(f) = 2\left(1+2\sum_{k=1}^{n} r(k)\, w(k) \cos 2\pi f k\right)$$

as an estimate of the spectral density at frequency $f$, where $r(k)$ is the correlation function of spatial lag $k$, and $w(k)$ is the lag window. This spectral formulation has the merit of not requiring regularly spaced data but, like the correlogram from which it is derived, it is conditional on the structure of the data collection units.

Questions about the relative merits in the analysis of geographical data of spatial correlograms—as in Section 6.2.1 of the paper—and spectral estimates based on spatial correlograms must be postponed. The statistical inference issues are formidable.

Mr PETER J. DIGGLE (University of Newcastle upon Tyne): I should like to begin by adding my thanks to the authors for a most interesting paper; in particular, I am indebted to them for an extensive list of references in a subject area previously unfamiliar to me.

My specific comments relate to Section 4 of the paper. I welcome the firm assertion of the difficulties inherent in attempts to infer process from pattern. For example, Harvey's (1968) remark that, in the generalized Poisson interpretation of the negative binomial distribution, "The spatial law governing the distribution of offspring around a centre appears to be circular normal in form" seems quite unjustified. In view of this apparent confusion, I would have preferred the authors to have stressed the distinction between descriptive *distributions* such as the negative binomial, and spatial stochastic *processes* such as the centre-satellite model.

On the subject of tests of randomness based on distance methods, I have recently investigated the effect of intensive sampling on the Clarke–Evans test, at the request of Dr I. Hodder, Department of Archaeology, University of Leeds. Dr Hodder's work concerns the spatial distribution of artefacts or pre-historic settlements for which, particularly after the elimination of edge-effects, the usable population size tends to be rather small. A complete census of nearest neighbour measurements is desirable on the grounds of full utilization of scarce data, but obviously raises the problem of dependent observations.

In this situation, one might expect the null variance of the test statistic to be increased by the pattern of spatial autocorrelation among the individual nearest neighbour distances, leading to spuriously significant results. However, the magnitude of this effect appears surprisingly small, presumably because of a balance between opposing "local" and "long-range" effects. The results of an extensive Monte Carlo experiment indicate that, for usable population sizes ranging from 25 to 80, the incorrect assumption of independent observations leads to a nominally 5 per cent equal-tailed test whose true size is of the order of 6 or 7 per cent.

A crucial point concerning tests of randomness is, of course, their power against plausible alternatives. In this respect, the test proposed by Hopkins (1954) in an ecological context is perhaps more attractive than any of those mentioned this evening. Hopkins' test incorporates the known value of $\lambda$ indirectly, in that it involves the random selection of individuals from the population; indeed, a number of the later tests were designed specifically to circumvent this often impracticable procedure, while admitting that a reduction in power would probably result. Comparative studies of the power of distance-based tests of randomness against various alternatives have usually confirmed the overall superiority of Hopkins' test (see, for example, Holgate, 1965a, b; Brown & Holgate, 1974; Diggle *et al.*, 1975).

Finally, the problem of inferring process from pattern appears particularly acute for the Whitworth–Cohen models of settlement size; it is not at all clear to me how one would here interpret a good fit to a particular model.

Dr Denis Mollison (Heriot-Watt University): I should like to draw attention to qualitative theoretical studies on spatial stochastic models which may be of interest to geographers, and at the same time to express a worry that too much emphasis on goodness-of-fit may lead to quantitative models, which while excellent for narrow purposes of prediction, lack the generality that one gets with a model whose structure genuinely reflects that of the real world.

Consider an example mentioned in Section 6·4, where "mixed autoregressive, moving average, regression models" are described as "very flexible": namely Tinline's fitting of a contact distribution to data of Hägerstrand on the spread of innovations (Hägerstrand, 1967). Tinline's contact distribution for infection to the 25 nearest areas, shows (1) wild variation suggesting too many parameters are being estimated; (2) nevertheless, a tendency to concentrate on the centre and edge of the $5 \times 5$ square of nearest neighbours at the expense of those in between. This to me strongly suggests that infection to greater distances is an important feature of the data. Since the fitting of a 25 point contact distribution appears to be over-ambitious, one cannot expect to fit an exact contact distribution allowing for infection to greater distances.

A more modest and perhaps more interesting aim is to look for a qualitative description of the sort of contact distribution which might provide an adequate fit. Now the theory and simulations of Mollison (1972) show how the qualitative spread for a similar (but alas one-dimensional) model, apart from the obvious linear dependence on a measure of dispersion such as the standard deviation of the contact distribution $v$, depends principally on which of three classes $v$ falls into. These classes are based on the asymptotic behaviour of $v$ for large distances; indeed contact distributions with finite cutoffs such as Hägerstrand and Tinline's produce very similar patterns of spread and, given equal standard deviations, nearly equal velocities. Contact distributions in the best behaved class—those with exponentially bounded tails—all show similar patterns of spread at a steady velocity.

In contrast, when $v$ is more spread than negative exponential, the simulations confirm that what is described by the speakers in Section 5.2 as the "neighbourhood effect" can occur even when $v$ decays regularly with distance $d$: for example, $v(d) \sim d^{-n}$, $n > 3$ (if $n \leqslant 3$, the velocity of spread is asymptotically infinite).

To sum up, one would not expect a fitting of Hägerstrand's data in terms of standard deviation and class (1, 2 or 3) of contact distribution to be better in terms of "goodness-of-fit", but I think it reasonable to claim that it might give more understanding of the underlying process.

I should add that my own work referred to is only part of a body of work on this topic of qualitative spread of stochastic processes, and that geographers might also be interested in the work on nearest-neighbour spread of Hammersley (1966) and Richardson (1973), and perhaps the application to tumour growth of Williams and Bjerknes (1972) and Mollison (1973); and the work of Kolmogorov *et al.* (1937), Kendall (1965) and Daniels (1975) on deterministic models.

Dr F. H. Hansford-Miller (I.L.E.A.): In the extensive list of references to the paper I was surprised to find that the paper to this Society by Tanner (1963) on car and motorcycle ownership was not included. This paper used a multiple regression analysis in investigation of spatial pattern. As this evening's speakers made only a passing reference to regression I assume that they are unsympathetic to this method.

