# THE SPATIAL STRUCTURE OF THE RURAL-URBAN FRINGE: A MULTIVARIATE APPROACH

DANIEL R. FESENMAIER,
MICHAEL F. GOODCHILD, and
SANDRA MORRISON

*University of Western Ontario*

The literature contains many definitions of the rural-urban fringe, but it is difficult to find consensus on how many zones it contains, or how they can be mapped. It is argued in this paper that transitions within the fringe and at its borders are continuous rather than abrupt. In a study area adjoining London, Ontario, cluster analysis is used, and it leads to the identification of three norms. Membership probabilities are then mapped on the basis of each areal unit's resemblance to these norms through discriminant analysis.

Dans la littérature, il y a beaucoup de définitions de la fringe péri-urbaine, mais on discute le numéro de zones dans la fringe et la méthode de sa cartographie. On raisonne ici que les transitions dans la fringe et sur ses bords sont continues plutôt qu'abruptes. Dans cette étude des environs de London, Ontario, on emploie de 'cluster analysis' qui donne l'identification de trois groupes normatifs. On en fait une carte des probabilités d'adhésion sur la base du correspondance de chaque unité aréale à ces groupes normatifs par l'analyse discriminatoire.

The term rural-urban fringe is generally used to describe the zone of transition between built-up urban areas and the surrounding countryside. Its inner and outer limits are difficult to define precisely, however, which has tended to reduce the utility of definitions of the fringe itself, as a distinctive element of metropolitan structure. Rodehaver, for example, identified it loosely as 'that area in which the land is utilized in an urban manner while at the same time certain attributes of the rural are present as well' (Rodehaver, 1946, p. 50). Russwurm is considerably more specific, but the empirical detail in his definition restricts its applicability: 'more than 50% of the population on the fringe should consist of non-farm people. For the large Canadian metropolitan areas in excess of 100,000 the fringe should have a width of more than 10 miles' (Russwurm, 1971, p. 108).

The internal structure of the fringe also has been described in many ways. Several observers have identified two zones, one more like the urban extreme and one more rural. Rememann (1960) called these the 'inner' and 'outer' fringes respectively. McKain and Burnight (1953) used the terms 'suburban' and 'extended,' while Myers and Beegle (1947) referred to the 'true' and 'potential' fringe. While these studies agree on the number of subzones within the fringe, there is no reason to suppose that they would agree on the limits of those subzones in a particular study area, since the accompanying definitions are general and often subjective. To add further confusion, Russwurm (1973) identified four zones rather than two (the suburban ring, urban fringe, urban shadow, and commuting shed).

There seems to be little prospect of a more adequate definition of the limits of the rural-urban fringe, since the question simply begs the far greater one of defining

urban and rural. The problem of identifying the fringe's zonal structure, however, is more tractable; that is, it appears that, by using certain multivariate techniques, one can identify the internal structure of the fringe without the need to define precisely the characteristics of the fringe (itself). Thus, the purpose of this paper is to develop methods capable of providing answers to several basic questions. First, it should be possible to distinguish a zone showing a more or less continuous transition from urban to rural, and second, it should be possible to test whether this zone consists of subzones which exhibit homogeneous characteristics and which extend over appreciable areas. In this case, it should be possible to determine the number of subzones and the archetypal characteristics of each. Finally, it should be possible to map the areal extent of the fringe zone. In this connection it is clearly unreasonable to expect a sudden transition between adjacent subzones: for example, if one were to decide that the fringe consisted of two parts, an inner fringe and an outer, there would be little point in trying to determine the precise location for a boundary between them. The approach taken here is to concentrate on identifying the archetypes of each zone, i.e., urban, rural and fringe, and then to map the internal structure of the fringe zone by expressing the degree of resemblance between each data unit of the study area and the various archetypes (hereafter referred to as norms). Under this approach it is assumed that the boundary of each of these zones is not clear, the zones instead forming fuzzy sets in the sense of Gale (1972). The proposed technique combines cluster and discriminant analyses and is illustrated in the paper by application to the fringe of London, Ontario.

