Urban Studies (1985) 22, 289–303
© 1985 Urban Studies

A Population-Density Approach to Regional Spatial Structure

John B. Parr

[First received February 1985; in final form April 1985]

Summary. The concept of the population-density function, which is usually applied within the context of an urban area, is extended to the scale of a metropolitan-area-based region or city region. A form for such a regional density function is proposed, and this is examined for selected regions of the UK and North America. It is demonstrated how such a region-wide density function may be related to other density functions which have been used to characterise the structures of the metropolitan and non-metropolitan parts of the region. Consideration is also given to problems associated with constructing and interpreting the regional density function and to its underlying theoretical basis. Finally, the application of the density-function approach in the analysis of regional structure is briefly discussed.

1. Introduction

The study of spatial structure has long been of interest to those concerned with regional analysis. The term 'spatial structure' is not readily defined, and is perhaps best viewed as comprising a series of interrelated perspectives on the social and economic organisation of a region and how this becomes modified over time. One such perspective emphasises the areal distribution of various regional aggregates which include population (both urban and rural), labour force, employment (and unemployment), capital stock, infrastructure, etc. (Boisvert, 1978; Friedmann, 1956). A second perspective deals with regional spatial structure in terms of a network of points which comprise a system of urban centres, and here the concern is with the size, spacing, frequency, and functional composition of centres, as well as their market-area and supply-area configurations. While the central-place sub-system frequently represents a significant component within this urban system (Christaller, 1933; Lösch, 1941; Philbrick, 1957; Berry, 1967), it is only one of a number of components and may on occasions be of secondary importance. A third perspective on spatial structure, and one which is most definitely implicit in the previous two, focuses on the patterns of interaction within the region, not only in terms of physical networks of communication (Kansky, 1963; Taaffe *et al.* 1963), but also in the more general sense of spatial interaction. This third perspective typically considers intraregional shopping patterns and commuting flows (Champion and Coombes, 1983), as well as intraregional commodity flows and flows of funds (Isard, 1960). Such interaction patterns may be 'circulatory' and unchanging in nature, and refer to the social and economic functioning of the region at given levels of intensity, or they may become modified over time (Termote, 1978). One particular type of interaction pattern is concerned with transmission of economic impulses within the region, e.g. the intraregional pattern of multiplier effects resulting from an impact at a particular location (Parr 1979), or the manner in which the unemployment rates of cities within a region are spatially and temporally linked (King *et al.* 1969; Cliff *et al.* 1975).

It is, of course, unreasonable to suppose that the analysis of regional spatial structure need be

John B. Parr is a reader in applied economics in the Department of Social and Economic Research, University of Glasgow.

The author is grateful to A. Findlay, M. Ingham, A. O'Sullivan and M. Partridge who were kind enough to comment on earlier drafts of this paper. The contents remain the sole responsibility of the author.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

confined to these perspectives, and other approaches clearly exist. One possibility which suggests itself involves the use of the population-density profile, particularly in the light of its widespread application to the analysis of spatial structure of cities and metropolitan areas. Within this setting the tendency for the gross density of residential (or night-time) population to display a systematic decline with distance from the centre of the metropolitan area has become an accepted empirical regularity. Such a distance-density decay, which has been shown to exist for metropolitan areas within different nations and at different points in time, conforms as a first approximation to the negative exponential function (Clark, 1951) which can be expressed as follows:

$$D_x = D_0 \exp(bx) \qquad (0 \leqq x \leqq x';\ b < 0) \qquad (1)$$

or, in logarithmic form,

$$\ln D_x = \ln D_0 + bx \qquad (1a)$$

where D_x is the population density at distance x from the centre, D_0 is the density at distance 0 (the centre of the metropolitan area), and b is the gradient of the density function, representing the rate at which the logarithm of density decreases with distance. The graph of the negative exponential function is shown in Figures 1a and 1b. In Figure 1a the density axis is scaled arithmetically, while in Figure 1b it is scaled logarithmically, indicating that the logarithm of density decreases with distance from the centre at a constant rate.

The pattern of population density decline associated with the negative exponential function is generally only observed out to a distance of x′, the distance from the centre of the metropolitan area at which some arbitrarily defined minimum urban density Dx′ is encountered (Parr and Jones, 1983). The question inevitably arises, however, as to what happens to the pattern of density decline beyond this distance x′. Does the negative exponential function continue beyond the confines of the metropolitan area, or does some other function exist? For example, does the logarithm of density decrease with distance at an increasing rate, or does it level off at some underlying rural population density? Such questions are seldom asked by the urban economist or the urban geographer, because their concerns are typically with the structure and functioning of the continuously built-up metropolitan area and perhaps its immediate rural margin, and usually not with the wider region which it can be said to dominate. A primary objective of the paper is to seek an answer to these questions by exploring the possibility that the spatial structure of a region may be viewed in terms of a density profile, centred on the dominant metropolitan area of the region.

2. The Metropolitan-area-based Region

The concept of the region has proven to be an extremely elusive one which has generated a substantial debate. No attempt will be made here to

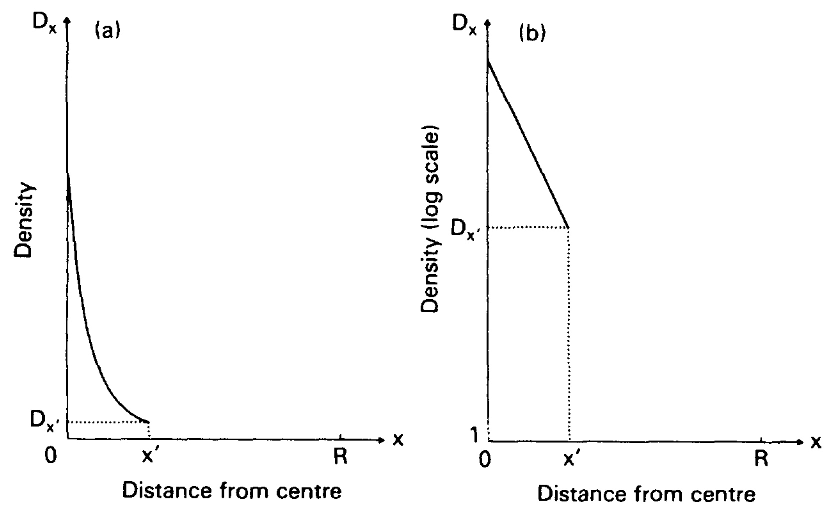

Fig. 1. Negative exponential function with density axis shown on (a) an arithmetic scale; (b) a logarithmic scale.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

review this debate, but it is necessary to specify the type of region under consideration. As is evident from what has already been said, the term 'region' is being used not in the sense of a natural region or a homogeneous region, but rather in the sense of a functional, nodal region, the focus of which is a major metropolitan area. A region of this type, which Dickinson (1947) termed a 'city region', consists of a large metropolitan area or regional metropolis, together with a territory over which it has a general social and economic dominance which is at least as great as that exerted by some neighbouring metropolitan area.

