Entropy and spatial geometry

Michael Batty, University of Reading

Summary. *The concept of entropy as used in explaining locational phenomena is briefly reviewed and it is suggested that the design of a zoning system for measuring such phenomena is a non-trivial matter. An aggregation procedure based on entropy-maximizing is suggested and applied to the Reading sub-region, and the resulting geometries are contrasted with certain idealized schemes.*

In the last decade, several researchers have suggested that the concept of entropy is a relevant statistic for measuring the spatial distribution of various geographic phenomena. For example, Leopold and Langbein (1962) use a measure of entropy in deriving the fact that the most probable longitudinal profile of rivers has a negative-exponential form. Curry (1964) has shown that the rank-size distribution of cities can be explained by considerations involving the definition of entropy, and more recently, Wilson (1970) has developed a procedure for maximizing a function of entropy which can be used to describe a host of locational phenomena ranging from distributions of trip-making behaviour to distributions of population. Furthermore, Mogridge (1972), in an excellent review of the concept, demonstrates that entropy ' is of great, indeed essential, use in understanding economic and spatial systems '.

Yet in all of this work, the spatial dimension in which the various phenomena are recorded and statistics computed is implicit rather than explicit. There is little concern for the way in which space is partitioned, for most of these entropy models appear to be based on the assumption that distributions are measured on spaces partitioned into equal areas or intervals. However, this assumption is not necessarily the most appropriate; geographers have long recognized that different and often conflicting statistical patterns can be interpreted from similar distributions measured on different areal systems. Although this problem has been quite widely studied in recent years under various guises such as that of spatial auto-correlation, there has been very little research into the design of optimal spatial systems for geographical analysis or for locational planning. For example, Neft (1966) in his classic book on spatial analysis, hardly broaches the problem which is surprising in view of its importance.

It is the purpose of this paper to investigate briefly the geometry of spatial systems in relation to the measurement of locational phenomena using the statistic of entropy. It is likely that there are ideal geometries associated with different spatial statistics, and in this paper, a geometry consistent with the definition of entropy will be outlined from both empirical and theoretical standpoints. In this quest, it is useful to begin by briefly reviewing the use of entropy in locational analysis.

The discrete entropy of spatial distribution

The definition of entropy most widely used in geographic research is due to Shannon (Shannon and Weaver, 1949) and it is also referred to as a measure of information or uncertainty. This measure can be written as

$$H(r) = -\sum_i p_i \ln p_i \qquad (1)$$

where p_i is the probability of the phenomena occurring in r_i. The summation is taken over all the discrete intervals notated by i and it is clear that equation (1) is normalized so that $\sum_i p_i = 1$. The problem of choosing an appropriate interval length, r_i, can best be illustrated using the entropy-maximizing procedure due to Wilson (1970). Wilson maximizes equation (1) where p_i is defined as a probability of location, subject to various constraints on the location and cost of locating some activity. In general, the solution to this problem implies that activity is located around some centre according to some decreasing monotonic function of travel cost or distance such as a negative-exponential or inverse-power function. This result, however, assumes that the interval size is constant. Broadbent (1969) has suggested that the discrete measure of probability in equation (1) be replaced by a probability normalized with respect to the size of interval or area (which may vary over the distribution). If this is done, then the entropy-maximizing model becomes a model for locating the density rather than the absolute amount of activity, and the model is consequently more general.

Although Broadbent's innovation suggests how interval size can be incorporated into the entropy formula, interval size is still exogenous to the formulation. Some light can be thrown on this problem, however, by recasting the solution procedure. Firstly, if equation (1) is maximized subject only to the normalization constraint, the solution implies that locational phenomena be distributed equally in every area or zone. This seems a sensible result, for the greatest amount of information is extracted from a population if each part of the population is described in the same detail. By substituting this result into the probability formula for density, then it is clear that the probability density varies only with the area or interval size. If entropy is now maximized subject to the locational constraints, then the solution implies that the appropriate area or interval size around some point varies in a negative-exponential fashion. In summary, then, with equal size areas the activity is distributed in a negative-exponential form, whereas with equal amounts of activity, the areal system varies in a negative-exponential form. If the emphasis is on describing phenomena in the best possible way, then there are clear reasons for preferring a negative-exponential distribution of zones and this links up with the work of Tobler (1963) who has derived a similar result from different considerations. A more formal presentation of this argument will not be pursued further here for this would be somewhat lengthy.