On the contrary, I have found regression analysis to be of considerable value and, as illustration, will cite an example from some work I did in the Geography of Religion (Hansford-Miller, 1968). I was concerned with the clergy deprivations in 1558–1564—somewhat earlier than some of the examples we have had this evening—at the beginning of the reign of Queen Elizabeth I. This is also an example of the data limitations rightly noted by the authors in Section 3.4. Data limitation here is of an extremely severe kind. It might be beneficial to some modern researchers who are used to a superfluity of data, in the wide range of statistics currently available, to sometimes consider spatial analysis problems in history when the data are so sparse and incomplete—and when the data have to be accepted and used because that is all that is available.

To give a little background, Queen Mary's reign had sent 282 people to the stake (Hansford-Miller, 1970) but Queen Elizabeth took a middle way in religion. She insisted on an oath of supremacy from all the clergy. Some of the clergy refused to give that oath. I was investigating those clergy who, as a result of this refusal, were deprived of their livings—not their lives.

In my research into the distributions of religious groups in England at that time this would indicate (a) where there was a strength of Catholicism, and (b) where there might possibly be extreme right-wing Protestantism. Members of both such groups were not satisfied with the religious settlement. The data available are in a list of dioceses and, in such a case, a simple $\chi^2$ analysis provides a lot of insight, indicating where there are significant areal variations from the nationwide norm.

I should point out here that some of our more basic elementary statistical methods, such as $\chi^2$, can illuminate very brightly on occasion and are not to be despised for their familiarity and simplicity. It revealed the anomalies in this particular situation, but could not explain them.

To assist in the explanation of the anomalies I used a multiple regression analysis—I must add that it did not fully explain them because I differ profoundly from the authors in their philosophy, which seems to me to be basically mechanistic. In Section 4·1 they equate human behaviour patterns with the random movement of molecules. Not only that but they—"hope that the underpinnings of the Second Law of Thermodynamics may apply to human activity". Nothing, in my view, could be more ghastly for humanity than for this hope to be fulfilled. Fortunately, I submit, it will not, for man's free spirit, with God's help, will never be quenched.

To get back nearer home, nor do I believe that the application of statistics to human geography is at all dependent on such a Marxist determinist view of human motivation. Statistical analysis is extremely valuable in pointing the way and in opening up new avenues of thought but, in my experience, it never tells the whole story, nor should it be expected to.

Fig. 1 shows that in the cited multiple regression analysis my dependent variables were standardized for area and number of parishes. I had three geographical variables and the others were religious and economic factors, which I considered might influence the distributions. These were regressed in two ways; by step-wise introduction in order of significance (Fig. 1) and by specified steps (not shown). In both regressions the $F$-values

| *Dependent variable* | *Step* | *Independent variable introduced* | *Square of multiple correlation* | *F-values significant at level of* 0.1% | 1% | 5% | *Dioceses with largest residuals* +ve | −ve |
|---|---|---|---|---|---|---|---|---|
| Clergy deprivations per area | 1 | 12 | 0·45 | 12 | — | — | Chichester | York |
| | 2 | 11 | 0·57 | — | 12 | 11 | Chichester | Ely |
| | 3 | 3 | 0·62 | — | 12 | — | — | — |
| | ⋮ | | | | | | | |
| Clergy deprivations per number of parishes | 1 | 17 | 0·43 | — | 17 | — | London | Ely |
| | 2 | 18 | 0·56 | — | 17 | 18 | Chichester | Ely |
| | 3 | 3 | 0·58 | — | 17 | 18 | — | — |
| | ⋮ | | | | | | | |

FIG. 1. Clergy deprivations 1558–1564. Multiple regression analysis with step-wise introduction of independent variables—latitude, 3; longitude, 4; Distance from London, 5. Diocesan wealth: Bishop, 6; other dignitaries, 7; Total, 8; related to area, 9, 10, 11; related to number of parishes, 15, 16, 17; Number of chantries and free chapels, 12, 18; Number of impropriated parishes, 14, 19. (Source, Hansford-Miller, 1968).

are important and answer the plea of Duncan *et al.* (1961, pp. 14 *et seq.*) for added rigour and precision in geographical research. A multiple regression analysis is extremely valuable also for its residuals. Quite often it is the anomalies which are of greater interest than the correlations as indications of the lines of possible fruitful future research. In the example the first independent variable to be introduced was number 12—number of chantries and free chapels per area. Chantry chapels were places where prayers were said for the dead, often endowed with a priest for such duties, and these were associated with Catholicism and High Church Anglicanism. It is thus a reasonable result for such chapels to correlate with the deprivations. I could go on to discuss the residuals—Chichester and York Dioceses—but I think I have made the point that a great deal of useful information can be obtained from this kind of investigation.

The authors in Section 3.3 rightly point out, with Yule and Kendall (1937, pp. 310–315), that in such an analysis the units chosen are modifiable units and affect the magnitude of the resulting correlations. I do not, however, see this as a serious disadvantage as it is often the relativity between different variables that is of importance.

It was my privilege to undertake geographical research under the late Professor S. W. Wooldridge, D Sc., F.R.S., at King's College, London, who did his major work before the statistics-cum-computer revolution—unfortunately. I wonder what he would have made of it?

What I have heard this evening is not my version of human geography; it is economic geography, it is urban geography, in part it is medical geography, but my feeling for human geography is completely different from what has been put before us this evening.

