Urban Studies, Vol. 42, No. 1, 141–161, January 2005

Routledge
Taylor & Francis Group

Quantifying Urban Form: Compactness versus 'Sprawl'

Yu-Hsin Tsai

[Paper first received, November 2003; in final form, June 2004]

Summary. This paper develops a set of quantitative variables to characterise urban forms at the metropolitan level and, in particular, to distinguish compactness from 'sprawl'. It first reviews and analyses past research on the definitions of urban form, compactness and sprawl, and corresponding quantitative variables. Four quantitative variables are developed to measure four dimensions of urban form at the metropolitan level: metropolitan size, activity intensity, the degree that activities are evenly distributed, and the extent that high-density sub-areas are clustered. Through a series of simulation analyses, the global Moran coefficient, which characterises the fourth dimension, distinguishes compactness from sprawl. It is high, intermediate and close to zero for monocentric, polycentric and decentralised sprawling forms respectively. In addition, the more there is more local sprawl, composed of discontinuity and strip development, the lower is the Moran coefficient.

1. Introduction

One fundamental issue in the debate over metropolitan compactness versus 'sprawl' is in quantitatively distinguishing different degrees of compactness and sprawl. In recent years, a number of quantitative variables have been developed to characterise urban sprawl. However, some gaps still exist in the definitions of compactness and sprawl, and in appropriate quantitative variables.

'Sprawl', a loose term representing certain types of urban form, is commonly defined as low-density, leapfrog, commercial strip development and discontinuity (Ewing, 1997; Weitz and Moore, 1998; Galster *et al.*, 2001; Hess *et al.*, 2001; Malpezzi and Guo, 2001). Among three archetypal urban forms (Group, 1990) at the metropolitan level (metropolitan forms for short),[1] monocentric forms and decentralised sprawling forms are recognised as compact development and haphazard growth respectively, but there is little consensus on whether polycentric metropolitan form represents compactness or sprawl. In addition, a technical issue exists in identifying metropolitan centres since this process inevitably involves arbitrary judgment through adoption of political boundaries or density-based criteria. Or, in a broader sense, quantitative variables to distinguish properly between the three archetypical forms are still lacking.

This article aims to characterise quantitatively metropolitan form in general and to distinguish compactness from sprawl in particular. It starts with a literature review of definitions of urban form and definitions of compactness and sprawl. Secondly, it defines theoretically different dimensions of metropolitan form, accompanied with appropriate

Yu-Hsin Tsai is in the Department of Urban Planning, Feng Chia University, 100 Wenhwa Road, Seatwen, Taichung, Taiwan 407. Fax: +011 886 4 2451 1248. E-mail: yhtsai@fcu.edu.tw. *The author is most grateful to Professors John D. Nystuen and Jonathan Levine at the University of Michigan, Ann Arbor, and Professor Robert Cervero and Chris Amado at the University of California, Berkeley, and Patricia Huang for their advice and support regarding this research. The author also acknowledges valuable comments received from the anonymous reviewers.*

0042-0980 Print/1360-063X Online/05/010141–21
DOI: 10.1080/0042098042000309748

quantitative indexes. Then, different types of sprawl and compactness are linked to the appropriate metropolitan-form dimension that may characterise them best. A simulation analysis follows of various hypothesised compact and sprawling forms, intended to identify the best quantitative compact/sprawl indexes from among the proposed indexes. Finally, an empirical analysis is conducted of 219 metropolitan areas[2] with populations of less than 3 million[3] in the US to reveal metropolitan forms and degrees of compactness and sprawl. The data source for the simulation analysis is the 1995 Census Transport Planning Package (CTPP)—Urban Element (Bureau of Transportation Statistics, 1997).

2. Literature Review of Urban Form: Compactness versus Sprawl

Urban form can be defined as the spatial pattern of human activities (Anderson *et al.*, 1996) at a certain point in time. In a general sense, it can be classified into three categories: density, diversity and spatial-structure pattern.[4] The spatial structure of a metropolitan area, possibly defined as the overall shape, may characterise such land use phenomena as monocentric versus polycentric forms, centralised versus decentralised patterns and continuous versus discontinuous developments. In a broader sense, urban form may involve design category, such as block or site design (Cervero and Kockelman, 1997). In addition to the above land use characteristics, urban form, in a still broader sense, may concern transport spatial structure such as miles of expressway.

Urban form can be viewed from various geographical scales and classified into such levels as metropolitan area, city and neighbourhood. The reason for this classification is twofold. First, some urban form variables operate only at certain levels, such as the jobs–housing balance variable.[5] Secondly, urban form variables (such as density)[6] may carry different meanings at different levels and may differently affect human activities, such as travel behaviour. Because far less research and knowledge exist on urban form at the metropolitan level than at intermediate and low geographical levels, this research primarily focuses on metropolitan forms.

2.1 Definitions of Compactness and Sprawl

Prior to the literature review of existing compact/sprawl indexes, this section reviews the definitions of compactness and sprawl. Definitions for each generally citing several urban form dimensions are not universally agreed upon. Despite a lack of agreement, sprawl is often defined by four land use characteristics: low density; scattered development (i.e. decentralised sprawl); commercial strip development; and, leapfrog development (Ewing, 1997). The last three are spatial-structure-based phenomena of sprawl, as opposed to density-based sprawl. Commercial strip and leapfrog developments often occur in particular parts of a metropolitan area, such that the degree of derived sprawl of a whole metropolitan area depends on such factors as the size and degree of discontinuity of these local sprawl conditions.

Compactness also does not have a generally accepted definition. Gordon and Richardson (1997) defined compactness as high-density or monocentric development. Ewing's definition (1997) was some concentration of employment and housing, as well as some mixture of land uses. Alternatively, Anderson *et al.* (1996) defined both monocentric and polycentric forms as being compact.[7] Other definitions are more measurement-based. Bertaud and Malpezzi (1999) developed a compactness index, rho—the ratio between the average distance from home to central business district (CBD), and its counterpart in a hypothesised cylindrical city with equal distribution of development. Galster *et al.* (2001) described compactness as the degree to which development is clustered and minimises the amount of land developed in each square mile. Despite various definitions, one common theme is the vague concept that compactness involves the concentration of development.

2.2 Dimensions of Metropolitan Form: Compactness versus Sprawl

A great many dimensions and quantitative indexes of metropolitan form or sprawl/compactness have been proposed.[8] For the purposes of understanding the similarity and differences of the dimensions and quantitative indexes proposed, they are classified according to distinguishable characteristics of metropolitan form. Their suitability for distinguishing sprawl from compactness is based on the definitions reviewed previously.

Metropolitan size. The land area of an urban area was proposed as an index of sprawl, based on the idea that sprawl causes the consumption of more land than compact development (Hess *et al.*, 2001), which may be problematic since overall land consumption is highly associated with population. Consequently, land area may be regarded as a dimension to characterise metropolitan size rather than as an index of sprawl.

Density. Density, as a distinct dimension of metropolitan form, can characterise density-based sprawling patterns by measuring land consumption per capita (Galster *et al.*, 2001; Malpezzi and Guo, 2001; Hess *et al.*, 2001). Numerous density-based measurements such as maximum tract density and percentile-based density have been developed, but empirically have proved highly correlated with density.

Unequal distribution. This dimension of metropolitan form is defined as the degree to which development is concentrated in a few parts of a metropolitan area, regardless of high-density sub-areas being clustered or sparsely scattered (i.e. the third dimension of Figure 1, as opposed to dimension 4). There are close to 50 indexes that characterise unequal distribution (or inequality), among which the Gini coefficient is perhaps the most well known. Galster *et al.* (2001) incorporate two equal-distribution dimensions—concentration represented by such indexes as the coefficient of variation[9] and Shannon's entropy (i.e. Theil's or Delta index), and compactness.[10] In addition, Hess proposed the Gini coefficient as a sprawl index. Among all these indexes, Shannon's entropy, however, is found to be superior to others since it is not affected by size, shape and number of sub-areas in calculating values (Smith, 1975). Inequality distribution is widely applied to characterise sprawl (see, for example, Yeh and Li, 2001); however, its lack of spatial relationship measures casts doubt on its suitability.[11] Consequently, unequal distribution may better be conceived as a dimension of metropolitan form, rather than sprawl.