Such a definition of the region is somewhat general and in attempting to render it operational, a number of difficult and interrelated questions arise. The first involves the problem of what constitutes a metropolitan area. Size, as measured by the level of population, obviously represents one criterion and a minimum figure might well be imposed. Bogue (1950), for example, arbitrarily selected a minimum size of 100,000 for metropolitan areas in the USA (250,000 in the northeastern states). Size alone, however, is probably an insufficiently precise indicator of metropolitan status. It is preferable, therefore, to think in terms of structural complexity and consider a range and level of economic activities that need to be present before metropolitan-area status can be said to have been attained, and in this connection the composition of the export base may be particularly revealing (Ullman and Dacey, 1962). The designation of centres as metropolitan areas in the definition of regions renders them in a sense equivalent, but in no way does this imply that they are structurally identical or that they are of equal hierarchical importance in the urban system. Some may be only important in a regional sense, while others will have a sectional or multi-regional importance, and one or possibly two metropolitan areas may have a definite national significance. The designation 'metropolitan area' is thus made with respect to centres at intermediate-to-high levels of the urban hierarchy.

A second question raised by the foregoing definition of the region involves the mode by which the metropolitan area dominates its region. This relates not only to the social and economic functions undertaken within the metropolitan area on behalf of the region, but to the role of the metropolitan area as a centre of administration, organisation, ownership and control, in both the private and public sectors, and more generally to the locational strength of the metropolitan area, relative to that of the rest of the region. A third question concerns the issue of regional delimitation and the criteria that may be employed to determine this. Obviously, the factors considered in the previous question are likely to have an important bearing here, but other criteria such as those involving patterns of spatial interaction may also be employed. It is sometimes the case that 'short-cut' methods are used for the purposes of regional delimitation, although not all of them are satisfactory. Bogue (1950), for example, defined a set of metropolitan-area-based regions by taking the perpendicular bisectors between pairs of neighbouring metropolitan areas, regardless of their size. Such an approximation ignores the fact that the region based on a relatively large metropolitan area is likely to extend beyond the perpendicular line between it and a smaller competing metropolitan area. More reliable approaches to regionalisation involve the use of graph theory (Nystuen and Dacey, 1961) and gravity models (Boudeville, 1966), the simplest form of which is the Law of Retail Gravitation (Reilly, 1954). It needs to be emphasized that the metropolitan-area-based regions are not self-contained entities or closed systems, and within each there is likely to be considerable social and economic interaction with other regions of the nation.

It must by now be apparent that any attempt at identifying and delimiting a set of metropolitan-area-based regions within a nation is likely to involve a measure of subjective judgement, whatever efforts are made to avoid this, and criticisms are likely to be made that there are too many or too few regions. A regionalisation scheme consisting of metropolitan-area-based regions will frequently cover the national territory in nations which have a compact shape and/or a fairly high level of population density. However, a good deal of blurring of regional boundaries can be expected, resulting from the differing patterns of the spatial organisation of supply among the various functions, which gives rise to cross-boundary interaction (Dickinson, 1947). In nations with relatively low levels of development or in nations with extended peripheries, a system of metropolitan-area-based regions may not cover the entire national territory, and part of this may be dominated by auxiliary centres which cannot be classified as metropolitan in status.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

The region, as defined above, thus consists of a metropolitan part (the metropolitan area) and a non-metropolitan part surrounding it, which contains a rural population, as well as a population located in a network of urban centres of varying size. Traditionally a good deal of attention has been devoted to analysing the spatial structure of the metropolitan part of the region, i.e. the metropolitan area (Evans, 1973; Mills, 1972; Alonso, 1964). By contrast, very little work has been devoted to the structure of the non-metropolitan part, the most significant having been undertaken by Bogue (1950) and discussed by Haggett (1965) and Vining (1955). Bogue's study, very much in the tradition of Gras (1922) and McKenzie (1933), was concerned with the manner in which a wide range of social and economic indicators (including population density) declined in intensity over the non-metropolitan part of the region with increasing distance from the metropolitan area. In order to demonstrate his principal thesis that transportation improvements had transformed the national territory of the USA into a series of 'metropolitan communities' (i.e. metropolitan-area-based regions), he initially worked with a composite structure which consisted of distance-density data for all regions. This procedure precipitated considerable criticism (Blumenfeld, 1950; Berry *et al.* 1968), much of which cannot be considered valid, in view of the fact that Bogue considered in his analysis, variations among the major geographical sections of the nation (i.e. the Northeast, the North Center, the South, and the West), among regions based on metropolitan areas of different size classes, and among sectors within regions.

By focusing on the spatial structure of the region as a whole, an attempt will be made here to draw together into a single framework the two parts of the region (the metropolitan and the non-metropolitan), which hitherto have tended to be treated as distinct and unconnected entities.

3. Regional Density Functions

A perusal of the available data on regional population density soon reveals that the overall pattern of density decline is not in accordance with the negative exponential function; see, for example, Berry and Horton (1970, p. 280) and Clark (1967, p. 345). For highly urbanised regions at least, the pattern resembles the one indicated in Figure 2a. This indicates that beyond x′ (the approximate boundary of the metropolitan area) the logarithm of density tends to decrease with distance at a decreasing rate. More specifically, the nature of the distance-density decay over this part of the region is such that the logarithm of density decreases with the logarithm of distance at a constant rate, as in Figure 2b, where both the density and distance axes are shown on logarithmic scales. Thus over the range x′ to R (the boundary of the region) the pattern of population

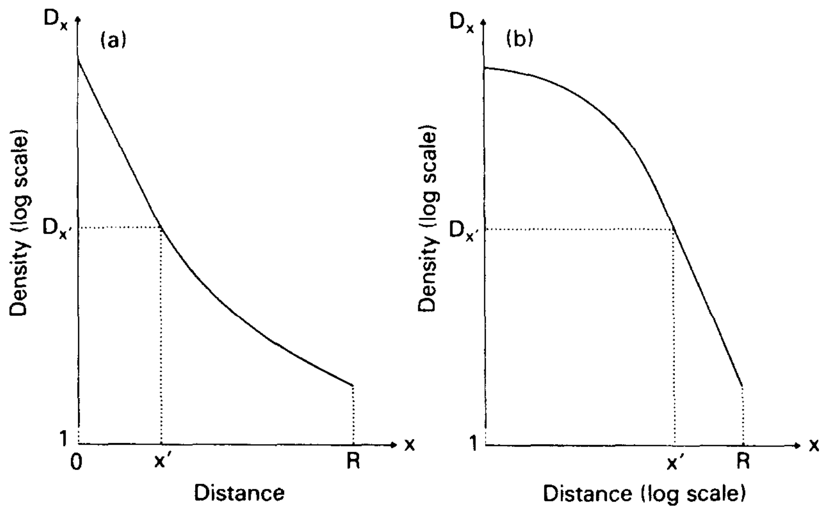

Fig. 2. Regional density profile with distance axis shown on (a) an arithmetic scale; (b) a logarithmic scale.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

density resembles a Pareto function, the form of which is as follows:

$$D_x = D_1 x^d \qquad (x' \leq x \leq R;\ d < 0) \tag{2}$$

or

$$\ln D_x = \ln D_1 + d(\ln x) \tag{2a}$$

where D_x is the density at distance x from the centre of the metropolitan area, D_1 represents the density at distance 1, and d is the gradient of the density function, i.e. the rate at which the logarithm of density decreases (beyond the boundary of the metropolitan area) with the logarithm of distance from the centre. Such a pattern is consistent with the evidence for composite regions presented by Bogue (1950). At first glance, then, it appears that the non-metropolitan part of the region possesses a population-density pattern which is significantly different from the pattern of the metropolitan part, a view proposed by Ajo (1965) and supported by Casetti (1969). It is entirely conceivable, however, that there exists a more general density function for the entire region which subsumes the distinct patterns of the metropolitan and non-metropolitan parts of the region. This represents a major working hypothesis of the paper.