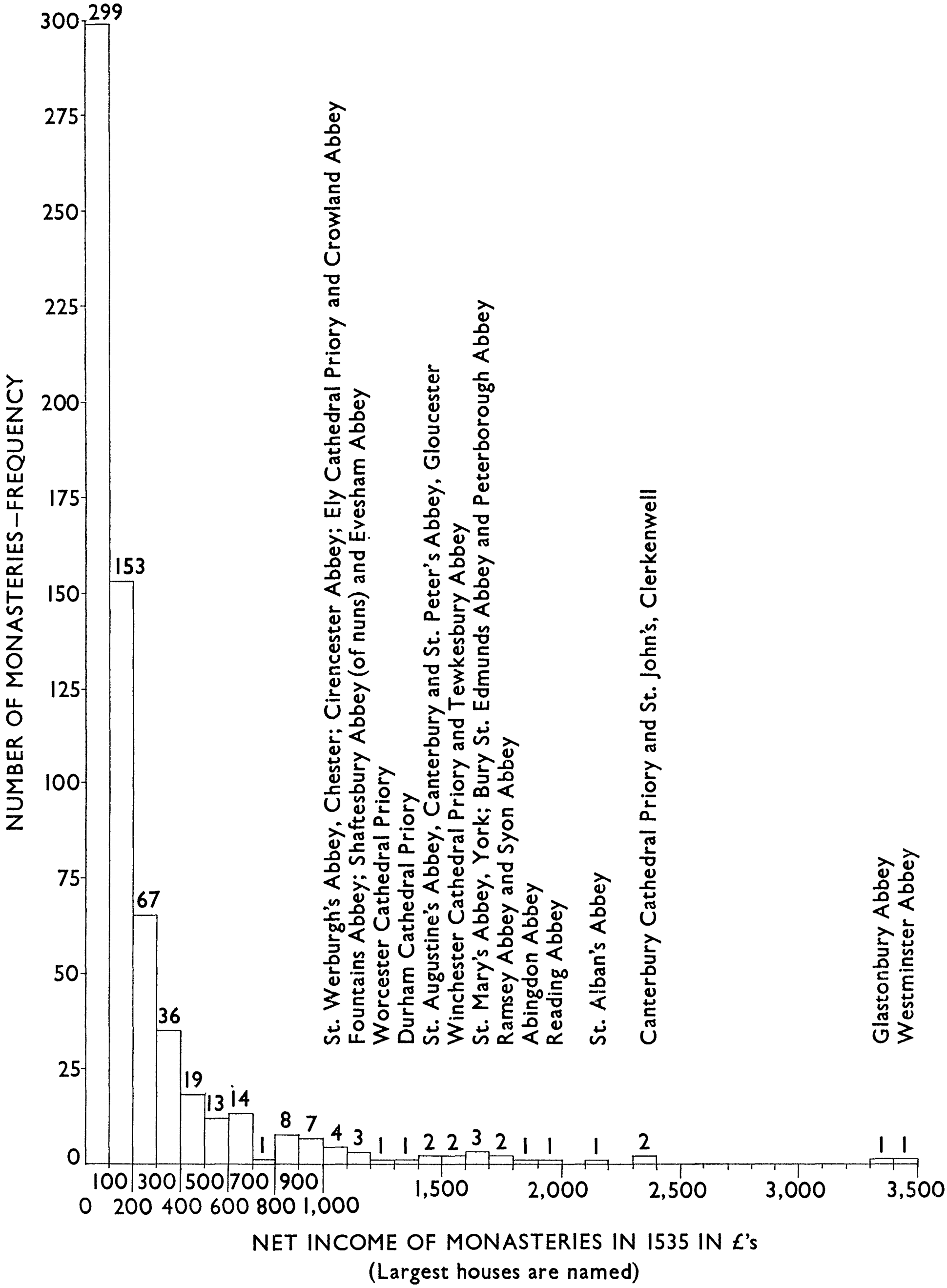


FIG. 2. Histogram of net income of monastic houses in England and Wales, 1535. (Source: Hansford-Miller, 1965.)

Wooldridge (1956, p. 18) wrote the following on the subject: "... it is in the field of human phenomena that it [geography] has its greatest opportunity, even if it there faces its greatest difficulties". C. Daryll Forde (1934, p. 465) amplified this when he said, "The geographer who is unversed in the culture of the people of the land he studies, or in the lessons which ethnology as a whole has to teach will, as soon as he begins to consider the mainsprings of human activity, find himself groping uncertainly for geographical factors whose significance he cannot truly assess. Human geography demands as much knowledge of humanity as of geography." This is my version of human geography. The French geographer, Vidal de la Blache, as long ago as 1911, wrote: "... country is a storehouse of dormant energies, laid up in the germ by Nature, but depending for employment upon man. It is man who reveals a country's individuality by moulding it to his own use ... till at length it becomes, as it were, a medal struck in the likeness of a people." Humanity has ethnic groups, nationalities, cultures and individualities which surely must find a place somewhere in the authors' molecular–man clusters. I appeal to geographers to attempt to do just this.

Finally, there is a feeling that everything must be complex, that nothing can be explained by simple means. Sometimes insight can be shown, even by histograms. Fig. 2 is a histogram of the net income of monasteries in England and Wales in 1535 (Hansford-Miller, 1965, pp. 124–216; 227–290). When people think of monasteries they usually think of huge ones such as Westminster Abbey but, in fact, a histogram shows that the mode monastery was a small house with an income of less than £100—not a great house like Westminster Abbey, with an income of £3,470, or Glastonbury Abbey, with £3,311, or even Fountains Abbey, with £1,115. No less than 299 of the 642 houses fall into this first income group and another 153 have incomes between £100 and £200. Only 25 of the monasteries are large enough to have incomes greater than £1,000. Much insight can be obtained from such a histogram.

If the authors wish to go further in their quest for spatial pattern how about delving into topology? Fig. 3 shows a topological transformation of the national boundary network of

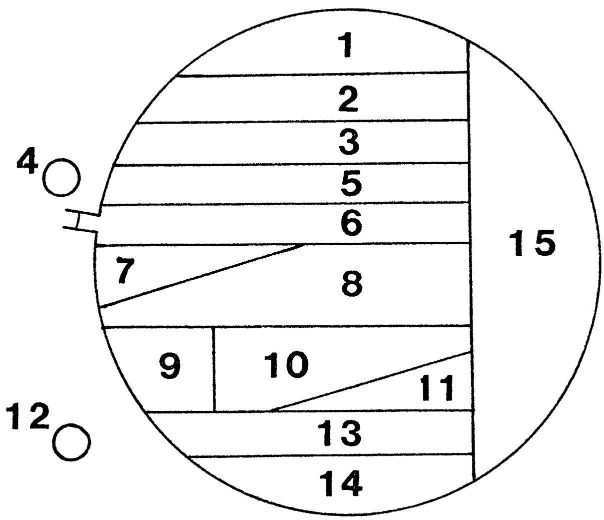


FIG. 3. A topological transformation of the National boundary network of South America. No. 15 is Brazil. (Source: Hansford-Miller, 1971.)

South America (Hansford-Miller, 1971). What do we get from this? Among other things that Brazil—which is 15—amazingly has common frontiers with no less than 10 of the 12 other countries in the South American continent.

Dr M. WESTCOTT (Imperial College): I, too, would like to thank the authors for a most interesting paper, and especially for their excellent presentation tonight. I shall comment on just a few points, the first of which at least is extremely mathematical.

