Michael Batty*

Spatial Entropy

Abstract

A major problem in information theory concerns the derivation of a continuous measure of entropy from the discrete measure. Many analysts have shown that Shannon's treatment of this problem is incomplete, but few have gone on to rework his analysis. In this paper, it is suggested that a new measure of discrete entropy which incorporates interval size explicitly is required; such a measure is fundamental to geography and this statistic has been called spatial entropy. *The use of the measure is first illustrated by application to one- and two-dimensional aggregation problems, and then the implications of this statistic for Wilson's entropy-maximizing method are traced. Theil's aggregation statistic is reinterpreted in spatial terms, and finally, some heuristics are suggested for the design of real and idealized spatial systems in which entropy is at a maximum.*

INTRODUCTION

In the physical sciences, there are few concepts which have more widespread applicability than the concept of entropy. Entropy appears to have that elusive but irresistible quality of generality which tempts researchers from very different fields to use the idea in defining the

*Geoff Hyman provided much of the inspiration for this work, and his many discussions on the concept of entropy with the author have served to clarify several problem areas. Thanks are also due to Andrew Broadbent for his ideas on continuous entropy in trip distribution modelling and to Roger Sammons for his comments on many of the mathematical details. This research has been partly financed by the Centre for Environmental Studies, London.

Michael Batty is a lecturer in geography at the University of Reading.

structure and behavior of their various systems of interest. Yet such widespread applicability is not without its problems. Georgescu-Roegen [*15*], writing on the application of the entropy law to economic systems, states that the meaning of entropy "varies substantially, at times even within the same domain of intellectual endeavour," and he continues by suggesting that the terminology of the concept "is probably the most unfortunate in the history of science."

The difficulties over the definition of entropy seem to originate from the fact that at one extreme, the concept is the basis of the second law of thermodynamics, which states that the entropy of a physical system must always increase, whereas at the other extreme, the concept is used in defining the amount of information contained in a probability distribution. In fact, there is a somewhat tortuous route linking these two interpretations through arguments involving statistical physics. Yet the implications of the concept are still difficult to trace in any definitive sense [*20*].

The last decade has seen several applications of the concept in geographical studies, and these applications can be conveniently divided into those concerned with the wider implications of entropy in a thermodynamic sense and those applications concerned with entropy from an information theory viewpoint. Leopold and Langbein [*22*] use entropy in deriving the most probable profile of a river system, whereas Curry [*12*] derives the rank-size frequency distribution of settlements from an application of the concept. There are, however, many more geographical applications in the context of information theory. For example, Chapman [*9*], Semple and Golledge [*29*], and Gurevich [*17*] use the concept to measure the amount of information in spatial probability distributions. Berry and Schwind [*6*] have followed Theil [*32*] in analyzing migration flows using an algebra of entropy based upon the algebra of probability. Curry [*13, 14*] has tentatively suggested that the entropy statistic may be of use in exploring spatial series, and Wilson [*36*] has laid a new foundation for geographical model-building by using the concept in analogy to problems in statistical mechanics.

There are already some useful reviews of the concept in geographical analysis. Anderson [*1*], Marchand [*24*], and Medvedkov [*25*] provide interesting summaries of the idea, and Mogridge [*26*] reviews the concept in depth in both spatial and economic contexts. However, running through much of this work is an impression of uncertainty and unfamiliarity with the concept which occasionally manifests itself in misinterpretations of the properties and terminology of the entropy law [*10*]. In this paper, the definition and implications of a *spatial entropy* will be developed from the information theory standpoint first developed by Shannon [*30*] and Wiener [*35*], and the terminology used is that which is generally accepted in the field of communication theory.

The first part of this paper will deal with the mathematics and derivation of a formula for spatial entropy, which will then be applied to some simple spatial aggregation problems. The entropy-maximizing model used by Wilson [*36*] in geographical analysis is reinterpreted in the light of the new formula, and this suggests certain design principles for the partitioning of spatial systems. In particular, the emphasis in this paper will be focused upon using spatial entropy to define cutoff points for boundary definition, to define a requisite number of zones for the analysis of spatial systems, and to explore questions concerning the optimal geometry of partitioned spatial systems.

THE MATHEMATICS OF ENTROPY

The amount of entropy in a probability distribution is also known as the information content of that distribution. Information is defined in terms of the prior probabilities of certain events occurring; the greater the prior uncertainty of such an occurrence, the greater the information gained if such an event occurs. Criteria for defining an information statistic suggest that the measure would vary from zero to infinity and that the measure would be additive between independent events. Hartley [*19*] was the first to define the entropy of a particular event, but it was Shannon [*30*] who first derived the general formula for measuring the amount of entropy in any set of probabilities. Shannon's well-known formula can be written as

$$H = \sum_i p_i \ln\left(\frac{1}{p_i}\right) = -\sum_i p_i \ln p_i, \tag{1}$$

$$\sum_i p_i = 1, \tag{2}$$

where p_i represents the probability of event i occurring and the summation is over the range $i = 1, 2, \ldots, n$ unless stated otherwise. In communications theory, the logarithms are taken to the base 2 whereas in this paper, the logarithms are natural logarithms to the base e.

Shannon also presented a formula for measuring the amount of information in a probability density which he suggested was the continuous form of equation (1).

$$H = -\int p(x) \ln p(x)\, dx, \tag{3}$$

$$\int p(x)\, dx = 1. \tag{4}$$

The integrations in equations (3) and (4) are implicit over the whole range. It is of interest to note that in Wiener's definition of information [*35*], the sign is the reverse of Shannon's and this early difference has undoubtedly been responsible for some of the ensuing confusion in the last two decades. However, the most important problem in both Shannon's and Wiener's work is the fact that equation (3) cannot be derived from equation (1) by letting the interval size in equation (1) tend to zero and passing to the limit. Shannon's analogy between discrete and continuous entropy lacks rigor. Furthermore, in problems where interval size is important, as in geographical studies, such lack of rigor is immensely disturbing.

Several researchers have been quick to point out the incorrectness of deriving equation (3) from equation (1). Woodward [*37*], in a useful book, develops a neat interpretation of the two formulas, and Jaynes [*21*] discusses the problem of discrete versus continuous entropies in some depth. To demonstrate that equation (1) does not converge to equation (3), consider that p_i in equation (1) is replaced by $p(x)_i\,\Delta x_i$ where Δx_i is the interval size. Then following Goldman [*16*],

$$H = \lim_{\Delta x_i \to 0} -\sum_i p(x)_i\,\Delta x_i\,\ln(p(x)_i\,\Delta x_i). \tag{5}$$

Expanding equation (5) gives

$$H = \lim_{\Delta x_i \to 0} -\sum_i p(x)_i \ln(p(x)_i)\,\Delta x_i + \lim_{\Delta x_i \to 0} -\sum_i p(x)_i \ln(\Delta x_i)\,\Delta x_i\,. \tag{6}$$

Equation (6) converges to

$$H = -\int p(x)\ln p(x)dx - \lim_{\Delta x_i \to 0} \sum_i p(x)_i \ln(\Delta x_i)\Delta x_i. \tag{7}$$

If it is assumed that Δx_i are equal for all i, then the second term on the right-hand side of equation (7) can be simplified and equation (7) becomes

$$H = -\int p(x)\ln p(x)\,dx - \lim_{\Delta x \to 0} \ln \Delta x. \tag{8}$$

Equation (8) demonstrates that the discrete entropy of equation (1) increases without bound for the case in which the Δx_i's are equal. This argument can be easily generalized to the case where each Δx_i is different.

Rearranging equation (7) and substituting the right-hand side of equation (1) for H, we can write the continuous entropy of equation (3) as

$$-\int p(x)\ln p(x)\,dx = -\sum_i p_i \ln p_i + \lim_{\Delta x_i \to 0} \sum_i p(x)_i \ln(\Delta x_i)\,\Delta x_i. \tag{9}$$

Since $p(x)_i\,\Delta x_i = p_i$, equation (9) becomes

$$-\int p(x)\ln p(x)\,dx = -\sum_i p_i \ln p_i + \lim_{\Delta x_i \to 0} \sum_i p_i \ln(\Delta x_i). \tag{10}$$

It is clear from (10) that the continuous entropy has been derived as the difference between the discrete entropy of equation (1) and a term reflecting the relationship between the set of probabilities and their intervals of measurement. Thus the formula for spatial entropy in which Δx_i represents the spatial interval size is written as

$$H = \lim_{\Delta x_i \to 0} -\sum_i p_i \ln\left(\frac{p_i}{\Delta x_i}\right). \tag{11}$$

Equation (11) is the most important in this paper, for it is the basis of much of the spatial analysis which follows from its application to geographical problems.

It is suggested here that equation (11) is more useful in spatial analysis than equation (1) because the effects of partitioning a spatial system in different ways can be compared in absolute terms using equation (11). Furthermore, the spatial entropy statistic can be used in comparisons between different regions.[1] There are several other possible entropy statistics which can be defined in discrete or continuous terms, and in the following section, two important alternative statistics are described.

Spatial Entropy Statistics

A well-known alternative measure of entropy is based on the concept of redundancy in communications theory. The redundancy Z involves measuring the ratio of actual entropy to the maximum entropy of a system and subtracting this ratio from 1. This measure is defined as

$$Z = 1 - \frac{H}{H_{\max}}. \tag{12}$$

Using the discrete entropy formula to measure H as in equation (1) and noting that the maximum entropy of equation (1) can easily be shown to be $\ln n$, we can rewrite equation (12) as

$$Z = 1 + \frac{\sum_i p_i \ln p_i}{\ln n}, \; 0 \leq Z \leq 1. \tag{13}$$

As the number of intervals n changes, the value of equation (13) changes, and comparisons between different systems are therefore difficult where

[1] Professor Curry's intriguing use of the entropy statistic to measure variation in a map which has been subjected to M-fold averaging by cascading [*14*] could be made more general by the use of equation (11). Thus it might be possible to account for irregular shapes and sizes of zones in averaging, and the information exhibited by different maps with different numbers of zones could be compared directly.

this measure of discrete redundancy is used.