Centrality. Ideally, the centrality dimension of metropolitan form may be expected to characterise the degree of centralisation and decentralisation in general and, in particular, to distinguish among monocentric, polycentric and decentralised sprawl forms as an index of sprawl. Galster *et al.* (2001) measure the degree to which development is located close to the CBD[12] and gauge the extent to which an urban area is characterised by a monocentric form as opposed to a polycentric form. Because they are based on the assumption of monocentricity, however, the indexes used for measuring centrality have an inherent fundamental problem. In terms of characterising nuclearity (or polycentric form), various methods have been developed to identify employment centres, including density (Galster *et al.*, 2001), size (Cervero and Wu, 1998), contiguity of development (Giuliano and Small, 1991; Song, 1992) and the employment–population ratio (McDonald, 1985). However, these methods alone do not reveal information on the spatial relationship of identified metropolitan centres. Negative exponential density function is another mainstream method for characterising centrality as well as the overall distribution pattern of development and it is capable of characterising both monocentric and polycentric forms (Griffith, 1981; Gordon *et al.*, 1986; Small and Song, 1992).[13] Other non-parametric regressions, such as spline density functions (Anderson, 1982; Craig and Ng,

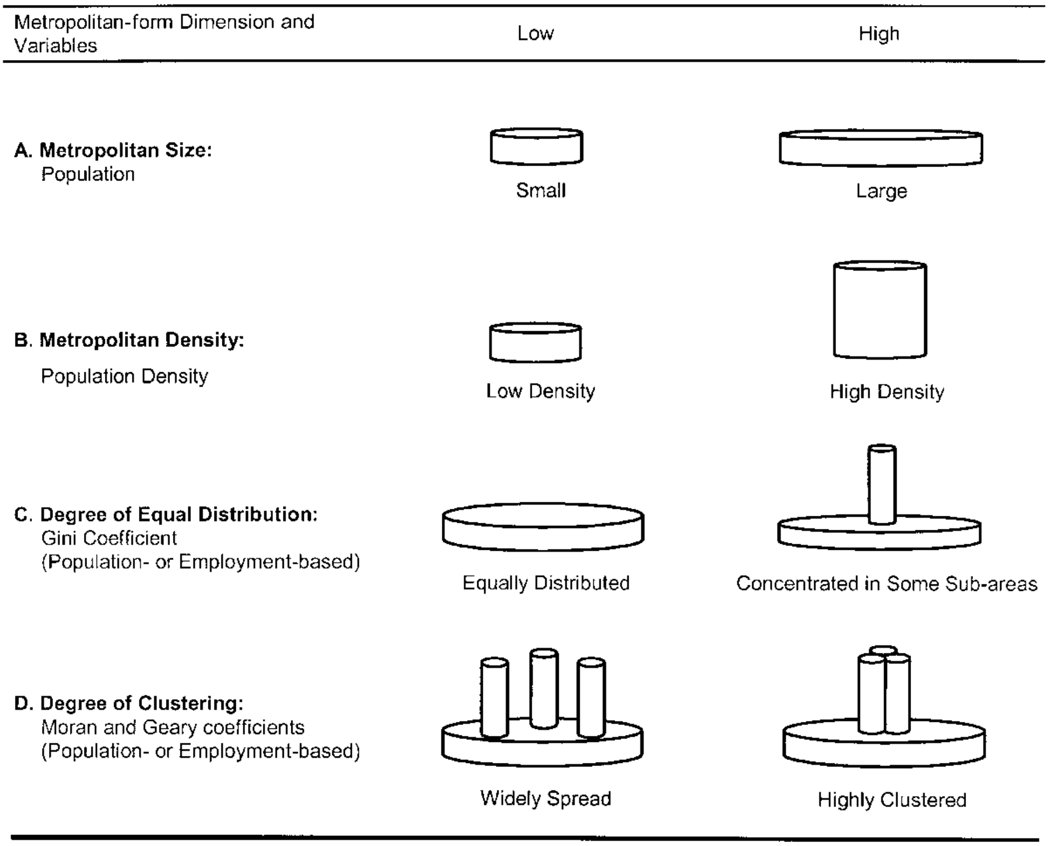


Figure 1. Four dimensions of metropolitan form.

2001; McMillen, 2001; Muñíz *et al.*, 2003) provide a more objective approach to identifying urban sub-centres than negative exponential density functions and can be applied to all metropolitan areas. Nonetheless, these models' capacity to distinguish compactness from sprawl is unclear since little research attempts to distinguish polycentric form from decentralised sprawl.

Continuity. The dimension of continuity, intended to characterise discontinuous developments, may need to be able to measure both the size of discontinuous developments and their distance from the main developed part of a metropolitan area. Galster *et al.*'s (2001) continuity dimension measures only the size of discontinuous developments. Malpezzi and Guo (2001) apply the R^2 of the exponential density function to measure the degree of discontinuity with R^2;[14] this is not totally correct since other development patterns, such as polycentricity and development radiating from metropolitan centres in different direction, may also contribute to a low R^2 value. The latter issue can be mitigated, but cannot be eliminated by segmenting the metropolitan area into axes or sub-regions (Zheng, 1991; Muñíz *et al.*, 2003), since the same problem still exists for each sub-region.

Other measures. The role of spatial autocorrelation (for example, Moran's I and Geary's C) to characterise metropolitan form in general, and to distinguish compactness from sprawl in particular, remains unclear.[15] Florida's anti-sprawl rule suggested using such accessibility indicators as average trip length, average commute time, VMT and vehicle hours travelled (Ewing, 1997). Ruck

(1993) applied an 'index of elasticity' measuring the extent to which cities are willing and able to expand their city boundaries. These measures do not directly characterise metropolitan form, or sprawl/compactness, but reflect effects influenced by not only urban form but also other factors such as transport infrastructure or arbitrary judgment.

3. Metropolitan-form Dimensions and Sprawl/Compactness Indexes

For the purposes of quantifying the three archetypal metropolitan forms and also accounting for local sprawl, this research undertakes a different approach from prior studies by first breaking down the metropolitan form into four distinct dimensions—metropolitan size, density, degree of equal distribution and degree of clustering—which together systematically portray a metropolitan form mathematically. Then, the dimension of degree of clustering potential to characterise the spatial-structure-based sprawling patterns, is examined in a simulation analysis to gauge its capacity to distinguish compactness from sprawl.

3.1 Four Metropolitan-form Dimensions and Quantitative Variables

A set of four dimensions of metropolitan form—metropolitan size, density, degree of equal distribution and degree of clustering—can be systematically identified in a metropolitan area, of which quantitative variables are developed for each dimension respectively. The first dimension, metropolitan size, is shown in Figure 1A with figurative examples of small and large metropolitan areas. Although land area might better characterise the examples in Figure 1A, population is more sensible in practical application since it is not affected by land consumption per capita, which is related to the second dimension (i.e. the reciprocal of density)—density; that is, in a statistical description, population is theoretically independent from density, but land area is not.

The second dimension, density, measures overall activity intensity in a metropolitan area (Figure 1B) and is the most commonly used variable in characterising urban form as well as *intensity-based* compactness/sprawl. Density by itself, however, does not address the pattern of activity distribution within a metropolitan area because it is unable to distinguish between different *spatial-structure-based* metropolitan forms.

Given metropolitan size and density, the third dimension probes the degree to which activities are equally or unequally distributed within a metropolitan area (Figure 1C). This dimension addresses the extent to which development is concentrated in a relatively small number of sub-areas. To characterise quantitatively the degree of equal distribution, indexes can be borrowed from those commonly used to measure inequality of income distribution, including the Gini coefficient. Among the many available indexes, research shows that relative entropy (an index derived from Shannon's entropy or Theil's index to rescale its values into the range from 0 to 1) is better than others because it is not affected by the number of sub-areas (Thomas, 1981). Shannon's relative entropy can be applied to measure inequality in population or employment distribution by spatial units in a metropolitan area, such as traffic analysis zones (TAZs). The relative entropy is defined as follows

$$\text{Relative Entropy} = \sum_{i=1}^{N} PDEN_i \times \log\left(\frac{1}{PDEN_i}\right) \Big/ \log(N)$$

where, $PDEN_i = DEN_i / \sum_{i=1}^{N} DEN_i$; $DEN_i =$ density of sub-area i; and $N =$ number of sub-areas.

The relative entropy, however, cannot be applied to data with a density value of zero, which in practice does exist (for example, parks). To overcome this problem, the sub-area boundary would need to be adjusted to avoid a zero density value. This adjustment, in practice, may not be appropriate since combining two different areas (such as residential areas and parks) may not make sense. Consequently, the Gini coefficient is selected for characterising this dimension. The Gini coefficient is applied to measure inequality of population or employment distribution by spatial units in a metropolitan area. Higher Gini coefficients (i.e. close to 1) mean that population or employment density is extremely high in fewer sub-areas. A Gini coefficient close to zero means that population or employment is evenly distributed in a metropolitan area. The Gini coefficient can be calculated as

$$\text{Gini} = 0.5 \sum_{i=1}^{N} |X_i - Y_i|$$

where, N is the number of sub-areas; X_i is the proportion of land area in sub-area i; and Y_i is the proportion of population or employment in sub-area i (Penfold, 2001).

Nonetheless, this dimension does not reveal the spatial relationship of high-density sub-areas—i.e. whether they are clustered or randomly distributed. This prevents the Gini coefficient from being an index of spatial-structure-based compactness/sprawl. For instance, given the same value of the Gini coefficient, it is still unclear whether a metropolitan form is more monocentric, polycentric or decentralised sprawl.

Given a metropolitan area with unevenly distributed population or employment revealed by the Gini coefficient, the fourth dimension, degree of clustering, is developed to estimate the degree to which high-density sub-areas are clustered or randomly distributed (Figure 1D). Theoretically aimed at measuring spatial relationship, this dimension potentially characterises spatial-structure-based sprawl and compactness—i.e. monocentric, polycentric and decentralised sprawling forms, and discontinuity and commercial-strip development. Whether it can achieve this goal essentially depends on the availability of quantitative indexes.

Theoretically, the global Moran and Geary coefficients, both measuring spatial autocorrelation, could estimate the level of clustering.[16] The Geary and Moran coefficients are similar, but slightly different in terms of mathematical definition and scaling of values. The Moran coefficient is defined as

$$\text{Moran} = \frac{N \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij}(X_i - X)(X_j - X)}{\left(\sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij}\right)(X_i - X)^2}$$

where, N is the number of sub-areas; X_i is population or employment in sub-area i; X_j is population or employment in sub-area j; X is the mean of population or employment; and W_{ij} denotes the weighting between sub-areas i and j.

The Moran coefficient ranges from -1 to $+1$, with a high positive value indicating that high-density sub-areas are closely clustered, a value close to zero meaning random scattering and a -1 value representing a 'chessboard' pattern of development.