At the end of Section 4.1.2 it is mentioned that doubly stochastic Poisson models and centre-satellite (Neyman–Scott) cluster models can in some cases produce probabilistically identical point processes. This raises the question of delineating the set of point processes which are both doubly stochastic Poisson and Poisson cluster models or, less ambitiously, deciding which members of one set are also members of the other. The last statement is more or less equivalent to asking, when is a doubly stochastic Poisson process infinitely divisible for (Kerstan and Matthes, 1965) all regular infinitely divisible point processes are Poisson cluster processes. It is a plausible conjecture that it is necessary and sufficient that the stochastic mean process also be infinitely divisible. Supporting evidence comes from naive consideration of a non-negative integer valued random variable $X$ which is both compound Poisson and infinitely divisible and hence (Feller, 1968, p. 290) also generalized Poisson.

The generating function of $X$ is

$$\Pi(z) = \int_0^\infty \exp\{-\lambda(1-z)\}\, dG(\lambda),$$

which of course is the Laplace transform of the compounding distribution $G$. Since this is also of the form $\exp\{P(z)-1\}$, $P$ some generating function, it is clear that $-\log \Pi(z)$ has a completely monotone derivative on $[0, 1]$ at least. If this could be extended to $[0, \infty)$ we would have established the infinite divisibility of $G$.

On a related point, it would have been nice to get some more details in the paper on how the two mechanisms mentioned which generate the negative binomial distribution can be distinguished if some extra information is available. It may be of interest that, in the related field of models for accident-proneness, where the negative binomial again arises from different models corresponding roughly to true and apparent contagion, Professor Violet Cane, in a Manchester research report, showed that these models were the extreme cases of a continuous spectrum of mechanisms all of which produced the negative binomial distribution of counts. If this phenomenon is relevant in the authors' situation it would perhaps suggest that attempts to unscramble equifinal processes are rather futile.

More generally, the investigation of models for spatial processes is clearly an important topic in view of the wide range of potential applications, as instanced in this paper. The authors appear to tacitly concur with Bartlett (1974) that only three types of spatial process have so far been extensively studied—doubly stochastic, cluster and lattice processes. So any new class of models is of potential interest. In a recent Ph.D. thesis at Imperial College Dr Valerie Isham has looked at several ways of building up spatial processes; there is no explicit temporal element though the processes are often implicitly built up temporally (see also Isham, 1975). Broadly, her models are:

(i) *Cartesian*: the two-dimensional Cartesian co-ordinates of points come from separate one-dimensional processes, with random shifts if necessary to remove lattice structure.

(ii) *Polar*: a generalization of the Poisson construction of Section 4.1.1 by taking the squared radial distances to be sums of fairly arbitrary independent and identically distributed variables, with one or more points on the circumference. Markovian extensions are also considered.

It turns out that nearly all these models are approximately Poisson far from the origin, a fact which is probably of little concern to geographers as they will rarely tend to infinity! However, it could be a useful warning if application of such models on a large scale were contemplated. Also, the processes do often have interesting non-Poisson local behaviour—a more relevant result—though the properties are often sufficiently complicated to deter all except writers of Ph.D. theses.

The following contributions were received in writing after the meeting:

Dr R. J. BENNETT (Department of Geography, University College London): I would first like to record my thanks to the authors for what is an extremely useful and timely paper. The paper is especially timely since it summarizes a stage of development in an expanding area of research in which co-operation between the geographer and the statistician will be of great value. May I, therefore, direct my comments at Section 6 of the paper since this is the section in which I have a particular interest, and in which there are important problems which might benefit from this co-operation.

My first comment relates to the problem area to which we should be directing attention. In this context, it would seem that in the long term we are not going to learn much from models of purely *spatial* processes. There are difficult methodological problems in the interpretation of spatial stochastic processes except in equilibrium conditions or in systems characterized by infinitely short relaxation times. Neither of these properties is common in geographical problems. The arguments surrounding this problem have been rehearsed in discussion of Besag's (1974) paper to the Society. No solution will be tractable, in general, without considerable *a priori* theoretical restrictions or a large amount of behavioural information. The incorporation of both of these sources of knowledge into spatial interaction models has been discussed by Wilson (1970) and in the appendix to Bennett (1975a) and allows a tractable and practical solution in this case. For more general models of mapped patterns present theory requires considerable development.

In the spatio-temporal context, the model proposed in equation (6.18) is extremely attractive, but would seem to require a high degree of over parameterisation. The authors note that the size of the $\mathbf{X}^T\mathbf{X}$ matrix precludes estimation in most practical situations. The size of this matrix arises from the inclusion of the spatial dependency of any zone $j$ on all other zones in the $n$-zonal region under study ($j = 1, 2, \ldots, n$). If estimation is actually undertaken with such a model there will be a marked loss of degrees of freedom, and one would expect in most applications, the introduction of sizeable degrees of collinearity between spatial zones. Considerable simplication could be achieved if model specification and identification criteria were incorporated as *a priori* restrictions. The principle of parsimony has been applied by Box and Jenkins (1970) to the reduction of the parameter space of time series models and an extension of this Yule–Walker solution to space–time problems has been given by Bennett (1975a) and since developed by Martin (1975). A second approach given in Bennett (1975a) is to use a canonical factorization of the parameter space given by the projection of the future onto the present and of the present onto the past. This results in the derivation of a minimal realization and would seem a fruitful area for future research.

The use of simultaneous equation estimation techniques proposed in Section 6.4 of the paper is also attractive in that the spatio-temporal model of the $n$-region system can be reduced to an $n$-variate equation system for one region. Such a reduction creates many new estimation problems, however. Such methods will be applicable only under certain conditions, the most important being that the $\mathbf{X}_j$ terms in equation (6.18) are each mutually independent for each variable in each region, and that the error sequence $\boldsymbol{\varepsilon}_j$ is also mutually independent for each observation in time and for each spatial region. It would seem to be very special conditions indeed in which $\mathbf{X}_j$ was not itself autocorrelated, and that the error sequence was not spatially and temporally dependent. Nevertheless, the 2*SLS* solution may still be attractive in practice and a recent study of temporal and spatial structure in the North West region of Britain (Bennett, 1975b) has adopted 2*SLS* instrumental variable estimators. Although this study ran into considerable problems of inefficiency and degeneracy of parameter estimates and made some rather gross assumptions about stationarity, it was still possible to identify meaningful patterns of spatio-temporal structure and policy effects. My point in raising this criticism is, therefore, to suggest that the solution to spatio-temporal estimation problems can be found in the application of simultaneous equation estimation techniques only as a first approximation to solve practical problems, or in very special cases.

A more general problem area connected with the models proposed in Section 6 of the paper is that each specification ignores the boundary and initial condition effects. The problem of specifying models along the boundary has been largely ignored in most geographic studies, except for the use of cordon surveys, etc. Moreover, the boundary value problem is only a special case of a further difficulty in that all the models proposed rely upon asymptotic estimation properties and sampling theory. The small sample properties of these estimators are unlikely to be as convenient and must be developed before the proposed models can have reliable application to practical problems, since small sample sizes are the rule in almost all spatio–temporal problems.