However, with the continuous entropy in equation (11) as a measure of H, comparisons are meaningful. Assuming that the interval size Δx_i is defined as A/n where A is the area of the system and n is the number of zones, we can write the redundancy as

$$Z = \frac{\ln n + \sum_i p_i \ln p_i}{\ln A}. \tag{14}$$

Equation (14) converges to a limiting value for Z as n increases. Another useful statistic is based on the concept of information gain, which has been studied in some depth by Theil [*32*]. Information gain I is defined as

$$I = H_{\max} - H. \tag{15}$$

The most important characteristic of equation (15) is that it converges to the same value for both discrete and continuous entropies. Information gain is calculated from equation (1) as

$$I = \ln n + \sum_i p_i \ln p_i. \tag{16}$$

Information gain is now calculated from equation (11) as

$$I = \ln A + \sum_i p_i \ln\left(\frac{p_i}{\Delta x_i}\right),$$

$$= \ln n + \sum_i p_i \ln p_i. \tag{17}$$

Equations (16) and (17) demonstrate the fact that the statistics for redundancy and information gain are related by

$$Z = \frac{I}{H_{\max}} = \frac{H_{\max} - H}{H_{\max}}. \tag{18}$$

The effect of changing the interval size Δx_i can be assessed using such statistics. For example, equation (11) provides a discrete approximation to equation (8) for given Δx_i, and a measure of information loss can be calculated by comparing such approximations with their continuous forms. Mathematically, the problem involves approximating an integral by a finite sum and in a geographical context, such interpretations are helpful in partitioning a spatial system into zones. Criteria for zoning can be fixed in terms of a difference ϵ between the integral and its finite sum, and such criteria could involve the choice of Δx_i by finding Δx_i which satisfy

$$\left| -\int p(x) \ln p(x)\, dx + \sum_i p_i \ln\left(\frac{p_i}{\Delta x_i}\right)\right| < \epsilon . \tag{19}$$

To demonstrate the application of this technique, some simple problems of spatial aggregation will be introduced. Such problems involve firstly the definition of regional boundaries and secondly the number of zones which provide an acceptable description of continuously variable spatial phenomena. A simple one-dimensional aggregation will be first introduced, followed by a more complex two-dimensional aggregation problem.[2] The data used in these examples are based upon the 1966 Census of Population for the Reading region.

One-Dimensional Aggregation Problems

A simple one-dimensional probability distribution based upon the well-known population density function [*8*] has been used as a basis for this analysis. The density equation is

$$p(r) = Ke^{-\lambda r}, \tag{20}$$

where $p(r)$ is the probability of location at distance r from the origin 0, λ is a parameter of the density function, and K is a normalizing constant defined from

$$\int_0^R p(r)\, dr = \int_0^R Ke^{-\lambda r} dr = 1. \tag{21}$$

Note that R is the boundary of the region. K is evaluated as

$$K = \frac{\lambda}{(1-e^{-\lambda R})}. \tag{22}$$

The parameter λ is related to the mean travel distance $\bar{R}$ to the origin 0 and is calculated from

[2]The subsequent analysis is of theoretical rather than practical significance, for the analysis has been restricted to monocentric radially symmetric systems. An extension to multicentric systems is mathematically cumbersome although essential if the results of this section are to have any practical import. Future research will be concentrated on this problem.

$$\bar{R} = \int_0^R K e^{-\lambda r} r \, dr,$$

$$= \frac{1}{\lambda} - \frac{e^{-\lambda R} R}{(1 - e^{-\lambda R})}. \tag{23}$$

The value of λ can be found'by solving (23) iteratively for given $\bar{R}$. If $R = \infty$, then $\lambda = 1/\bar{R}$, and this value for λ serves as a useful first approximation in the iterative scheme.

The problem of choosing a value for R which is the regional boundary can be approached by considering the difference between the entropy for $R = \infty$ and the entropy for finite R. Then for any R, the entropy is calculated as

$$H = -\int_0^R p(r) \ln p(r) \, dr,$$

$$= -\frac{\lambda}{(1 - e^{-\lambda R})} \int_0^R e^{-\lambda r} (\ln \lambda - \lambda r - \ln(1 - e^{-\lambda R})) dr. \tag{24}$$

Equation (24) simplifies to

$$H = -\ln \lambda + 1 - \frac{\lambda e^{-\lambda R} R}{1 - e^{-\lambda R}} + \ln(1 - e^{-\lambda R}). \tag{25}$$

When $R = \infty$, it is clear that $H = -\ln \lambda + 1$, and therefore the difference in entropy $H(e)$ between R and ∞ is easily evaluated as

$$H(e) = -\int_R^\infty p(r) \ln p(r) \, dr,$$

$$= \frac{\lambda e^{-\lambda R} R}{1 - e^{-\lambda R}} - \ln(1 - e^{-\lambda R}). \tag{26}$$

In Figure 1, equations (25) and (26) are plotted against different values of R. With 0.05 as an acceptable value for the ratio of equations (26) to (25), it is clear from the graph that the radius of a region centered on the origin 0 would be in the order of about 25 minutes travel time. Different ratios would lead to different radii, and this analysis could be thus used to determine the boundaries of any monocentric region.

To illustrate the problem of describing a probability distribution using the discrete and continuous entropy statistics, equation (1) is first plotted against the number of intervals n as in Figure 2. A better illustration of the fact that the discrete entropy increases without bound as n increases,

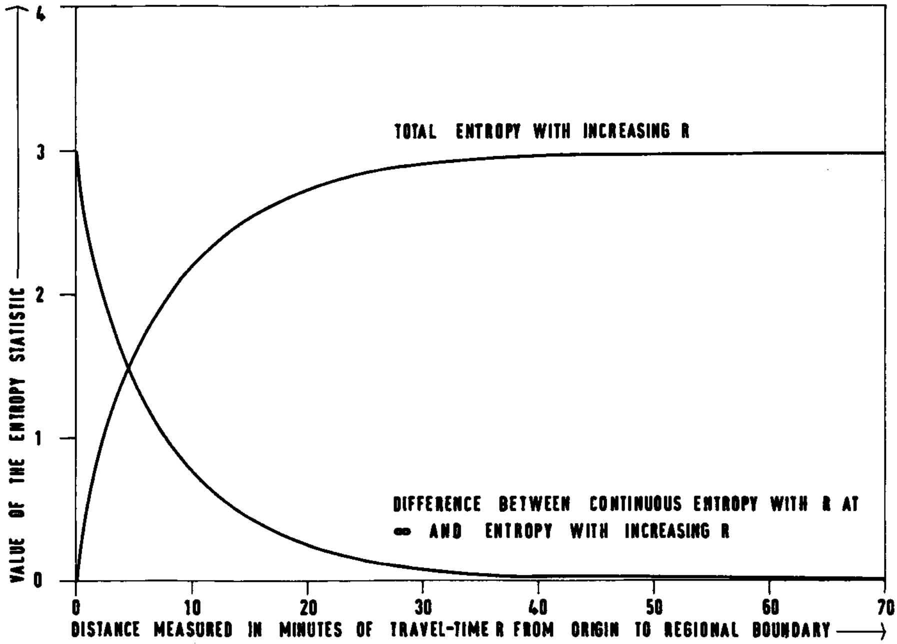


FIG. 1. Variation in one-dimensional continuous entropy with different regional boundaries.

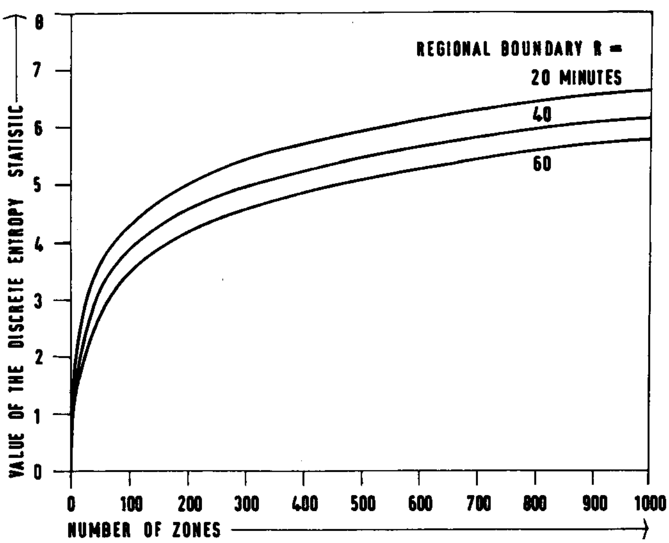


FIG. 2. Variation in one-dimensional discrete entropy with different numbers of zones.