The problem of delimiting the entire fringe area is not tackled explicitly in this paper. Instead, the study area is drawn far enough into the city and far enough out into the countryside to ensure that it contains areas which strongly resemble the urban and rural norms, as well as any intermediate norms typical of the fringe. In this way, the results will be insensitive to the precise limits of the study area. One may then choose to define the fringe from the results of the application; however, it is crucial that no prior definition is required.

## CASE STUDY: LONDON, ONTARIO

The city of London, Ontario, is situated in the Windsor-Quebec axis, the area of the most intensive urbanization in Canada. The metropolitan area had a population of 286,000 in 1971, following a 26.3 per cent increase in the previous decade. However, southern Ontario is also one of Canada's three major agricultural regions and over 90 per cent of the soils in the surrounding area are classified high in capability for agricultural production (Canada Land Inventory, 1968). Thus, London's hinterland is a particularly interesting location for the study of the rural-urban interface.

To the south of London, the city of St Thomas complicates the identification of London's fringe by exerting its own urban influence on the surrounding rural area. A study area north of London was therefore selected (see Figure 1) to minimize the implications of nearby competing centres and to allow London's fringe to be more clearly observed. The study sector consists of London and Biddulph townships and the northern portion of the city of London.

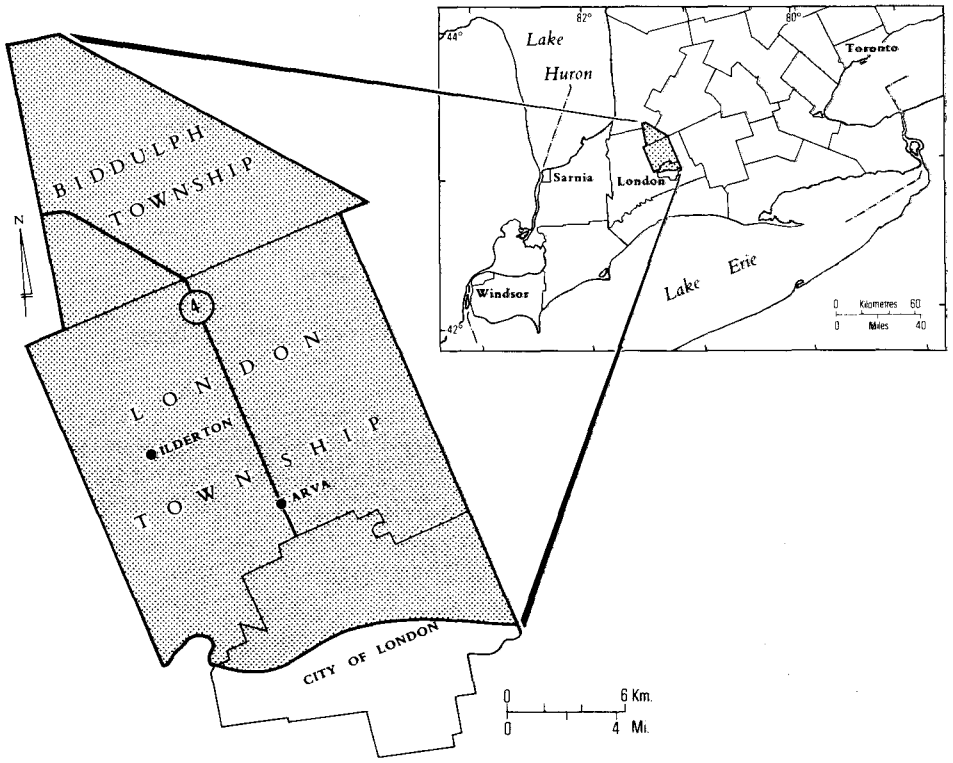


FIGURE 1. Location and details of study area.