In making these comments I do not intend to understate the contribution of this paper but to define those areas in which significant statistical problems remain to be solved. What is required for the future is the derivation of practical small sample estimators (including estimators on the boundary), and the determination of appropriate minimal forms and representations for space–time forecasting problems.

IAN S. EVANS (Department of Geography, University of Durham): The authors are to be congratulated on providing statisticians with an authoritative and well-written outline of work on spatial analysis in human geography. Such an ambitious work inevitably leaves omissions, of which the largest concerns the correlation of spatial distributions. This is mentioned only briefly, in relation to the comparison of actual and simulated patterns. Hence the review has a univariate emphasis, which could be broadened to a multivariate one.

Difficulties and indeterminacies in fitting statistical models to single human distributions stem largely from correlations between the latter and other distributions such as rainfall, altitude or soil fertility which follow quite different models. Reasonable levels of prediction are achieved only by combining several such models. For the ensuing assessment of goodness-of-fit between the real spatial pattern of a dependent variable and its predicted or simulated counterpart, the authors would probably favour an autocorrelation (mainly contiguity) approach.

We should also consider spatial cross-spectral correlation (Rayner, 1971; this is surely an improvement on the aggregation approach of Section 3.3 and of Gehlke and Biehl, 1934), the correlation of trend surfaces (Krumbein and Jones, 1970) and the correlation of generalized surfaces or potential surfaces (Warntz, 1956). Criticism of the latter approach (Chow, 1961) suggests a relationship between spatial correlation and linear programming; one might draw an analogy between optimum relative location and a correlation of $+1$. Pielou (1965) provided a further technique for lateral spatial association in qualitative maps.

Little attention is paid to the nature of data in human geography. Most such data are expressed in percentage or ratio form, yet human geographers have as yet paid little attention to the experiences of geologists in transforming (Krumbein, 1957) or correlating (Chayes, 1971) these. The first generation of "statistical geographers" has, furthermore, been overawed by the null hypothesis of statisticians. In general, the nature of systematic deviations from randomness is of much greater interest than the question of whether or not randomness is present: this implies more attention to confidence intervals for observed measures of spatial structure. (The null hypothesis that *residuals* are random is, of course, important in the latter context.)

Finally, three minor qualifications may be offered to points in the review. First, the rank-size rule is a *non*-spatial model: in no way does it take account of relative spatial position. It may have an important spatial expression, and the work of Chapman (1970) is interesting in combining settlement size and settlement spacing with other scales of spatial differentiation in pattern.

Second, like Hepple (1974), the authors mention stationarity requirements especially in relation to spectral analysis. This is unfair, since most of the techniques discussed involve averaging over space: hence they imply stationarity of the relevant properties. I believe that geographers will welcome further work by statisticians on testing for stationarity, on robustness to deviations from stationarity, and on non-stationary models.

Third, data on a regular spatial grid are, fortunately, more common than implied in Section 3.4. From the 1971 census 1,571 social or economic variables are slowly becoming available for every 1 km square in Britain. The great mass of data provided via satellites is relevant to human as well as to physical geography. Human geographers would for many purposes prefer data for units equal in population rather than area, but unit-area data are preferable to data for irregular spatial divisions which hopelessly confound population scales as well as spatial scales.

This is not to denigrate the importance of the authors' attempt to make the best of a bad job by developing techniques for irregular data; such data will remain of primary importance in historical geography. But where we wish to test new models, it might be easier to start with gridded data, before moving on to the extra complications of irregular data.

The field considered by the authors is a very broad one and further partial reviews are provided by King (1969), Bassett (1972), Greer-Wootten (1972), Bartlett (1974), Hepple (1974) and Rogers (1974).

Mr R. MEAD (University of Reading): I found the whole paper interesting, particularly since my experience of the use of statistical methods in geography is rather limited. I want to comment on Section 4 on the study of pattern. I think it is important to recognize that most of the methods discussed in Section 4.1 were developed initially for ecological research and it is important therefore to ask how suitable these methods are in geography. The authors note the difficulty of drawing inferences about the underlying process from the distribution of quadrat counts. However, methods of detecting pattern from quadrat counts were not evolved for any purpose of process identification but rather, simply to detect departures from randomness; in particular, the Greig-Smith method was designed to detect the existence of several scales of pattern. It is perhaps worth noting that an extension of Greig-Smith methods now provides tests for patterns at several scales, each test being independent of the existence of pattern at other scales (Mead, 1974). The effects of quadrat size and interquadrat distance are important in the sense that inferences from data can be misleading if the precise methods of data collection are not considered but, I suggest, no more than that, and the main effect is on inferences about the underlying process.

The other point that needs emphasizing is that methods of pattern detection based on nearest-neighbour distances are designed for samples of independent observations, not for complete populations. The comment of the authors about the dangers of using nearest-neighbour distances from all individuals is most important. (I personally have found this necessary assumption of independence being ignored by research workers in geography, physiology and astronomy.) In this context it is worth-while pointing out that for most of the situations where it can be assumed that $\lambda$ is known (sentence of equation (4.4)) the independence assumption will not hold and therefore the tests discussed, assuming $\lambda$ to be known, will not be valid. Conversely for most situations in which independent observations can be taken, $\lambda$ will not be known and the tests quoted will not be appropriate. In such situations the test of Hopkins which was not referred to in the paper might be appropriate, possibly in the form proposed by Besag and Gleaves.

Dr S. OPENSHAW (Department of Town and Country Planning, Newcastle University): Most of the models described justify the adjective "spatial" only in the sense that they use geographical data. It is not surprising, therefore, that the major results of recent years have been to foster advances of a methodological nature rather than of a geographical one.

I would like to have your comments on the proposal that in order to become geographically relevant, spatial modelling should be seen as involving three stages, whereas at present it is a one step all or nothing statistical process. These stages are as follows.

(i) The first stage is concerned with the specification of model form using statistical and mathematical model building techniques. This is covered by your paper.

(ii) The second stage concerns the introduction of scale and aggregation effects. I think it must be recognized that spatial models are unlikely ever to produce consistent and comparable results until the scale and aggregation component which is implicit in all spatial data can be explicitly controlled. Empirical research has shown how it is possible to rearrange the scale and aggregation of data in such a way that the performance of a model can be made to vary between quite wide limits. This must have tremendous implications for building spatial models, indeed it is very doubtful whether this spatial freedom of the data can ever be removed, but it can be used.

