Geographic Patterns of Urban Residential Development

Jay Lee

Department of Geography, Kent State University, Kent, Ohio 44242-0001. Email: jlee@kent.edu

1. Introduction

In an urbanized environment, the geographical pattern of residential development is a complex phenomenon to model quantitatively. This is because it is a comprehensive and dynamic phenomenon that involves a wide spectrum of social, economic, cultural, and geographical variables. Conventional approaches to studying this phenomenon have been focusing on using one or just a few variables, while holding others constant, to obtain a sketchy impression of how the development of urban residential land use has changed over space and in time. Some examples can be found in Morcombe (1984), Donovan and Neiman (1993), Hitt (1994), Fulford (1996), Fader (2000), and Levia and Page (2000). As such, results from many laborious research and literature on this topic offer only partial understandings of how urban residential lands develop geographically and temporally. For practical application, we would need to develop integrated models with definable quantitative measures.

Urbanization expands the size of a city or a population settlement until the city becomes too crowded for further development at its core. At that time, sub-urbanization occurs. Both population and economic activities move from the city's core to its peripheral areas. With advancement in transportation technology, sub-urbanization actually expands urbanized areas and connects the surrounding scattered developments into metropolitans. During such process, the development of residential lands takes a variety of geographic patterns at different stages of urbanization.

In the early stage of urbanization, residential lands develop in a more compact form, as new developments tend to occur adjacent to existing ones within and surrounding the urban core. During sub-urbanization, the population settlement continues to expand, new residential developments may occur in a more haphazard manner and in a faster rate -- a phenomenon often referred to as urban sprawl. The sprawling of urbanized area surrounding its core typically occurs by first spilling new developments over peripheries and then followed by in-filling those vacant lands in between. As peripheries become saturated, the next stage of expansion begins to ignite another cycle.

We illustrate here a set of models that can be used to measure the geographical patterns of how residential lands develop spatially and temporally. The actual pattern can be measured to fit into one of the defined models to see if it is closely related to a compact expansion, a sprawling growth, or an in-fill process. Specifically, we use Join-Count Statistics, a simple spatial statistic, to measure the geographical patterns of the development of residential lands. Coupled with temporal trends, these models can be used to quantitatively study residential developments.

The parcel-level data from Geauga County, Ohio, USA, were used to show how a simple statistical method could be applied to quantitatively model the geographical pattern of residential development over time. With similar data at parcel level, modeling of urbanization process can be carried out in other regions in the same way.

2. Join-Count Statistics as a Measure of Geographical Patterns

We suggest that the Join-Count Statistics can be used to measure quantitatively the clusterness and dispersion of the geographical patterns of residential lands. Join-Count Statistic is the simplest form of spatial autocorrelation. While not being the most powerful statistic for measuring spatial patterns, it is appropriate for this study because of its ability to handle polygonal binary nominal data.

In statistical concepts, autocorrelation is the relationship between successive values of residuals along a regression line. In most cases, a strong autocorrelation indicates successive values are strongly related, which implies that data values being regressed may have a systematic trend among them. Spatial autocorrelation is a simple extension of the autocorrelation concept into two dimensions:

- A strong, positive spatial autocorrelation means that the characteristics of geographic objects are very similar to those of nearby objects. This is normally referred to as a ***clustered*** pattern.
- Alternatively, a strong negative (or inverse) spatial autocorrelation suggests that geographic objects may have very distinctive properties between adjacent objects. This is often known as a ***dispersed*** or ***uniform*** pattern.
- When there is no measurable spatial autocorrelation, the geographic objects are said to be in a ***random*** pattern.

The three patterns, clustered, random, and dispersed, serve as three mileposts on a spectrum along which many other possible patterns may exist. Because of this, a value of spatial autocorrelation coefficient by itself is not useful unless it is tested for its statistical significance of how different it is from a coefficient value indicating a particular pattern. For example, a spatial autocorrelation coefficient measured from an observed pattern will need to be tested to see if it is statistically significantly different from the coefficient value of a random pattern – giving the same spatial structure of the geographic objects.

Join-Count statistics provides a simple and efficient way of quantitatively measuring the degree of clusterness or dispersion among a set of spatially adjacent polygons. It bases its measurements on how different the geographical pattern being observed is different from a theoretically constructed random pattern, given the same number of geographical areas and the same spatial structure. This statistic allows users to measure and test if a geographic pattern is statistically significantly different from a random pattern, and if so, whether the geographic pattern is more clustered than a random pattern or it is more dispersed than a random pattern.

Join-Count Statistics only work with binary nominal data associated with polygons. Given a geographic phenomenon with a yes/no, presence/absence, black/white, or other forms of dichotomy characteristics, each geographic area, or termed as a *polygon*, is associated with either one of the two possible characteristics. In our case, a land parcel can be classified as built or vacant. A *built* parcel is one where it has been built with a construction unit while a *vacant* parcel is without one and is available for future development.

Once the polygons of land parcels are defined as either built or vacant, the Join-Count Statistic counts the numbers of various types of Joints between adjacent polygons. These numbers are then compared with those of a random pattern to determine if the pattern being observed is significantly different from a random pattern or if it is more clustered or dispersed than a random pattern.

A *joint* is a segment of shared boundary between two adjacent polygons. It can be a *BV* joint if the joint connects a built polygon and a vacant polygon. A joint can be a *BB* joint if it is shared by two adjacent polygons that both have been built for residential use. Similarly, a *VV* joint is one that is shared by two vacant parcel polygons.

Following Lee and Wong (2001, pages 147-156, also in Cliff and Ord 1981, Upton and Fingleton 1985, Goodchild 1986, Griffith 1987), let O_{BV} be the number of observed *BV* joints, E_{BV} be the number of expected *BV* joints from a random pattern, and σ_{BV} be the estimated standard deviation of E_{BV}, a Z score can be calculated as:

$$Z = \frac{O_{BV} - E_{BV}}{\sigma_{BV}} \quad \text{………………….………. (1)}$$

If the observed number of *BV* joints is greater than the number of expected *BV* joints, it means that the observed pattern has more *BV* joints than that of a random pattern with same polygonal structure. In this case, the observed pattern is likely more dispersed than a random pattern because it has more occurrences of built polygons adjacent to vacant polygons. Alternatively, a pattern whose number of *BV* joints is less than that of a random pattern would imply that it is a more clustered pattern because the built polygons tend to be located next to built polygons. A word of caution should be given here regarding the appropriate use of Z score in testing the statistical significance of Join-Count Statistics. There should be at least 30 (or more) parcels in each data set and that the ratio between built and vacant parcels should not be too close.