Entropy as a measure of spatial grouping

From the previous argument, what is required is a method for aggregating areas into zones with equal probabilities of location. Using a method due to Theil (1967), equation (1) can be expressed for the case in which the original intervals or zones are aggregated into larger sets called S_k. Entropy now becomes the sum of a *between-set entropy* and a *within-set entropy*.

$$H(r) = -\sum_k P_k ln P_k - \sum_k P_k \left[\sum_{i \varepsilon S_k} \frac{p_i}{P_k} ln \left(\frac{p_i}{P_k} \right) \right] \quad (2)$$

The first term on the right-hand side of equation (2) is between-set entropy whereas the second term represents the within-set entropy. Equation (2) is subject to the constraints $\sum_k P_k = 1$, and $\sum_{i \varepsilon S_k} p_i = P_k$. In terms of equation (2),

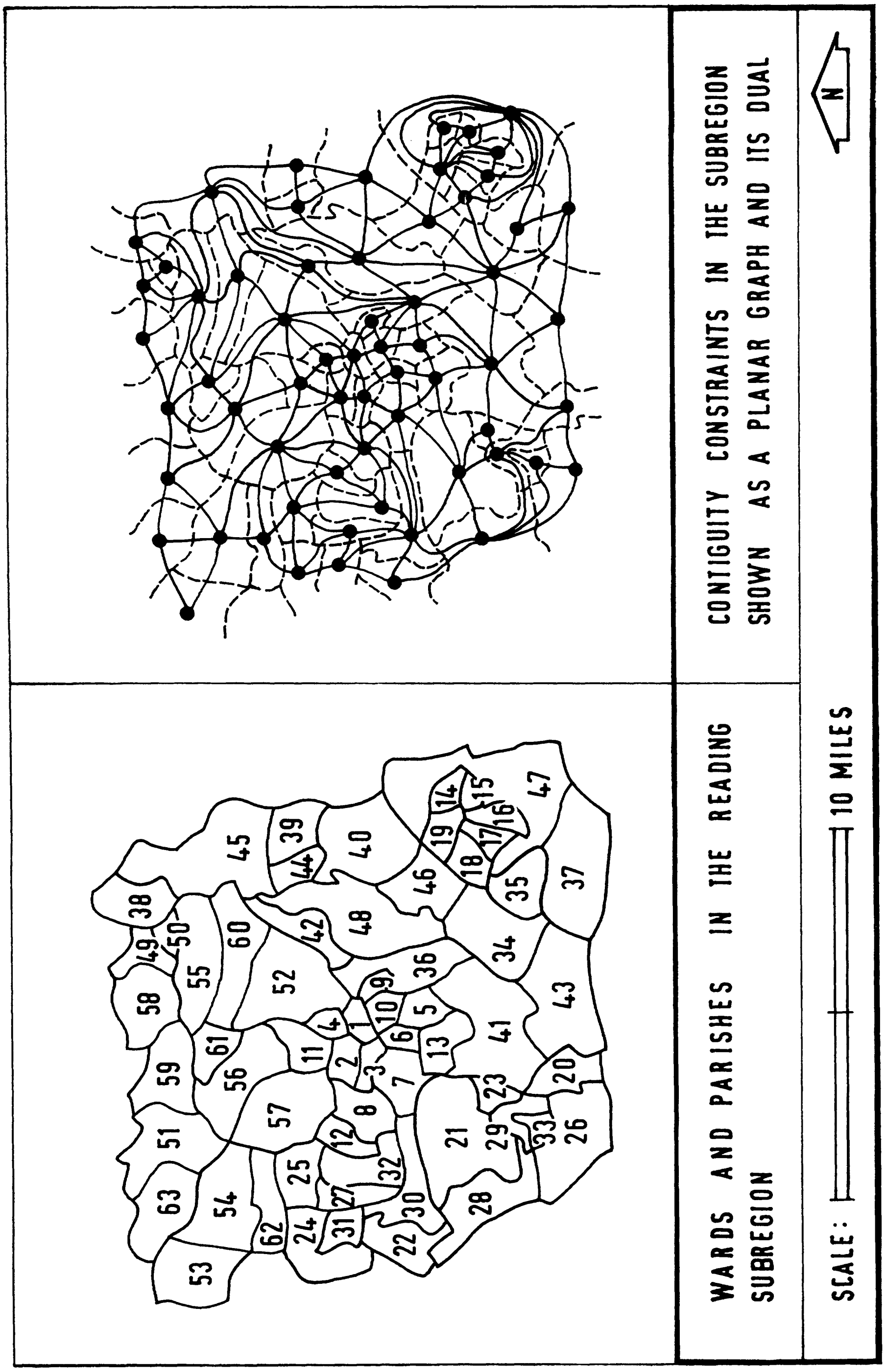

Figure 1. Existing geometry of the Reading sub-region.

an optimal zoning system is one in which between-set entropy is maximized (which implies that within-set entropy is minimized). A formal demonstration of this can be achieved by maximizing $\sum_{k} P_k \ln P_k$, subject to $\sum_{k} P_k = 1$, using the method of Tribus (1969) or Wilson (1970). This yields the well-known result that $P_1 = P_2 = P_3 \ldots.$, and this measure of grouping clearly satisfies the argument of the previous section.