(iii) The third stage is of comparable geographical importance to the second one. The earth's surface is not a flat featureless plain but a surface which displays strong spatial variations in geographical characteristics. Unfortunately, much of this information cannot be handled by either the current stage one models or the model building techniques used to provide theoretical justification for them. However, models can be developed to incorporate this kind of geographical information. For example, a global function can be replaced by a set of functions each of which is defined over a limited spatial domain so as to reflect aspects of macro spatial structure.

Clearly, these three stages cannot be independent of each other as the type of stage one model developed must also reflect stages two and three in a recursive manner. Nevertheless, in terms of this framework, the models you review hold the second stage constant, or try to remove it, and ignore stage three completely. In fact, in order to use these stage one models we have to pretend that we know far less about a study area than we actually do and as a result we often produce complicated statistical generalizations which are largely devoid of any really substantive geographical interpretation. I would suggest that this state of affairs is a direct result of ignoring stages two and three which is precisely where geographers could be expected to make their greatest contribution, and where the larger part of a geographical interpretation would have to come from.

The authors replied in writing after the meeting, as follows:

We are most grateful to the various discussants for their careful and thought-provoking comments on the paper. The even split between geographers and statisticians encourages us in the belief that the meeting has drawn the challenging problems of the former to the notice of the latter!

Professor Cormack goes straight to the heart of the matter in asking "sample or population"? We are reminded of an investigation into shoppers' behaviour patterns, described in the S.S.R.C. handbook on *Research in Human Geography*, where certain conclusions are drawn which hold "at least in Swansea". As with economic time series (perhaps more so) the question of population (essentially a unique set of observations) or sample (often from a conceptual super-population) must be faced. Elsewhere (Cliff and Ord, 1973, p. 8), we have drawn a distinction between situations where (a) the set of random permutations of the data may be regarded as the reference set, or (b) the data are regarded as a set of drawings from a parent population. As with other branches of statistics we can only draw inferences if we are willing to consider our data set as a realization of an underlying stochastic process. Whether this superstructure is useful must depend upon the judgement of the geographer. While recognizing that an isolated analysis does not produce a theory, we feel that the development and testing of new theories will usually proceed in an inferential framework—possibly a better co-ordinated framework than exists at present.

We agree that mean squares, rather than sums of squares, should be considered in Section 3.3.

We recognize that our proposed resolution of the equifinality of the negative binomial distribution is scantily presented. The bones of the argument are as follows:

*Case* 1

Let $X_1$ and $X_2$ be independent Poisson random variables with parameter $\lambda$, denoting the number of clusters per cell. If the size of each cluster follows a log series distribution with parameter $\alpha$ then the argument of Section 4.1 shows that the total number of individuals per cell, $Y_i$, follows a negative binomial distribution with parameters $k = \lambda\{-\log(1-\alpha)\}^{-1}$ and $\alpha$. That is,

$$Y_i \sim \text{NBD}\,(k, \alpha).$$

Further,

$$Y_1 + Y_2 \sim \text{NBD}\,(2k, \alpha).$$

*Case* 2

Let $Y_1$ and $Y_2$ be independent Poisson random variables denoting the counts in two *adjacent* cells. If the parameter $\lambda$ varies according to the gamma law, with index $k$ and scale parameter $\omega = (1-\alpha)/\alpha$ then

$$Y_i \sim \text{NBD}\,(k, \alpha).$$

Suppose now that the mean level of cases, $\lambda$, is similar in the two adjacent cells (implying a positive value for a simple spatial autocorrelation coefficient). In the limit, when the two cells have identical $\lambda$ values,

$$Y_1 + Y_2 \sim \text{NBD}\,\{k, 2\alpha/(1+\alpha)\}.$$

Thus, with the additional assumption that adjacent cells are similar, discrimination is possible. The general class of Cane (1973), referred to by Dr Wescott, considers mixtures of negative binomials with the "clustering" and heterogeneity hypotheses as opposite extremes. If two adjacent spatial units (or two time periods in accident studies) are considered, as above, the two models still give distinct distributions for $Y_1 + Y_2$.

As Mr Diggle has noted, the negative binomial distributions described here are limiting cases of spatial stochastic processes. Future spatial modelling must involve a more genuine regard for the effects of distance, even though this often leads to grave analytical difficulties. Dr Mollison's reinterpretation of Tinline's results and his work on the speed of spread of epidemics show both the need for a spatial dimension and the problems facing the builder of analytical models. Also Mr Openshaw's comments serve as a timely reminder of the need for such a change of direction.

Another recent development not covered in the paper is the study of the topological properties of drainage basins (Dacey, 1975). We agree with Dr Hansford-Miller that the use of topological methods needs fuller exploitation.

Like Professor Cormack, we feel that spatial stationarity may prove a rather elusive property. However, it is well worth pursuing because there are potential advantages from modelling in the frequency domain, as Professor Bartlett has indicated. Dr Evans feels that we have been less than fair to spectral methods in stressing the need for stationarity. We agree that time domain methods also require averaging, but these are, perhaps, more readily adapted to handle trends in the mean and evolutionary behaviour (although, see Priestley and Tong, 1973, and references therein). In general, we would agree with Mr Davies that space and frequency domain methods are complementary rather than competitive, and feel that insights can be gained from each mode of analysis.

On re-reading Medvedkov's paper, we agree that his description of "$x$ towns uniform $y$ towns random" is misleading. Professor Cormack's modified Poisson law seems a much clearer way of modelling the process, and is in the spirit of the Bernoulli–Poisson mixture used by Dacey (1971).

Professor Haggett and Dr Evans are both critical of existing models for settlement size. While accepting that models such as the rank-size rule are very naive, we would like to draw attention to the work of Haran and Vining (1973) who show how the theory of reversible Markov processes (Kingman, 1969) may be used to derive the rank-size rule as an "aspatial" limit to a spatial process. See, also, the review by Ord (1975). In addition, the Whitworth–Cohen models can be justified by such an approach although we would not pretend that this is conclusive. In passing, we note that Dr Hansford-Miller's histogram has the *J*-shape typical of the Pareto and rank-size rules.

We agree with Dr Evans that our paper was primarily concerned with univariate methods (although see Section 3.2.1) and we hope that Professor Bartlett's contribution to the discussion will stimulate further research to redress the imbalance in the literature.

