

Intra-Urban Analysis of Commercial Locations: A GIS-Based Approach

Christopher D. Storie, Department of Geography, University of Winnipeg, Winnipeg, MB, Canada

ABSTRACT

The urban landscape is an interspersed mixing of residences, shops, theatres, parks, natural areas, and a multitude of other uses. From the early days of the central markets, to the planned downtown, to the heavily planned super-regional shopping complexes, the commercial areas within this urban landscape have evolved. There has been considerable research conducted on analyzing the commercial structure of urban environments in an attempt to better understand the nature of retailing and its resultant impacts on the geography of the city. This paper details the development of a GIS-based semi-automated method to detect commercial structure. The approach generated nearest commercial neighbour statistics as a measure of proximity between commercial locations. These were used as the foundation for clustering commercial operations into commercial areas.

Keywords: Commercial Clustering, Geographic Information System, Retail Geography, Retail Structure, Urban Environment

INTRODUCTION

Over the last two decades there has been a resurgence of interest among social scientists in the cultures of consumption and their associated spaces of retail capital (Lowe & Wrigley, 2000). The biggest reason for this resurgence is that shopping is one of the most common place human activities and that these locations in which consumption takes place have dramatically shaped the urban environment. Shops and shopping areas are common characteristics of virtually all settlements in the developed world. Following the development of central place theory as a means of studying service provision in rural areas, Proudfoot (1937) examined a

series of urban environments and developed the first classification of urban commercial location. Subsequent commercial models examined the spatial patterns formed by the development of commercial activities within the urban area.

The commercial system is one of the major forces shaping the geography of our cities. The vitality, either actual or perceived, of a city is largely connected to the health of the central business district (CBD), otherwise known as the downtown core. Since this core tends to command a high concentration of commercial activity, it has become a surrogate for a city's economic health (Bunting & Millward, 1998). The introduction of regional-sized malls into the urban fabric over the last half century has created highly visible and highly accessible

DOI: 10.4018/jagr.2013010101

nodes outside of the CBD thereby changing the overall form of the commercial system. The continuing result has been the suburbanization of commercial development. These nodes shape our daily behaviour and our perceptions of the city, but more importantly, they affect location decisions, other land uses, and shape planning choices (Simmons, 1991). The end result is an urban form that is dominated by the commercial system. The commercial system in the context of this research is defined as “the size, variety, and location of commercial clusters that collectively service an urban region” (Jones & Simmons, 1993, p. 26)

Over time, cities change and so does the commercial system. This changing system is a result of complex and often disruptive forces, both internal and external, such as geographical constraints and economic influences including a slumping or booming economy (Adolphe, 2001; Bourne, 1997). In addition, as the economy of many urban areas continues to move towards the service sector and beyond, the nature of our cities will continue to change and reflect the nature of the economy at that time. Furthermore change can also be caused, or accelerated, by competition between retailers. The single greatest noticeable difference to the Canadian commercial environment has been its rapidly changing commercial structure. Commercial services exert considerable influence on the morphology and function of the Canadian city. The present-day city has to a large degree developed in relation to its commercial facilities. This relationship is the direct result of the increasing dependence on the automobile. As Jones (2000, p. 406) states, “when mobility is low, retail activities concentrate, when mobility is high retail activity disperses.” Current studies have shown that cities are dispersing and, in many instances are becoming multi-nuclei as the size of the city continues to grow (Batty, 2001; Bunting & Filion, 1999; Millward & Bunting, 1999). This signifies an era where mobility is very high and therefore dispersion is equally high.

Enclosed mall development reached a pinnacle in the late 1980s as a result of long

term planning and increased capital investment prior to the recession. Since then, little to no shopping centre growth has occurred. Only the most profitable and growing urban centres experienced any new mall development specifically Toronto and Vancouver (Jones & Simmons, 1993).