The Geary coefficient is similar to the Moran coefficient, but instead of focusing on deviations from the mean, it examines deviations of each observation area relative to another. It is defined as

$$\text{Geary} = \frac{(N-1)\left[\sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij}(X_i - X_j)^2\right]}{2\left(\sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij}\right) \sum_{i=1}^{N} (X_i - X)^2}$$

In order to have a similar scaling with that of the Moran coefficient, the Geary coefficient, ranging between 0 and 2, can be transformed as

$$\text{Adjusted Geary} = -(\text{Geary} - 1)$$

Other than the typical spatial patterns revealed by the values of $+1$, 0 and -1, prior studies provide little knowledge on the connection between the Moran and Geary coefficients and the three archetypal metropolitan forms, as well as the two local sprawling patterns.

4. Simulation Analysis on Sprawl/Compactness Indexes

Based on the findings and issues identified in the previous section, this simulation analysis has two goals: first, to determine if the quantitative variables of the fourth dimension (i.e. the Moran and Geary coefficients) can distinguish between spatial-structure-based compactness and sprawl at both metropolitan and local scales, and, secondly, to discuss the relationship between the third (i.e. the Gini coefficient) and fourth dimensions with more examples. This simulation analysis concludes that the Moran coefficients are high, intermediate and low for monocentric, polycentric and decentralised sprawling metropolitan forms respectively, which will be lowered, if strip development and discontinuity occur.

4.1 *Differentiating Compactness from Sprawl*

Compactness may contain three situations: monocentric, polycentric (i.e. the spatial-structure-based definition) and high-density (i.e. the intensity-based definition) metropolitan forms. Sprawl may be composed of two conditions at the metropolitan level—decentralised sprawl and low density—and two conditions at the intermediate level—discontinuity and commercial-strip development. This simulation analysis centres on the spatial-structure-based compactness/sprawl since intensity-based compactness/sprawl can be easily measured by density. The Moran and Geary coefficients can potentially differentiate spatial-structure-based compactness from sprawl, but their interpretations are sometimes complicated (Anselin, 1995). To fill in this gap, several sets of hypothesised metropolitan forms are created to represent typical compact and sprawling forms. In each set, the related coefficients are predicted based on their theoretical definitions. Then the coefficients are calculated to find out if the prediction agrees with the calculated values and to determine if these variables can differentiate compactness from sprawl.

The hypothesised metropolitan forms discussed below are presented in both three-dimensional and two-dimensional charts; the height or tone represents activity intensity in each cell. Each set has two to three hypothesised forms representing different degrees of change in one dimension of metropolitan form, or similar forms with minor differences.

Monocentric, polycentric and decentralised sprawling metropolitan forms. The first simulation is intended to identify whether the Moran and Geary coefficients can distinguish among the three archetypal metropolitan forms. In Figure 2, A, B and C represent hypothesised monocentric, polycentric and decentralised sprawling forms respectively, given the same population, population density and degree of equal distribution of activities (i.e. the same Gini coefficient); they differ only in their clustering patterns (i.e. the fourth dimension).

Based on the definitions, the Moran and adjusted Geary coefficients are theoretically expected to be high, medium and low for the monocentric, polycentric and decentralised sprawling forms respectively, with the decentralised sprawling form close to zero.[17] This is because high-density cells are completely clustered in the monocentric form; they are randomly distributed in the decentralised sprawling form; and the polycentric form has some concentrations. The simulation results agree with the prediction, which shows that the Moran coefficients of the monocentric, polycentric and decentralised sprawl forms are high, intermediate and low respectively (i.e. 0.13, 0.06 and -0.01).[18] The Geary coefficient is not as good as the Moran coefficient in distinguishing among these three forms primarily because the values may lead to an incorrect interpretation of urban forms. For example, the monocentric form has an adjusted Geary coefficient close to zero, which indicates random dispersion. Also, the polycentric form's value is -0.04, which represents negative spatial autocorrelation.

Finally, the weighting in calculating both the Moran and Geary coefficients is the inverse distance between the centroids of two cells, rather than the most commonly used contiguity criteria (i.e. 0 for discontinuous cells and 1 for continuous cells). Distance-based criteria are more sensitive and accurate in characterising metropolitan forms than contiguity criteria, as demonstrated by the two hypothesised forms in Figure 3 (A and B). The different degree of discontinuity of the two forms can be distinguished by an inverse-distance-based weighting where both the Moran and Geary

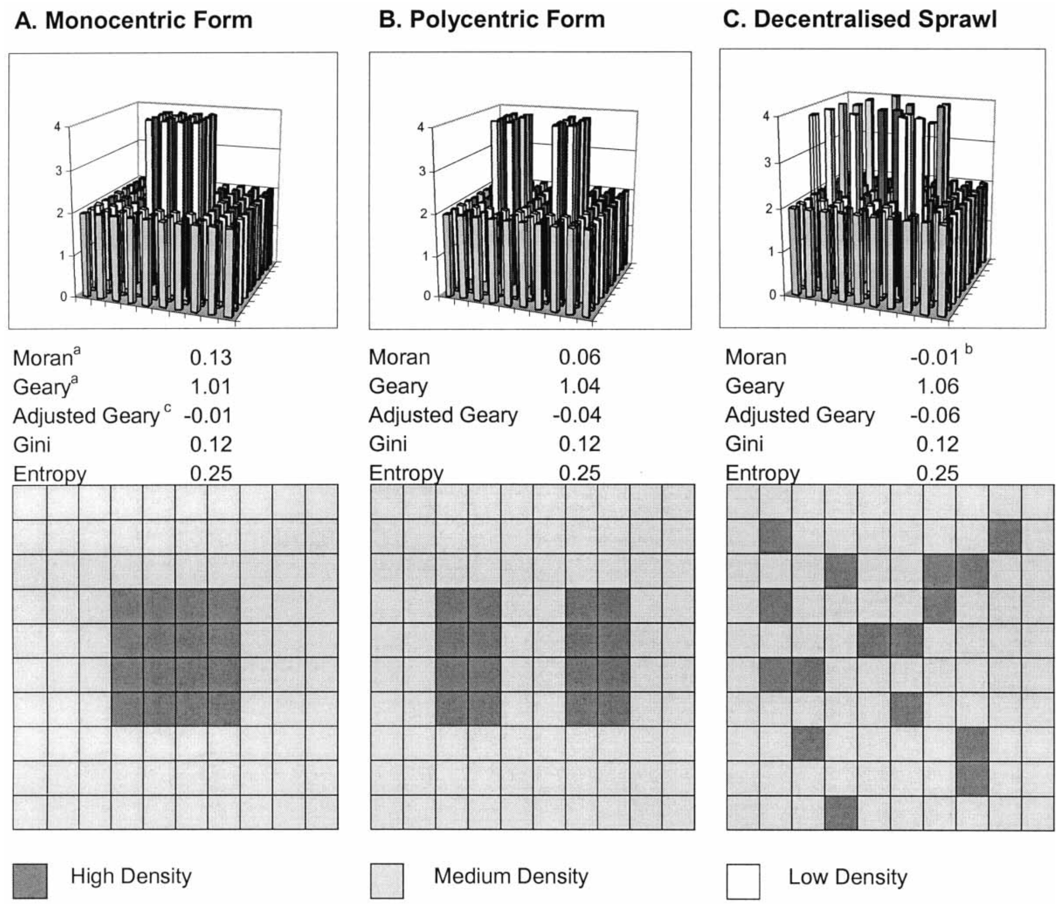
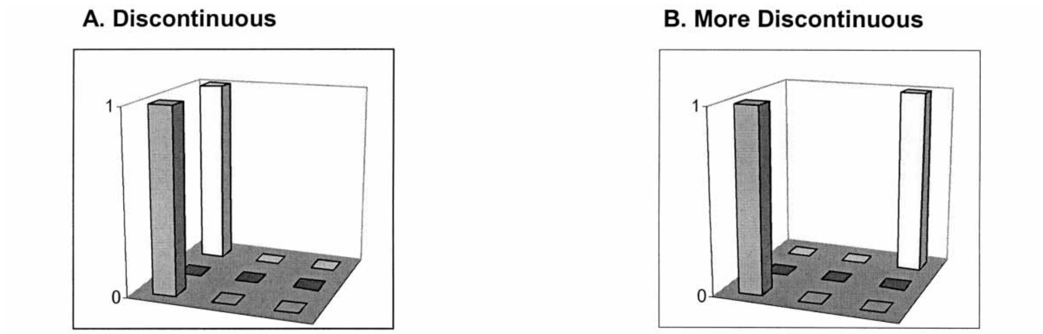


Figure 2. Hypothesised monocentric, polycentric and decentralised sprawling forms. *Notes*: a. The weighting is the inverse distance between the centroids of two cells; b. (-1/number of cells) = randomly scattered. In this case, the Moran coefficient = -0.01 means randomly scattered; c. The scale of the Geary coefficient is adjusted to become similar to the Moran coefficient. That is, values close to $+1$ mean high clustering; values close to zero mean random scattering; and negative values mean a chessboard pattern.

coefficients have different values. In contrast, connectivity-based weighting leads to the same values for both coefficients for these two hypothesised forms due to its consideration of only immediate, neighbouring cells.

Local sprawl. This simulation is intended to examine how the Moran coefficient measures the two local sprawling patterns—i.e. strip development and discontinuity. In Figure 4, A, B and C represent a compact form, a compact centre with two extending strips of development and four strips of development respectively, given the same population, density and Gini coefficient. The Moran coefficients are expected to be high, medium and low for the compact, some-strip and more-strip developments respectively, due to high, intermediate and low levels of clustering. The simulation results agree with the prediction, with the Moran coefficients equal to 0.16, 0.14 and 0.12 respectively. This simulation shows that the value of the Moran coefficients is lowered due to local strip developments and, the more strip developments, the lower is the Moran coefficient.