The erratic boundary behaviour of polynomial trend surfaces described by Professor Haggett seems inherent in the method itself, as consideration of a regular grid with orthogonal polynomials reveals. The computing problems for higher order surfaces can be severe also. The spatial models of Section 6.3 may represent a reasonable alternative approach, possibly with spatial differencing to handle non-stationarities. The use of Fourier surfaces leads us back to a consideration of spectral methods and more detailed comparisons of these different methods would be useful. However, we do not agree with Dr Evans that spectral methods are necessarily better. Neither do we agree with Dr Hansford-Miller that we have ignored regression methods, as both Sections 3 and 6 depend heavily on this approach, and regression is a tool widely used by geographers. With regard to the interesting data concerning clergymen deprived of their livings, we wonder whether strong spatial autocorrelation might not be present as a result of local leadership effects?

We welcome the comments of Mr Besag and Mr Diggle relating to nearest-neighbour and quadrat methods. It is good news that the use of dependent nearest-neighbour distances does not seriously affect the inferences drawn about the spatial pattern. The improved tests mentioned should supplant those described in our paper, provided that the dependence between observations can be properly handled in these cases also.

Both Mr Openshaw and Dr Hansford-Miller stress the difficulties associated with the size of spatial data-recording units. Of course, spatial aggregation can always conceal major local variations and highly aggregated data cannot be used to answer questions about local variations. Nevertheless, it is reasonable in many cases to postulate some degree of regularity in spatial behaviour. One of the many advantages of data collection on fine regular grids, mentioned by Dr Evans, is that the loss of information caused by aggregation will be open to analysis.

While we recognize that purely spatial models have a restricted role to play in the study of spatial–temporal processes, we regard them as being rather more important than, apparently, does Dr Bennett. Our earlier comments on alternatives to polynomial trend surface methods indicate one potentially valuable application area. We agree with Dr Bennett's suggestions for reducing the size of the $\mathbf{X}^T\mathbf{X}$ matrix and only regret that his own work appeared too late for inclusion in the paper. We are, however, puzzled by his statement that the $\mathbf{X}_j$ should be mutually independent. While this may simplify the numerical details there is no theoretical reason to impose this requirement. The questions of boundary effects and the small sample performance of asymptotically efficient procedures are too broad to be considered here. However, we note that several Monte Carlo studies of small sample econometric estimators are reviewed by Johnson (1972, pp. 408–420) while Mr Ross-Parker of Reading University has carried out an investigation for the purely spatial model of Section 6.3.1 (abstract to appear in *Advances in Applied Probability* later this year). Clearly, further work is needed in these areas.

Professor Bartlett's comment on the importance of stochastic local extinction effects for measles epidemics in rural areas confirms our own tentative views (Cliff *et al.*, 1975a, p. 179) and this seems an interesting area for further research.

Many statistical investigations glibly assume normality without any justification (we are not exempt from this criticism either!) and we agree with Dr Evans that more care needs to be taken in this respect. The recent book by Fleiss (1973) describes existing statistical methods for ratio and proportions, while the generalized linear model of Nelder and Wedderburn (1972) is likely to be widely used in future regression studies. The relative emphasis of estimation against testing is very much a matter for the investigator, although one of us (J. K. O.) has strong views on the matter (see Ord and Patil, 1975).

We are more sanguine about progress on Mr Openshaw's stages two and three than he appears to be. We feel that some progress has already been made on stage two (Section 3.3 and Huijbregts, 1975) while the validation of methods for stationary processes allows us to move on to the non-stationary landscapes of the real world, perhaps following the approach of Rayner and Golledge (1972).

We are grateful to Professor Gregory for his kind remarks and take to heart his reminder that not everyone would agree with our view of human geography. Indeed, a comment by Hind (1864, p. 99) from the Society's *Journal* may be noted, even though quoted grossly out of context:

> "Hypothetical geography has proceeded far enough in the United States. In no country has it been carried out to such an extent, or been attended with more disastrous consequences."

We cannot agree with Dr Hansford-Miller that our attempts to model human behaviour smack of Marxist dictatorship. Surely the statistician relies on the freedom and independence of individuals to justify his limit theorems for aggregates!

Finally, we should like to thank contributors to the discussion for the additional references they have brought to our attention. We would also like to mention recent copies of *The Statistician*. Parts 3 and 4 (1974) are entirely devoted to statistical methods in geography, while Part 3 (1975) is devoted to comments on these papers by statisticians.

## References in the Discussion


BARTLETT, M. S. (1974). The statistical analysis of spatial pattern. *Adv. Appl. Prob.*, **6**, 336–358.

BASSETT, K. (1972). Numerical methods for map analysis. *Progr. Geogr.*, **4**, 217–254.

BENNETT, R. J. (1975a). The representation and identification of spatio-temporal systems: an example of population diffusion in N.W. England. *Transactions and Papers*, Institute of British Geographers (to appear).

—— (1975b). Dynamic systems modelling of the North-west region. *Environment and Planning*, **7**, Nos 2, 3, 4 and 5.

BESAG, J. E. (1974). On spatial–temporal models and Markov fields. *Trans. European Meeting of Statisticians* (to appear).

BROWN, S. and HOLGATE, P. (1974). The thinned plantation. *Biometrika*, **61**, 253–262.

CANE, VIOLET R. (1973). The concept of accident proneness. *Bull. Inst. Math.* (Bulgaria), **XV**, 183–189.

CHAPMAN, G. P. (1970). The application of information theory to the analysis of population distributions in space. *Econ. Geogr.*, **46**, 317–331.

CHAYES, F. (1971). *Ratio Correlation*. Chicago: University of Chicago Press.

CHOW, G. C. (1961). Review of *Toward a Geography of Price*, by W. Warntz. *J. Amer. Statist. Ass.*, **56**, 209–210.

DACEY, M. F. (1971). Regularity in spatial distributions. In *Statistical Ecology* (G. P. PATIL, ed.), Vol. 1, pp. 287–309. University Park, Pennsylvania: Pennsylvania State University Press.

—— (1975). Probability laws for topological properties of drainage basins. In *Statistical Distributions in Scientific Work*, (G. P. PATIL *et al.*, eds), Vol. 2, pp. 327–342. Dordrecht and Boston: Reidel.

DANIELS, H. E. (1975). The deterministic spread of a simple epidemic. In *Perspectives in Probability and Statistics*. (Papers in honour of M. S. BARTLETT on the occasion of his 65th birthday.) London: Academic Press (for Applied Probability Trust).