A graphic example for a clustered pattern (Figure 1a), a random pattern (Figure 1b), and a disperse pattern (Figure 1c) is shown in Figure 1. In this hypothetic polygonal structure, black squares denote built parcels while white squares denote vacant parcels.

Over time, a geographical pattern of residential land use may change from one form to another by the added new development. For example, Figure 2 shows two examples for changing geographical patterns by adding a new development. In the first example (Figure 2a), the newly added residential parcel reduces the number of BV joints thereby making the pattern more clustered. In the second example (Figure 2b), the newly added residential parcel increases the number of BV joints thereby making the pattern more dispersed.

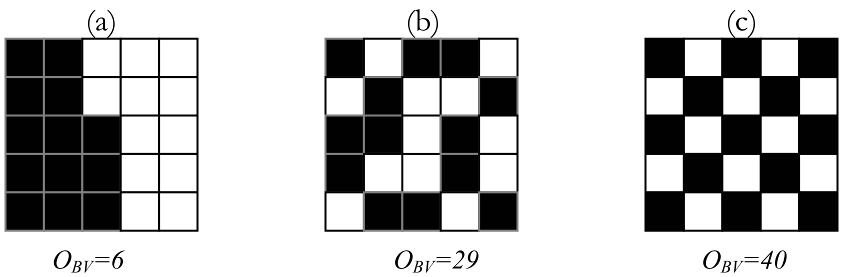

FIGURE 1: Hypothetic polygonal structure of 13 built parcels and 12 vacant parcels

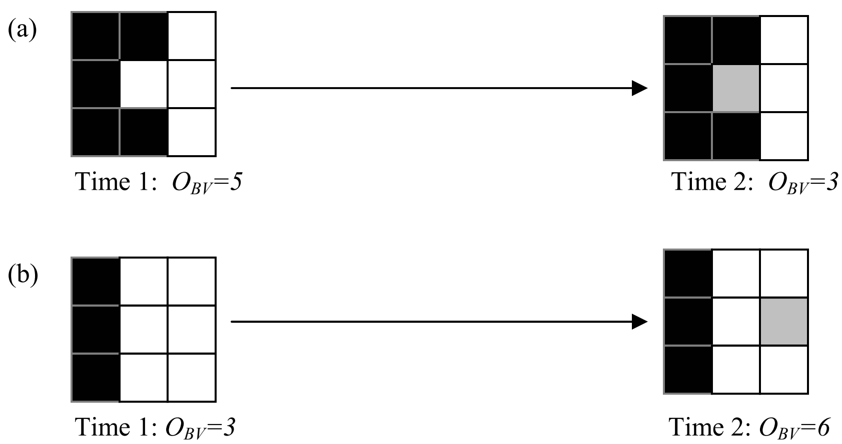

FIGURE 2: Changes of geographical patterns with new residential development

The changes of O_{BV} will be reflected in the Z scores. When O_{BV} increases, Z score tends to increase. Similarly, Z score decreases when O_{BV} decreases. Given this relationship between the numbers of BV joints and geographic patterns, it is possible, then, to examine how geographic patterns change over time.

The spatial statistic, Join-Count Statistic, has been implemented in ESRI's Avenue scripting language as a part of the accompanying package in Lee and Wong (2001). With this computer script, a total of two steps are needed to calculate the Z scores of the Join-Count Statistics for any data set. The first step is to calculate a distance matrix to record the adjacency between polygons. In this step each parcel was examined in turn to determine its adjacent neighbor polygons. These neighbor polygons are then recorded by their identifiers in the distance matrix in a binary form: 0 for non-neighboring polygons and 1 for neighboring polygons.

The second step calculates the Join-Count Statistics based on *B* (built parcels) and *V* (vacant parcels). The testing of the statistical significance can be based on a free sampling or a non-free sampling hypothesis. The *free-sampling* hypothesis assumes that the probability of a parcel being built or remain vacant is known without reference to the study region. In other words, the built-vacant division is not based on local condition but rather on the trend from a much larger region that contains the study area. The *non-free sampling* hypothesis assumes that the numbers of built parcels and vacant parcels remain the same but their arrangement can be different in the study area. This hypothesis is based on local conditions. The statistical testing based on this hypothesis allows analyst to assume that the observed pattern is only one of many possible arrangements using the same numbers of built parcels and vacant parcels, without making any reference to outside factors.

3. Quantitative Models with Temporal Trends

There are three general categories of geographical patterns conventionally used as benchmarks to describe how polygons (as areas) structure spatially. The first category is cluster patterns when polygons of similar properties cluster together. Next, the disperse pattern is when polygons of similar properties are apart from each other. In the extreme case of disperse pattern, a uniform pattern is a pattern in which every polygon is surrounded by polygons of different property. Finally, the third category is the random pattern in which there does not seem to be any structural pattern.

Using the three categories as benchmarks and together with Z scores of Join-Count Statistics calculated from a geographical pattern of polygons, it is possible to construct a number of patterns that characterize different types of urbanization process. Specifically, by calculating Z scores of Join-Count Statistics of a polygonal structure over time, we would plot the changes of Z scores to see how they change so to gain detailed understanding of its temporal trends.

First, a *compact expansion* of residential development is a growth pattern in which new residential lands are developed next to existing built parcels. The growth, in a compact expansion pattern, tends to be slow and gradual whereby the overall geographical pattern shows a compact form of increasing territory. When a population settlement starts as a small village and gradually grows into a city, it often takes the form of the compact expansion. Quantitatively, this pattern is characterized by negative Z scores of Join-Count Statistics because of its apparent clusterness, *i.e.*, $O_{BV} < E_{BV}$. Over time, a compact expansion tends to have decreasing Z scores, but only in a slow rate. In Figure 3a, a gentle decrease of Z scores over time gives the temporal trend of a compact expansion as measured by Join-Count Statistics.

Next, a *sprawling growth* is a development pattern that newly built residential lands occurred in a leap-frogging, haphazard manner. In this pattern, the developers typically look for lands that are the most economical, efficient to build, opting for the most profit. The sprawling pattern is also the results of a fast-pace growth and, to some degree, a growth pattern that is not under tight control or according to a defined plan. In a sprawling growth, it is often the case that many small, hard-to-use vacant lands are created by the leap-frogging developmental pattern. In quantitative terms as shown in Figure 3b, a sprawling growth is a pattern where Z scores of Join-Count Statistics increase over time. The rate of increase in Z scores depends on the rate of urban sprawl.