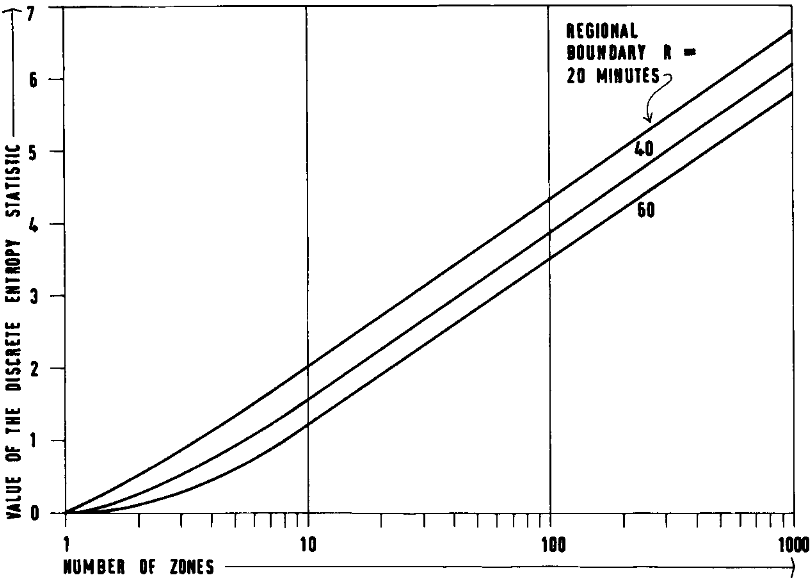


FIG. 3. One-dimensional discrete entropy against the logarithm of the number of zones.

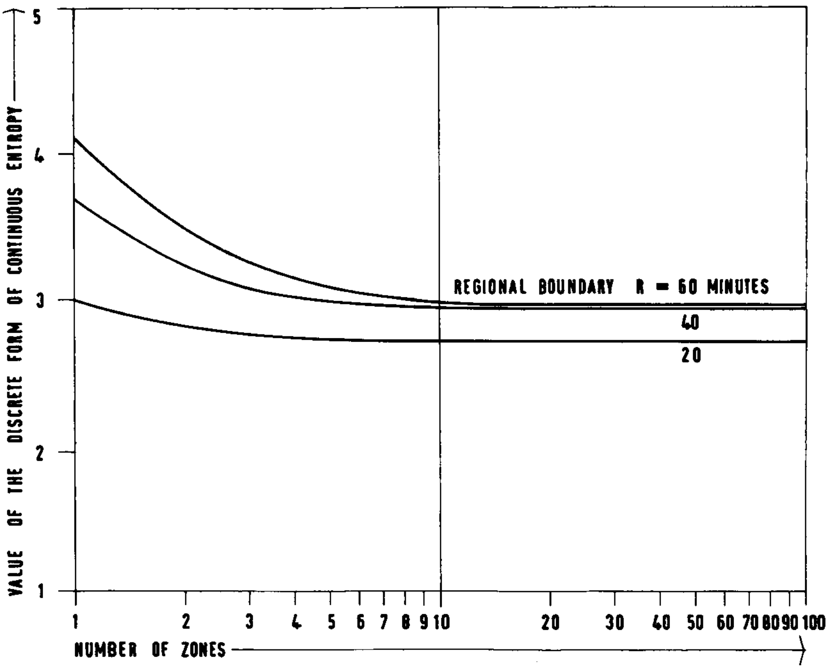


FIG. 4. One-dimensional continuous entropy against the logarithm of the number of zones.

which has already been deduced, is provided in Figure 3, where the entropy is plotted against the logarithm of the number of zones. It is also clear from Figure 3 that for large n, the increase in entropy tends to be linearly related to the logarithm of the number of zones.

The effect of increasing the number of zones (or equivalently reducing the interval size Δx_i) on the discrete approximation to continuous entropy given in equation (11) is plotted in Figure 4.[3] It is apparent from this graph that with a radius R set at 40 minutes travel time, about 10 zones ensure that the discrete entropy is within 2 percent of the value for the continuous entropy. Although a one-dimensional region is highly unrealistic, some insight into the number of zones required in a two-dimensional region is provided by squaring this number of zones. This gives as a first approximation 100 zones, which is certainly in excess of the number of zones used recently in building operational urban models of this region [*3*]. However, a more realistic number of zones can only be calculated when the more complex two-dimensional problem is considered; the next section outlines such an analysis.

Two-Dimensional Aggregation Problems

A two-dimensional probability distribution based on the population density model [*8*, *23*], is used as the basis for extending this analysis. This distribution measures the probability of locating at a given distance r from the origin 0 in a radially symmetric density field of the type developed by Angel and Hyman [*2*]. Location in the model is defined by polar coordinates where θ is the angle of variation and r is the distance from the pole. The density function is

$$p(r,\theta) = Ke^{-\lambda r}, \tag{27}$$

where λ is a parameter and K is a normalizing constant which ensures that

$$\int_0^{2\pi} \int_0^R p(r,\theta)\, r\, d\theta\, dr = 1. \tag{28}$$

K is evaluated from equation (28) as

$$K = \frac{\lambda}{2\pi\left(\dfrac{1}{\lambda}(1 - e^{-\lambda R}) - e^{-\lambda R}R\right)} \tag{29}$$

[3]Comparing Figures 3 and 4, it is apparent that in Figure 3 for a given number of zones, discrete entropy decreases as the regional boundary increases, while in Figure 4, there is a reverse effect. This difference is due to the way in which Δx_i is measured in the discrete approximation to the continuous case.

The parameter λ, as in the one-dimensional function, is related to the mean travel distance $\bar{R}$ to the pole which is

$$\bar{R} = \int_0^{2\pi} \int_0^R Ke^{-\lambda r} r^2 \, d\theta dr. \tag{30}$$

Evaluating equation (30) gives

$$\bar{R} = K2\pi \left\{ \frac{2}{\lambda^3} - \frac{e^{-\lambda R}}{\lambda} \left(R^2 + \frac{2R}{\lambda^2} + \frac{2}{\lambda} \right) \right\}. \tag{31}$$

A value for λ can be found by iteration of equation (31) with a first approximation for λ calculated from

$$\bar{R} = \int_0^{2\pi} \int_0^{\infty} Ke^{-\lambda r} r^2 \, d\theta dr = \frac{2}{\lambda}. \tag{32}$$

As a digression, in comparing equation (32) with equation (23), it is interesting to note that for a one-dimensional system, $\lambda = 1/\bar{R}$, whereas for a two-dimensional system, $\lambda = 2/\bar{R}$.

As in the previous analysis, the first stage in determining a value for R involves relating the entropy for finite R to the entropy in which $R = \infty$. The entropy H is defined as

$$\begin{aligned} H &= -\int_0^{2\pi} \int_0^R p(r,\theta) \ln p(r,\theta) \, r \, d\theta dr \\ &= -K2\pi \int_0^R e^{-\lambda r} (\ln K - \lambda r) \, r \, dr \\ &= -K2\pi \ln K \int_0^R e^{-\lambda r} r \, dr + \lambda K2\pi \int_0^R e^{-\lambda r} r^2 \, dr. \end{aligned} \tag{33}$$

The entropy H in equation (33) is evaluated as

$$H = -\ln \lambda + \ln 2\pi + \ln\left(\frac{1}{\lambda}(1 - e^{-\lambda R}) - e^{-\lambda R} R \right) + \frac{\left(\frac{2}{\lambda} - \lambda e^{-\lambda R} \left(R^2 + \frac{2R}{\lambda^2} + \frac{2}{\lambda} \right) \right)}{\left(\frac{1}{\lambda}(1 - e^{-\lambda R}) - e^{-\lambda R} R \right)} \tag{34}$$

If $R = \infty$, then $H = -2 \ln \lambda + 2 + \ln 2\pi$, and therefore the error

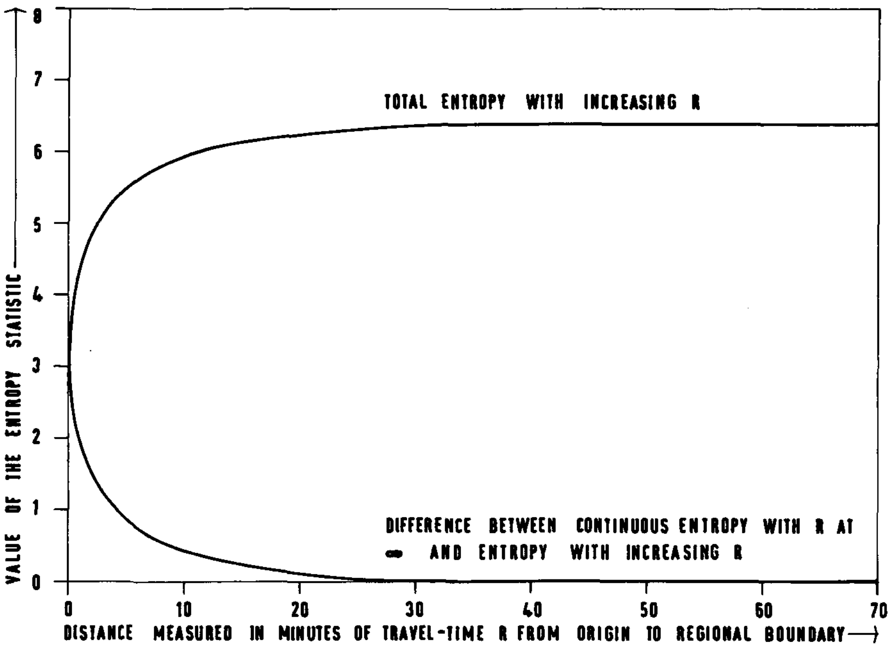


FIG. 5. Variation in two-dimensional continuous entropy with different regional boundaries.

term $H(e)$ can easily be deduced by subtracting equation (34) from the value of entropy when $R = \infty$. The equation for $H(e)$ has similar terms to equation (34) and has therefore been omitted because of its somewhat cumbersome nature.