DE LA BLACHE, V. P. (1911). *The Personality of France*, p. 14. (Translated by H. C. Brentnall, 1928.)

DIGGLE, P. J., BESAG, J. E. and GLEAVES, J. T. (1975). On the statistical analysis of spatial point patterns by means of distance methods. *Biometrics* (to appear).

DUNCAN, O. D., CUZZORT, R. P. and DUNCAN, B. (1961). *Statistical Geography*. New York: Free Press of Glencoe.

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*, 3rd ed., Vol. 1. New York: Wiley.

FLEISS, J. L. (1973). *Statistical Methods for Rates and Proportions*. New York: Wiley.

FORDE, C. D. (1934). *Habitat, Economy and Society*.

FRENCH, A. S. and HOLDEN, A. V. (1971). Alias-free sampling of neuronal spike trains. *Kybernetik*, **8**, 165.
GEHLKE, C. E. and BIEHL, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census-tract material. *J. Amer. Statist. Ass.*, **29**, 169–170.
GREER-WOOTTEN, B. (1972). A bibliography of statistical applications in Geography. Technical paper No. 9, Ass. Amer. Geogr. Commission on College Geography.
HAMMERSLEY, J. M. (1966). First-passage percolation. *J. R. Statist. Soc.* B, **28**, 491–496.
HANSFORD-MILLER, F. H. (1965). The distribution of religious groups in England and Wales, 1350–1550. M.Sc. Thesis, University of London.
—— (1968). A geographical and statistical analysis of religious nonconformity in England and Wales, 1550–1720. Unpublished Thesis. Vol. I, 86–107; Vol. II, 23–33.
—— (1970). *The* 282 *Protestant Martyrs of England and Wales*, 1555–1558. London: Educational Publishers.
—— (1971). *Quantitative Geography for Schools, Book One*. 36–40. London: Educational Publishers.
HARAN, E. G. P. and VINING, D. R. (1973). On the implications of a stationary urban population for the size distribution of cities. *Geogr. Anal.*, **5**, 296–308.
HEPPLE, L. (1974). The impact of stochastic process theory upon spatial analysis in human geography. *Progr. Geogr.*, **6**, 89–142.
HOLGATE, P. (1965a). Tests of randomness based on distance methods. *Biometrika*, **52**, 345–353.
—— (1965b). Some new tests of randomness. *J. Ecol.*, **53**, 261–266.
HOPKINS, B. (1954). A new method of determining the type of distribution of plant individuals. *Ann. Bot.*, **18**, 213–226.
ISHAM, V. S. (1975). On a point process with independent locations. *J. Appl. Prob.*, **12**, (in the press).
KENDALL, D. G. (1965). Mathematical models of the spread of infection. In *Mathematics and Computer Science in Biology and Medicine*. London: Medical Research Council.
KERSTAN, J. and MATTHES, K. (1965). Stationare zufallige Punktfolgen. II. *J-ber. Deutsch. Math. Verein.*, **66**, 106–118.
KING, L. J. (1969). The analysis of spatial form and its relation to geographic research. *Ann. Ass. Amer. Geogr.*, **59**, 573–595.
KINGMAN, J. F. C. (1969). Markov population processes. *J. Appl. Prob.*, **6**, 1–18.
KOLMOGOROV, A. N., PETROVSKY, I. and PISCOUNOV, N. (1937). Etude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Bull. Univ. d'Etat à Moscou*, A1 Fasc. 6, 1–25.
KRUMBEIN, W. C. (1957). Comparison of percentage and ratio data in facies mapping. *J. Sediment. Petrol.*, **27**, 293–297.
—— (1966). A comparison of polynomial and Fourier models in map analysis. Technical Report No. 2, Northwestern University, Department of Geography.
KRUMBEIN, W. C. and JONES, T. A. (1970). The influence of areal trends on correlations between sedimentary properties. *J. Sediment. Petrol.*, **40**, 656–665.
MEAD, R. (1974). A test for spatial pattern at several scales. *Biometrics*, **30**, 295–307.
MOLLISON, D. (1972). The rate of spatial propagation of simple epidemics. *Proc. 6th Berk. Symp. Math. Statist. & Prob.*, **3**, 579–614.
—— (1973). Percolation processes and tumour growth. *Adv. Appl. Prob.*, **6**, 233–235.
NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. R. Statist. Soc.* A, **135**, 370–384.
ORD, J. K. (1975). The size of human settlements. In *Statistical Distributions in Scientific Works* (G. P. Patil *et al.*, eds), Vol. 2, pp. 141–150. Dordrecht and Boston: Reidel.
ORD, J. K. and PATIL, G. P. (1975). Statistical modelling: an alternative view. In *Statistical Distributions in Scientific Work*, (G. P. Patil *et al.*, eds) Vol. 2, pp. 1–9. Dordrecht and Boston: Reidel.
PIELOU, E. C. (1965). The concept of randomness in the pattern of mosaics. *Biometrics*, **21**, 908–920.
PRIESTLEY, M. P. and TONG, H. (1973). On the analysis of bivariate non-stationary processes. *J. R. Statist. Soc.* B, **35**, 153–166.
RICHARDSON, D. (1973). Random growth in a tesselation. *Proc. Camb. Phil. Soc.*, **74**, 515–528.
ROGERS, A. (1974). *Statistical Analysis of Spatial Dispersion: The Quadrat Method.* London: Pion.
SMITH, J. M. (1968). *Mathematical Ideas in Biology*. Cambridge: University Press.
STEELE, J. H. (1974). Stability of plankton ecosystems. In *Ecological Stability* (M. B. Usher and M. H. Williamson, eds), Part 4. London: Chapman & Hall.

TANNER, J. C. (1963). Car and motorcycle ownership in the counties of Great Britain in 1960. *J. R. Statist. Soc.* A, **126**, 274–284.

TINBERGEN, J. (1968). The hierarchy model of the size distribution of centres. *Papers and Proc. Regional Sci. Ass.*, **20**, 65–68.

WARNTZ, W. (1956). Measuring spatial association with special consideration of the case of market orientation of production. *J. Amer. Statist. Ass.*, **51**, 597–604.

WILLIAMS, T. and BJERKNES, R. (1972). A stochastic model for abnormal clone spread through epithelial basal layer. *Nature, Lond.*, **236**, 19–21.

WOOLDRIDGE, S. W. (1956). *The Geographer as Scientist*. London: Nelson.

YULE, G. U. and KENDALL, M. G. (1937). *Introduction to the Theory of Statistics*. London: Griffin.