Suburban malls dominated the commercial structure at this time. Approximately 84 regional malls (those malls with more than 500,000 square feet of retail space) existed in Canada accounting for more than 10% of all retail sales by 1990 (Simmons, 1991). Even though the suburban malls dominated the commercial landscape, both the CBD and the arterial ribbons still played a major role in the provision of commercial services within the overall commercial environment. As the forces of decentralization continued the importance of the CBD began to diminish in virtually all major Canadian cities except those within the high-growth centres. Two large contributions to this decline were the increased distance and time one would have to travel to reach the core and the cost of parking once there. This distance fuelled the development and encouraged the expansion of arterial ribbons connecting the main areas of the city to one another.

During the early 1990s the major force of change came in the form of the new-format retailer, more commonly known as the big-box store. The arrival of these behemoth retail ventures, and their associated power centres or clusters precipitated a shift in consumer behaviour away from the traditional department store and regional mall to these new free-standing discount superstores. These large destination retailers, led by Wal-Mart, Home Depot, Price Costco, Business Depot, and others stressed aggressive pricing and merchandising practices (Jones, 2000). According to Jones and Doucet (2000) 175 big-box retailers entered the Greater Toronto Area market from 1990 to 1997 adding over 6.5 million square feet of retail space compared to the 1.0 million square feet of retail space added through new shopping centre development or re-development of existing centres.

These new power centres were typically brownfield (and later greenfield) developments on the periphery of the city and in close proximity to major highways or roads. In many instances, they were strategically located to serve a super-regional market. Brownfield developments were required because of the amount of land needed for these store-types which were typically big, single-storied, free-standing superstructures with a spacious asphalt parking lot (Arnold, 2000) and as a result of abandoned industrial locations were readily available in many urban environments. This resulted in a continued and advancing level of commercial decentralization. The historical remnants of the older structure still existed, but in many cases these new developments created new commercial areas. In other situations, they were additions to existing commercial areas, therefore, changing the function and composition of these areas.

As Jones and Simmons (1993) conclude “the [commercial] environment, including our patterns of consumption, and the facilities that we use in the process of consumption, has always played a central role in social, economic, and geographic processes” (p. 26). The commercial system is one of the major forces shaping the overall spatial organization of our cities. Yet, commercial growth also contributes negatively in terms of urban sprawl through the continued development of low density commercial clusters (Bunting & Filion, 1999), and the inequitable distribution of goods and services across the entire urban environment, sometimes referred to as ‘deserts’ (Wrigley, 2002; Wrigley, Guy et al., 2002). The result is that there needs to be way of analyzing this change in a manner that is replicable, and can be operationalized.

Geographical information systems (GIS) provide a flexible and efficient platform for conducting commercial analysis especially when large spatial databases are involved. According to Bond (2002) advances in algorithm development and a decrease in the cost of the software packages and data means the probability of using GIS for more advanced urban analysis is much greater. Before the introduc-

tion of computer-assisted methods, researchers developed and performed similar functions manually. The initial works of Berry (1963) and Simmons (1964) involved maps which were made by hand and patterns that were identified based on visual observation. The researchers examined commercial locations and groupings of these locations to determine the form of each cluster. The process of collecting, mapping, and analyzing these commercial locations and the consequent commercial zones was a tedious and time-consuming process as the commercial structure was built up node-by-node and strip-by-strip.

Building on these pioneering works this paper presents a unique method for analyzing the commercial environment. Using a case study approach this research has two broad goals: (a) to develop a technique that makes operational, in a systematized and objective manner, an approach to analyzing the structure of the commercial environment; and, (b) to apply the approach within a GIS environment. Simply stated this approach seeks to semi-automate what has traditionally been done manually. Academics over the years have been trying to capture a snap-shot of an urban environment’s commercial structure and to compare that commercial structure over time to reveal changes. The commercial system is highly dynamic and has gone through several eras of development which have left distinct imprints on the ground. By semi-automating the analysis of commercial structure a more objective means of detecting and comparing commercial structure can be conducted.

STUDY AREAS

The primary area of focus for this research is the Census Metropolitan Area (CMA) of Kitchener, sometimes referred as the Region of Waterloo, or simply Kitchener-Waterloo. It is located within the Province of Ontario, and is the eleventh largest urban area in Canada with a population in 2006 of 451,235 residents (Statistics Canada, 2006). Located approximately

110 km from Downtown Toronto it is within the outer edge of the Toronto commuter zone (Figure 1).