This finding also applies to the local sprawl of discontinuous developments. In Figure 5,



A. Discontinuous

	Weighting Method	
	Connectivity	Inverse Distance
Moran	-0.143	-0.109
Geary	0.857	0.919

B. More Discontinuous

	Weighting Method	
	Connectivity	Inverse Distance
Moran	-0.143	-0.143
Geary	0.857	0.949

Figure 3. Moran and Geary coefficients, connectivity-based vs inverse-distance-based weighting.

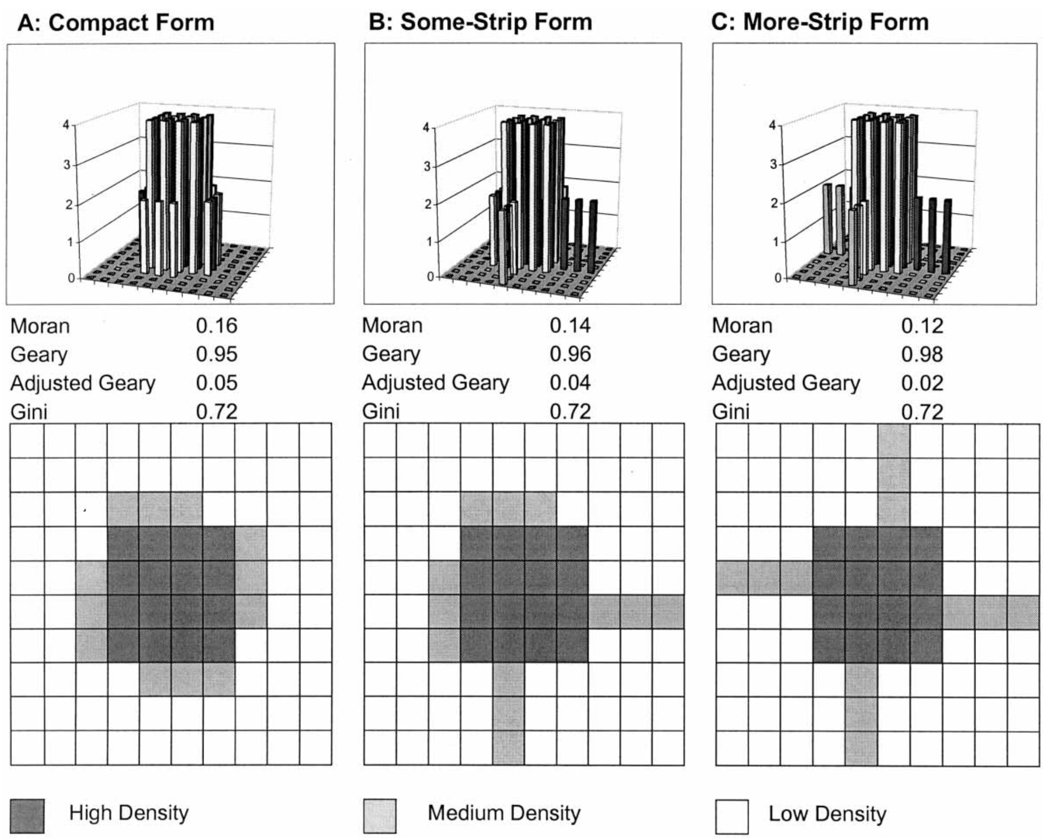


Figure 4. Hypothesised compact, some-strip and more-strip forms.

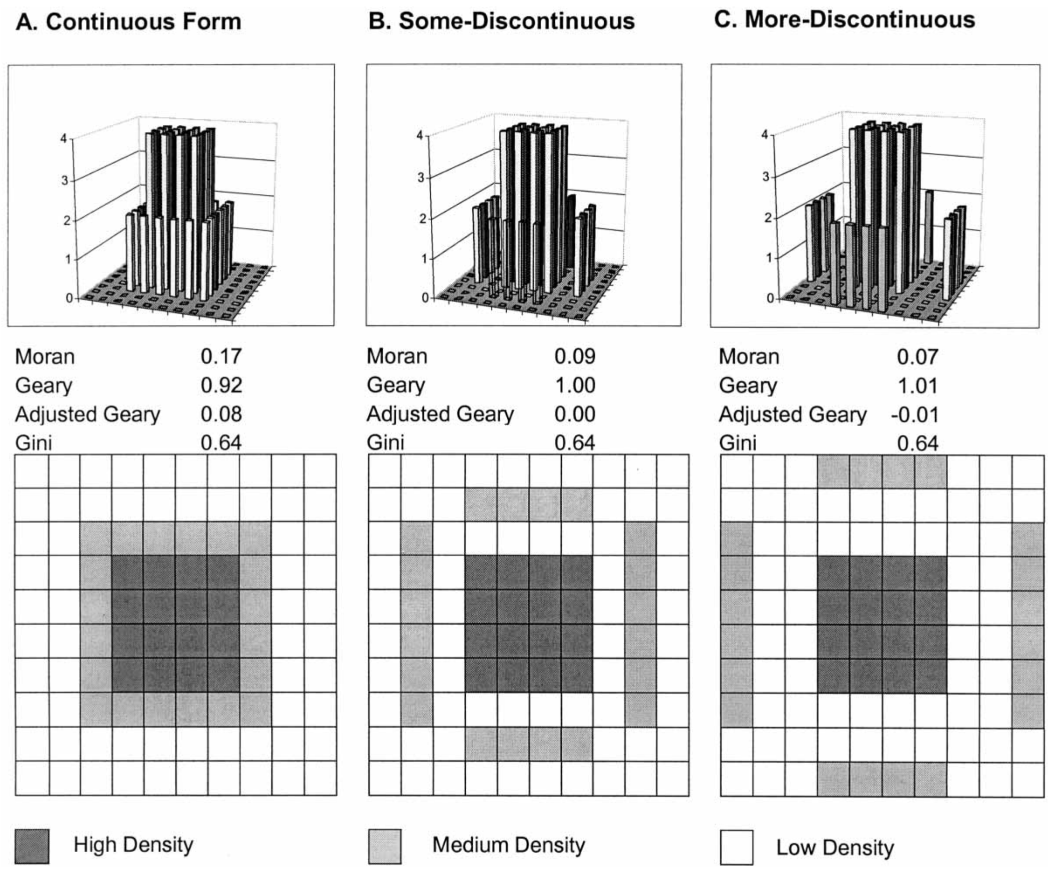


Figure 5. Hypothesised urban forms with different degree of continuity.

A, B and C represent a continuous monocentric form, a form with discontinuous development and a form with the same size of discontinuous development located farther away from the urban core. Their Moran coefficients are expected to be high, intermediate and low respectively and the simulation results correspond (i.e. 0.17, 0.09 and 0.07 respectively). It is noticeable that the gap between continuous and some-discontinuous development is larger (0.17 → 0.09) than that of the above strip-development cases (0.16 → 0.14). The Geary coefficient can also distinguish between the three forms, but again the values of the adjusted Geary coefficients for the two discontinuous forms (i.e. 0.00 and −0.01) may mislead the interpretation of the urban forms.

Based on the above simulation, the Moran coefficient can distinguish compactness from sprawl; the Geary coefficient, although having a similar definition to the Moran coefficient, proves less useful. From the perspective of overall metropolitan structure, the Moran coefficients are high, intermediate and low for monocentric, polycentric and decentralised sprawling forms. Given an overall metropolitan form, the Moran coefficient will be lowered if local sprawl occurs; the more strip developments and/or the more discontinuous the discontinuous developments, the lower the Moran coefficients will be. The Moran coefficient characterises the different components of compactness/sprawl with one index, which differs from previous research—characterising all components separately and then mathematically synchronising all the indexes into one sprawl index. The Moran coefficient could be superior to other methods since it does not need to weight arbitrarily the various sprawl indexes to be combined into one final sprawl index.

More about the Gini and Moran coefficients. The previous section shows how the Moran coefficient can distinguish compactness from sprawl; however, there are limitations in this regard. To help describe the limitations, this section first shows how the Moran coefficient can be predicted visually, based on three-dimensional charts (to a certain extent) as opposed to two-dimensional charts for the Gini coefficient. Based on these findings, the limitations of the Moran coefficient can be better described.

Figure 6 represents monocentric forms with the same populations and population densities, but with various population distributions in the urban centres and surrounding areas. The Gini coefficients are predicted to be low, intermediate and high respectively, because A in Figure 6 has the most evenly distributed population, and C the least, as observed in the three-dimensional charts. These predictions are supported by the simulation analysis.