Sources of Data and the Construction of Profiles

A major problem encountered in any examination of population density patterns on a region-wide basis is the paucity of data, although for an increasing number of nations, comparable data on population densities for the metropolitan and non-metropolitan parts of regions are becoming available. The profiles to be examined here (involving selected regions of the UK and North America) have been derived from published data on population density or total population. A conservative course was adopted in the choice of regions, and only those regions were selected in which the centres could by any standard be considered of metropolitan-area status. It was also necessary to guard against the overbounding of regions, particularly with UK regions, but in most cases each region extends well beyond the confines of its metropolitan area. The regional density profiles indicated in Figure 3 are based on data for population densities within concentric rings based on the centre of the metropolitan area. This is not the case with the profile for the New York region, however. Here the profile is simply a representation of the best-fitting curve (presented by Hoover and Vernon, 1962, p. 5) of a scatter diagram showing the relationship between county population density and the distance from the country's centre of population to Mid-Town Manhattan.

For the London region and the North American regions (excluding New York), the profiles were based on distance-density data from previous studies: London (Transport Studies Unit, 1984); Chicago (Berry and Horton, 1970, p. 280, adapted from Rees, 1968); Montréal (Ville de Montréal,

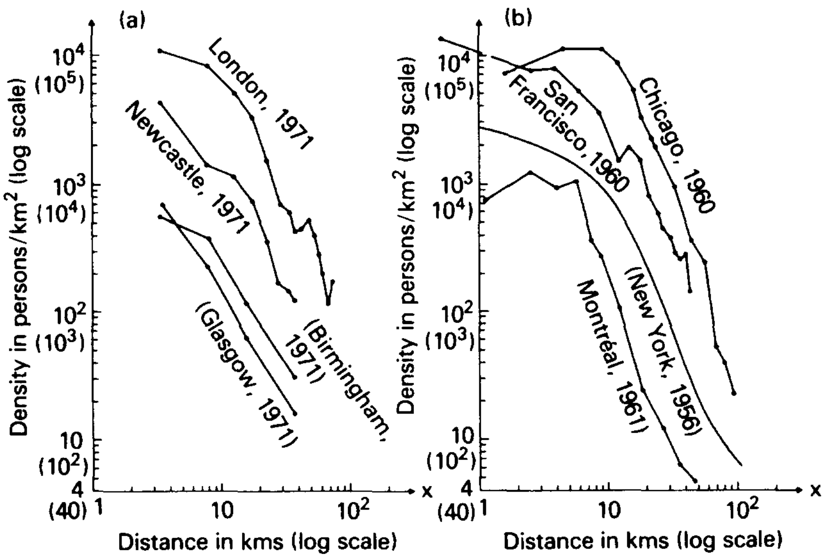

Fig. 3. Distance-density profiles for (a) U.K. regions; (b) North American regions. N.B. a region shown in parentheses has its density scale indicated in parentheses.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

1964); San Francisco (Clark, 1968, p. 345). In the case of the London region the density data were available by 5 km rings out to a distance of 75 km, and for the Montréal region the density data were available according to the following ring structure: 1.61 km rings over the range 0–9.66 km; a single ring over the range 9.66 to 14.49 km; 8.05 km rings over the range 14.49–30.59 km; a single ring for the range 30.59–40.25 km; and a single ring for the range 40.25–53.11 km. In the case of the Chicago and San Francisco regions the ring structure was less regular. For the Birmingham, Glasgow and Newcastle regions the profiles were derived by calculating densities from population data assembled by concentric rings (Rhind *et al.* 1977, p. 20 and p. 36). In the case of the Birmingham and Glasgow regions, the concentric-ring structure was as follows: 0–5 km, 5–10 km, 10–20 km, 20–50 km, 50–100 km and 100–200 km, but in the case of the Newcastle region the data were available by 5 km rings out to a distance of 200 km.

In plotting the distance in each profile of Figure 3, the distances employed were the same as those used in the published studies for the Chicago, New York and San Francisco regions. For the other regions, however, it was possible to plot the population density of each ring against the mean distance from within that ring to the centre of the region, rather than against the distance to the centre from the midway point of the ring. Thus, for the 5–10 km ring, the distance to the centre was taken as 7.78 km, rather than 7.5 km.

It is important to comment on the maximum extent of the density profiles for the various regions presented in Figure 3. Table 1 indicates for each region the radii of the most-distant ring, as well as the distance from within this ring to the centre, i.e. the distance of the last observation point in the profile. The profiles for the London, Chicago, Montréal, New York and San Francisco regions clearly do not cover the full extent of the regions and simply reflect the availability of data. The profiles for the other UK regions also do not reflect the full extent of the regions. It was not possible to extend each profile beyond the distance indicated by considering an additional ring (the 50–100 km ring for the Birmingham and Glasgow regions and the 40–45 km ring for the Newcastle region), since this would have involved the inclusion of parts of one or more adjoining regions: the regions based on Bristol, Liverpool, London, Manchester and Nottingham in the case of the Birmingham region; the region based on Edinburgh in the case of the Glasgow region; and a possible region based on Teesside (centred on Middlesbrough) in the case of the Newcastle region.[1] Thus for reasons of data availability or

Table 1

Data on the extent of regions

	Radii of furthest ring (km) inner radius	outer radius	Distance of furthest ring to centre (km)
UK Regions:			
London (1971)	70.00	75.00	72.53
Birmingham (1971)	20.00	50.00	37.14
Glasgow (1971)	20.00	50.00	37.14
Newcastle (1971)	35.00	40.00	37.56
North American Regions:			
New York (1956)	[a]	[a]	103.00[b]
Chicago (1960)	102.41	107.41	97.41
San Francisco (1960)	40.50	46.50	43.50
Montréal (1961)	40.25	53.11	46.98

[a]Not applicable; see text.
[b]Distance to centre of region from the centre of population of Dutchess County, NY.