The Kitchener CMA is comprised of three cities (Kitchener, Waterloo, and Cambridge) and four smaller townships (Wellesley, Woolwich, Wilmot, and North Dumfries). The region was established in 1973 by the Province of Ontario to govern this highly interconnected and rapidly urbanizing area (Muncaster, 1998). The result was a political structure that merged the former cities of Galt, Preston, and Hespeler into the one new city, Cambridge and placed this new city under the new regional authority along with Waterloo and Kitchener. Four peripheral townships were added because of their increasing levels of urbanization as well as the linkages to the core cities. However, each township had its own central service community which still exist to this day (Storie et al., 2002).

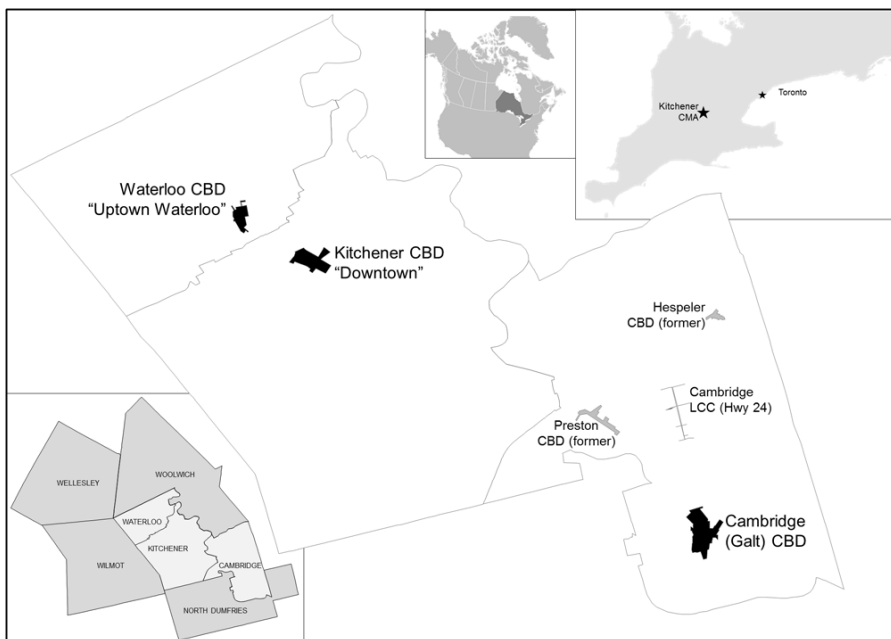
The general commercial structure of the region of Waterloo reflects the amalgamation of five cities into three. Each of the former

cities within Cambridge have the remnants of their former central business districts, with the former city of Galt's CBD functioning as the central business district for Cambridge while the other two are now neighbourhood level business districts. Given the proximity of Kitchener and Waterloo to one another, Kitchener has traditionally had the downtown which served the regional market, while Waterloo had the "uptown." The Kitchener CBD functions are the dominant core area for the entire region, while Waterloo's "uptown" tends to serve a more upscale niche market and directly caters to both the high-tech industry and the university sector of the community (e.g., Research In Motion [RIM] the maker of Blackberry mobile phones is located in Waterloo).

DATA

Many authors have discussed the data requirements for analyzing urban areas and have concluded that regardless of the size and extent

Figure 1. Location of study area and the Kitchener Census Metropolitan Area



of the database, a truly accurate representation is not generally possible (Martin et al., 1994). However, detecting basic or generalized structures is possible. This conclusion fits strongly within this research because it is the basic or generalized commercial structures that are being “detected.” Furthermore, the commercial structure is an “outcome” or “result” of a series of complex interactions with many other aspects of the urban environment. Consequently, the analysis focuses on the visible structures and not on the underlying processes which become much more challenging to model with any level of accuracy.

The data required for this research is comprised of three components; (a) commercial location data; (b) land use planning data; and; (c) ancillary support data.