In terms of the Moran (and Geary) coefficients, conventional knowledge seems incapable of predicting the values for these cases. This simulation finds that the Moran coefficients are the same for these three cases because the clustering patterns look the same in the two-dimensional charts in terms of the

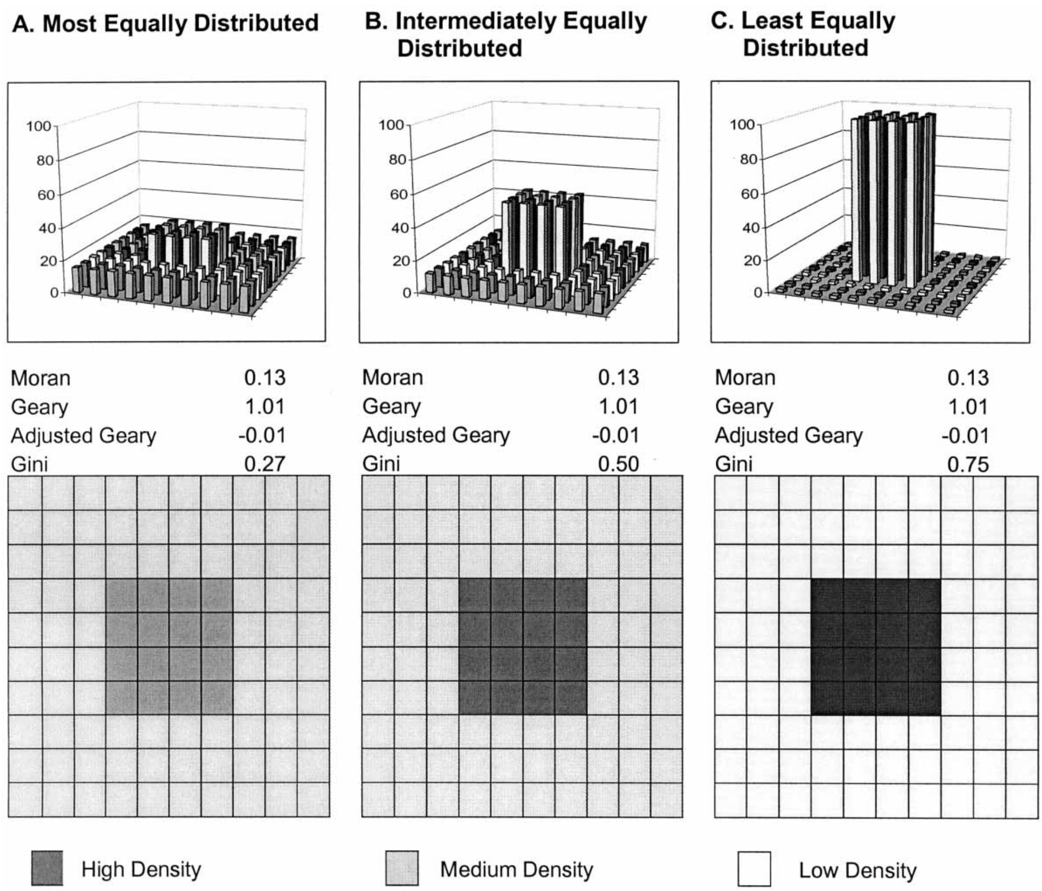


Figure 6. Hypothesised urban forms with different degrees of distribution.

boundaries separating the urban centres and surrounding areas (regardless of the difference between them in terms of density).[19]

The simulation in Figure 7 shows that the Moran coefficient cannot distinguish the hypothesised polycentric form from leapfrog developments. Based on the same clustering patterns observed in the two-dimensional charts, the polycentric form (A in Figure 7) and leapfrog development (B in Figure 7) are predicted to have the same Moran coefficient, which is proved true by the simulation analysis. This simulation shows the Moran coefficient's inability to distinguish between these hypothesised polycentric and leapfrog developments. These examples, in fact, highlight the needs for density and the Gini coefficients to distinguish their metropolitan-form differences.

5. Empirical Data Analysis

Following the preceding simulation analysis, this section first addresses the third and fourth metropolitan-form variables (i.e. the Gini and Moran coefficients) for US metropolitan areas. Unlike the first and second variables (i.e. size and density), they have barely been explored. Secondly, the relationship between the four variables is explored. The final part presents empirical, figurative examples of high, intermediate and low Moran coefficients.

5.1 Gini and Moran Coefficients of US Metropolitan Areas

This section presents the Gini and Moran coefficients for 219 metropolitan areas—selected for their data availability—with populations of less than 3 million. The two coefficients

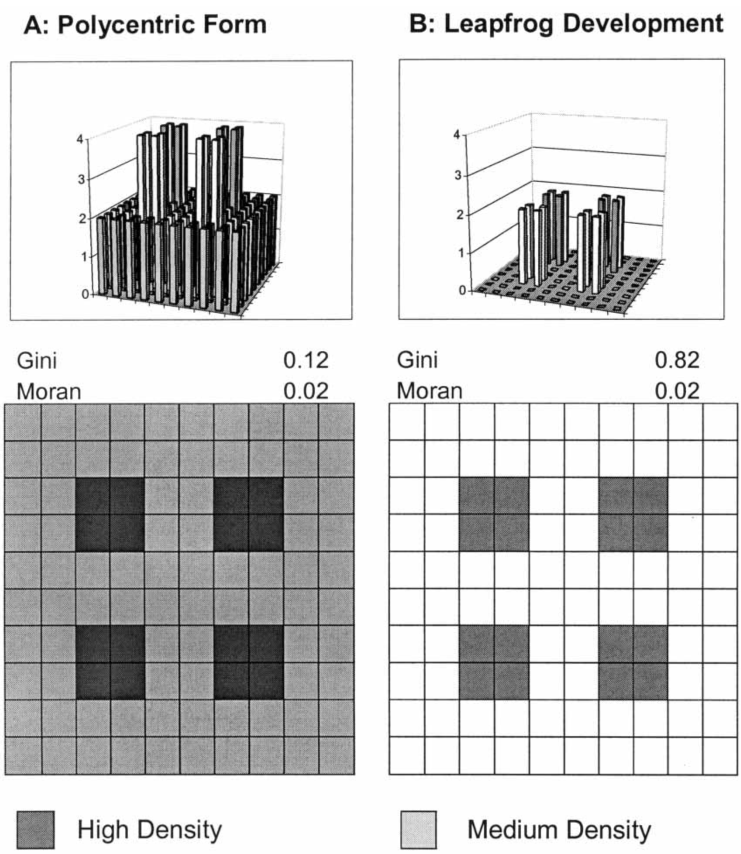


Figure 7. Polycentric and leapfrog development.

are calculated for both population and employment distribution. Population and employment data were obtained from the 1995 Census Transportation Planning Package (CTPP)—Urban Element, disaggregated to the level of Traffic Analysis Zones (TAZs) or blocks. For calculating the Moran coefficient, an inverse distance-based weighting is applied with distance defined between the centroids of two TAZs or blocks.

The Gini coefficients of the 219 metropolitan areas show that employment is generally more concentrated (i.e. less evenly distributed) than population; the average employment-based Gini coefficient is 0.87, which is higher than the 0.67 for population. And the concentration of employment varies less across different metropolitan areas than population. The employment Gini coefficient for the majority—90 per cent—of the metropolitan areas (B in Figure 8) spans a smaller range (0.80–0.99) than does the Gini coefficient for population: 80 per cent between 0.50 and 0.79 (A in Figure 8).

The population Moran coefficients span between −0.1 and 0.4, with an average of 0.11 (A in Figure 9). The majority—60 per cent—of metropolitan areas have Moran coefficients between 0.10 and 0.19, meaning fairly compact development. Only 5 per cent of development is highly compact (between 0.20 and 0.39). Another large proportion (33 per cent) has a more sprawling form with Moran coefficients between 0 and 0.09. The clustering pattern of employment is quite split. The average employment Moran coefficient is 0.11. Slightly more than half of the metropolitan areas exhibit more compact forms (i.e. the Moran coefficient is no less than 0.1), of which 6 per cent are very compact (i.e. 0.20 ~ 0.39) (B in Figure 9) while the remaining 45 per cent are more sprawling with the Moran coefficient between 0 and 0.09.

It is hard to define the exact range of Moran coefficients of monocentric and polycentric forms for two reasons. First, this product can be obtained only if the real metropolitan forms were as distinguishable as the three archetypical forms. In addition, two forms might need to have exactly the same cores and surrounding areas to have the same Moran coefficients. Secondly, local sprawl will reduce the Moran coefficients, which increase the overlapped ranges of the Moran coefficients of different forms. Under these circumstances, what can be derived from the Moran coefficient is that the ideal monocentric, polycentric and decentralised sprawling forms may have high, intermediate and close to zero Moran coefficients respectively.

5.2 *The Relationship between Four Metropolitan-form Variables*

From the geometric perspective alone, the four metropolitan-form dimensions are independent of each other except that clustering of high-density sub-areas (i.e. high Moran coefficients) occurs only when development is unevenly distributed (i.e. Gini coefficients are not equal to zero). However, considering urban development phenomena, they could be associated. For example, a population

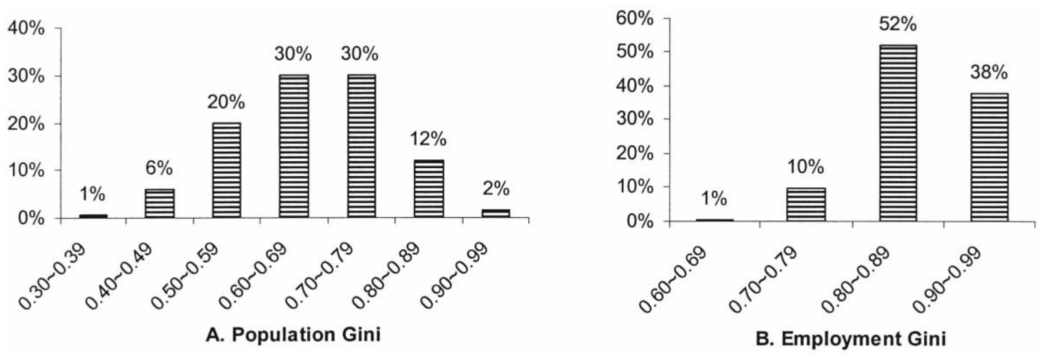


Figure 8. Gini coefficients, by population and employment.