*Variables Selected to Describe the Fringe*

The variables selected for the study fall into three categories: socio-economic, land use, and accessibility characteristics. They have all been used in previous studies of the fringe, and can be determined fairly readily from available sources of information. The reader must be cautioned that the results of this study are to some extent dependent on the choice of variables; however, these results would be affected radically only if a variable could be found which was both an important indicator of the rural-urban transition and which varied across the study area in a substantially different manner from each of the variables used in this study. No suitable candidates are suggested in the literature.

*Socio-economic variables: population density* Although a broad range of social measures have been examined in the literature, population density has been used most often in studies of the fringe. It has been noted that gross population density can be used as a surrogate for the type of agriculture in an area (Coleman, 1976) and also for the incidence of part-time farming, as demonstrated in Found's (1971) agricultural intensity model.

Population densities could only be estimated for the study area, since data were not available below the township level. Using addresses provided by the assess-

ment rolls, the exact location of each dwelling was plotted with the aid of cadastral base maps available for each township and the 1:25,000 National Topographic map series. A square mile grid was superimposed on the study area and an estimate of population density obtained by counting dwellings in each square and assuming a ratio of four persons per dwelling, following Russwurm (1971). Although recent evidence indicates that this ratio is dropping, the use of a different value would not affect the results.

*Land use variables: proportion of non-farm residents* An index of the proportions of farm and non-farm residents in the total population was included to give a more detailed picture of the structure of population across the study area. It provides an indirect measure of the amount of invasion from the adjacent urban community.

The classification into 'farm' and 'non-farm' was made according to the occupation information recorded in the assessment rolls for the owner or resident tenant of each dwelling. A resident was categorized as non-farm if his occupation was not recorded as farmer or farm-hand. Part-time farmers were therefore classified as non-farm residents, because their non-farm employment was cited in the assessment rolls. The proportion of households classified as 'non-farm' was calculated for each square mile in the study area.

*Land use variables: percentage of land in non-farm ownership* This criterion also provides a measure of indirect urban influence. It is derived from a set of variables employed by Gertler and Hind-Smith (1962) to identify the 'urban shadow.' On the basis of the previous classification for the occupation of the household head, land in non-farm ownership was calculated as a percentage of the total land area within each grid square.

*Accessibility variables* Accessibility to the central city has also been considered an important factor, as it delimits the commuting zone which has been considered largely responsible for the arrangement of land use in the fringe (Wissink, 1962). For example, Troughton (1977) observed a decline in the incidence of part-time and hobby farmers with distance from the city.

In this study accessibility was equated with distance and measured according to a 'city block' metric between the centroid of each grid square and the approximate location of the London central business district.

*Multivariate Analysis*

If, as the literature suggests, the fringe can be divided into a number of relatively homogeneous zones, this should be reflected in the spatial patterning of the four variables. There should, for example, be zones where the non-farm population has values typical of predominantly urban areas, of predominantly rural areas, and of various stages of transition. Each zone should resemble a 'fuzzy set,' in the sense that individual areal units will be similar to the norm for the zone to a greater or lesser degree. Resemblance to the urban norm will decrease with distance from the CBD, while resemblance to the rural will increase, and changes will also occur in the resemblance of each unit to any transitional norms which may exist. Thus, the first

stage of analysis was concerned with identifying the number of norms present in the study area, and with characterizing them through the four variables. In the second stage, the pattern of spatial variation was analysed in terms of the identified norms by assigning probabilities to each areal unit which described its resemblance to the norms, and thus identified the transition from urban to rural.

*Cluster analysis* The first stage of analysis used a variant of cluster analysis (Anderberg, 1973; Johnston, 1968). The algorithm is designed to identify a fixed number of norms, each characterized by a set of values for each of the four study variables. The criterion to be satisfied in selecting the norms is that when each areal unit is allocated to the norm it most resembles, the groupings so formed are as homogeneous as possible. Resemblance is defined as the generalized distance used in many forms of cluster analysis:

$$D_{ij}^2 = \sum_k (z_{ik} - z_{jk})^2$$

where $i$ is an areal unit, $j$ is a norm, $k$ is a variable, $z_{ik}$ is the standardized score of unit $i$ on variable $k$, and $z_{jk}$ is the value of variable $k$ for norm $j$.