[1]Even with such a restriction it is possible that the Birmingham and Glasgow regions may be slightly overbounded in certain directions. The need to avoid this problem of overbounding made it impossible to utilise the population data arranged by concentric rings of 0–5 km, 5–10 km, 10–20 km, 20–50 km for the regions based on Leeds, Liverpool, Manchester, Nottingham and Sheffield (Rhind *et al.* 1977, p. 20). In each case the centre of the region lies within 51 km (i.e. less than 100 km) of the centre of one of the other regions. For any of these regions, therefore, to have included the 20–50 km ring would have involved encroachment on one or more of the other regions. If this ring had been excluded, however, the regional profile would have only been based on the density within each of three rings, and this would have obviously been unsatisfactory for the purposes of description. Even the use of four rings, for both the Glasgow and Birmingham regions, was not desirable, but was nevertheless thought to be acceptable in view of the preliminary nature of the study.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

because of the nature of the arrangement of population data by concentric rings, the profiles indicated in Figure 3 refer to 'incomplete' regions.

Patterns of Regional Population Density

The density profiles are shown in Figure 3 where both the distance and density axes are scaled logarithmically, and as will be seen later there are certain advantages with such a convention. In general, it is apparent that except for an obvious flattening in the direction of the centre of the metropolitan area the profiles all display a downward linearity, similar to the generalised pattern of Figure 2b. As a crude approximation, therefore, the population density pattern for each region over the range 0 to L (the last observation point) can be said to conform to the Pareto function of equation (2). In view of this flattening of the profiles, the application of equation (2) results in unrealistically high density values in the vicinity of the centre. Moreover, the nature of the Pareto function is such that as x approaches zero, the value of D_x tends to infinity. The tendency for the profile to flatten is, however, reflected in the three-parameter Pareto function which has the form

$$D_x = C(x+z)^h \qquad (0 \leqq x \leqq L;\ h < 0) \qquad (3)$$

or

$$\ln D_x = \ln C + h[\ln(x+z)] \qquad (3a)$$

where z is the third parameter, h is the density gradient (indicating the rate at which the logarithm of density declines with the logarithm of modified distance), and C is a constant. In applying such a function to a regional density pattern, that value of z is selected which maximises the goodness of fit. The three-parameter form of the Pareto function has the advantage that D_0 has a finite value. It nevertheless suffers from the drawback that the parameter z, besides being difficult to interpret, varies from region to region, thus making it impossible to compare regions in terms of the h parameter.

There exists, of course, a number of other functions which are able to reflect the major features of the profiles in Figure 3. The one to be employed here is the square-root negative exponential function which can be expressed as follows:

$$D_x = D_0 \exp(ax^{0.5}) \qquad (0 \leqq x \leqq L;\ a < 0) \qquad (4)$$

or

$$\ln D_x = \ln D_0 + ax^{0.5} \qquad (4a)$$

where a is the rate at which the logarithm of density decreases with the square root of distance from the centre.[2] This function was employed by Ajo (1965) to describe population densities over the non-metropolitan part of the London region (the range 31–112 km). Here, however, it is considered for the entire region, or more accurately, the range 0 to L. Ajo applied the function to the values of 'average' density (e.g. the density within the area of a circle of radius 10 km), rather than to values of the more usual 'marginal' or 'local' density (e.g. the density within the concentric ring of radii 5 km and 10 km), as used in this paper. In Figure 4 the profiles of the London and San Francisco regions are presented, together with the best-fitting square-root negative exponential (SNE) functions, both axes being scaled logarithmically as in Figures 2b or 3.

The SNE function is able to reflect the two salient characteristics of the density profiles presented in Figure 3: the flattening of the profiles in the direction of the centre; and the tendency for the logarithm of density to fall with the logarithm of distance at a roughly constant rate over a large part of the distance range. What the SNE function fails to capture is the tendency for a profile to reach a crest in the vicinity of the centre, giving rise to the so-called 'density crater'. This cresting of densities has been observed in many larger cities which have reached a mature stage of development (Newling, 1969). Such a phenomenon will not be considered here, since it only occurs in two of the eight regions, although it would probably have been observed for most of the other regions, had the plots been based on narrower distance rings in the inner parts of the metropolitan areas. It is possible to incorporate this phenomenon of cresting in the analysis of regional density patterns, by employing more complex func-

[2] A more general version of the SNE function is the negative exponential function of degree i. This function has the form $D_x = D_0 \exp(kx^i)$ where $k < 0$ and i is a third parameter. Again, the third parameter i is not readily interpreted and is also likely to vary from region to region, thus precluding comparison among regions in terms of the slope parameter k.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

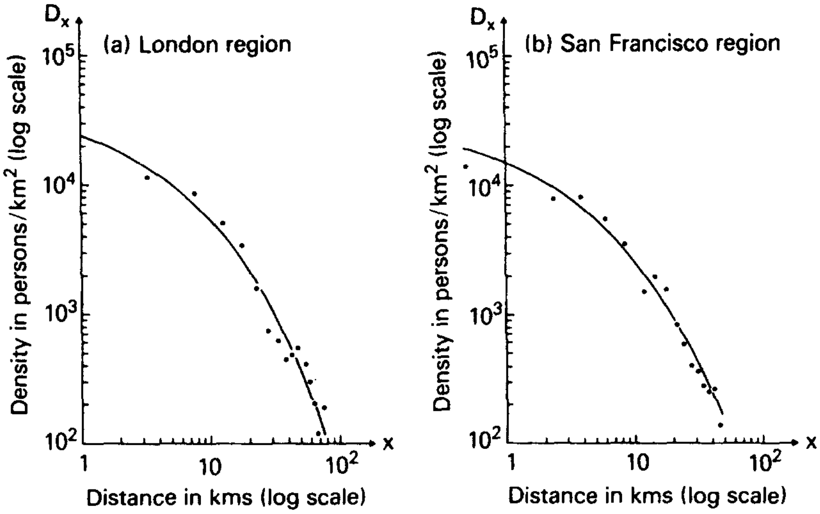

Fig. 4. Density profile and best-fitting SNE function for (a) London region; and (b) San Francisco region.

Table 2

Best-fitting SNE functions for regional profiles shown in Figure 3

	D_0	D_1	a	R^2
UK Regions:				
London (1971)	44,113 (44.42)	22,370	−0.679 (−17.32)	0.958
Birmingham (1971)	22,573 (37.84)	11,101	−0.710 (−10.70)	0.983
Glasgow (1971)	29,750 (29.78)	12,309	−0.883 (−10.19)	0.981
Newcastle (1971)	19,048 (39.09)	8,305	−0.830 (−14.80)	0.973
North American Regions:				
New York (1956)	62,900 (20.81)	28,791	−0.781 (−9.04)	0.803
Chicago (1960)	83,458 (33.00)	37,165	−0.809 (−13.94)	0.942
San Francisco (1960)	30,829 (77.31)	14,090	−0.783 (−26.31)	0.980
Montreal (1961)	61,108 (25.82)	20,012	−1.116 (−9.64)	0.932

Values for D_0 and D_1 are given in persons/km². The t-ratios are shown in parentheses.

tional forms (Parr, 1985a).[3] For the present purposes, however, it suffices to use the SNE function, recognising that this is a simplification of actual conditions in the extreme central part of the region.