In Figure 5, the entropy from equation (34) and the entropy error $H(e)$ are plotted against different values for R. With 0.05 as the ratio of $H(e)$ to the entropy where $R = \infty$, the graph suggests that this ratio is met when the radius of the region is about 15 minutes travel time. This radius is smaller than the one-dimensional radius for obvious reasons, but this value gives a useful indication of the cut-off point for defining the regional boundary. Equation (11) has been calculated for various values of Δx_i in the quest to determine the optimum number of zones for describing the population density field and this graph is shown as Figure 6. As in the previous analysis, a 2 percent difference between equation (3) and equation (11) is reached when the region is partitioned into about 100 zones with a radius R equal to 20 minutes travel time. This value is close to the value of R calculated for the one-dimensional case, which is encouraging.[4]

[4] At present, some research is being carried out into the differences in computing equation (11) using polar or Cartesian coordinates. A measure of this difference can be derived analytically using the continuous form of entropy by finding the transformation between the coordinate systems [*16*, *33*].

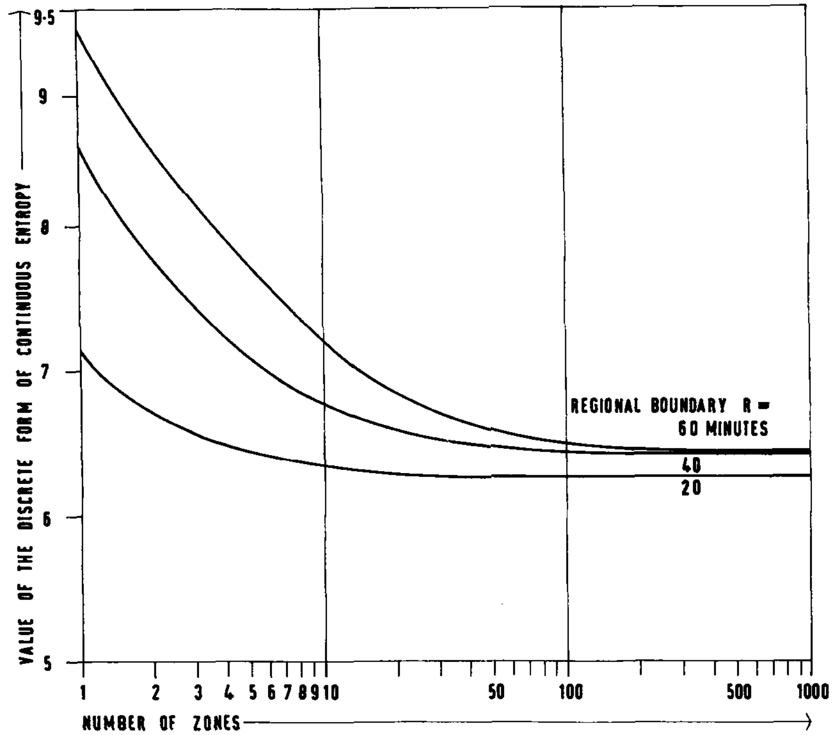


FIG. 6. Two-dimensional continuous entropy against the logarithm of the number of zones.

The same type of analysis can be carried out using the redundancy and information-gain statistics. Figures 7 and 8 show that plots of redundancy Z from equation (14) and information gain I from equation (17) behave similarly and converge to limiting values as the number of zones n is increased. However, the rate of convergence of both these statistics is slower than the convergence of equation (11), and the 2 percent criteria used above suggest that number of zones required would be nearer to 500 than to 100. Precise theoretical criteria are somewhat arbitrary to determine, but a summary of the pertinent features of this analysis is given in Table 1 in terms of specific limits. Having described a procedure establishing rules for boundary definition and zone size, it is now necessary to take the analysis further by exploring the implications of this work for geographical model-building.

A REINTERPRETATION OF THE ENTROPY-MAXIMIZING MODEL

One of the most important conceptual and technical advances in theoretical geography and model-building in recent years is the applica-

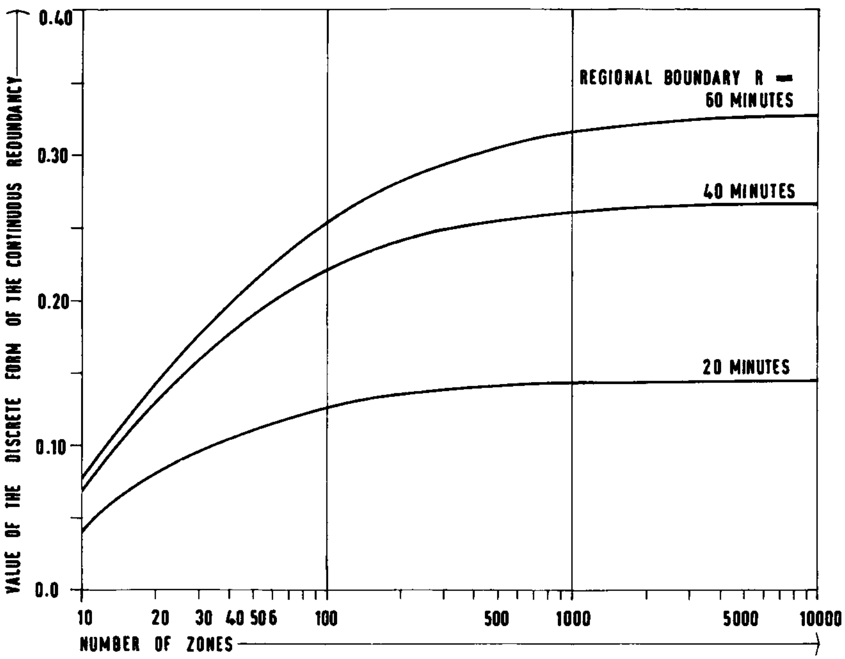


FIG. 7. Two-dimensional continuous redundancy.

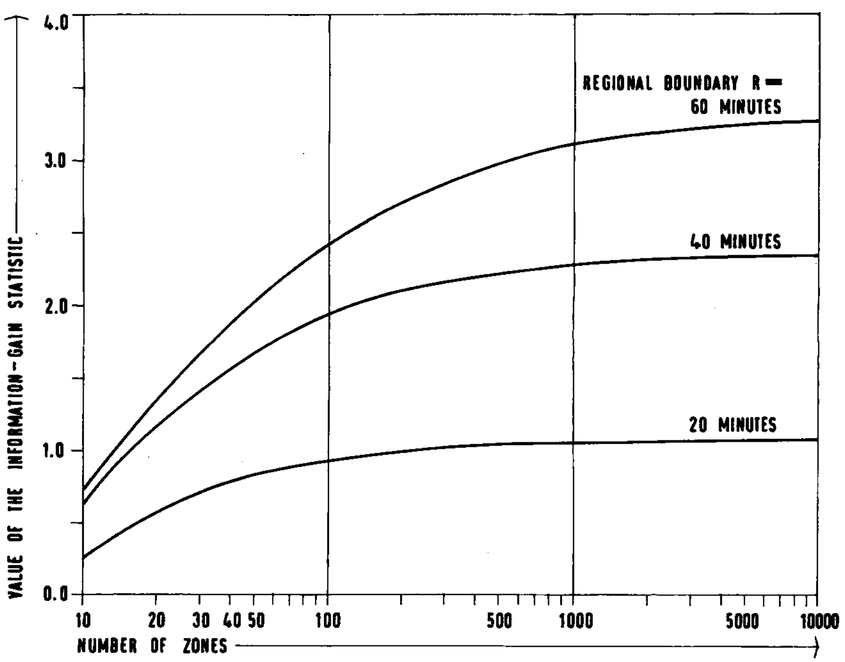


FIG. 8. Two-dimensional information gain.

TABLE 1

VALUES FOR THE ONE- AND TWO-DIMENSIONAL ENTROPY STATISTICS IN A MONOCENTRIC REGION WITH A RADIUS OF 20 MINUTES TRAVEL TIME

Entropy Statistic	Actual Value	Limiting Value
One-dimensional continuous entropy	2.8332	2.9773
One-dimensional approximation to continuous entropy	2.7303	2.7271
Two-dimensional continuous entropy	6.1594	6.4062
Two-dimensional approximation to continuous entropy	6.2677	6.3064
Two-dimensional approximation to continuous redundancy	0.1341	0.1451
Two-dimensional information gain	0.9598	1.0711

NOTE: The one-dimensional statistics are based on 10 zones; the two-dimensional statistics are based on 100 zones.

tion of entropy-maximizing models used in statistical mechanics to the modelling of spatial phenomena. Although this approach was suggested as far back as 1959 by Cohen [*11*], Wilson [*36*] has been mainly responsible for developing the idea in a geographical context. To introduce the technique, consider a problem of finding the set of probabilities for locating population in a bounded region. This location problem will be subject to constraints of which the two most important are likely to be

$$\sum_i p_i = 1, \tag{35}$$

$$\sum_i p_i r_i = \bar{R}. \tag{36}$$

Note that here r_i is a measure of location cost at i and $\bar{R}$ is the mean location cost in the system. In the context of the population density models introduced previously, r_i could be the travel cost between origin 0 and location i. By maximizing equation (1) subject to equations (35) and (36), Wilson [*36*] shows that the probability of locating in any i is given by

$$p_i = \frac{e^{-\lambda r_i}}{\sum_i e^{-\lambda r_i}}. \tag{37}$$

The model of equation (37) has been derived by Wilson for locating a wide variety of spatial phenomena ranging from trip-making activities to population.