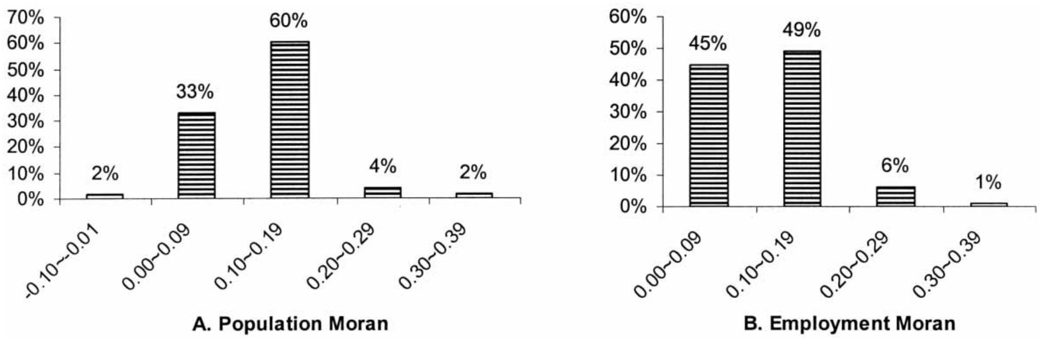


Figure 9. Moran coefficients, by population and employment.

increase, all else being equal, will cause a rise in density (Alonso, 1964); rail transit, normally existing in large metropolitan areas, may lead to more dense transit-oriented development; and high-density business sub-areas may cluster spatially to a certain degree due to agglomeration effects. The following addresses these relationships based on empirical correlation analysis.

First, correlation analysis shows that the larger a metropolitan area is, the higher are the density and degree of clustering of high job density sub-areas. Table 1 shows that population has a statistically positive relationship with density—the Pearson correlation coefficients of population with population and employment densities are 0.51 and 0.472 respectively—and the job-based Moran coefficient (the Pearson correlation coefficient is 0.174). In addition, larger metropolitan areas tend to have a relatively even distribution of employment possibly due to decentralisation of employment. Table 1 shows that population is negatively associated with the job-based Gini coefficient (the Pearson correlation coefficient is −0.375). The above findings together may explain why more dense metropolitan areas tend to have relatively low concentrations of employment in certain sub-areas; the Pearson correlation of population density with the employment-based Gini coefficient is −0.607. However, the size of a metropolitan area is not statistically associated with the spatial distribution

Table 1. Correlation between the four metropolitan-form variables ($N = 219$)

	Size		Density		Gini		Moran	
	Population	Job	Population	Job	Population	Job	Population	Job
Size								
Population								
Job	0.979***							
Density								
Population	0.510***	0.520***						
Job	0.472***	0.524***	0.943***					
Gini								
Population	−0.022	0.002	−0.097	−0.028				
Job	−0.375***	−0.413***	−0.607***	−0.620***	0.347***			
Moran								
Population	0.013	0.026	0.039	0.038	−0.093	−0.109		
Job	0.174***	0.187***	0.188***	0.167**	−0.052	−0.207***	0.132	

*** indicates a significant correlation coefficient at the level of 0.01; ** indicates a significant correlation coefficient at the level of 0.05.

of population (i.e. the third and fourth dimensions).

Secondly, more dense metropolitan areas are moderately associated with a higher degree of clustering of high-density sub-areas; the Pearson correlation coefficients of population and employment densities with the employment-based Moran coefficients are 0.188 and 0.167 respectively. Finally, the relationship between the concentration of population or employment in some sub-areas and the degree of clustering in high-density sub-areas is either moderately negative or statistically insignificant, as reflected by the correlation between the Gini and Moran coefficients.

5.3 *Practical Cases of High, Intermediate and Low Moran Coefficients*

The preceding simulation analysis displayed how the Moran coefficient distinguishes among some hypothesised, simplified compact and sprawling forms which are not likely to exist in the real world. Hence, the following practical cases have been selected to demonstrate the application of the Moran coefficients. For purposes of simplicity, all cases had 1990 populations of less than 20 000 persons. The results show that a metropolitan area with a higher Moran coefficient will tend to have a more consolidated, high-density core and less local sprawl, and vice versa. It is difficult, however, to predict the Moran coefficient by observing a map, primarily due to minor differences between, and complexity of, the clustering patterns of metropolitan areas.

The first case is Athens, Georgia, which displays a high employment-based Moran coefficient (0.25). This high coefficient seems to be a reflection of its monocentric form with very high employment density—greater than 7000 jobs per square mile—in the core (Figure 10).

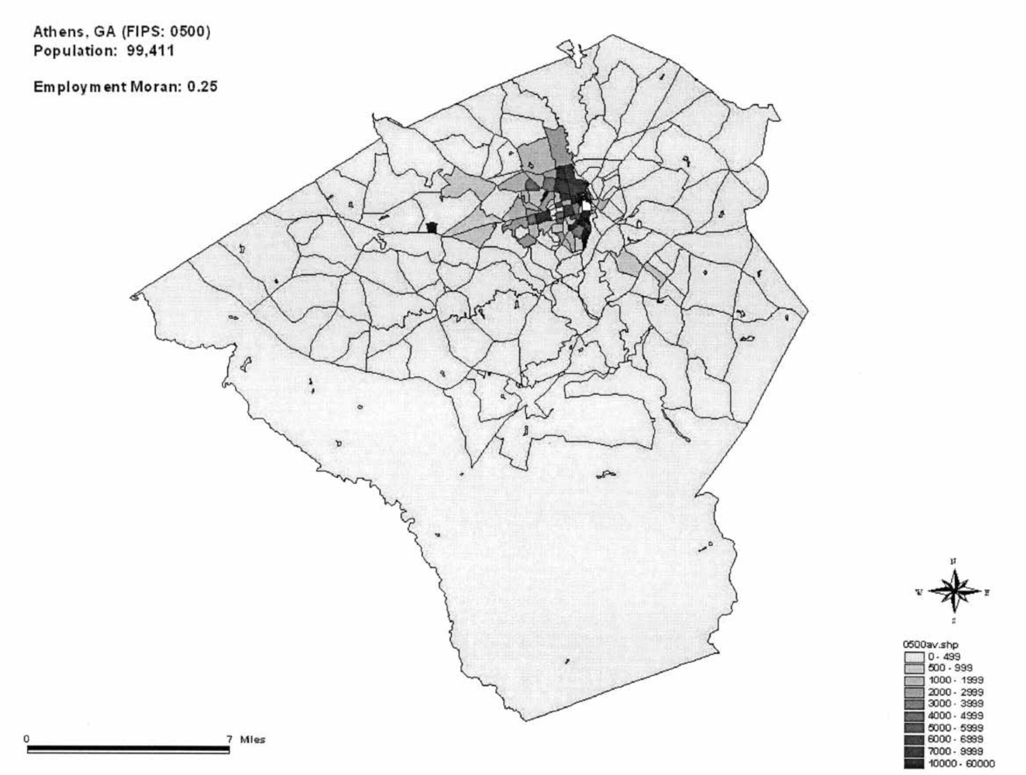


Figure 10. Metropolitan area with a high employment-based Moran coefficient: Athens, GA.

There are some TAZs with low-to-intermediate density (1000–2000 jobs per square mile) in places located discontiguously from the centre; this form of development may slightly decrease the Moran coefficient.

The second case is Albany, Georgia, with an intermediate employment-based Moran coefficient (0.10). It has a high-density centre with one commercial strip along the north-west corridor and some scattered developments in the east (Figure 11). Some of the employment centres are discontiguous from the core. This metropolitan form looks less compact than that of Athens, GA.

The final case is Kokomo, Indiana, with a low employment-based Moran coefficient equal to 0.02, which indicates a no-pattern form. However, its metropolitan form looks like a monocentric pattern (Figure 12). Although commercial development in the core is not very clustered and density may be relatively low compared with the above cases, its low Moran coefficient may still be beyond expectation. This case may demonstrate the merit of a quantitative measure for capturing differences between metropolitan forms in terms of compactness and sprawl that may be too complex to estimate by observation alone.

6. Summary and Conclusion

Metropolitan form can be analysed as four distinguishable dimensions—size, density, degree of equal distribution and degree of clustering—represented by population, population density, the Gini and Moran coefficients respectively. The Moran coefficient is capable of distinguishing compactness from sprawl: for overall metropolitan form, the more compact the metropolitan area, the higher is

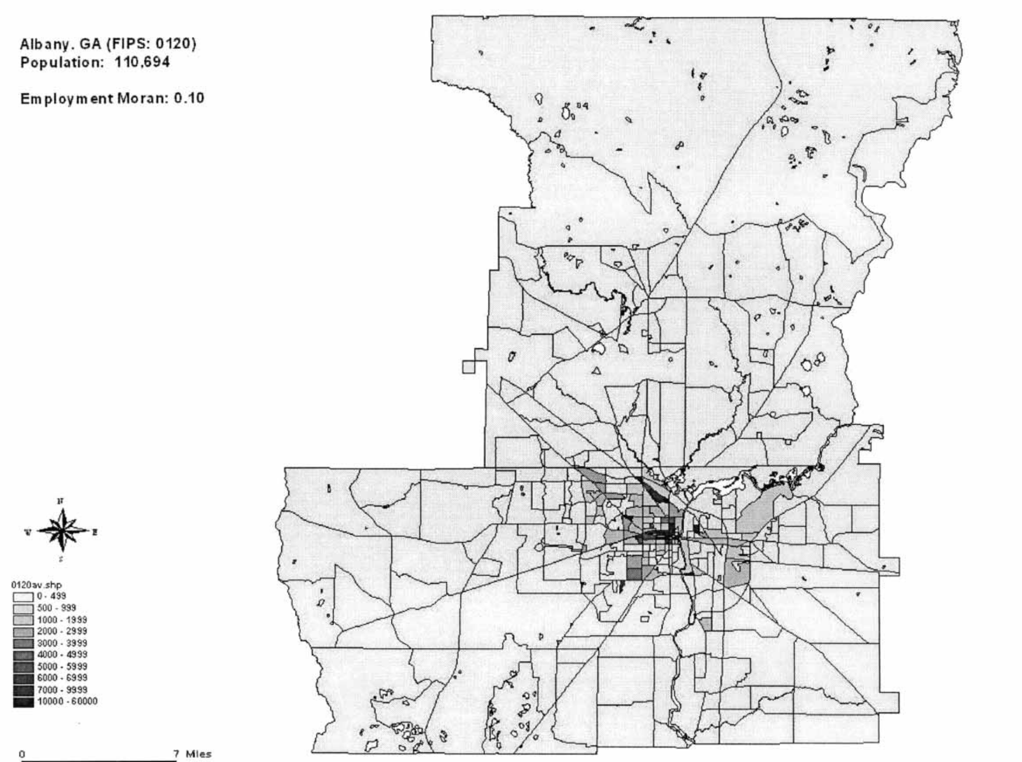


Figure 11. Metropolitan area with an intermediate employment-based Moran coefficient: Albany, GA.