Table 2 presents the best-fitting values of D_0 and a, when equation (4a) is applied to the profiles shown in Figure 3. In the case of the New York region equation (4a) has been applied to the scatter diagram on which the profile in Figure 3b was based (Hoover and Vernon, 1962, pp. 5–6). The goodness of fit is generally high, although little reliance can be placed on the R^2 values for the Birmingham and Glasgow regions, since these were each based on only four observations. The value of D_0, the density at distance 0, is merely an extrapolated value which indicates the level to which densities are tending, and if a density crater is present the SNE function overestimates the density levels in the vicinity of the centre. Since the value of D_0 for each region cannot be indicated in Figure 3, Table 2 also includes the best-fitting values for D_1, the density at distance 1. The value of a, the slope parameter, can be regarded as an index of regional population concentration, summarising the unevenness with which population is distributed over the region or that part of it under investigation. A low value of a indicates a small degree of metropolitan/non-metropolitan differentiation, implying that the non-metropolitan part of the region is well developed, relative to the metro-

[3]The particular form of the quadratic exponential function employed by Newling (1969) to take account of the cresting of densities within the metropolitan area is not suitable at the regional scale, mainly because beyond the crest the logarithm of density decreases with distance at an approximately constant rate. Although such a function may represent a satisfactory description of densities within metropolitan areas, it is unrealistic for the non-metropolitan part of the region, where, as we have seen, the logarithm of density tends to decrease at a decreasing rate with distance, or at a constant rate with the logarithm of distance. Newling, himself, was clearly aware of this limitation of the quadratic exponential function.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

politan part. By contrast a high value of *a* would indicate that the non-metropolitan part of the region is undeveloped, relative to the metropolitan part.

It is not possible to compare the regions of the UK with those of the USA since the periods of time are sufficiently different. One finding that does emerge in both nations, however, is the tendency for an inverse relationship between the population size of the metropolitan area and the level of regional concentration, as measured by the value of *a*, so that the greater the population of the metropolitan area, the lower the value of *a*. This may be regarded as consistent with (and related to) a similar inverse relationship between the size of metropolitan area and the value of the density gradient within the metropolitan area (Mills, 1972). Unfortunately, there is a scarcity of time-series data, thus making it difficult to compare the structures of individual regions over time. It may be reasonably assumed, however, that the patterns of parameter variation over time, which are shown in Table 3, are typical of highly-urbanised regions. The value of D_0 appears to have declined in recent years, after having earlier attained a peak value. The value of *a* displays a decline over a longer period, reflecting a continued reduction in the overall level of regional concentration.

4. Features of the SNE Function

The argument so far has been that population densities throughout a metropolitan-area-based region tend to adhere to an SNE function. This raises a number of questions involving the relationship of the SNE function to other better-known functions, as well as certain practical and conceptual problems that arise in the use of population-density functions at the regional scale.

Table 3

Temporal variation of parameters of SNE function for three regions

Region	D_0	a
Chicago:		
1940	75,607	−0.918
1950	85,033	−0.890
1960	83,458	−0.809
Montréal:		
1941	50,869	−1.272
1951	62,561	−1.242
1961	61,108	−1.116
London:		
1971	44,113	−0.679
1981	37,780	−0.614

Note: Values for D_0 are given in persons/km^2.

SNE Function and other Population-Density Functions

While the SNE function appears to describe fairly accurately the pattern of population density throughout the region, the negative exponential function is generally regarded as an appropriate characterisation of densities within the metropolitan area (Berry and Horton, 1970; Mills, 1972). Yet the negative exponential function differs substantially from the SNE function, as is obvious from a comparison of equations (1) and (4). This apparent inconsistency can be explained by the fact that quite distinct statistical functions may approximate each other over their tails. This point is demonstrated graphically in Figure 5a, which shows the best-fitting SNE function for the Chicago region. The distance axis is scaled on an arithmetic scale, the density axis being scaled logarithmically. Now with this scaling of the axes a negative exponential function would appear as a straight line, as in Figure 1b. It can be seen that over the range 0 to 30 km (the approximate extent of the metropolitan area) the SNE function has relatively little curvature and can thus be taken as a reasonable approximation of a negative exponential function. Clearly, such an approximation could not be said to exist over the entire range of the function, i.e. from 0 to L. For relatively small values of x, therefore, the SNE function tends to display the characteristics of the negative exponential function.[4]

Turning to the non-metropolitan part of the region, it was argued earlier that here population densities generally conform to the Pareto function of equation (2). If, however, equations (2) and (4) are compared, it is obvious that the Pareto function differs from the SNE function. As above, it is a case of two distinct functions approximating each other over a particular range, and this can be seen from Figure 5b. Here the best-fitting SNE function for the

[4]It is of some interest to note that the profiles for certain metropolitan areas (based on the axes of Figures 1b or 5a) even exhibit a slight curvature or convexity to the origin, as in Figure 5a (e.g. Clark, 1951, pp. 492–93). Such a tendency is, of course, in keeping with the existence of a region-wide SNE function.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

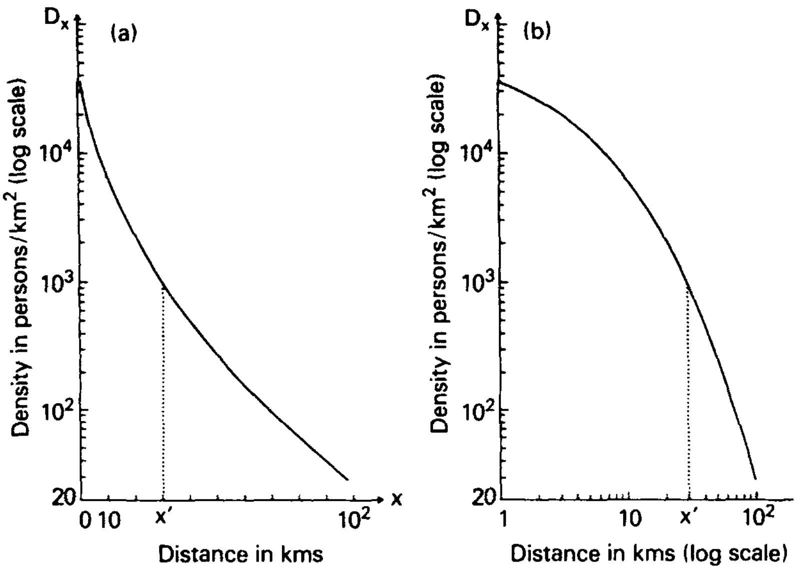

Fig. 5. Best-fitting SNE function for the Chicago region with the distance shown (a) on an arithmetic scale; (b) on a logarithmic scale.

Chicago region is again displayed. However, both axes are scaled logarithmically, so that linearity of a curve indicates adherence to the Pareto function of equation (2). Beyond 30 km, the boundary of the metropolitan area, the SNE function displays a pronounced linearity, and over this range, therefore, it can be regarded as an approximation of the Pareto function.