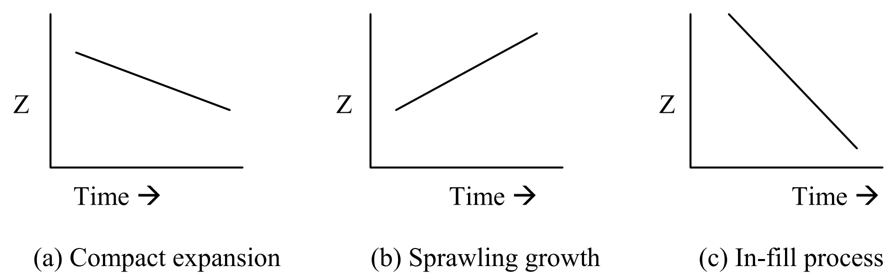

FIGURE 3: Three temporal trends of residential development

Finally, an in-fill process would occur after sprawling growth begins to slow down and the growth of the urbanized area expands into filling up the vacant lands left behind by urban sprawl. Using the Z scores of Join-Count Statistics as shown in Figure 3c, the in-fill process is characterized by steep slope for fast decrease of Z scores over time. This is because the number of O_{BV} is quickly reduced with each new residential development that fills in a hole created by the surrounding built parcels. Compared Figure 3c with Figure 3a, the slope of an in-fill trend is much steeper than that of a compact expansion.

In spatial terms, Figure 4 shows the three temporal trends by adding new residential developments. Figure 4a shows that the three new residential developments (shown in gray) impact on the Z scores of Join-Count Statistics only modestly since O_{BV} only decreases slightly. The number of observed BV joints decreases slightly from 9 to 8.

With also three new residential developments (shown in gray), all detached from existing developments (shown in black), Figure 4b shows that O_{BV} increases dramatically in a leap-frogging development. In this case, O_{BV} increases significantly from 9 to 19. In turn, this increase will be translated to a significant increase in the Z scores of Join-Count Statistics.

The last of the three trends is shown in Figure 4c. Also being added with three new residential developments, the number of *BV* joints decreases from 9 to 7. This is a decrease that is faster than that of a compact expansion. It is because for every infill land parcel, there are three or four *BV* joints converted to either *BB* or *VV* joints. Thus, we would expect a faster decrease of O_{BV} if new land parcels are being developed during the infill stage.

It should be pointed out that it is possible to combine the three temporal trends to model long-term urbanization process. For example, Figure 5 shows that a population settlement started out as a small village going through a slow growth of compact expansion (segment a), then following up with a fast growth before a urban sprawl (segment b), growing in a sprawling manner (segment c), and finally ending up with a slower growth when sprawling is over (segment d).

In summary, the changes in Z scores of Join-Count Statistics as calculated from geographical patterns of residential lands at different time period allow us to construct a curve with the vertical axis being the Z scores and the horizontal axis being the time to show temporal trends. The directions of changes and their slopes help to define both the geographical patterns and their temporal trends. With these models, it is now possible to examine how they perform with real world data sets.

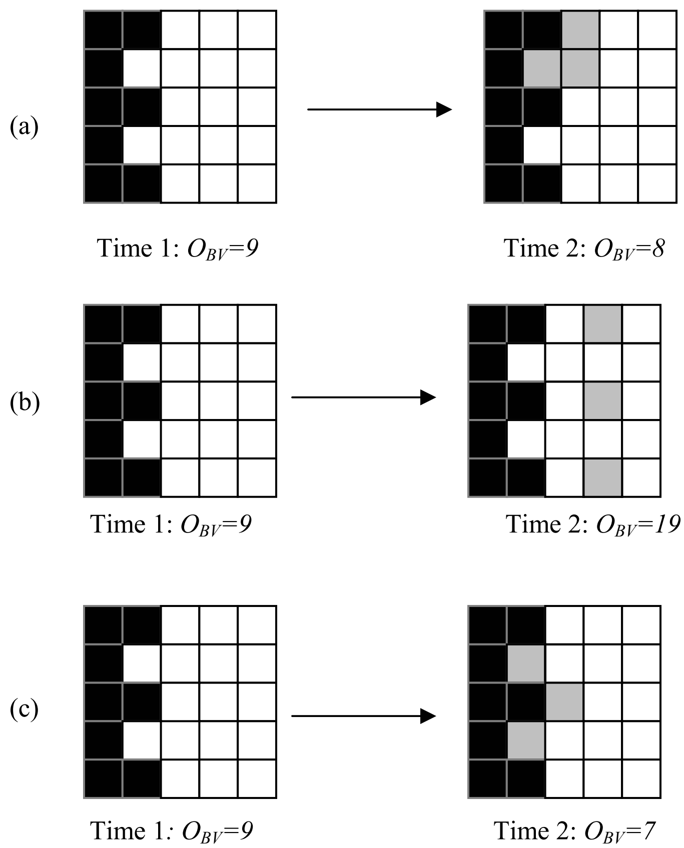

FIGURE 4: Three temporal trends. (a) a compact expansion, (b) a sprawling growth, (c) an in-fill process.

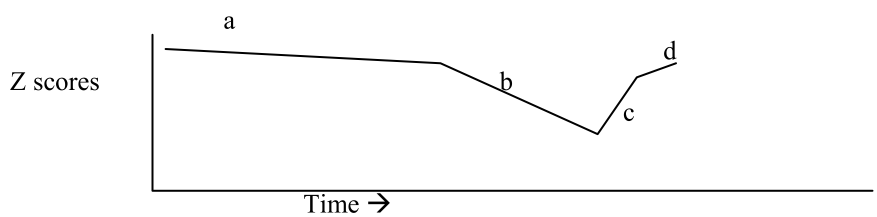

FIGURE 5: Long-term trend of urbanization modeled by Z scores from Join-Count Statistics

4. Case Study

We use the parcel level data set of Geauga County in Ohio of the United States. Geauga County is located in Northeast Ohio with the greater Cleveland metropolitan area to its west. Due to the process of sub-urbanization occurring in Cleveland over the last several decades, there has been significant urban sprawl observed in Geauga and nearby counties at the same time.

While the sprawling built-out in Geauga County is by no means an isolated phenomenon, the reasons and patterns of residential development displayed in it are similar to many other regions in the country, and perhaps in other countries as well. As such, it is suitable to be used as an example here for demonstrating the quantitative models established earlier for the study of urbanization processes. We expect that similar application of these models to other geographic regions is possible if data are available in similar forms.

In Figure 6, land parcels that were built earlier are shown in darker colors while parcels built recently are shown in lighter color. The four categories are separating parcels into: pre-1900, 1900-1950, 1950-1970, and 1970-2000.