To demonstrate the use of this formula in reality, an algorithm is needed to solve the maximization problem. It is unlikely that this problem can be solved in any reasonable time using mathematical programming methods and therefore the hierarchical heuristic devised by Ward (1963) has been adopted. This procedure optimizes the grouping by beginning at the top of the hierarchy and reducing the number of groups by one at each level. The algorithm has been applied to grouping 63 wards and parishes in the Reading sub-region, and at each level in the hierarchy the between-set entropy has been maximized. The probabilities in equation (2) have been based on the observed distribution of population in the sub-region. Figure 1 shows the distribution of the basic zones in the sub-region, and the contiguity constraints necessary to ensure realistic spatial aggregation are represented as a planar graph. In Figure 2, the different

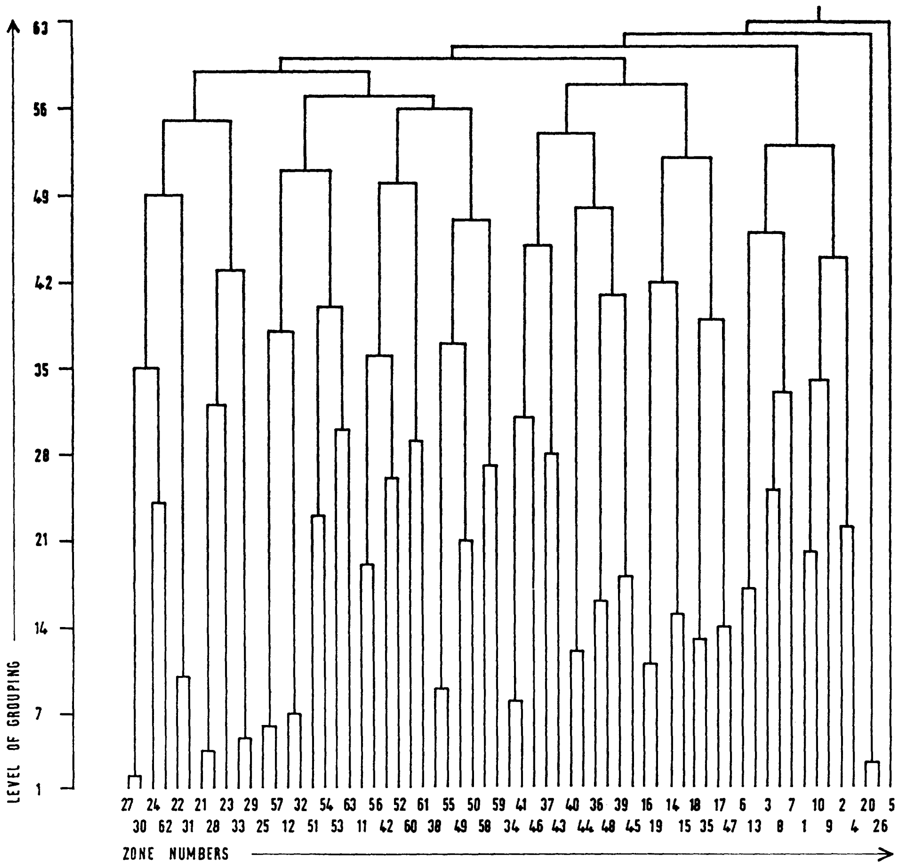

Figure 2. Levels of spatial clustering.

levels of aggregation which maximize between-set entropy are displayed by a tree-like graph.