The algorithm alternately identifies norms and allocates units to them in a direct adaptation of the alternating heuristic proposed by Cooper (1964) for location-allocation problems. Each norm is first given a set of random values for each variable. Second, each areal unit is allocated to the most similar (minimum $D^2$) norm. Third, each norm is then redefined by reassigning values so as to minimize the total $D^2$ for the units assigned to it; this is simply achieved by assigning to the norm the mean value of each variable computed over the units allocated to it. The second and third stages are alternated until a stable solution develops. Because the algorithm is heuristic, five runs were made with different starting positions and the best solution adopted.

This approach to cluster analysis requires the researcher to specify the number of groups or norms in advance. Six different solutions were therefore obtained, to cover the range from one to six groups. They were then compared through the total within-group dissimilarity (pooled within group sum of squares) as shown in Figure 2. The relatively small increases from six to three groups, and the sharp jump between three and two, are strong support for the proposition that three norms exist in the data, or that the study area can be divided into three zones on the basis of resemblance to these norms.

Table 1 shows the values of each of the four variables for the three norms. Norm I shows values for each of the four variables which are typical of the most rural part of the study area; norm III has values which can be expected for the most urban portion; while norm II is intermediate in all variables, and so will be referred to as the transitional norm.

*Discriminant analysis* It would be possible at this stage to allocate each areal unit to the norm it most resembles, and make a map showing three areas labelled 'urban,' 'transitional,' and 'rural.' This would be unrealistic, however, because it would convey an impression of sharp changes at zone boundaries, and a greater degree of homogeneity within zones than actually exists. The second stage of the analysis

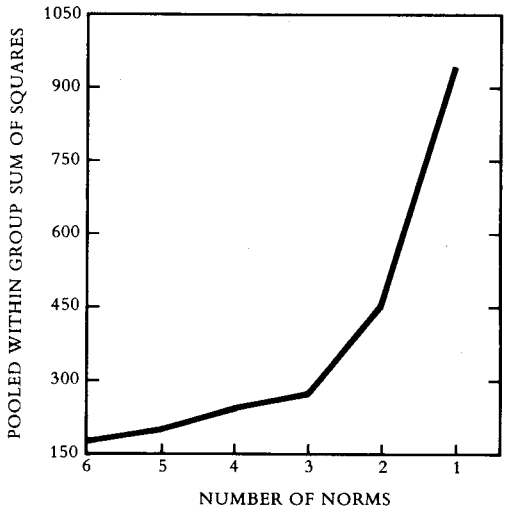


FIGURE 2. Pooled within group sum of squares for given number of norms.

TABLE 1

STANDARDIZED VALUES FOR EACH VARIABLE FOR EACH NORM

| Variable | I (rural) | II (transition) | III (urban) |
|---|---|---|---|
| Non-farm ownership | −0.5615 | −0.197 | 2.244 |
| Non-farm households | −0.988 | 0.489 | 1.204 |
| Population density | −0.344 | −0.332 | 2.067 |
| Distance | 0.791 | −0.220 | −1.528 |

was therefore designed to identify the pattern of continuous variation among zones, and to assign degrees of zone membership to each areal unit by using discriminant analysis.

Discriminant analysis (Tatsuoka, 1971; Klecka, 1975) identifies that linear combination (or combinations) of a set of independent variables which best discriminates among a number of previously defined groups. In this study, the groups consisted of the areal units assigned to the norms they most resembled, and the independent variables were the four previously described. Table 2 shows the results. Since three groups were present, it was possible to define a second combination of variables, or discriminant function, independently of the first. However, the relative magnitude of the respective eigenvalues shows that most of the variation between groups was accounted for by the first discriminant function.