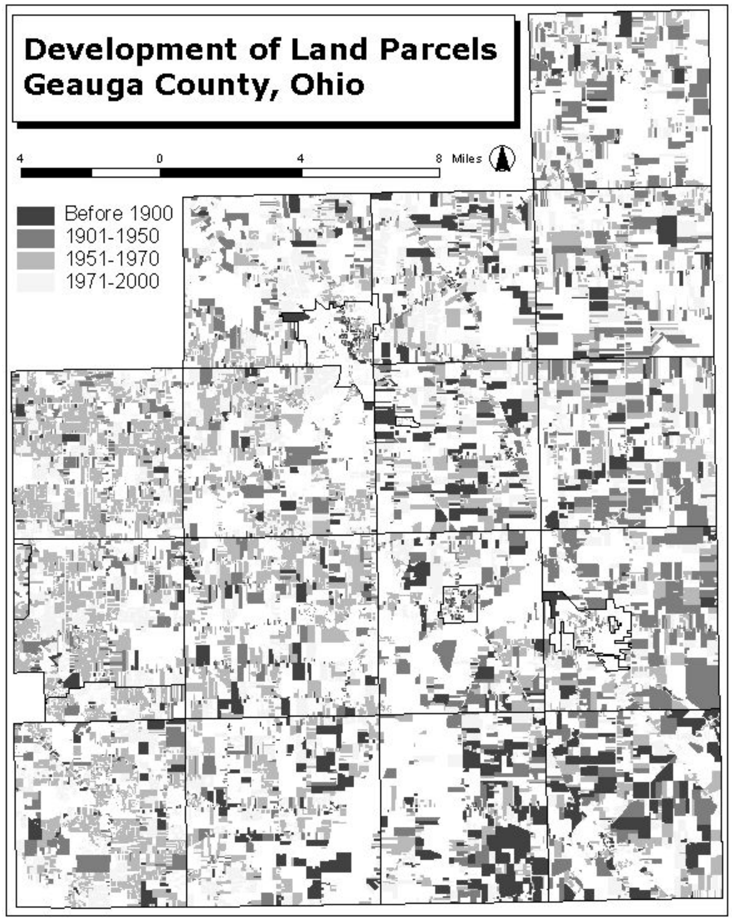

FIGURE 6: Development of land parcels in Geauga County, Ohio, USA.

As shown in Figure 6, Geauga County has seen much of its lands been developed within just the last several decades. The sprawling of developed parcels is certainly to a great extent being influenced by the growth of the greater Cleveland metropolitan from the west. However, it is difficult to gain a detailed understanding just by visually examining this map. A quantitative measure that describes the spatial pattern of how the parcel development proceeded will allow us to fully explore the trend of land development.

For Geauga County, we obtained from the county's Auditor's Office a data set that contains boundaries of all land parcels, their assessed land values, the years parcels were built (or developed), land use types, and many other administrative attributes. For the purpose of this study, we use only the attribute that contains the years when the residential parcels were built. With this attribute information, we are able to construct, dating back to pre-1900 time, the spatial patterns of developed lands in the county for any year between 1800 and 200, or any duration of the time periods.

With this study, we show that Join-Count Statistics can be used to measure the geographical and temporal trends of residential development. This method is appropriate because the land parcels are represented by polygons with a built/vacant binary form of attribute information. What we are measuring is, for any given year, the statistical significance of its geographic pattern being different from a random pattern. With this, we will also examine any significant difference regarding its direction and strength as indicated by the Z score of the Join-Count Statistics.

There are a total of 23 townships and villages in Geauga County. Built/Vacant data are available from 1800 to 2000. For reasons of annexation and simplifying data processing, 23 townships/villages were integrated into 21 areal units. Along the temporal axis, yearly data were integrated to every 10^{th} year, starting with pre-1900, then 1910, 1920, …, and 2000.

Specifically, we performed the following steps when calculating the Z scores for the Join-Count Statistics:
For each township/village,

1. Count the actual number of shared boundaries (*BV* joints) between neighboring built and vacant parcels (polygons), O_{BV}.
2. Calculate the expected number of built-vacant joints for a random pattern, given the number and the structure of polygons, E_{BV}.
3. Calculate the standard deviation of the number of built-vacant joints, (σ_{BV}).
4. Calculate the Z score using the observed number of joints, expected number of joints, and the standard deviation of joints, ($Z=(O_{BV}-E_{BV})/\sigma_{BV}$).

In this way, the value of Z score is an indication of:

- If Z approximates 0, it means that the observed pattern is not significantly different from a random pattern of the same polygonal structure.
- If Z > 0, it means that the observed pattern has more built-vacant joints than a random pattern, implying that the observed pattern is likely a dispersed pattern. For this study, a high positive Z score would mean an urban sprawl pattern because very few newly built parcels would be adjacent to existing built parcels.
- If Z < 0, it means that the observed pattern has less built-vacant joints than a random pattern, implying that the observed pattern is likely a clustered pattern. For our purpose here, a strong negative Z score may mean that the built-out pattern to be similar to a compact expansion where newly built parcels are adjacent to existing built parcels.

For the issue of statistical significance level, we used the conventional $\alpha=0.05$, which translates to a critical value of ± 1.96 for Z scores:

- If Z < -1.96, it means that the observed pattern is a more clustered pattern than a random pattern. Typically, a more contagious expansion of residential development will yield such Z values. The more negative the Z value is, the more clustered the observed pattern would be.

- If Z > *1.96*, it means that the observed pattern is a more dispersed pattern than a random pattern. Similarly, a more leap-frogging development will yield such Z values. The larger the Z value is, the more severe the dispersion is.

With the consolidated 21 townships, Figure 7 shows the relative position of these townships and villages. Note that Geauga County is located east of Cleveland. Therefore, the influence by Cleveland is apparently coming from the west (left) side of the map.

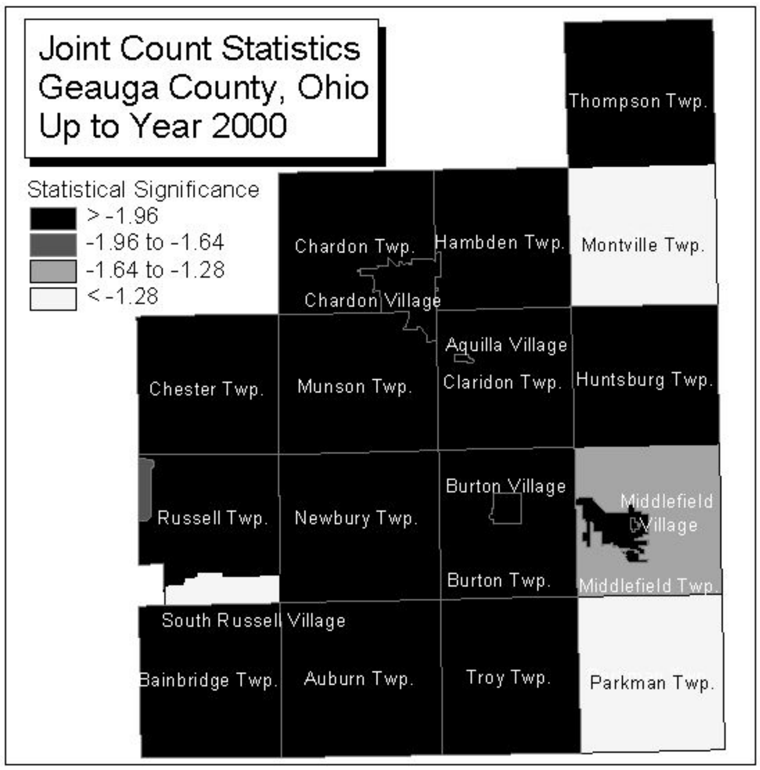

FIGURE 7: Z scores of Join-Count Statistics for townships and villages in Geauga County, Ohio, USA.