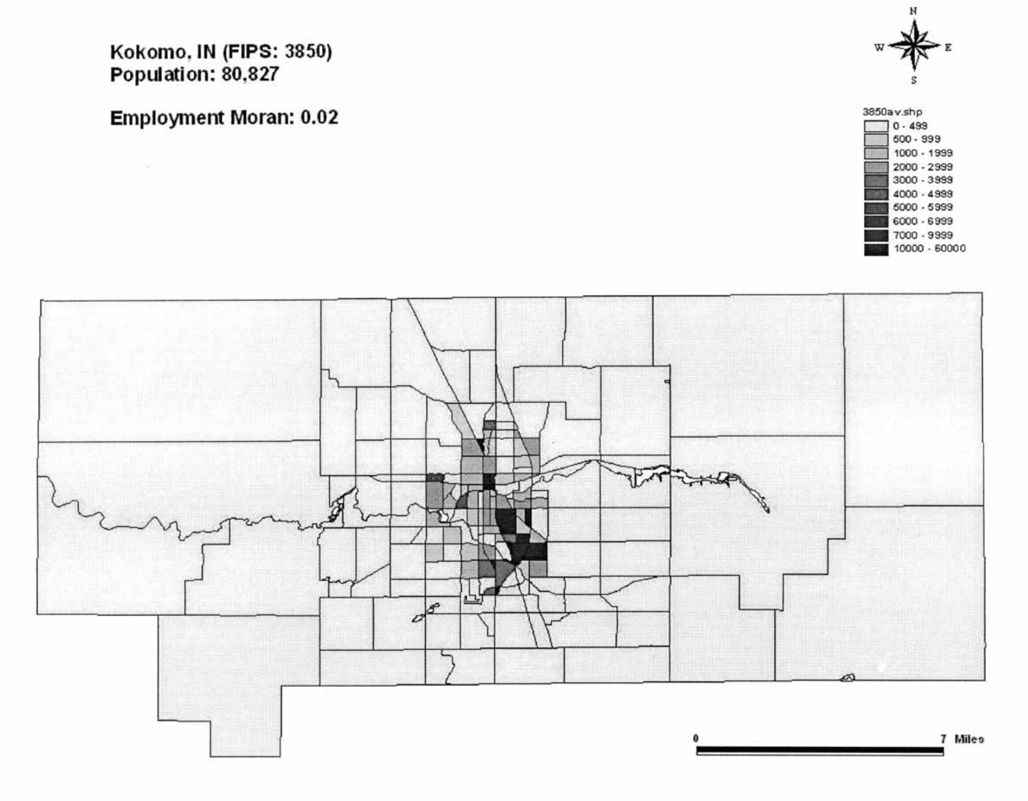


Figure 12. Metropolitan area with a low employment-based Moran coefficient: Kokomo, IN.

the Moran coefficient; the Moran coefficients of monocentric, polycentric and decentralised sprawling forms are high, intermediate and close to zero respectively; and, the local sprawl, comprising leapfrog and strip developments, will lower the value of the Moran coefficients.

The empirical data on the 219 metropolitan areas with populations of less than 3 million reveal that population and employment are concentrated in some sub-areas in a metropolitan area (represented by the Gini coefficient); the concentration of employment is higher than the concentration of population. Further considering the spatial relationship between the high-density sub-areas (characterised by the Moran coefficient), two-thirds of the metropolitan areas have populations that are spatially fairly clustered (i.e. compact), but only half have compact employment distribution.

The advantages of the quantitative metropolitan-form variables are many, including their convenience in the application of quantitative analysis, providing a summary spatial description of metropolitan areas without recourse to maps and the capability of capturing minor differences between, and the complexity of, urban forms. The disadvantages of the quantitative metropolitan-form variables are, first, the variables do not provide as clear an image as can maps. Secondly, it is hard to catch the original condition based on metropolitan-form variables. For example, low Moran coefficients may imply either a discontinuous development or decentralised sprawl. In other words, two metropolitan areas with the same Moran coefficient are likely to represent two very different forms.

This paper also finds that, for calculating the Moran coefficients, the distance-based weighting criteria serve better than contiguity-based criteria in terms of the capability of characterising different forms. In addition, inconsistency in these coefficients may occur due to the inconsistency of partitioning metropolitan areas involving scale effects,[20] zonal effects[21] and the shape of sub-areas. Although these effects could not be avoided altogether, the scale effects could be diminished to a certain extent by partitioning metropolitan areas at the same level, say TAZs or census tracts. However, scale effects are still likely to exist since the ways of defining sub-areas may vary from one metropolitan area to another; the averages and variances of sub-area sizes are likely to be different between metropolitan areas. It may be impractical to repartition metropolitan areas since the quality of repartitioned data may become worse. A more efficient way to diminish scale effects is through statistical adjustment such as regression analysis.[22]

Another fundamental issue in calculating metropolitan-form variables is the definition of a metropolitan area. First, since the census-bureau-defined metropolitan areas are county-based, different metropolitan areas contain rural areas to different degrees. This inconsistency of metropolitan boundaries may bias the metropolitan-form variables. To reduce this bias, metropolitan boundaries may be redefined with density and contiguity criteria (Galster *et al.*, 2001; Tsai, 2001). In addition, within a metropolitan area, undeveloped sub-areas (such as rivers, mountains) may be excluded in calculating the metropolitan-form variables to reflect only the land use policy; in this way, ‘urban sprawl’ caused by natural constraints can be excluded. In contrast, if the purpose of characterising metropolitan form is to evaluate the effect of metropolitan form, say, on travel behaviour, the natural landscape may need to be included since both man-made and natural settings may affect travel behaviour.

Notes

1. Central, nodal and spread are used for monocentric, polycentric and decentralised sprawling in the IBI report.
2. Only those metropolitan areas with complete data were adopted.
3. Larger metropolitan areas are excluded due to the limited computational capacity of the personal computer.
4. Density measures the degree of activity intensity. Diversity refers to spatial scale or grain at which different land uses interact, such as land-use mixing (Cervero, 1996; Cervero and Kockelman, 1997; Douglas, 1998) and jobs–housing balance (Cervero, 1996; Levine, 1998).
5. Jobs and housing are by definition balanced at the metropolitan level and possibly extremely imbalanced at a more local level such as the block.
6. The meaning of density at the metropolitan level, as opposed to the neighbourhood level, is less clear (Miller and Ibrahim, 1998; Burchell *et al.*, 1998). Most previous research has focused on the latter, and only a few have applied the former, such as Gordon *et al.* (1989) and Newman and Kenworthy (1989).
7. Anderson *et al.* (1996) used different terms. They used compact centralised for centralised compact (i.e. monocentric) development; and compact decentralised for decentralised compact (polycentric) development.
8. Research with more comprehensive typologies of metropolitan form or sprawl includes studies employing eight distinct dimensions of sprawl with corresponding quantitative indexes for each (Galster *et al.*, 2001), seven categories of quantitative metropolitan-form variables (Malpezzi and Guo, 2001) and seven indexes of four dimensions of sprawl (Hess *et al.*, 2001).
9. The coefficient of variation is a general statistical index measuring the spread of the distribution.
10. Compactness is a similar dimension but focuses on a smaller geographical area, in which the index is calculated by taking the average of the values of the inequality distribution indexes of each sub-area in an urbanised area.
11. For instance, three archetypal forms—monocentric, polycentric and decentralised—will share the same values of entropy if their population distribution are the same (a figurative example is available in Figure 2).
12. Two quantitative variables are proposed to characterise centrality: one is the average

distance of development to the CBD; the other is the ratio of accumulated actual development in the CBD to accumulated development in a conceptual concentric format where development moves progressively outwards from the CBD. Other similar indexes include fringe-to-centre ratio of development, fringe-to-centre land per capita ratio (Hess *et al.*, 2001) and rho (i.e. a compactness index) (Allen *et al.*, 1993; Malpezzi and Guo, 2001).