It now becomes evident that the SNE function of equation (4) combines in an approximate manner the characteristics of a negative exponential function of equation (1) for the metropolitan part of the region, with the characteristics of Pareto function of equation (2) for the non-metropolitan part of the region. The SNE function does not, of course, correspond exactly to either the negative exponential function or the Pareto function over the relevant ranges, and there is a smooth transition (i.e. no kink or abrupt change of slope) in the pattern of distance-density decay from the metropolitan to the non-metropolitan part of the region. The fact that such a transition is common at the margins of metropolitan areas in the absence of strong land use regulations strengthens the case for the SNE function as a general description of population densities on a region-wide basis. It must be stressed, however, that this smooth transition of densities may be related to the averaging of densities within the concentric rings (particularly in the case of irregularly-shaped metropolitan areas), a question which will now be discussed.

The Effect of Averaging and the Existence of Local Crests

The characterisation of regional structure in terms of a density function inevitably implies a considerable degree of generalisation. Obviously such a one-dimensional representation of the spatial structure of the entire region is only achieved at the cost of considerable loss of detail, and the regional density function can cast little light on the structure and development of a particular locality within the metropolitan area or of a district in the non-metropolitan part of the region. Clearly, then, this is a case of generalisation of the observed variance of a phenomenon (population density), having a possible validity at the macro scale but being of little relevance at the micro scale. Here the generalisation involves the aggregation of sometimes highly differentiated areal units into concentric rings and then a calculation of the average densities within each ring. This form of generalisation, which tends to produce relatively high R^2 values, has been criticised by Muth (1969). Some of the criticism can, of course, be overcome by applying (data permitting) the density function to the individual sectors of the region, each sector extending from the centre toward the

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

regional boundary. Such additional sector-based analysis may be particularly illuminating where the region has extensive areas occupied by lakes, parks, forests, etc. in the metropolitan as well as non-metropolitan parts of the region. The various methods of applying density functions to the data have been reviewed by Zielinski (1979).

The effect of averaging is to produce a smoothness in the distance-density profile, and the wider (and therefore the fewer) the concentric rings, the smoother will be the profile, as can be observed when comparing the profile for Glasgow and San Francisco regions. It might be expected that the presence of an urban centre at some distance from the metropolitan area would show up as a local crest, as it would in the relevant sector or uni-directional traverse from the centre of the metropolitan area to the boundary of the region. Unless extremely narrow concentric rings are being employed, this is usually not the case, since the distance-density profile represents an average of sectors or traverses in all directions and this militates against the presence of a local crest. Such a tendency is reinforced by the regularity, existing in many regional urban systems, by which the larger the size class of urban centre, the lower its frequency, and the greater its distance from the dominant metropolitan area. When this regularity does not hold, a local crest tends to appear, as would be the case with a large urban centre located relatively close to the metropolitan area.

A local crest in the profile might also occur as a result of a ring of cities located at approximately similar distances from the metropolitan area. For the Chicago region in 1940 and 1950, Berry and Horton (1970, p. 280) identified a local crest at 50 km which was based on a ring of satellite urban centres. By 1960, however, the development of the overall profile was such that this local crest had given way to a protrusion (Figure 3b). The profile for the London region in Figure 3a reveals a local crest at around 50 km. This is in part due to the existence of a ring of urban centres: Chelmsford, Luton, High Wycombe, Aldershot, Guildford, East Grinstead, Tunbridge Wells, as well as the New Towns of Crawley and Bracknell. However, the crest may also be due to the operation of a green belt policy which discouraged residential development within the green belt and led to a 'leapfrogging' of development to areas and urban centres beyond the outer limits of the green belt (Parr, 1985b). The profile for the London region also indicates an upturn in the profile at around 70 km. Since the region extends beyond this distance, it is reasonable to assume that the upturn is, in effect, part of a local crest, based on a ring of urban centres which include Colchester, Cambridge, Bedford, Milton Keynes, Oxford, Basingstoke, Worthing, Brighton and Ashford.

The Density Profile and the Extent of the Region

It was pointed out earlier that because of data limitations, the profiles indicated in Figure 3 refer to underbounded regions. The underbounding is not severe inasmuch as the SNE function extends well beyond the metropolitan area. Nevertheless, because of this underbounding, we have to entertain the possibility that beyond the last observation point the profile will not conform to the SNE function. The evidence for composite regions presented by Bogue (1950, p. 62), however, suggests that the function will continue toward the boundary of the region.

At the other extreme there exists the problem of overbounding, which typically arises when concentric-ring data are available but a satisfactory regionalisation scheme is lacking. We may briefly examine the effect of extending the density profile for one region 'too far', i.e. into the neighbouring region or regions. Let us consider the case of a region which is bordered by one or more similar metropolitan-area-based regions. The effect of exceeding the regional boundary is to cause a flattening of the density profile, as is indicated by the broken line in Figure 6a. In the case of the profile for the Newcastle region (Figure 3a) the slackening of the slope at around 30 km probably reflects the presence of the Teesside region. If the plot is extended further, this slackening of the slope may be followed by an upturn, also indicated in Figure 6a. As already noted, however, the upturn of the profile (at around 70 km) for the London region (Figure 3a) is due to a local crest, rather than to a regional boundary being exceeded.

A related problem of overbounding also arises in the case of a region which extends relatively far in certain directions but in other directions is bordered by other regions at a shorter distance. The inclusion of part of the territory of one or more adjoining regions at an intermediate distance will cause the gradient of the profile to flatten, starting at the intermediate distance where the adjoining regions

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

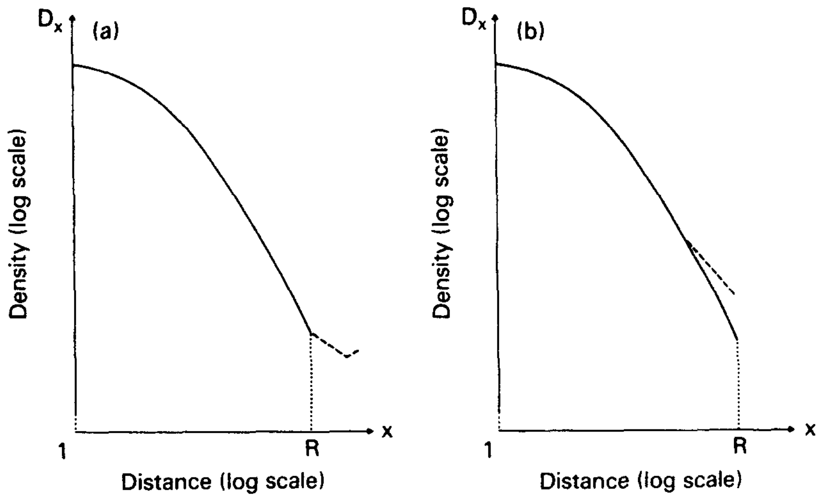

Fig. 6. Regional density function showing the effects of (a) extension beyond the regional boundary; (b) inclusion of parts of one or more adjoining regions.

are first encountered and continuing to the maximum reach of the region R, as indicated by the broken line in Figure 6b. Thus, if the density function can be regarded as an acceptable summary of regional structure, then abrupt changes in the slope of the profile of the type described may indicate that the regional boundary has been exceeded, at least in certain directions. It follows from this that the accurate definition of a regional density profile may be of assistance in the delimitation of the region, in the same way as other representations of regional spatial structure have been employed for this purpose.