The coefficients shown for each function identify the relationship between it and each variable. The major dimension is directly related to accessibility and inversely to the other variables. The most urban parts of the study area, the urban norm, score toward the low end on this dimension, while the rural parts score at the high end. The relative insignificance of the second discriminant function indicates that all three norms are approximately centred on the first dimension, where the 'transi-

TABLE 2

RESULTS OF DISCRIMINANT ANALYSIS

| Discriminant function | Eigenvalue | Percentage of trace | Chi square | Significance level |
|---|---|---|---|---|
| I | 9.57058 | 86.3 | 737.089 | 0.00 |
| II | 1.52230 | 13.7 | 207.701 | 0.00 |

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

| Variable | Function I | Function II |
|---|---|---|
| Non-farm ownership | −1.79751 | 1.01866 |
| Non-farm households | −0.59847 | −1.50987 |
| Population density | −0.81166 | 0.72455 |
| Distance | 0.52813 | 0.66436 |

tional' norm is situated between the 'rural' and 'urban' norms. The statistical significance of the second function, however, indicates that this dimension does increase the discrimination among the three norms.

The score of each areal unit on these two discriminant functions can be converted into a probability of group membership by assuming that each group's members have a multivariate normal distribution. When the grouping of areal units made prior to the discriminant analysis was compared with these probabilities, it was found that only one case had not been previously assigned to the group to which it most likely belonged.

For each areal unit, probabilities of membership in each of the three groups (urban, rural, and transitional) were computed. In all cases, one of the three probabilities was extremely low, thus allowing all three probabilities to be converted into a single index. When the two significant probabilities were for membership in groups I and II, rural and transitional, the areal unit was assigned a score between 1.0 and 2.0; similarly units between the transitional and urban norms were assigned scores between 2.0 and 3.0. A score of 2.2, for example, indicates an 80 per cent probability of being a member of the transitional group and 20 per cent probability of membership in the urban group; a score of 2.8 indicates the reverse.

These scores have been mapped in Figures 3 and 4. In Figure 4 the visual effect has been enhanced by squaring each value. The scores show the 'fuzziness' which has been the theme of this paper, although three plateaux are clearly evident in areas where resemblance to the three norms is strongest.

Three zones are apparent, as follows:

1. The urban zone. The city limits are illustrated on Figure 3 and the sharp distinction between the urban area (group III) and the next zone (group II) is clearly evident on both diagrams.

2. The rural-urban zone. From the base of the urban cliff a general plateau extends northwards in the central portion of London township. Within this plateau are depressions of less urbanized areas and a peak of higher values which corresponds to a settlement node with associated urban characteristics. The peak and pits