However, the effect of zone size or shape is subsumed within the logic of this model, and at best one can only assume that zone size is unimportant in the model if the zones are all of equal size. It is possible to build zone size explicitly into the model if the spatial entropy function, rather than the discrete entropy function, is maximized. The spatial entropy function for fixed Δx_i is repeated here for convenience.

$$H = -\sum_i p_i \ln\left(\frac{p_i}{\Delta x_i}\right). \tag{38}$$

To maximize equation (38) subject to the constraints in equations (35) and (36), first construct a Lagrangean L where the undetermined multipliers α and λ refer to the constraint equations (35) and (36) respectively.

$$L = -\sum_i p_i \ln\left(\frac{p_i}{\Delta x_i}\right) - \alpha\left(\sum_i p_i - 1\right) - \lambda\left(\sum_i p_i r_i - \bar{R}\right). \tag{39}$$

Differentiating equation (39) with respect to p_i and setting the result equal to zero gives the first-order conditions for a maximum.

$$\frac{\partial L}{\partial p_i} = -\ln p_i + \ln \Delta x_i - \alpha - \lambda r_i = 0. \tag{40}$$

Rearranging equation (40) and taking antilogs makes clear that

$$p_i = \Delta x_i e^{-\alpha - \lambda r_i}. \tag{41}$$

To find the value of α, substitute equation (41) into equation (35). Then

$$\sum_i p_i = e^{-\alpha} \sum_i \Delta x_i e^{-\lambda r_i} = 1, \tag{42}$$

and from equation (42),

$$e^{-\alpha} = \frac{1}{\sum_i \Delta x_i e^{-\lambda r_i}}. \tag{43}$$

The probability location model can now be written as

$$p_i = \frac{\Delta x_i e^{-\lambda r_i}}{\sum_i \Delta x_i e^{-\lambda r_i}}, \tag{44}$$

and equation (44) demonstrates that zone size Δx_i has explicitly entered the model. Wilson's result in equation (37) can easily be derived from equation (44) by assuming that $\Delta x_1 = \Delta x_2 = \ldots = \Delta x_n$. A derivation similar to the above analysis has been made by Broadbent [7] in constructing a maximum-entropy model dealing with trip density rather than trip distribution, and it is also worth noting that Cohen [*11*] briefly discussed the influence of zone size in such models.[5]

Although a maximum entropy model incorporating zone size has been derived, zone size is still exogenous to the model. It should, however, be possible to explore the question of zone size and shape in order to assess the effect of varying size and shape on the model's conceptual and technical form. For example, several location theorists have suggested that zones should be constructed so that there are equal amounts of activity, rather than equal areas in each zone. This conclusion has been reached in problems concerned with political districting [*18*], and this fact demonstrates that theorists intuitively feel that optimum zoning systems are those in which populations are described in equal terms. If, therefore, the probabilities of location are assumed to be equal in each zone, the maximum entropy model can be recast so that some insight is obtained into zone shape. In equation (44), assume that the value of p_i is set equal to $1/n$ in each zone. Then

$$\frac{1}{n} = \frac{\Delta x_i e^{-\lambda r_i}}{\sum_i \Delta x_i e^{-\lambda r_i}}. \tag{45}$$

From equation (45),

$$\Delta x_i = \frac{\left(\sum_i \Delta x_i e^{-\lambda r_i}\right) e^{\lambda r_i}}{n}. \tag{46}$$

Equation (46) could be used as the basis for some iterative scheme for fixing zone sizes which give equal location probabilities. At equilibrium, equation (46) suggests that such zone sizes would vary directly with the exponential distribution of location costs. In fact, in a later section of this paper, a continuous form of equation (46) is used in constructing an idealized zoning system. It is also interesting that Tobler [*33*] has suggested a class of map transformations based on equal-area projections for translating the variations in spatial phenomena into geometric terms. The model above in equation (46) involves a similar concept of geometric distortion.

[5]Bussière and Snickars [*8*] derive equation (44) using an entropy maximizing scheme based on the continuous entropy of equation (3). However, zone size is implicit in their model and can only be defined by making the appropriate integrations between given limits.

Entropy-Maximizing Cluster Analysis

In problems of spatial aggregation and regionalization, one of the most important criteria involves aggregating activities in such a way that the heterogeneity of the system is preserved as far as possible. Various clustering schemes which incorporate this notion have been used in geographical studies [*31*]. A particularly attractive aggregation procedure has been suggested by Theil [*32*], and this scheme is based upon Shannon's discrete entropy measure given in equation (1). This entropy function can be expressed as the sum of a *between-set* entropy and a *within-set* entropy

$$H = -\sum_k P_k \ln P_k - \sum_k P_k \left(\sum_{i \epsilon S_k} \frac{p_i}{P_k} \ln \frac{p_i}{P_k} \right), \tag{47}$$

where
$$P_k = \sum_{i \epsilon S_k} p_i, \tag{48}$$

and
$$\sum_k P_k = \sum_k \sum_{i \epsilon S_k} p_i = 1. \tag{49}$$

Equation (47) is constant for any aggregation of the spatial units notated by i into larger sets S_k. The first term on the right-hand side of equation (47) is the measure of between-set entropy whereas the second term is the within-set entropy. On aggregating the basic spatial units into sets S_k, it can easily be shown that the between-set entropy is monotonically-decreasing as the size of sets S_k increases and that the within-set entropy is monotonically-increasing. The proof of this assertion has been formally demonstrated by Ya Nutenko [*38*] and here it is sufficient merely to indicate the main lines of his proof. Consider an initial aggregation of the spatial units into $n + 1$ sets where the set S_{n+1} contains a single spatial unit. On the next level of aggregation, set S_n and set S_{n+1} are combined to form set S_m. The difference J between the entropy of the new m sets and the old $n+1$ sets is written as

$$J = \sum_{k=1}^{n+1} p_k \ln P_k - \sum_{k=1}^{m} P_k \ln P_k . \tag{50}$$

Equation (50) can be simplified to

$$J = P_n \ln P_n + P_{n+1} \ln P_{n+1} - P_m \ln P_m , \tag{51}$$

which in turn can be expressed solely in terms of the first $n+1$ sets.

$$J = P_n \ln P_n + P_{n+1} \ln P_{n+1} - (P_n + P_{n+1}) \ln(P_n + P_{n+1}). \tag{52}$$

Grouping like terms in equation (52) gives

$$J = P_n(\ln P_n - \ln(P_n + P_{n+1})) + P_{n+1}(\ln P_{n+1} - \ln(P_n + P_{n+1})). \tag{53}$$

As equation (53) is clearly negative, the proof is complete and can easily be extended to demonstrate the increasing value of within-set entropy at higher levels of aggregation.

In using equation (47) as a basis for aggregation, it is suggested that the grouping procedure should maximize between-set entropy, which in turn implies that within-set entropy is minimized. Maximizing between-set entropy subject to the probability constraint in equation (49) leads to a solution in which the location probabilities in each set are equal. This result is completely consistent with the suggestion of the previous section that an optimum zoning system in terms of the smallest information loss is one in which all location probabilities are equal. In fact, equation (47) has already been used by the author [*4*] in aggregating zones in the Reading region.

The theme of this paper however suggests that the discrete entropy function should be replaced by a spatial entropy function which is the discrete equivalent of the continuous function. Therefore, it is necessary to explore whether or not the discrete aggregation formula of equation (47) can be replaced by a continuous equivalent, thus bringing zone size explicitly into the aggregation procedure. First, expand equation (11) as follows:

$$H = -\sum_i p_i \ln p_i + \sum_i p_i \ln \Delta x_i. \tag{54}$$

Then, the second term on the right-hand side of equation (54) is rewritten as

$$\begin{aligned}\sum_i p_i \ln \Delta x_i &= \sum_k \sum_{i \in S_k} p_i \ln \Delta x_i,\\ &= \sum_k P_k \sum_{i \in S_k} \frac{p_i}{P_k} \ln \Delta x_i.\end{aligned} \tag{55}$$

The term $\ln \Delta x_i$ in equation (55) can be written as

$$\ln \Delta x_i = \ln \frac{\Delta x_i}{\sum_{i \in S_k} \Delta x_i} + \ln \sum_{i \in S_k} \Delta x_i, \tag{56}$$

and since $\Delta X_k = \Sigma_{i \epsilon S_k} \Delta x_i$, equation (55) can be expressed as

$$\sum_i p_i \ln \Delta x_i = \sum_k P_k \left(\sum_{i \epsilon S_k} \frac{p_i}{P_k} \left(\ln \frac{\Delta x_i}{\Delta X_k} + \ln \Delta X_k \right) \right),$$

$$= \sum_k P_k \left(\sum_{i \epsilon S_k} \frac{p_i}{P_k} \ln \frac{\Delta x_i}{\Delta X_k} \right) + \sum_k P_k \ln \Delta X_k \qquad (57)$$

The first term on the right-hand side of equation (54) can be expanded as equation (47), and adding equations (47) and (57) produces the spatial form of Theil's aggregation formula.

$$H = -\sum_k P_k \ln P_k + \sum_k P_k \ln \Delta X_k - \sum_k P_k \left(\sum_{i \epsilon S_k} \frac{p_i}{P_k} \ln \frac{p_i}{P_k} \right)$$

$$+ \sum_k P_k \left(\sum_{i \epsilon S_k} \frac{p_i}{P_k} \ln \frac{\Delta x_i}{\Delta X_k} \right),$$

$$= -\sum_k P_k \ln \left(\frac{P_k}{\Delta X_k} \right) - \sum_k P_k \left(\sum_{i \epsilon S_k} \frac{p_i}{P_k} \ln \left(\frac{p_i \Delta X_k}{\Delta x_i P_k} \right) \right). \qquad (58)$$

Maximizing the first term on the right-hand side of equation (58) subject to the usual probability constraints leads to

$$L = -\sum_i P_k \ln \left(\frac{P_k}{\Delta X_k} \right) - \alpha \left(\sum_k P_k - 1 \right), \qquad (59)$$

$$\frac{\partial L}{\partial P_k} = -\ln P_k + \ln \Delta X_k - \alpha = 0. \qquad (60)$$