An idealized geometry of spatial systems

In contrast to the real geometry of the space which is heavily influenced by the basic zoning system, it is possible to construct an idealized but simplified geometry for a given number of zones. Assuming that the probabilities of location can be represented in some radially symmetric fashion around a pole using a mathematically tractable function, then zones can be constructed which have equal probabilities of location. As an example, the probability of population

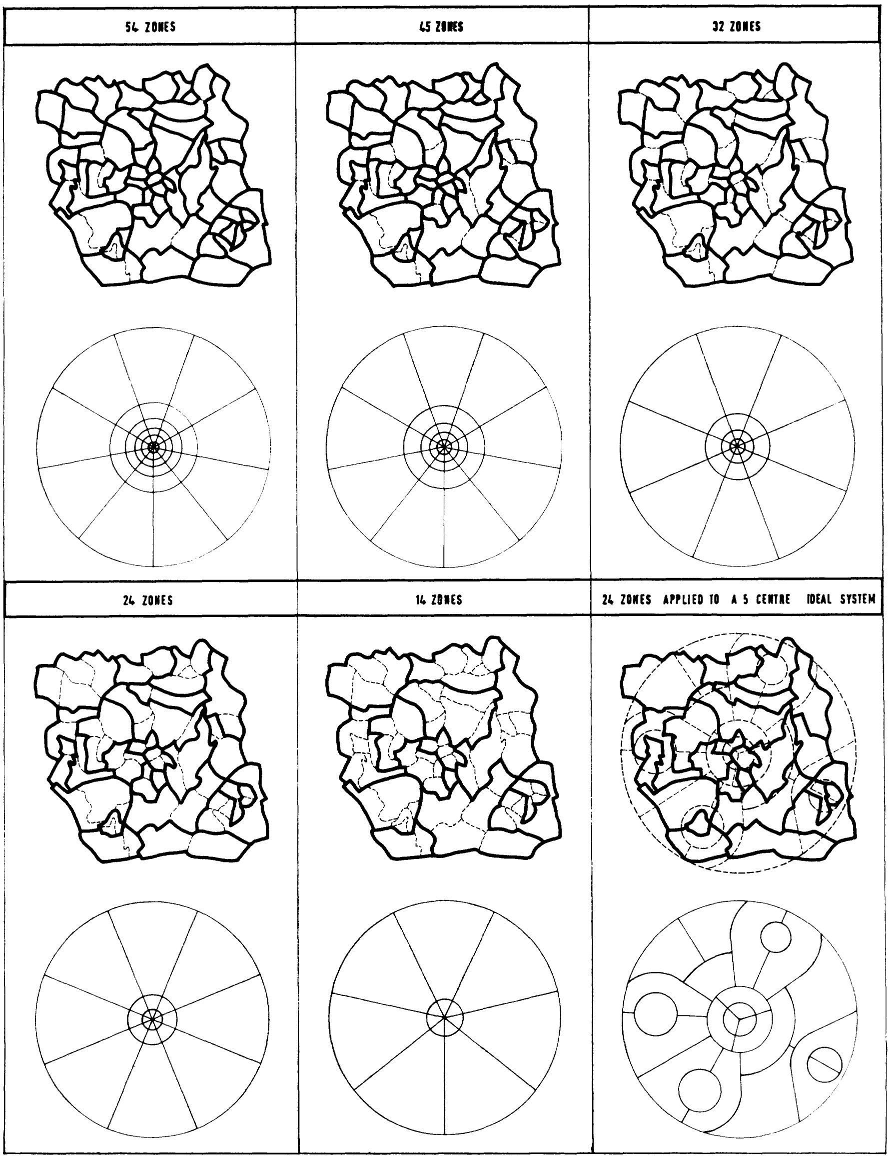

Figure 3. Real and idealized zoning systems.

location is often represented using a negative-exponential density. In polar co-ordinates, this probability is

$$p(r, \theta) = A \exp(-\lambda r) \tag{3}$$

where A is the normalizing factor, θ is the angular variation and r is the distance from the pole. The factor A can be evaluated from

$$\int_0^R \int_0^{2\pi} A \exp(-\lambda r)\, r\, d\theta\, dr = 1 \tag{4}$$

R is the radius of the region and λ is a parameter of the distribution which has an approximate value of $2/\bar{R}$ where $\bar{R}$ is the mean travel distance to the pole (Angel and Hyman, 1971). The discrete probability p_i can be calculated from

$$p_i = \int_{r^1}^{r^2} \int_{\theta^1}^{\theta^2} A \exp(-\lambda r) r\, d\theta dr \tag{5}$$

where $0 \leqslant r^1 < r^2 \leqslant R$ and $0 \leqslant \theta^1 < \theta^2 \leqslant 2\pi$.