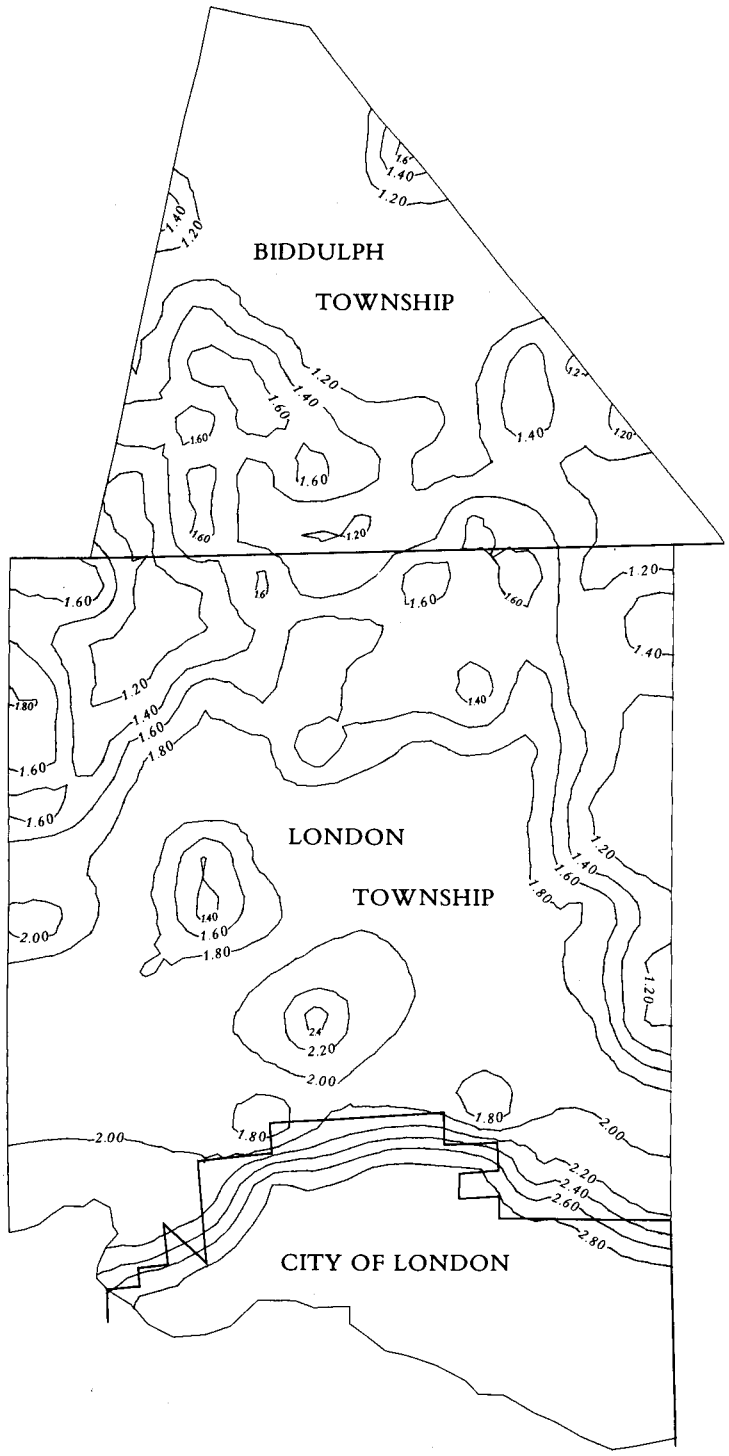


FIGURE 3. Contour map of membership scores.