Equation (60) implies that the probabilities of location P_k are determined according to the geometry of the system. In algebraic terms

$$P_k = \frac{\Delta X_k}{\sum_k \Delta X_k}. \qquad (61)$$

The spatial entropy function in equation (58) has been used in aggregating zones in the Reading area where the population has been used to determine the probabilities of location. The data is recorded in 130-kilometer grid squares [*5*], and the method of aggregation is based on the hierarchical heuristic devised by Ward [*34*]. At each level of the hierarchy, between-set entropy is maximized by computing the

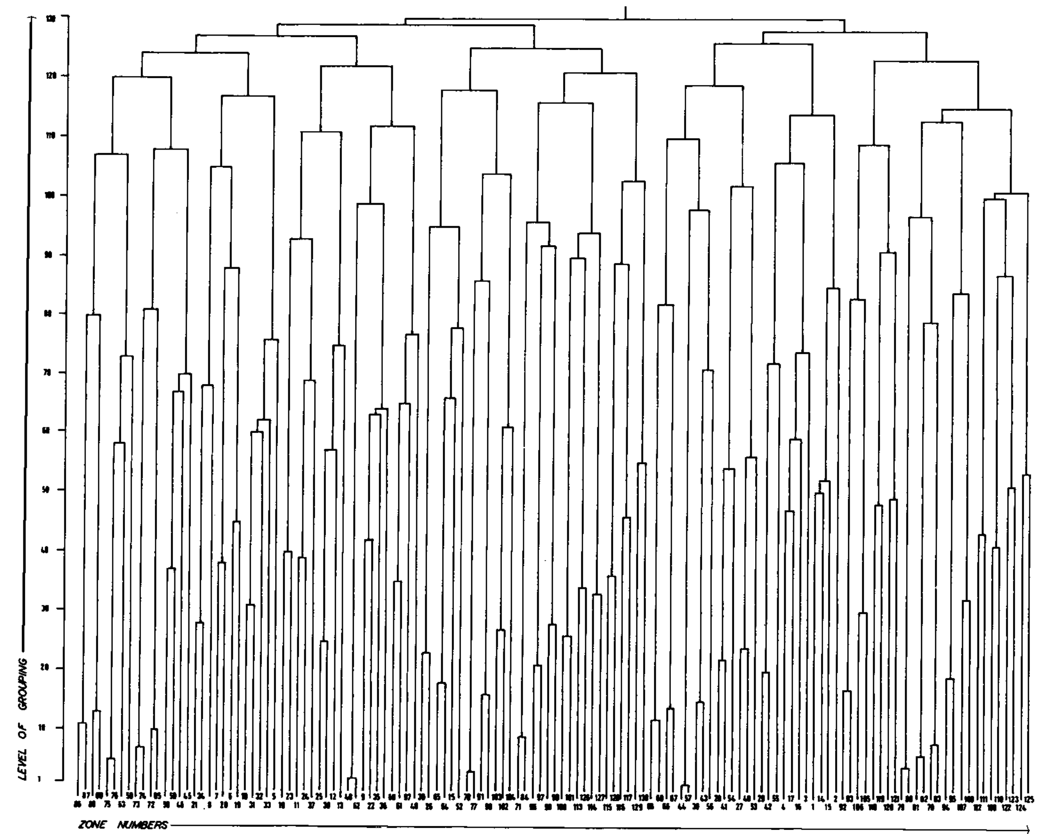
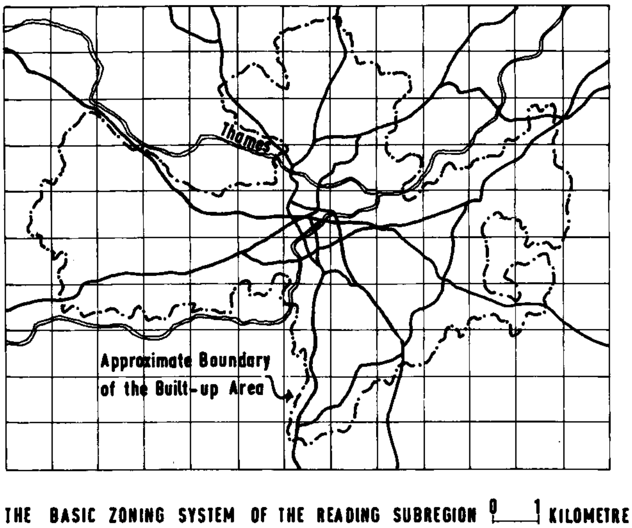


FIG. 9. A dendrogram summarizing hierarchical aggregation-clustering based on Theil's formula and Ward's algorithm.

measure for every possible aggregation of single spatial units to their spatially adjacent sets. There are also fairly severe limitations on the process imposed by the contiguity constraints. In Figure 9, the hierarchical aggregation scheme is illustrated by a dendrogram; and in Figure 10, the initial partition of the town into 130-kilometer squares is depicted together with the aggregated solution containing 20 sets. It is clear from both Figures 9 and 10 that although the method might optimize the



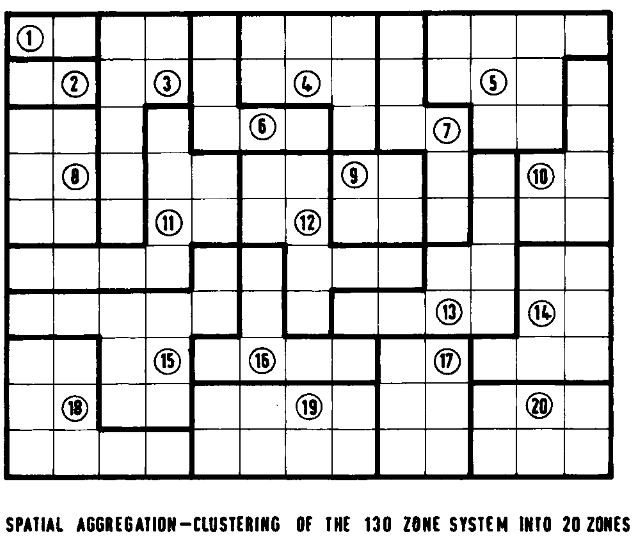


FIG. 10. The original zoning system and its aggregation into 20 sets.

entropy measure at each level in the hierarchy, the overall solution at any level is probably suboptimal. This is a limitation of Ward's algorithm which is particularly serious in problems of spatial aggregation; and in the quest to find optimal solutions to this zoning problem, a simple algorithm which can be applied at any level of the hierarchy, has been developed.

The Design Of Maximum Entropy Spatial Systems

An algorithm designed to minimize the within-set entropy given by equation (58) is illustrated by the flow chart in Figure 11. At any particular

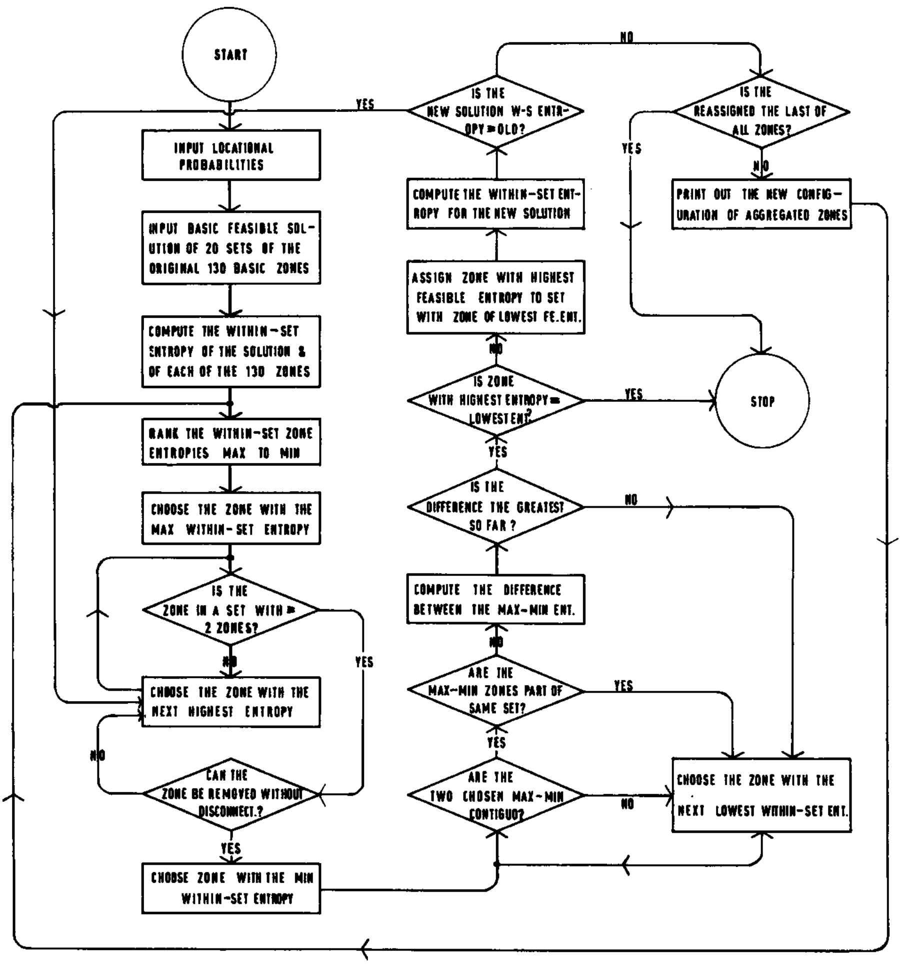


FIG. 11. A flow chart outlining an heuristic procedure for finding an optimal aggregation into a fixed number of sets.

level of aggregation, the algorithm operates by attempting to improve a basic feasible solution in a trial-and-error fashion. First, the within-set entropy of the basic feasible solution is calculated and the contribution of each zone to this entropy is ranked. Then, a search is initiated to discover the greatest difference in entropies between any two zones; if these two zones are contiguous and if the zone with the largest entropy belongs to a set of two or more zones, then a new set is formed by combining these two. If such combination is impossible, then the search continues to discover the next highest difference in entropies and so on.