13. The models of exponential density functions are developed based on the economic model that density function is the trade-off behaviour between accessibility and density (or housing values), assuming that employees are concentrated in the CBD or employment centres, where polycentric models are additions of layers of concentric rings of all CBD and employment centres. The parameters of polycentric models measure the peak densities and declining rate of density from metropolitan centres to the surrounding areas which, altogether, characterise the overall shape of a metropolitan area. On the other hand, the R^2 of the exponential density models revealing the degree of conformity of actual development to the conceptual concentric metropolitan forms bears the potential to characterise certain local sprawling patterns since its unexplained part (i.e. $1-R^2$) is, in part, caused by local sprawling phenomena, such as discontinuity (Malpezzi and Guo, 2001).
14. $R^2 = 1$ indicates complete continuity; $R^2 = 0$ represents extremely discontinuous development.
15. To date, besides the knowledge of measuring the influence of neighbouring areas on each other, and the popular examples of Moran's I equal to 1, 0 and −1, the understanding of spatial autocorrelation is limited. What is known, taking Moran's I for example, is that positive values mean neighbouring areas are similar to each other in terms of development intensity, and vice versa; higher values mean a higher degree of similarity and lower values mean a lower degree of similarity.
16. The *global* Moran and Geary coefficients are distinct from the *local* Moran and Geary coefficients.
17. A random scattering spatial distribution will have a value of Moran coefficient equalling (−1/number of cells).
18. Theoretically, the range of the Moran coefficients is between +1 and −1; however, it is relatively small (i.e. $-0.01 \sim 0.13$) in these cases. A possible reason is that they use different weighting methods. The former uses contiguity (for example, the Rook criterion) weighting method; while the latter uses distance (inverse distance for these cases) weighting method. For details on comparative analysis of these two weighting methods, see the following paragraph in the main text.
19. This finding, in fact, can also be proved mathematically and it also proves true in two other simulation sets. See Tsai (2001) for more details about this analysis.
20. Scale effect is defined as the variability of analysis results caused by data being recorded at different levels of partitioning (such as city, census tract, block) for the same area (Wong, 1996).
21. Zonal effect is defined as the inconsistency of analysis results caused by data for the same region, but divided in different ways, with the same number of sub-areas in different partitioning schemes (Wong, 1996).
22. This statistical process may involve two steps. First, apply regression analysis to examine the effects caused by the mean land area of sub-areas and variation of the land area of sub-areas (as independent variables), on the Moran (or the Gini) coefficient (as the dependent variable)

$$\text{Moran coefficient} = a + b\,(\text{average area of sub-areas}) + c\,(\text{variation of the areas of sub-areas})$$

Secondly, exclude the effects on the Moran coefficient caused by the discrepancy of average area and variation from their means of all metropolitan areas.

$$\text{Adjusted Moran} = \text{Moran coefficient} - b\,(\text{discrepancy from mean of average area of sub-areas}) - c\,(\text{discrepancy from mean of variation of the areas of sub-areas})$$

References

ALLEN, W. B., LIU, D. and SINGER, S. (1993) Accessibility measures of U.S. metropolitan areas, *Transportation Research B*, 27(6), pp. 439–449.

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

ANDERSON, J. E. (1982) Cubic spline urban density functions, *Journal of Urban Economics*, 12, pp. 155–167.

ANDERSON, W. P., KANAROGLOU, P. S. and MILLER, E. J. (1996) Urban form, energy and the environment: a review of issues, evidence and policy, *Urban Studies*, 33(1), pp. 7–35.

ANSELIN, L. (1995) *SpaceStat version 1.80 user's guide*. University of Illinois, Urbana Champaign, IL.

BERTAUD, A. and MALPEZZI, S. (1999) *The spatial distribution of population in 35 world cities: the role of markets, planning and topography*. Center of Urban Land Economics Research.

BURCHELL, R. W., SHAD, N. A., LISTOKIN, D. *ET AL.* (1998) *The Costs of Sprawl—Revisited*. TCRP Report 39. Washington, DC: National Academy Press.

BUREAU OF TRANSPORTATION STATISTICS (1997) *Census Transportation Planning Package (CTPP)—Urban Element*. US Department of Transportation, Washington, DC.

CERVERO, R. (1996) Jobs–housing balance revisited: trends and impacts in the San Francisco Bay Area, *Journal of the American Planning Association*, 62(4), pp. 492–511.

CERVERO, R. and KOCKELMAN, K. (1997) Travel demand and the 3Ds: density, diversity, and design, *Transportation Research D*, 2(3), pp. 199–219.

CERVERO, R. and WU, K.-L. (1998) Sub-centring and commuting: evidence from the San Francisco Bay Area, 1980–90, *Urban Studies*, 35(7), pp. 1059–1076.

COULTER, P. B. (1989) *Measuring Inequality: A Methodological Handbook*. Boulder, CO: Westview Press.

CRAIG, S. G. and NG, P. T. (2001) Using quantile smoothing splines to identify employment subcenters in a multicentric urban area, *Journal of Urban Economics*, 49, pp. 100–120.

DOUGLAS, R. P. (1998) Transit-focused development: a progress report, *Journal of the American Planning Association*, 64(4), pp. 475–487.

EWING, R. (1997) Is Los Angeles-style sprawl desirable? *Journal of the American Planning Association*, 63(1), pp. 107–126.

GALSTER, G., HANSON, R., RATCLIFFE, M. R. *ET AL.* (2001) Wrestling sprawl to the ground: defining and measuring an elusive concept, *Housing Policy Debate*, 12(4), pp. 681–717. (http://www.fanniemaefoundation.org/programs/hpd/pdf/HPD_1204_galster.pdf).

GIULIANO, G. and SMALL, K. A. (1991) Subcenters in the Los Angeles region, *Regional Science and Urban Economics*, 21, pp. 163–182.

GORDON, P. and RICHARDSON, H. W. (1997) Are compact cities a desirable planning goal? *Journal of the American Planning Association*, 63(1), pp. 95–106.

GORDON, P., KUMAR, A. and RICHARDSON, H. W. (1989) The influence of metropolitan spatial structure on commuting time, *Journal of Urban Economics*, 26, pp. 138–151.

GORDON, P., RICHARDSON, H. and WONG, H.-L. (1986) The distribution of population and employment in a polycentric city: the case of Los Angeles, *Environment and Planning A*, 18, pp. 161–173.

GRIFFITH, D. A. (1981) Modeling urban population density in a multi-centered city, *Journal of Urban Economics*, 9, pp. 298–310.

GROUP, I. (1990) *Greater Toronto area urban structure concept study*. Prepared for the Greater Toronto Coordinating Committee.

HESS, G. R., DALEY, S. S., DERRISON, B. K. *ET AL.* (2001) Just what is sprawl, anyway? *Carolina Planning*, 26(2), pp. 11–26.

LEVINE, J. (1998) Rethinking accessibility and jobs–housing balance, *Journal of the American Planning Association*, 64(2), pp. 133–148.

MALPEZZI, S. and GUO, W.-K. (2001) *Measuring 'sprawl:' alternative measures of urban form in U.S. metropolitan areas*. The Center for Urban Land Economics Research, University of Wisconsin, Madison, WI. (http://www.bus.wisc.edu/realestate/docs/docs/Alternative%20Measures%20of%20Urban%20Form.doc).

MCDONALD, M. K. (1985) The intensity of land use in urban employment sectors: Chicago 1956–1970, *Journal of Urban Economics*, 18, pp. 261–277.

MCMILLEN, D. P. (2001) Nonparametric employment subcenter identification, *Journal of Urban Economics*, 50, pp. 448–473.

MILLER, E. J. and IBRAHIM, A. (1998) Urban form and vehicular travel, *Transportation Research Record*, 1617, pp. 18–27.

MUÑÍZ, I., GALINDO, A. and GARCÍA, M. Á. (2003) Cubic spline population density functions and satellite city delimitation: the case of Barcelona, *Urban Studies*, 40(7), pp. 1303–1321.

NEWMAN, P. W. and KENWORTHY, J. R. (1989) Gasoline consumption and cities, *Journal of the American Planning Association*, 55(1), pp. 24–37.

PENFOLD, R. (2001) Lecture note for elementary statistics (http://www.cquest.toronto.edu/geog/ggr270y/notes/not04c.htlm).

RUCK, D. (1993) *Cities without Suburbs*. Washington, DC: The Woodrow Wilson Center Press.

SMALL, K. A. and SONG, S.-F. (1992) 'Wasteful' commuting: a resolution, *Journal of Political Economy*, 100, pp. 888–898.

SMITH, D. M. (1975) *Patterns in Human Geography*. Newton Abbot: David and Charles.

SONG, S.-F. (1992) *Monocentric and polycentric density functions and their required commutes*. Working Paper 198, The University of California Transportation Center, University of California, Berkeley, CA.

THOMAS, R. W. (1981) *Information Statistics in Geography*. Norwich: Geo Abstracts.

TSAI, Y.-H. (2001) *Travel-efficient urban form: a nationwide study on small metropolitan areas*. Dissertation, University of Michigan, Ann Arbor, MI.

WEITZ, J. and MOORE, T. (1998) Development inside urban growth boundaries: Oregon's empirical evidence of contiguous urban form, *Journal of the American Planning Association*, 64(4), pp. 424–440.

WONG, D. (1996) Aggregation effects in geo-referenced data, in: S. L. ARLINGHAUS, D. A. GRIFFITH, W. C. ARLINGHAUS *ET AL.* (Eds) *Spatial Statistics*, pp. 83–106. New York: CRC Press.

YEH, A. G.-O. and LI, X. (2001) Measurement and monitoring of urban sprawl in a rapidly growing region using entropy, *Photogrammetric Engineering and Remote Sensing*, 67(1), pp. 83–90.

ZHENG, X. P. (1991) Metropolitan spatial structure and its determinants: a case study of Tokyo, *Urban Studies*, 28, pp. 87–104.