With Geauga County divided into 21 areal units and the year-built attribute information integrated to every 10^{th} year, we have compiled the following table (Figure 8) that shows the Z scores for testing the significance level of computed Join-Count Statistics. In the case when Z scores are greater than +*1.96*, they are listed in bold font. For Z scores that are smaller than –*1.96*, they are listed in italic font.

Pre1900	~1910	~1920	~1930	~1940	~1950	~1960	~1970	~1980	~1990	~2000
Thompson Twp.	-0.07	0.42	0.30	0.29	0.01	-0.23	-1.29	*-3.07*	*-3.13*	*-4.08*
Montville Twp.	**2.55**	**2.76**	**2.65**	1.69	0.87	-0.43	*-2.40*	*-3.93*	*-5.05*	*-5.67*
Hambden Twp	**2.81**	**2.36**	**2.26**	1.69	0.82	-1.19	*-5.20*	*-7.82*	*-10.50*	*-9.32*
Chardon Twp.	1.23	1.06	1.04	-0.01	0.05	-1.64	*-12.10*	*-16.83*	*-18.51*	*-15.12*
Chardon Village	*-5.83*	*-9.84*	*-11.53*	*-14.86*	*-15.79*	*-18.23*	*-25.03*	*-28.72*	*-32.50*	*-29.85*
Huntsburg Twp.	**2.39**	**2.71**	**3.21**	**3.40**	**2.98**	**3.00**	0.04	-1.68	*-4.13*	*-5.61*
Claridon Twp.	0.67	0.53	0.67	-1.10	-1.24	*-2.15*	*-5.10*	*-6.04*	*-7.21*	*-6.00*
Munson Twp.	-0.65	-0.89	-0.44	-0.74	-1.35	*-2.46*	*-7.99*	*-13.43*	*-17.66*	*-19.76*
Chester Twp.	0.76	0.22	0.00	-0.20	-0.76	*-3.45*	*-21.03*	*-34.64*	*-22.04*	*-14.78*
Aquilla Vlg.	0.67	0.53	0.67	-1.10	-1.24	*-2.15*	*-5.10*	*-6.04*	*-7.21*	*-6.00*
Middlefield Twp.	**4.12**	**2.69**	**2.63**	**5.11**	**4.68**	1.13	-0.75	*-3.54*	*-5.60*	*-3.86*
Burton Twp.	-0.08	-0.19	-0.19	-0.97	-0.93	-2.04	*-7.69*	*-10.03*	*-15.35*	*-15.20*
Newbury Twp.	0.38	-0.06	-0.31	-1.49	-1.71	*-2.43*	*-9.01*	*-9.41*	*-17.44*	*-19.54*
Russell Twp.	-0.06	-0.36	-0.38	-0.56	-1.38	*-4.95*	*-15.14*	*-22.03*	*-16.72*	*-9.90*
Hunting Valley V.	-1.15	-1.45	-0.98	-1.03	-1.26	*-1.94*	*-0.99*	*-2.36*	*-2.63*	-1.71
Burton Vlg.	*-4.82*	*-5.77*	*-6.95*	*-8.91*	*-10.14*	*-9.50*	*-9.82*	*-9.67*	*-8.18*	*-7.18*
Middlefield Vlg.	-1.33	*-2.71*	*-3.41*	*-9.28*	*-9.83*	*-11.74*	*-14.70*	*-15.36*	*-17.60*	*-18.55*
South Russell Vlg.	0.78	0.13	0.14	-0.54	-0.80	-1.46	*-8.73*	*-27.74*	*-30.27*	*-16.51*
Parkman Twp.	1.85	1.63	1.41	1.51	1.15	-0.88	*-2.78*	*-3.58*	*-5.72*	*-5.94*
Troy Twp.	**3.22**	**2.65**	**2.81**	**1.90**	**2.02**	-0.10	*-2.83*	*-4.38*	*-5.75*	*-5.96*
Auburn Twp.	-0.03	0.07	-0.01	-0.53	-1.45	*-3.86*	*-6.68*	*-7.77*	*-14.93*	*-22.78*
Bainbridge Twp.	-0.65	-0.89	-1.14	-1.49	-1.74	*-3.15*	*-11.62*	*-20.95*	*-37.35*	*-47.49*

FIGURE 8: Z scores of Join-Count Statistics for townships and villages in Geauga County, Ohio, USA.

Among the 21 areal units, it can be seen that they display different geographic patterns of development of residential parcels over time. To simplify the description, the townships and villages are grouped to the following categories:

Old, intensively developed areas – This group includes Chardon Village, Burton Village, and Middlefield Village. The three villages were developed the earliest in the county and were urbanized the earliest. Their Z scores are statistically significant dating back to pre-1900 era. This indicates that these villages are matured urbanized areas with few vacant parcels for new development.

Mature, developed areas – This group includes Montville Township, Humbden Township, Huntsburg Township, Middlefield Township, and Troy Township. They are mostly in the eastern part of the county. These areas were sparsely developed in the earlier part of the century but began to develop since 1950 with influence from the old, intensely developed areas.

Young, developing areas – This group includes the rest of the county. As can be seen in Figure 7 and Figure 8, these townships did not display any significant spatial pattern until 1950's when they began receiving influences from Cleveland metropolitan area. Due to space limit, the rate of growth has slowed down somewhat in recent year but the dramatic growth experienced in the county is apparent as the Z scores indicated in Figure 8.