In a system of n zones with equal probabilities in each zone, equation (5) can be evaluated for $p_i = 1/n$. For different n, equation (5) has been solved in two stages, first by finding annuli of equal probability starting from $r^1 = 0$ and second by dividing the radially symmetric system into equal sized sectors. This is only one possible method for solving equation (5) and therefore the idealized system is somewhat arbitrary. Nevertheless, interesting contrasts are provided at each level between the realistic aggregation of zones computed by Ward's algorithm and the idealized system. In Figure 3, the real and ideal geometries are compared and it is clear that the real geometry is markedly constrained by the original zoning. The last diagram on Figure 3 shows one possible ideal zoning system of 24 zones and 5 centres which approximate the size and location of the centres at Reading, Wokingham, Henley, Pangbourne and Burghfield. This zoning system was constructed from the successive application of equation (5) to each centre followed by iteration and adjustment. In this case, the correspondence between the real and ideal systems is much closer.

In conclusion, it appears that the use of the concept of entropy in explaining geographic phenomena can help in designing spatial patterns relevant to the distribution of such phenomena. Such measures of entropy also suggest ways in which zones can be aggregated and the resulting patterns can then be compared with an idealized geometry. This concept leads to a promising method of describing and constructing spatial geometries and it seems that the measure of entropy could be further developed to test the efficiency of different zoning systems.

Acknowledgements

The author wishes to thank Jane Read for her skilful unravelling of the dendrogram from some disorganized computer print-out, and Roger Sammons, Geoff Hyman and Andrew Broadbent who have all helped in clarifying some of the concepts presented above.

References

Angel, S. and Hyman, G., 1971. Urban spatial interaction. *Working Pap.*, No. 69, Centre Environmental Stud., London.

Broadbent, T A., 1969. Zone size and singly-constrained interaction models. *Working Note*, No. 132, Centre Environmental Stud., London.

Curry, L., 1964. The random spatial economy: an exploration in settlement theory. *Ann. Ass. Am. Geogr.*, 54, 138–46.

Leopold, L. B., and Langbein, W. B., 1962. The concept of entropy in landscape evolution. *US Geol. Surv. Prof. Pap.* 500-A.
Mogridge, M. J. H., 1972. The use and misuse of entropy in urban and regional modelling of economic and spatial systems. *Working Note*, No. 320, Centre Environmental Stud., London.
Neft, D. S., 1966. Statistical analysis for areal distributions. *Monogr. Ser.* No. 2, Reg. Sci. Res. Inst., Philadelphia.
Shannon, C. E., and Weaver, W., 1949. *The mathematical theory of communication.* University of Illinois Press, Urbana.
Theil, H., 1967. *Economics and information theory.* North Holland Publishing Company, Amsterdam.
Tobler, W., 1963. Geographic area and map projections. *Geogrl Rev.* 53, 59–78.
Tribus, M., 1969. *Rational descriptions, decisions and designs.* Pergamon Press, Oxford.
Ward, J. H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. statist. Ass.*, 58, 236–44.
Wilson, A. G., 1970. *Entropy in urban and regional modelling.* Pion Press, London.

Agricultural Geography Study Group

Suggestions have been made from time to time that the Group should broaden its interests to include all aspects of the use of rural resources. It is clear to the Committee that, while there is a considerable number of members with an interest in agricultural geography, there are not sufficient actively engaged in research in this field to sustain a satisfactory programme of meetings. Furthermore, the nature of agriculture is changing, particularly in developed countries. The wider impact of urbanization, especially the growth of outdoor recreation, the problems posed for the countryside by technical developments in agriculture and the withdrawal of land from agricultural use in problem rural areas are obvious examples. It is proposed to discuss this issue at the business meeting to be held during the annual conference at Birmingham; members who are unable to attend and other members of the Institute who would be interested in such a widening of the Group's interests and have views they wish to express should write to the Group's secretary, Dr J. Henderson, Dept. of Geography, University College Swansea, before 31 December, 1972.

It is intended to devote the Group's discussion following the business meeting to an examination of the use of agricultural census data in geographical research; it will be introduced with a paper by E. K. Anderson and J. T. Coppock, University of Edinburgh, on the progress of the Type of Farming research project, now in its final year.

J. T. Coppock

Proposed Biogeography Research Group

A number of geographers interested in biogeography have been considering the possibility of setting up a formal Group with this special focus, possibly within the IBG. Within the general field we would wish to include not only pure biogeography and ecology, but possibly also conservation, soil science, land resource and recreation studies, and land use ecology.

The heads of university departments of geography have already been circulated and have provided the names of many members of the Institute who embrace these interests. Those whose names have been received will already have received notice of the proposed Symposium to be held at the Annual Conference in Birmingham in January 1973, at which it is hoped to inaugurate such a Biogeography Group. Any members who have not so far received a notice and who are interested in the formation of such a group are asked to write to: **Dr R. P. Moss**, Department of Geography, The University, PO Box 363, Birmingham, B15 2TT. They will then be sent the relevant notices.