One further point relating to the extent of a region involves the calculation of total regional population. If a region is roughly circular in shape with a radius R, and if densities are known to conform approximately to the SNE function, the total regional population P_R (the population living within radius R) can be derived by integrating equation (4) over the range 0 to R, and rotating this through 360°. Thus

$$P_R = 2\pi D_0 \int_0^R x \exp(ax^{0.5})\, dx \qquad (a<0) \qquad (5)$$

Naturally, some adjustment would have to be made if the region has a missing sector (as a result, for example, of a coastal or estuarine location), in which case the rotation will take place through something less than 360°. It becomes clear, therefore, that if approximate magnitudes of D_0 and a are available, the regional density function can be used to provide an estimate of the total population of a region. Furthermore, if the population of the region is known, then the estimated population derived from equation (5) can be regarded as an additional test of the extent to which the SNE function is able to summarise the overall pattern of regional population distribution. The fact that the profiles were based on incomplete regions, for which the total population was not always known, precluded such an experiment here.

Bases for the SNE Function

It is not at all clear why the structure of regional population densities should conform, as it appears to, to the SNE function, and the question naturally arises as to what theoretical rationale might explain the emergence of such a structure. Although the extensive literature on location theory and regional structure can provide a number of insights, nowhere does there appear an explanation for the existence of a population density structure at the regional scale which conforms to the SNE function or similar ones. It might be expected that a regional density function could be derived from the urban-system models proposed by Christaller and Lösch, since these models are specifically concerned in their different ways with the size, spacing and frequency of urban centres of a region. However, serious difficulties arise with the use of such models. First, in the regional systems of developed nations, many urban centres

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

have significant non-central place elements in their populations, these being based on certain types of manufacturing and on resource ex ploitation. Second, while the frequency and spacing of centres are specified in models of the Christaller-Lösch type, the sizes of urban centres are only arranged by order or relative size, so that the size distribution of centres is not explicit. Third, the distribution of rural population is not specified, and these models assume a uniform rural population density, which is obviously at odds with actual conditions. It is possible, of course, to introduce certain reasonable assumptions with respect to each of these three factors, perhaps as part of a simulation exercise designed to generate a regional density function. But what is 'reasonable' in this connection? If the conditions assumed are based on observed regularities, there is the great danger of a circular argument. That is, by incorporating particular observed regularities into a model, the density function is not derived independently, and merely reflects those actual conditions which were built into the model.

In the case of the metropolitan part of the region, a number of explanations have been proposed for the negative exponential function (Bussière and Snickers, 1970; Muth, 1969), and as already argued, the upper tail of the SNE function can be regarded as an approximation of this. Such approaches, which put considerable emphasis on accessibility to the centre of the metropolitan area, particularly as a location for employment, can probably be extended, over the non-metropolitan part of the region, to consider the distribution of the populations dependent on commuting and also on agriculture, if the latter is closely related to the centre in a von Thünen sense. However, the commuting-dependent population becomes relatively unimportant at distances beyond 25–35 km from the centre of the region, while the agriculture-dependent population may be a relatively small proportion of the regional population. Furthermore, while it may be desirable to strive for a region-wide perspective, it is perhaps unrealistic to approach the spatial structure of the non-metropolitan part of the region in terms of a theoretical framework used to explain the spatial structure of the metropolitan part. In the latter case the distribution of population is essentially determined by principles of 'allocation', while in the former case the distribution of population is dependent more on principles of 'location'.

It is possible that some insight into the regional density function might be gained by segmenting it into its metropolitan and non-metropolitan parts. The metropolitan part could be analysed along the lines discussed above, while the non-metropolitan part could be analysed in terms of its constituent sub-populations. This would involve disaggregating the density function for the non-metropolitan part of the region into a series of sub-population density functions. Many types of disaggregation based on social or economic criteria are possible. One such disaggregation of the non-metropolitan population would involve a breakdown into the following elements: (a) the population dependent on commuter travel to the metropolitan area; (b) the population based on agriculture; (c) the population based on the processing of agricultural output; (d) the population based on natural-resource exploitation (including recreational and resort activity); (e) the population based on manufacturing and service activities which supply markets beyond the non-metropolitan part of the region; (f) the population involved in service activities for households and firms in the non-metropolitan part of the region (this may include a population linked to certain types of market-oriented manufacturing); (g) a residual element which would include the retired population. The advantage of such a disaggregation is that the various elements of the population can be examined in terms of the fundamentally different sets of forces which influence their distributions, rather than in terms of a common set of forces. This is not to imply that the various elements of population co-exist in an independent manner, and careful consideration needs to be given to the interdependencies which govern the distributions of these sub-populations. For example, the population dependent on commuting and the population based on agriculture tend to be related to each other in a competitive sense, at least in the absence of land-use regulation. By contrast, the population based on service provision can be expected to be related in a complementary manner to the other elements of the population.

This segmentation of the regional density function into its metropolitan and non-metropolitan parts and the subsequent analysis of each must not be allowed to obscure the fact that certain housing-market and labour-market forces operate throughout the region. Particular attention would therefore need to be given to drawing together the separate

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

analyses for the two parts of the region, by considering the influence of these region-wide forces. Otherwise, the case for employing the concept of a regional density function is greatly weakened. The proposals outlined here are merely suggestive of one of a number of ways of proceeding in the search for a theoretical basis for the SNE function. For the moment at least, the explanation for a function of this type remains as something of a research challenge, particularly in view of the need to ensure a generality of explanation across regions with widely differing internal characteristics.

5. Concluding Comments

It has been argued here that the spatial structure of a metropolitan-area-based region can be described in terms of a population density function of the square-root negative exponential type. Such a region-wide function is able to reconcile the distinct patterns of population density which have been observed in the metropolitan and non-metropolitan parts of the region. The cases considered have all referred to highly urbanised nations in which the regional metropolis represents a well-developed feature of the urban system, and it is possible that the form of the function would be different for nations with significantly lower levels of urbanisation. Obviously, the reduction of regional structure to a one-dimensional characterisation represents a definite limitation in the descriptive ability of the density function, and it can be of little analytical value in any study concerned with a particular area of the region, be the scale sub-metropolitan, urban, rural or sub-regional. Nevertheless, by the same token that the concept of density function has been applied extensively and to good advantage in analysing the internal structure of the metropolitan area, so may it prove useful in the broader context of a metropolitan-area-based region. Within this wider, regional setting the concept can be readily applied to the analysis of changing regional structure and to the examination of policies of land use regulation (Parr, 1985b), as well as to the analysis of such interrelated contemporary phenomena as metropolitan decline, regional deconcentration, and counter-urbanisation.

A major advantage of the density function approach to regional structure lies in its ability to override the distinctions between metropolitan and non-metropolitan and between urban and rural, which are so common in regional analysis. A number of conceptual and policy-oriented questions do not lend themselves to analysis in terms of these kinds of classification and the stark differentiations which they inevitably imply, but require a more general, region-wide perspective which is able to de-emphasise such distinctions. There is, of course, a need for further empirical testing and for the development of a satisfactory theoretical foundation, before the concept of a regional density function can be accepted, but at this stage at least it offers the promise of being an addition to our existing stock of frameworks for understanding the spatial structure of regions.