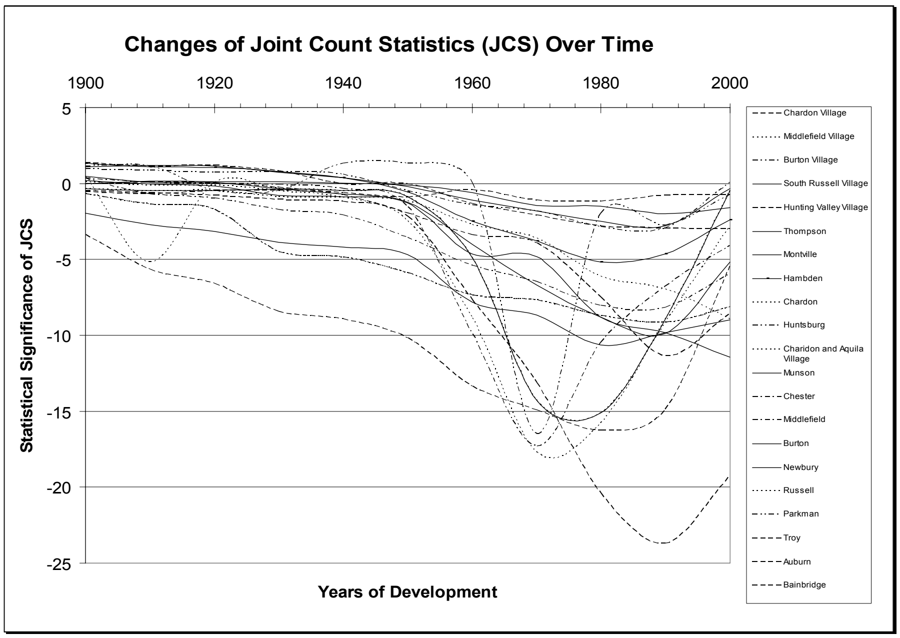

FIGURE 9: Temporal trend lines of Z scores

To examine the temporal trends of how Z scores change over time, they were translated to trend lines in Figure 9. In graphic form as shown in Figure 9, most townships and villages showed decreasing Z scores until the 1950's. After that time, townships and villages developed in very different rates because the slopes of the trend lines show very different steepness. This is because different townships/villages developed with different degrees of compactness. Different townships and villages arrived at the lowest Z scores at different times that were also the times they began to experience the rebound of Z scores. Note that the upward trend of Z scores shows fast growth of the township or village, often to a sprawling pattern.

Finally, in graphic form, the maps in Figure 10 show a temporal account of the changes in Z scores by the townships and villages in Geauga County. It can be seen that the earlier development in the county was mostly generated from regional urban cores within the county but the recent development was clearly influenced by Cleveland to the west (left).

Based on the temporal trends showed in Figure 9, we chose to present the changes of Z scores by townships and villages for up to 1910, 1940, 1960, 1970, 1990, and 2000. These years were chosen to reflect the shape of trend lines in Figure 9. As can be seen in Figure 10, a wave-like shift of strong negative Z scores go through the county from west to east. In the early years up to 1910, most townships and villages showed only weak negative Z scores, then the whole county showed stronger Z scores. In the most recent years, some old intensively developed area began to show changes of Z scores from strong negative to weak negative.

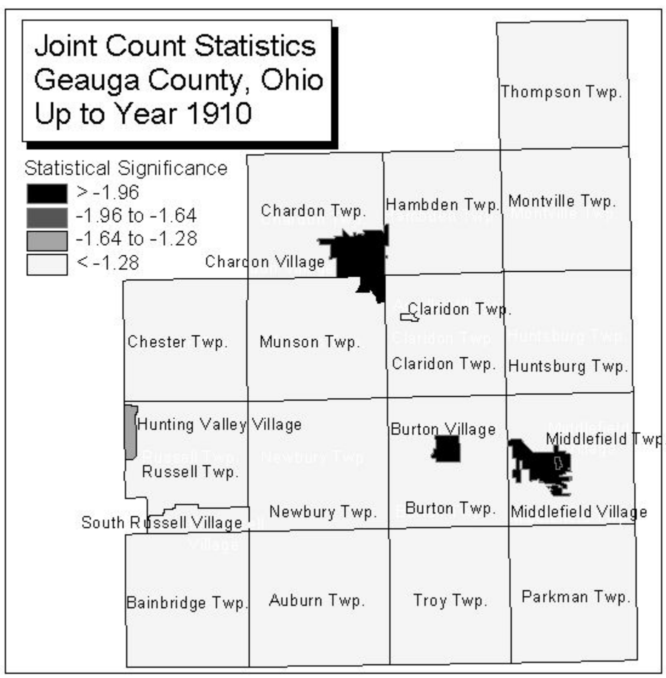

FIGURE 10a: Statistical significance of Join-Count Statistics, up to 1910.

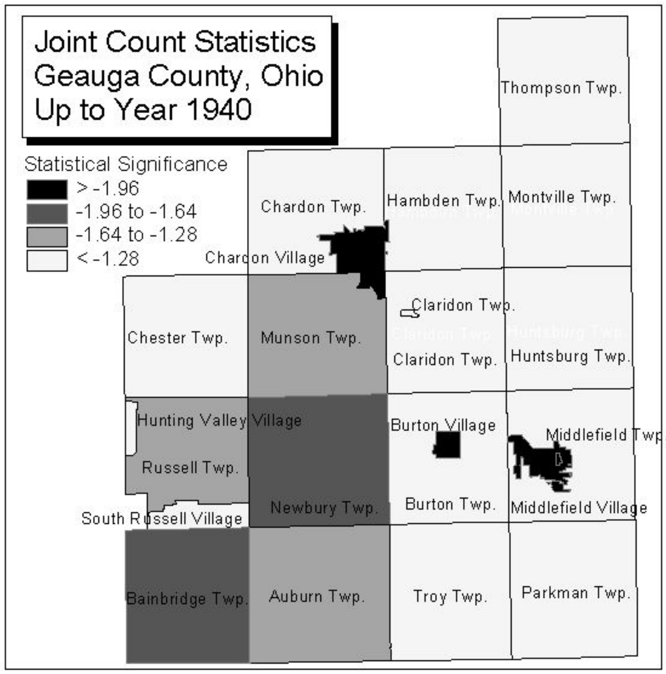

FIGURE 10b: Statistical significance of Join-Count Statistics, up to 1940.