When a new set has been formed, it is necessary to find out whether or not the new solution produces a lower within-set entropy. If not, the search is continued as above; but if the entropy is lower, the solution is accepted and the zones are ranked as before. The process is continued by searching for an improvement in the new solution which minimizes within-set entropy. The graph in Figure 12 shows the improvement in the minimum value for within-set entropy set against the cumulative number of solutions defined and tested by the heuristic. After about 1,500 solutions further improvement becomes much more difficult to achieve, and to all intents and purposes, this suggests that the procedure has converged. Figure 12 also presents a graph which shows that the average number of solutions required to produce a new solution is fairly

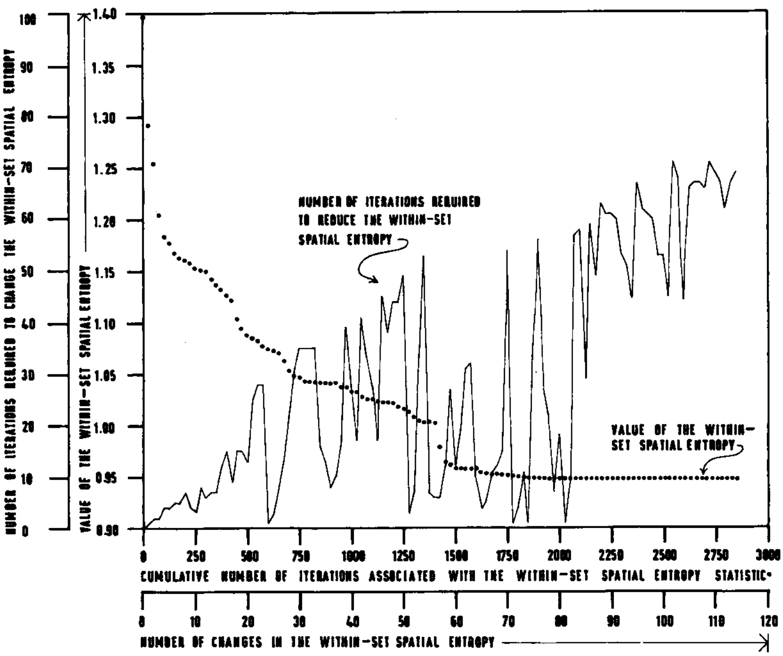
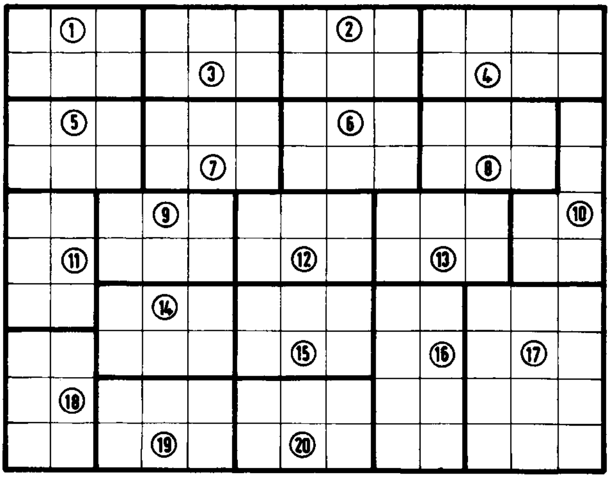
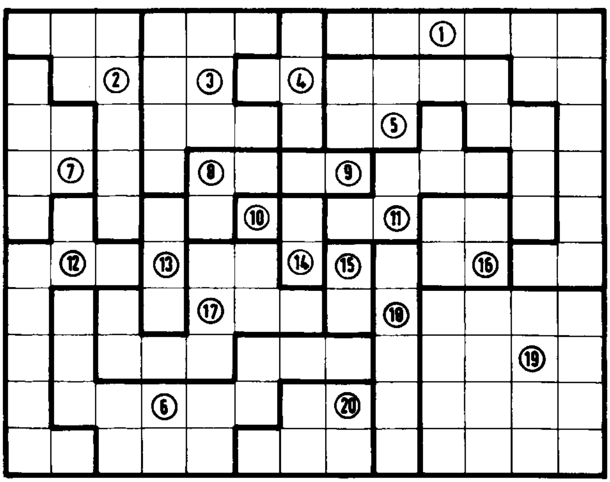


FIG. 12. The cumulative improvement in the value of within-set entropy based on the aggregation heuristic.

constant. In Figure 13, the basic feasible solution and the final solution are presented; and in comparing this solution with the solution in Figure 10, it is clear that this algorithm is considerably more efficient than the cluster analysis in producing a near optimal solution. It is possible that there are heuristics or formal programming procedures such as some of those suggested by Scott [28] which are better suited to this kind of minimization problem; such methods are at present being considered.



THE INITIAL BASIC FEASIBLE SOLUTION USED IN THE OPTIMISATION



THE FINAL SOLUTION BASED ON 2032 ITERATIONS OF THE OPTIMISATION

FIG. 13. The initial and final basic feasible solutions.

Of particular interest is the relationship between this algorithm and the transformation algorithm used by Rushton [27]. Rushton's procedure is applied to a system of central places which are systematically distorted due to varying population density, and it appears that this algorithm could be applied to the problem considered here.

Apart from the design of realistic spatial systems by aggregating basic spatial units, it is possible to design idealized systems using the concept of maximum entropy. For example, consider a radially symmetric monocentric population density field where it is required to partition the field into zones in which location probabilities are equal. Clearly there are many solutions to this problem [33], although fairly realistic procedures can be defined. Using the density field previously specified in equations (27) to (29), and noting that $p_i = 1/n$ where n is the number of zones, an inner zone of radius R_1 can be calculated from

$$p_i = \frac{1}{n} = \int_0^{2\pi} \int_0^{R_1} Ke^{-\lambda r} r \, d\theta dr. \tag{62}$$

R_1 can be found by using the following iterative scheme derived from equation (62)

$$R_1^{(m+1)} = \frac{\ln\left(\frac{2\pi K}{\lambda}\right) - \ln\left(\frac{2\pi K}{\lambda^2} - p_i\right) + \ln\left(\frac{1}{\lambda} + R_1^m\right)}{\lambda}. \tag{63}$$

Having found R_1, it is possible to define other zones which are segments of annuli around the pole of the system, by attempting to meet a condition that the radial distance of the segment be as close as possible to the circumferential distance. In general, on iteration (k) of this procedure, the probability of locating in the annulus (k) is

$$p(k) = K2\pi \int_0^{R_k} e^{-\lambda r} r \, dr - \sum_{j=1}^{k-1} p(j). \tag{64}$$

A trial value for R_k can be based on previous distance measures, and the number of zones N which would be partitioned in the annulus is calculated as

$$N = \text{integer}\,[np(k)]. \tag{65}$$

Then the circumference of the segment d can be calculated as

$$d = \frac{2\pi R_k}{N}. \tag{66}$$

If $|d - (R_k - R_{k-1})| > \epsilon$, then dN is substituted for R_k and equations

(64) to (66) are recalculated until the number N is unchanged or until the difference ϵ is reached.

The actual distance $R_k^{(m+1)}$ is found by substituting N/n for p_i in equation (63) and solving for $R_k^{(m+1)}$ by iteration. In this way, a solution can be built up progressively by moving from the pole to the boundary of the field. Because of difficulties in meeting the distance condition exactly, this method may overshoot the boundary and finish with too many zones or with a peculiarly shaped final set of zones. To avoid such problems, if the final value of N computed on any iteration of the procedure is greater than the number of zones required to complete the solution, this value of N is disregarded and a value for N which completes the solution is fixed. It is possible that the solution might be easier to achieve if the procedure were operated in a reverse fashion, from the boundary to the pole. In such a case, the probability $p(k)$ for the annulus on iteration k would be calculated from

$$p(K) = K2\pi \int_{R_k}^{R_{k-1}} e^{-\lambda r} r \, dr$$

$$= \frac{K2\pi}{\lambda} \left[e^{-\lambda R_k} \left(R_k + \frac{1}{\lambda} \right) - e^{-\lambda R_{k-1}} \left(R_{k-1} + \frac{1}{\lambda} \right) \right]. \tag{67}$$

The number of zones N and the distance d could be found in a manner similar to the above procedure, and iteration on equation (67) would take place until N is unchanged or until ϵ is met. The final value for R_k is found by fixing $p(k)$ as N/n and finding R_k from equation (67). Overshoot is less likely to occur using this method, but the final value for N might have to be disregarded and arbitrarily fixed if overshoot does occur. In Figure 14, the form of an idealized zoning system based on the Reading data is presented, thus illustrating the problem of overshoot. Comparing Figures 13 and 14, some correspondence between the real and idealized systems is evident.