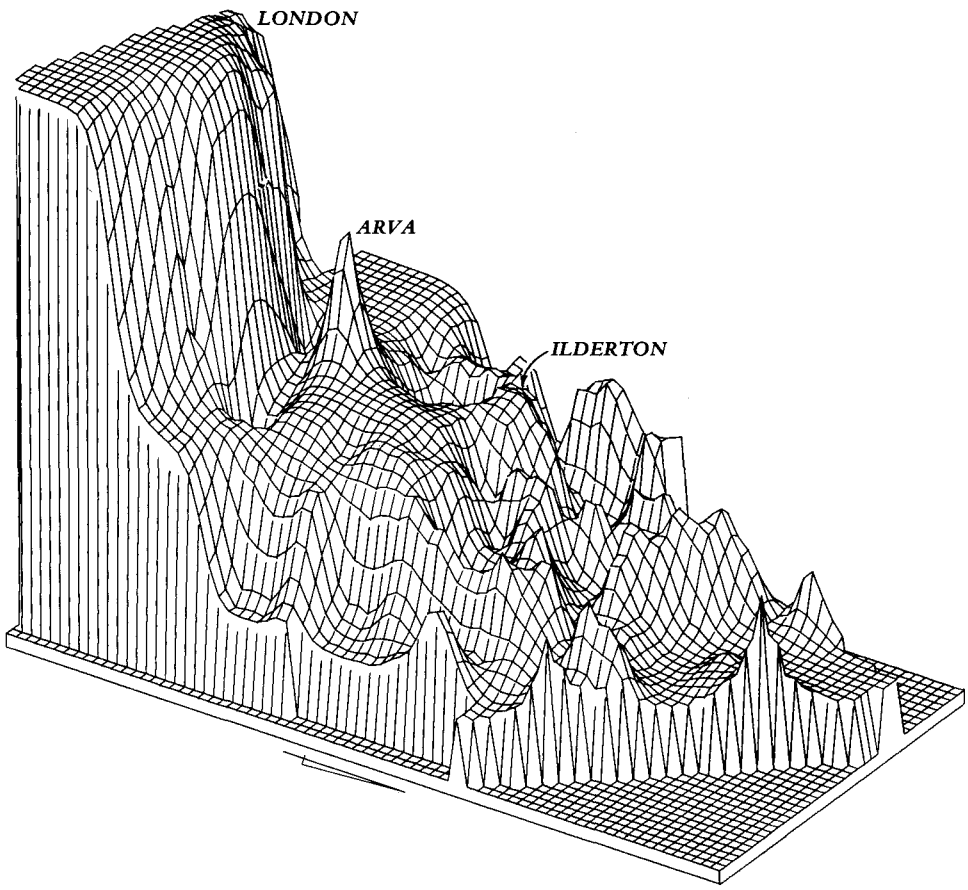


FIGURE 4. Three-dimensional surface of membership scores.

are most clearly distinguished in Figure 4. This plateau grades into an undulating zone which, as well as representing the boundary to the rural zone (group I) can, in itself, be recognized as a distinct transitional area. The subdivision of the rural-urban zone into an inner urbanized subzone (urban fringe) and an outer rural subzone can tentatively be made here. The undulating transitional zone contains urbanized village-based pockets mixed with more 'rural' areas and so constitutes the rural fringe, while the plateau area possesses characteristics of the 'urban' subzone. Although these subzones clearly exist, it must be recognized, in the context of the entire study area, that the differences between the urban and rural subzones within the fringe area are statistically small. They must therefore be considered together as *one* larger zone which lies spatially and characteristically between the urban and rural zones.

3. The rural zone. The final area, typified by a low incidence of urban characteristics, is restricted to the northern third of Biddulph township. It appears with two associated peaks of relatively low value. These can be interpreted as outliers of the advancing rural-urban frontier which lies to the south in the transitional zone.

## CONCLUDING REMARKS

The rural-urban fringe is a zone of more or less continuous transition between extremes. It follows, then, that it cannot be delimited or segmented by sharp boundaries, but that it can be mapped through the use of continuously varying measures of resemblance. Four variables were selected in this paper as independent indicators of the rural-urban transition, and it was found through cluster analysis that the study area tended to exhibit three general norms – rural, urban, and transitional. Discriminant analysis showed that for each areal unit resemblance to these norms could be determined by the unit's scores on two independent discriminant functions, which were then used to calculate probabilities of group membership. Finally, these probabilities were mapped in the form of a single score.

The results of the study can be affected by changes in several aspects of the research design. It was noted earlier that the results are sensitive to the choice of variables, although it seems unlikely that any variables could be found which would produce substantially different results. The delimitation of the study area is also important. If the study were carried further into the urban zone, the urban norm would slowly tend to become more extreme; conversely, if the inner boundary were moved outward, the urban norm would eventually be lost altogether. However, such changes would not markedly alter the mapping of resemblances, or the major conclusions.

The techniques used are relatively free of assumptions and limitations. Although the effects of sampling on cluster analysis are poorly understood, the problem does not arise here because the technique is not being used inferentially. The same is true of discriminant analysis, since again there is no question of generalizing the results of a sample to a larger population. The only important statistical assumption is in the assignment of probability scores in the discriminant analysis, where a multivariate normal distribution is assumed for the distribution of cases around each norm. An effective test of the validity of this assumption, and the severity of its effect on the results, would require a much larger number of cases, and must be left for future research. Even if the assumption proves severely incorrect, however, the score can still be used as a measure of resemblance, even if it is not strictly a probability. The term 'membership score' is suggested for this reason.