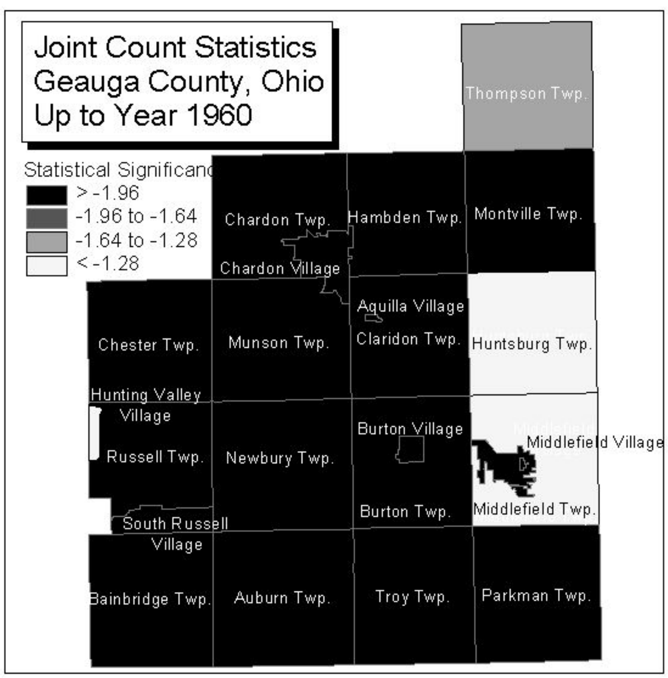

FIGURE 10c: Statistical significance of Join-Count Statistics, up to 1960.

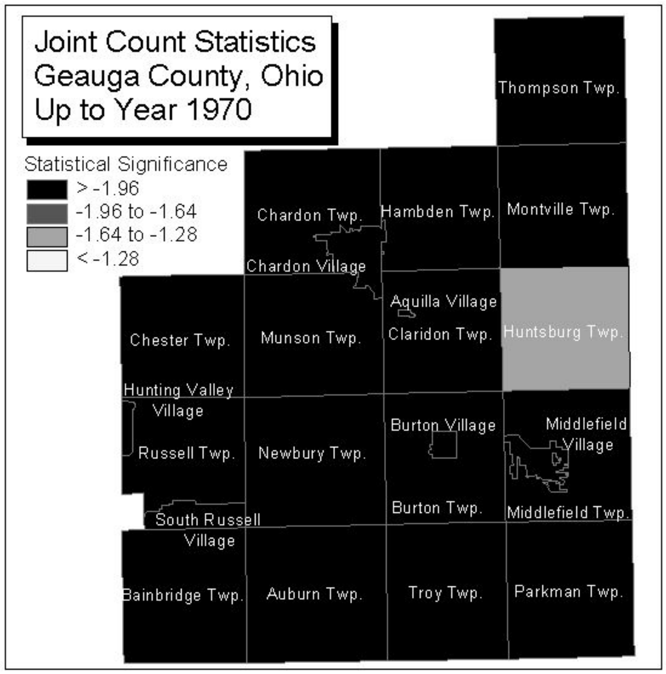

FIGURE 10d: Statistical significance of Join-Count Statistics, up to 1970.

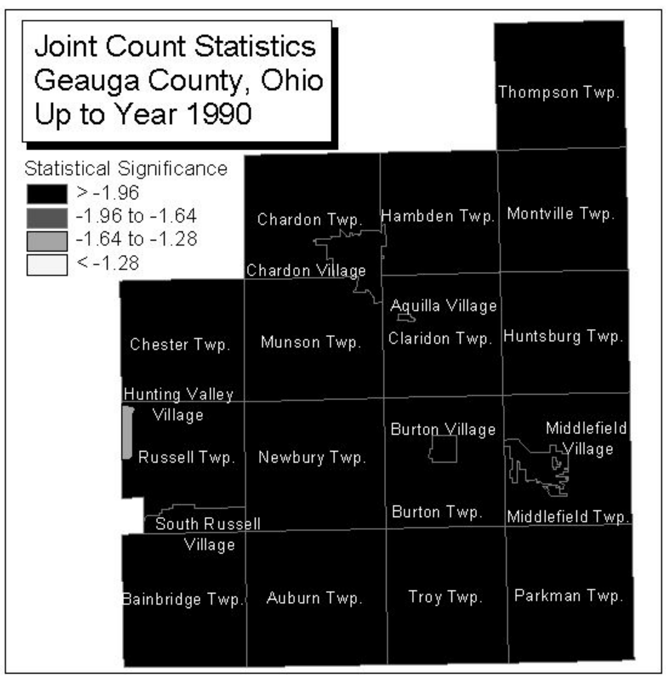

FIGURE 10e: Statistical significance of Join-Count Statistics, up to 1990.

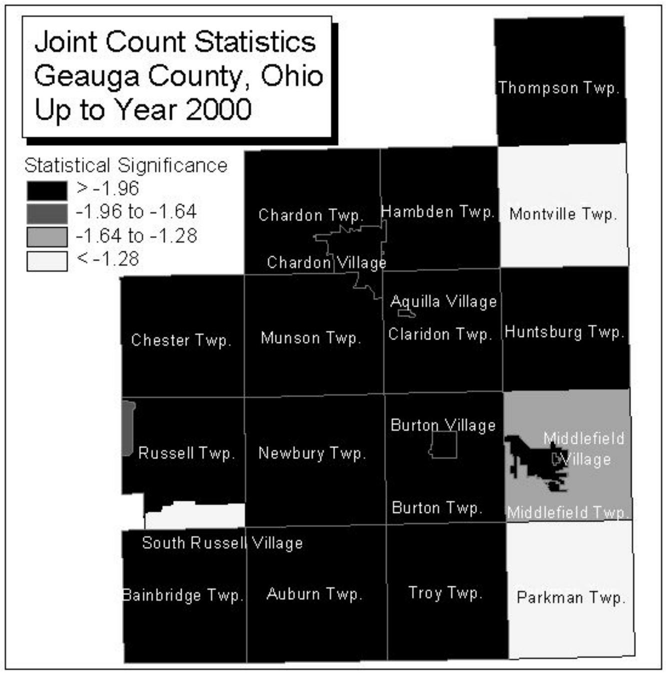

FIGURE 10f: Statistical significance of Join-Count Statistics, up to 2000.

5. Discussion and Concluding Remarks

Study of urban form and the process of urbanization is by no means a new topic. As early as in 1939, Hoyt examined the geographic patterns of the growth of the built-up areas in five American cities (Hoyt 1939). With vivid graphics showing how residential developments expanded over time for San Francisco, California, New Orleans, Louisiana, Baltimore, Maryland, Salt Lake City, Utah, and Dallas, Texas, He showed how the industrial revolution prompted the dramatic growth in population and territorial extent.

Following that, there had been models proposed to describe internal structure of urban land use patterns. These include the well known concentric zone land use model originated from von Thuüen's land rent theory, Hoyt's own sectoral model of urban structure (Hoyt 1939), and the more recent multiple-nuclei model of urban form by Harris and Ullman (1945) – all helped to explain some aspects of how urban land use changed geographically with various variables and over time.