Conclusions and Further Research

If there is any message in this paper, it is that space is not a trivial matter in geographical analysis. This may appear to be somewhat naïve but it is astonishing how much analysis, especially in the field of urban modelling, either treats space implicitly rather than explicitly or ignores its influence altogether. Much of the statistical analysis in recent years, such as that concerned with spatial regression and with spatial spectra [*13, 14*], has confronted the influence of space directly, and the time now seems ripe to begin to synthesize much of this material into a coherent body of analysis. It is possible that a fruitful area of synthesis is in the context of the aggregation problem, and it appears that some

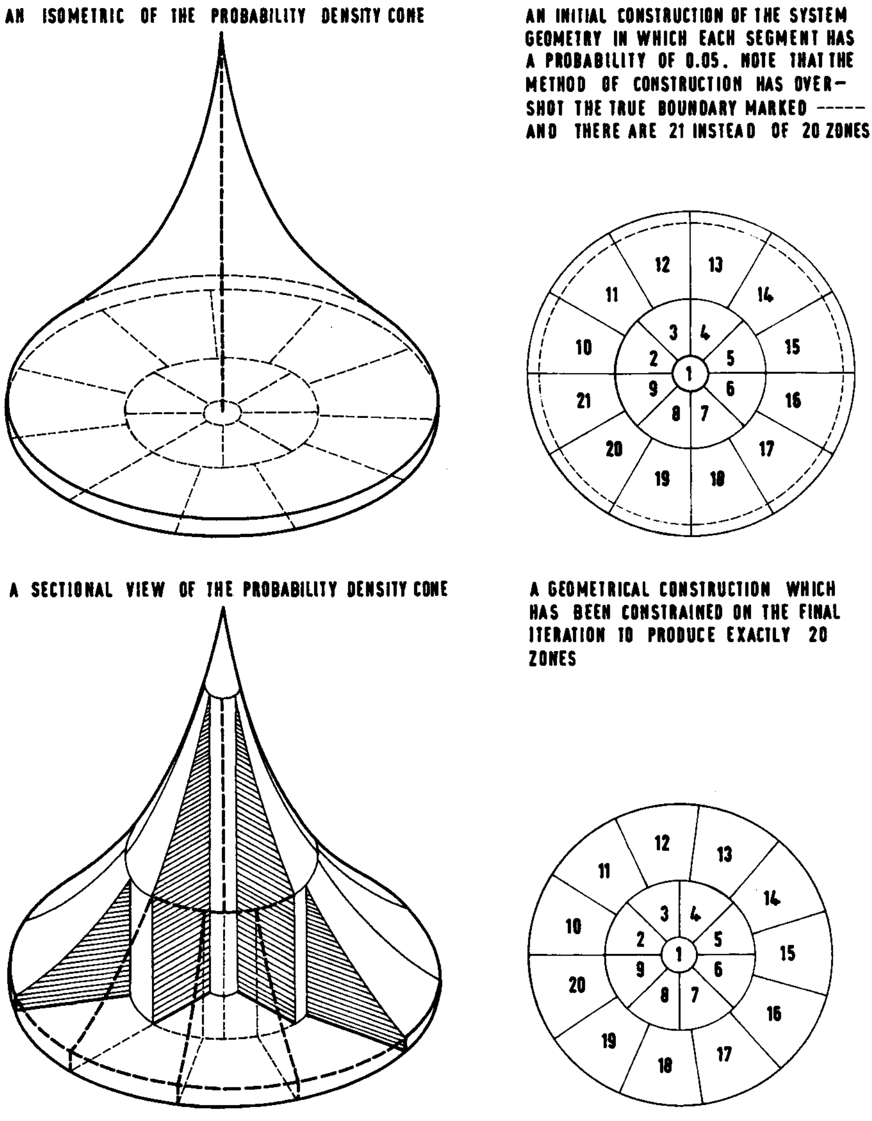


FIG. 14. The construction of idealized spatial geometries.

of the ideas explored here may be of interest, if not of use, in this area.

The problem of treating phenomena discretely or continuously has never been formally treated in theoretical geography, yet this area appears extremely promising from the work of Tobler [33], and it may be possible to extend some of the design methods for spatial systems using such work. Furthermore, more research is required in problems of regionalization and the statistical implications of such analysis, which has largely been neglected to date. It is hoped to extend some of the ideas of this paper and to assess their relevance in the context of mainstream spatial analysis in the quest to synthesize the concept of entropy with more well-defined and widely applied methods of spatial analysis.

LITERATURE CITED

1. ANDERSON, J. "On General System Theory and the Concept of Entropy in Urban Geography," Research Paper No. 31, London School of Economics, Department of Geography, 1969.
2. ANGEL, S. and G. HYMAN. "Urban Spatial Interaction," Working Paper No. 69, London: Centre for Environmental Studies, 1971.
3. BATTY, M. "An Experimental Model of Urban Dynamics," *Town Planning Review*, 43 (1972), 166-86.
4. ———. "Entropy and Spatial Geometry," *Area*, 4 (1972), 230-36.
5. BAXTER, R. "An Urban Atlas: Reading," Working Paper No. 48, Centre for Land Use and Built Form Studies, University of Cambridge, 1971.
6. BERRY, B. and P. SCHWIND. "Information and Entropy in Migrant Flows," *Geographical Analysis*, 1 (1969), 5-14.
7. BROADBENT, T. A. "Zone Size and Singly-Constrained Interaction Models," Working Note No. 132, London: Centre for Environmental Studies, 1969.
8. BUSSIÈRE, R. and F. SNICKARS. "Derivation of the Negative Exponential Model by an Entropy Maximizing Method," *Environment and Planning*, 2 (1970), 295-301.
9. CHAPMAN, G. P. "The Application of Information Theory to the Analysis of Population Distributions in Space," *Economic Geography*, Supplement, 46 (1970), 317-31.
10. CHORLEY, R. and B. KENNEDY. *Physical Geography: A Systems Approach.* London: Prentice-Hall International Inc., 1971.
11. COHEN, M. H. "The Relative Distribution of Households and Places of Work," in R. Herman, ed., *Proceedings of the Symposium on Traffic Flow, Detroit 1959.* Amsterdam: Elsevier, 1961.
12. CURRY, L. "The Random Spatial Economy: An Exploration in Settlement Theory," *Annals of the Association of American Geographers*, 54 (1964), 138-46.
13. ———. "Applicability of Space-Time Moving-Average Forecasting," in M. Chisholm, A. Frey and P. Haggett, eds., *Regional Forecasting*, Colston Papers No. 22, London: Butterworths, 1971.
14. ———. "A Spatial Analysis of Gravity Flows," *Regional Studies*, 6 (1972), 131-47, summarized as P1204, "Spatial Entropy," in W. P. Adams and F. M. Helleiner, eds., *International Geography 2.* Toronto and Buffalo: University of Toronto Press, 1972.
15. GEORGESCU-ROEGEN, N. *The Entropy Law and the Economic Process.* Cambridge, Mass: Harvard University Press, 1971.
16. GOLDMAN, S. *Information Theory.* London: Constable and Co. Limited, 1953.
17. GUREVICH, B. L. "Geographical Differentiation and Its Measures in a Discrete System," *Soviet Geography: Reviews and Translations*, 16 (1969), 387-413.
18. HAGGETT, P. *Locational Analysis in Human Geography.* London: Edward Arnold, 1965.
19. HARTLEY, R. V. L. "Transmission of Information," *Bell System Technical Journal*, 7 (1928), 535-63.
20. JAYNES, E. T. "Information Theory and Statistical Mechanics," *Physics Review*, 106 (1957), 620-80.
21. ———. "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics*, SSC-4 (1968), 227-41.
22. LEOPOLD, L. B. and W. B. LANGBEIN. "The Concept of Entropy in Landscape Evolution," U.S. Geological Survey, Professional Paper 500-A, 1962.

23. MARCH, L. "Urban Systems: A Generalized Distribution Function," *London Papers in Regional Science*, 2 (1971), 157-70.
24. MARCHAND, B. "Information Theory and Geography," *Geographical Analysis*, 4 (1972), 234-57.
25. MEDVEDKOV, Y. "Entropy: An Assessment of Potentialities in Geography," *Economic Geography*, Supplement, 46 (1970), 306-16.
26. MOGRIDGE, M. J. H. "The Use and Misuse of Entropy in Urban and Regional Modelling of Economic and Spatial Systems," Working Note No. 320, London: Centre for Environmental Studies, 1972.
27. RUSHTON, G. "Map Transformations of Point Patterns: Central Place Patterns in Areas of Variable Population Density," *Papers of the Regional Science Association*, 28 (1972), 111-29.
28. SCOTT, A. J. *Combinatorial Programming, Spatial Analysis and Planning*. London: Methuen, 1971.
29. SEMPLE, R. K. and R. G. GOLLEDGE. "An Analysis of Entropy Changes in a Settlement Pattern over Time," *Economic Geography*, 46 (1970), 157-60.
30. SHANNON, C. E. "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27 (1948), 379-423 and 623-56.
31. SPENCE, N. A. and P. J. TAYLOR. "Quantitative Methods in Regional Taxonomy," *Progress in Geography*, 2 (1970), 1-63.
32. THEIL, H. *Economics and Information Theory*. Amsterdam: North Holland Publishing Company, 1967.
33. TOBLER, W. "Geographic Area and Map Projections," *Geographical Review*, 53 (1963), 59-78.
34. WARD, J. H. "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, 58 (1963), 236-44.
35. WIENER, N. *Cybernetics*. New York: The M.I.T. Press and John Wiley, 1961.
36. WILSON, A. G. *Entropy in Urban and Regional Modelling*. London: Pion Press, 1970.
37. WOODWARD, P. M. *Probability and Information Theory, With Applications to Radar*. Oxford: Pergamon Press, 1953.
38. YA NUTENKO, L. "An Information-Theory Approach to the Partitioning of an Area," *Soviet Geography: Reviews and Translations*, 11 (1970), 540-44.