With these qualifications, the method is objective. It provides a clear indication of the internal structure of the fringe in terms of norms and resemblance zones, so that two independent analyses of the same area will always yield the same results, and so that the basis for an analysis can be communicated in a clear and precise manner.

More specifically, the study indicates the presence of a zone with characteristics intermediate between urban and rural. Within this zone, an area exists where transition is very slow if it occurs at all (the plateau in Figure 4), whereas there are also areas of rapid transition between this and the rural norm. The alternative, the possibility of more than one intermediate norm, is specifically denied. The study is essentially inductive, and it remains for research into the processes occurring in the fringe to provide a deductive rationalization.

## REFERENCES


ANDERBERG, M.R. (1973), *Cluster Analysis for Applications* (New York: Academic Press).

CANADA LAND INVENTORY (1968), *Soil Capability for Agriculture, 1:25,000 Series, Kitchener Sheet (40 p–0)* (Ottawa: Surveys and Mapping Branch). Department of Energy, Mines and Resources

COLEMAN, A. (1976), *Canadian Settlement and Environmental Planning* (Ottawa: Ministry of State for Urban Affairs).

COOPER, L. (1964), 'Heuristic methods for location-allocation problems,' *SIAM Review*, 6: 37–53.

FOUND, W.C. (1971), *Rural Settlement and Land Use Patterns* (Toronto: Macmillan).

GALE, S. (1972), 'Inexactness, fuzzy sets and the foundations of theoretical geography,' *Geographical Analysis*, 4: 337–49.

GERTLER, L.O. and HIND-SMITH, J. (1962), 'The impact of urban growth on agricultural land: a pilot study,' in *Resources for Tomorrow*, Supplementary Volume, Background Papers (Ottawa: Queen's Printer).

JOHNSTON, R.J. (1968), 'Choice in classification: the subjectivity of objective methods,' *Annals of the Association of American Geographers*, 58: 575–89.

KLECKA, W.R. (1975), 'Discriminant analysis,' in Nie, N.H., et al., *Statistical Package for the Social Sciences* (New York: McGraw-Hill).

MCKAIN, W.C. and BURNIGHT, R.G. (1953), 'The sociological significance of the rural-urban fringe from the rural point of view,' *Rural Sociology*, 18: 109–16.

MARTIN, L.R.G. (1975), *A Comparative Urban Fringe Study Methodology* (Ottawa: Environment Canada, Lands Directorate, Occasional Paper No. 6).

MYERS, R.R. and BEEGLE, J.A. (1947), 'Delineation and analysis of the rural-urban fringe,' *Applied Anthropology*, 6: 14–22.

REMEMANN, M.W. (1960), 'The pattern and distribution of manufacturing in the Chicago area,' *Economic Geography*, 36: 139–49.

RODEHAVER, M.W. (1946), 'The Rural-Urban Fringe: An Interstitial Area' (unpublished Ph.D. Dissertation, Department of Geography, University of Wisconsin).

RUSSWURM, L.H. (1971), 'Urban fringe and urban shadow,' in R.R. Krueger and R.C. Bryfogle (eds.), *Urban Problems: A Canadian Reader* (Toronto: Holt, Rinehart and Winston).

— (1973) *The Urban Fringe in Canada: Problems, Research Needs, Policy Implications* (Ottawa: Ministry of State for Urban Affairs).

TATSUOKA, M.M. (1971) *Multivariate Analysis* (New York: Wiley).

TROUGHTON, M.J. (1977), 'The rural-urban fringe: a challenge to resource management?,' *Proceedings, Geography Inter-University Resource Management Seminar*, 7: 43–75.

WISSINK, G.A. (1962), *American Cities in Perspective: With Special Reference to the Development of Their Fringe Areas* (Assen, Netherlands).