Specific to the urban core, the core-frame concept of central business district (CBD) by Horwood and Boyce (1959) helped to establish the relationship between urban core and its surrounding areas. For sub-urbanization, Erickson (1983) proposed the three-stage process of spillover and specialization, dispersal and diversification, and infilling and multinucleation. Together, these models provided the literature of urban geography a solid foundation pointing clear directions for more in-depth understanding of the urbanization and suburbanization processes.

With the proliferation of geographic information systems (GIS), researchers began to use them to examine changes of urban form with an integrated and operational approach to understand urbanization process. With parcel data for the city of Buffalo, New York, Batty (1997) showed ways to simulate urban expansion for practical applications. In similar way, Clarke *et al*. (1997) integrated raster data layers to examine urban sprawl for large geographic regions. It should be noted that contemporary efforts in studying geographic forms of urban growth have geared toward dynamically simulating urban built-out scenarios by incorporating various geographic factors, planning practices, growth management strategies, and newly available technology in GIS. Examples for this may be found in Landis' work for California future land use modeling (Landis and Zhang 1998), Lee's GIS-based simulation of built-out scenarios for Northeast Ohio counties (Lee *et al*. 1999, also the Urban Growth Simulator at http://empact.geog.kent.edu) and many others.

In light of proposing an effective and applicable quantitative approach to studying geographic patterns of residential developments, we presented the method of using the Join-Count Statistics whose Z scores may be used to formulate various models to describe different urban growth patterns. The models presented here suggest a possible taxonomy of three growth patterns to help explain how residential temporal developments evolved in different localities. These three growth patterns, to be used to characterize localities with different developmental patterns, are: old intensely developed areas, mature developed areas, and young developing areas.

In many ways, this is a practical approach to understanding how any given locality developed. This is because the approach allows for the incorporation of the established urban land use models and the sub-urbanization processes of spillover, sprawling growth, and infill. As a result, urbanization and sub-urbanization can be better understood both conceptually and operationally.

On the technical side, the studies of urban form have been conducted mostly using raster data structure that really did not capture the spatial configuration of land parcels in a city. Examples for raster-based studies on urban sprawl/form can be found in Couclelis (1997), Clarke et al (1997), Batty (1997), and Landis and Zhang (1998). The cell size in a raster-based analysis of this kind is almost always difficult to define or to be justified. Analyzing land parcels in their original form of vector data structure is the only way to allow researchers to take full accounts of the geographic relationship among land parcels. With the methods and models described here, this geographic relationship can finally be measured and tested. Of course, this is assuming that the parcel-level data and the temporal data of when parcels were built are available to researchers – an assumption sometimes does not hold in places where such data are not available. Also, the success and usefulness of the analysis greatly depend on the types of data available. For example, if parcel data record the history of land ownerships and their transactions, it will be possible to extend studies to examine how such process impacts spatial patterns over time.

Finally, the many spatial analytical methods that geographers developed have not been incorporated into modern GIS software packages. The use of these spatial analytical methods is very much limited to only academia and perhaps a small circle of researchers who are knowledgeable in spatial statistics. Although some scattered efforts have attempted to widen the usage of spatial statistics for more practical applications, the few software packages developed did not gain much popularity because of their long learning curves and the requirements of user's computer literacy. To this end, Lee and Wong (2001) developed and bundled a suite of computer scripts that work directly within ArcView GIS, a widely used GIS system, to allow even novice users to take full advantage of the powerful tools in spatial statistics. With these scripts, the calculation of Join-Count Statistics and the associated Z scores can be easily carried out for parcel level data in any geographical locations.

6. Acknowledgement

This study was funded by the US Environmental Protection Agency under Grant #985989-01-0. The funding and support from US EPA Chicago, Illinois and Westlake, Ohio offices are greatly appreciated.

7. References

Batty, M., 1997. Cellular Automata and Urban Form: A Primer, *Journal of the American Planning Association, **63***(2): 216-274.

Clarke *et al*. 1997. A Self-Modifying Cellular Automaton Model Of Historical Urbanization In The San Francisco Bay Area, *Environment and Planning B*, ***24***: 237-61.

Cliff, A. and Ord, J. (1981) *Spatial Processes: Models and Applications*. London: Pion.

Couclelis, H. (1997) From Cellular Automata to Urban Models: New Principles for Model Development and Implementation, *Environment and Planning B, **24***: 165-174.

Donovan, T. and Neiman, M. (1993) Community Social Status, Suburban Growth, and Local-Government Restriction on Residential Development, *Urban Affair Review*, ***28***(2): 323-336.

Fader, S. (2000) *Density By Design: New Directions in Residential Development*. Washington, DC: Urban Land Institute. 137p.

Fulford, C. (1996) The Compact City and the Market: the Case of Residential Development in M. Jenks, E. Burton, and K. Williams (eds.) *The Compact City: A Sustainable Urban Form?,* pages 122-134. London, NY: E & FN Spon.

Erickson, R. A. (1983) The Evolution of the Suburban Space Economy, *Urban Geography **4***: 96.

Goodchild, M. (1986) *Spatial Autocorrelation*. CATMOG 47. Norwich: Geo.

Griffith, D. (1987) *Spatial Autocorrelation: A Primer*. Washington, DC: Association of American Geographers.

Harris, C. D. and Ullman, E. L. (1945) *The Nature of Cities*: *Annals, American Academy of Political and Social Science **CCXLII**:* 7-17.

Hitt, K. (1994) *Refining 1970's Land-Use with 1990 Population to Indicate New Residential Development*. Reston, VA: U. S. Geological Survey.

Horwood, E. M. and Boyce, R. R. (1959) *Studies of the Central Business District and Urban Freeway Development*, Seattle, Washington: University of Washington Press, p. 21.

Hoyt, H., 1939. *The Structure and Growth of Residential Neighborhoods in American Cities*. Washington, DC: Federal Housing Administration, pp. 158-9.

Landis, J. and Zhang, M. 1998. The Second Generation of the California Urban Future Model. Part I: Model Logic and Theory, *Environment and Planning B: **25***(5): 651-666.

Lee, Jay and Wong, David W. S., 2001. *Statistical Analysis with ArcView GIS*. New York, New York: Wiley and Sons. 208p.

Lee, J., Erickson, L., Kulikowski, T., and Tian, L. (1999) Managing Urban Sprawl with Geographic Information Systems, *Environment and Planning B,* ***25***: 865-879.

Levia, D. and Page, D. (2000) The Use of Cluster Analysis in Distinguishing Farmland Prone to Residential Development: A Case Study of Sterling, *Massachusetts. Environmental Management,* ***25***(5): 541-548.

Morcombe, K. (1984) *The Residential Development Process: Housing Policy and Theory*. Brookfield, VT: Gower. 148p.

Upton, G. and Fingleton, B. (1985) *Spatial Data Analysis by Examples*. New York: Wiley and Sons.