REFERENCES

AJO, R. (1965). On the structure of population in London's field. *Acta Geographica*, Vol. 18: 1–17.

ALONSO, W. (1964). *Location and Land Use*. Cambridge, Mass: Harvard University Press.

BERRY, B. J. L. (1967). *Geography of Market Centers and Retail Distribution*. Englewood Cliffs N.J: Prentice-Hall.

BERRY, B. J. L., GOHEEN, P. G. and GOLDSTEIN, H. (1968). *Metropolitan Area Definition: a Re-evaluation of Concept and Statistical Practice* (Bureau of the Census, Working Paper No. 28). Washington, D.C.: Bureau of the Census.

BERRY, B. J. L. and HORTON, F. E. (1970). *Geographic Perspectives on Urban Systems*. Englewood Cliffs, N.J.: Prentice-Hall.

BLUMENFELD, H. (1950). The dominance of the metropolis. *Land Economics*, Vol. 26: 194–96.

BOGUE, D. J. (1950). *The Structure of the Metropolitan Community*. Ann Arbor, Mich.: Horace H. Rackham School of Graduate Studies.

BOISVERT, M. (1978). *The Correspondence Between the Urban System and the Economic Base of Canada's Regions*. Ottawa: Economic Council of Canada.

BOUDEVILLE, J. R. (1966). *Problems of Regional Economic Planning*. Edinburgh: Edinburgh University Press.

BUSSIÈRE, R. and SNICKARS, F. (1970). Derivation of the negative exponential model by an entropy-maximising method. *Environment and Planning*, Vol. 2: 295–301.

CASETTI, E. (1969). Alternate urban population density models: an analytical comparison of their validity. In A. J. Scott (ed.), *Studies in Regional Science*. London: Pion. 105–16.

CHAMPION, A. and COOMBES, M. G. (1983). Functional regions: definitions, applications, advantages. Factsheet No. 1, Centre for Urban and Regional Development Studies, University of Newcastle-upon-Tyne.

CHRISTALLER, W. (1933). *Die zentralen Orte in Süddeutschland*. Jena: Gustav Fischer. Translated by C. Baskin as *Central Places in Southern Germany*. Englewood Cliffs, N.J.: Prentice-Hall (1966).

CLARK, C. (1951). Urban population densities. *Journal of the Royal Statistical Society*, Series A, Vol. 114: 490–96.

CLARK, C. (1968). *Population Growth and Land Use*. New York: Macmillan.

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

DICKINSON, R. E. (1947). *City, Region and Regionalism*. London: Routledge and Kegan Paul.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016

EVANS, A. W. (1973). *The Economics of Residential Location*. London: Macmillan.

FRIEDMANN, J. R. P. (1956). Locational aspects of economic development. *Land Economics*, Vol. 32: 213–27.

GRAS, N. S. B. (1922). *An Introduction to Economic History*. New York: Harper.

HAGGETT, P. (1965). *Locational Analysis in Human Geography*. London: Edward Arnold.

HOOVER, E. M. and VERNON, R. (1962). *Anatomy of a Metropolis*. Garden City, NY: Anchor Books.

ISARD, W. (1960). *Methods of Regional Analysis*. Cambridge, Mass: The MIT Press.

KANSKY, K. J. (1963). Structure of transport networks: relationships between network geometry and regional characteristics. Research Paper No. 84, Department of Geography, University of Chicago.

KING, L., CASETTI, E. and JEFFREY, D. (1969). Economic impulses in a regional system of cities. *Regional Studies*, Vol. 3: 213–18.

LÖSCH, A. (1941). *Die räumliche Ordnung der Wirtschaft*. Jena: Gustav Fischer. Translated by Woglom, W. H. and Stolper, W. F. as *The Economics of Location*. New Haven, Conn.: Yale University Press (1954).

McKENZIE, R. D. (1933). *The Metropolitan Community*. New York: McGraw-Hill.

MILLS, E. S. (1972). *Studies in the Structure of the Metropolitan Economy*. Baltimore: Johns Hopkins Press.

MUTH, R. (1969). *Cities and Housing*. Chicago: University of Chicago Press.

NEWLING, B. E. (1969). The spatial variation of urban population densities. *Geographical Review*, Vol. 59: 242–52.

NYSTUEN, J. D. and DACEY, M. F. (1961). A graph theory interpretation of nodal regions. *Papers and Proceedings of the Regional Science association*, Vol. 7: 29–42.

PARR, J. B. (1979). Spatial structure as a factor in economic adjustment and regional policy. In D. Maclennan and J. B. Parr (eds.), *Regional Policy: Past Experience and New Directions*. Oxford: Martin Robertson. 191–210.

PARR, J. B. (1985a). The form of the regional density function. *Regional Studies*, Vol. 19: forthcoming.

PARR, J. B. (1985b). The regional density function: applications of the concept. Unpublished manuscript, Department of Social and Economic Research, University of Glasgow.

PARR, J. B. and JONES, C. (1983). City size distributions and urban density functions: some interrelationships. *Journal of Regional Science*, Vol. 23: 293–307.

PHILBRICK, A. K. (1957). Principles of areal, functional organization in regional human geography. *Economic Geography*, Vol. 33: 299–336.

REES, P. (1968). The functional ecology of metropolitan Chicago. Master's thesis, University of Chicago.

REILLY, W. J. (1953). *The Law of Retail Gravitation*, 3rd ed. New York: Pilsbury Publishers.

RHIND, D. W., STANNESS, K. and EVANS, I. S. (1977). Population distribution in and around selected British cities. Census Research Unit Working Paper No. 11, Department of Geography, University of Durham.

TAAFFE, E. J., MORRILL, R. L. and GOULD, P. R. (1963). Transport extension in underdeveloped countries: a comparative analysis. *Geographical Review*, Vol. 53: 503–29.

TERMOTE, M. G. (1978). Migration and commuting in Lösch's central-place system. In R. Funck and J. B. Parr (eds.), *The Analysis of Regional Structure: Essays in Honour of August Lösch*. London: Pion. 83–90.

TRANSPORT STUDIES UNIT (1984). Transport, land use and energy interaction. Unpublished note, Transport Studies Unit, University College London.

ULLMAN, E. L. and DACEY, M. F. (1962). The minimum requirements approach to the urban economic base. *Lund Studies in Human Geography*, Series B, No. 24: 121–143.

VILLE DE MONTRÉAL (1964). *La Vague d'Expansion Metropolitaine* (Bulletin Technique, No. 1). Montréal: Service d'Urbanisme.

VINING, R. (1955). A description of certain spatial aspects of an economic system. *Economic Development and Cultural Change*, Vol. 3: 147–95.

ZIELINSKI, K. (1979). Experimental analysis of eleven models of urban population density, *Environment and Planning A*, Vol. 11: 629–41.

Downloaded from usj.sagepub.com at PENNSYLVANIA STATE UNIV on May 9, 2016