Commercial Location Data

The primary data component is the commercial locations file obtained from the Centre for the Study of Commercial Activity, Ryerson University. This data comprised 355,431 individual commercial location records, excluding automotive dealerships for the twenty-five CMAs for the year 2000 (Tenenbaum & Zsigovics, 1999; Simmons et al., 2000). From this file 6,848 commercial locations were extracted for Waterloo, Kitchener, and Cambridge. All commercial locations in the dataset have the following variables; street address, city, province, postal code, and NAICS (North American Industry Classification) codes. The geocoded file also contains the latitude/longitude coordinate pairs for each location. This allows for a wider range of spatial analytical functions to be performed such as cluster and nearest neighbour analyses.

Land Use Planning Data

The second component of the analysis requires the use of land use planning data. The data identifies the planning designation of each of the parcels in the urban area. More importantly, it allows for the extraction of areas that have been specifically planned as commercial and consequently can be used as either a part of

the analysis or as a means of validating the analysis. The geographic coverage database used included land use for the entire region of Waterloo in 2000 at the parcel level. Given the nature of land use controls and zoning regulations, this dataset was generalized to only show broad classifications such as commercial, residential, industrial, etc.

Ancillary Data Requirements

The ancillary data requirements for this research were minimal. An up-to-date street network file was needed for each study area to provide the necessary contextual information. This ancillary data was used to gain a better understanding of the ‘nature’ of each study area. Secondly, high-resolution aerial imagery was acquired to aid in the visual confirmation of commercial zones and to act as a point of reference when assessing the accuracy of the model.

METHODS

Conceptually and structurally commercial elements within the urban area can be easily represented within a vector-based GIS environment. The concepts of stores, strips, and downtowns can be directly related to points, lines, and polygons. Therefore the visual representation of commercial location data is immensely aided by using a GIS approach. Before any detailed discussion of methods can begin a major issue must first be addressed; namely the ability to reduce a three-dimensional geographic commercial entity, such as a Home Depot retail store down to a point on a map. It is very common to represent 3D entities as simple points because of data availability – commercial addresses which generate point locations are more readily available than building footprints. One must recognize that this representative centroid is an artificial point of reference meant to illustrate the approximate centre of a location (Longley et al., 2001). The problem facing commercial structure research is that the very nature of the entities being studied have dimensionality greater than that of a simple point. In other

words, by representing a given Home Depot store as a single point, you cannot then extrapolate its true spatial extent. When trying to detect the spatial extent of a commercial zone, this becomes an even greater consideration. Given this constraint, the approach outlined should be viewed as a generalized method of detecting commercial structure. This research used a two-phase analytical approach. First, objective measures of the nearest commercial neighbour were generated to assess proximities. Second, clusters of commercial neighbours were detected to determine ‘meaningful’ groupings of commercial locations. The following section outlines the mechanics of the approach.

The Nearest Commercial Neighbour

The generation of the nearest commercial neighbour was used to determine the spatial proximity of commercial locations to one another. The nearest commercial neighbour was generated by calculating the crow-flight distance from one commercial location to its closest neighbour. The underlying premise here is that “shopping areas” or commercial zones function as a coherent unit because they are in close proximity to each other and nothing else. Therefore by generating each location’s nearest neighbour, and through a series of basic descriptive statistics, a generalized measure of proximity can be calculated of which two specific types are calculated:

1. **Average Nearest Commercial Neighbour (*aNCN*):** This descriptive measure uses the average of the distances generated by the nearest neighbour algorithm. The benefit to this measure is that each city will have its own *aNCN* thereby allowing the analysis to be specifically tailored to the city. The major drawback is that outliers and single stores outside of commercial zones have the potential to skew the average value. However, given the large number of points being addressed this skewing effect should not be significant. When this presents a problem, outliers can be mitigated by removing them from subsequent iterations once the general commercial zones are identified.
2. **Median Nearest Commercial Neighbour (*mNCN*):** This value represents the middle value in the dataset and therefore represents a balance between the very small and very large values without being directly influenced as an average value would be. Unlike the *aNCN*, it is not influenced by outliers and other anomalies. Furthermore it can be used as a surrogate density measure as it provides a generalized proximity value.

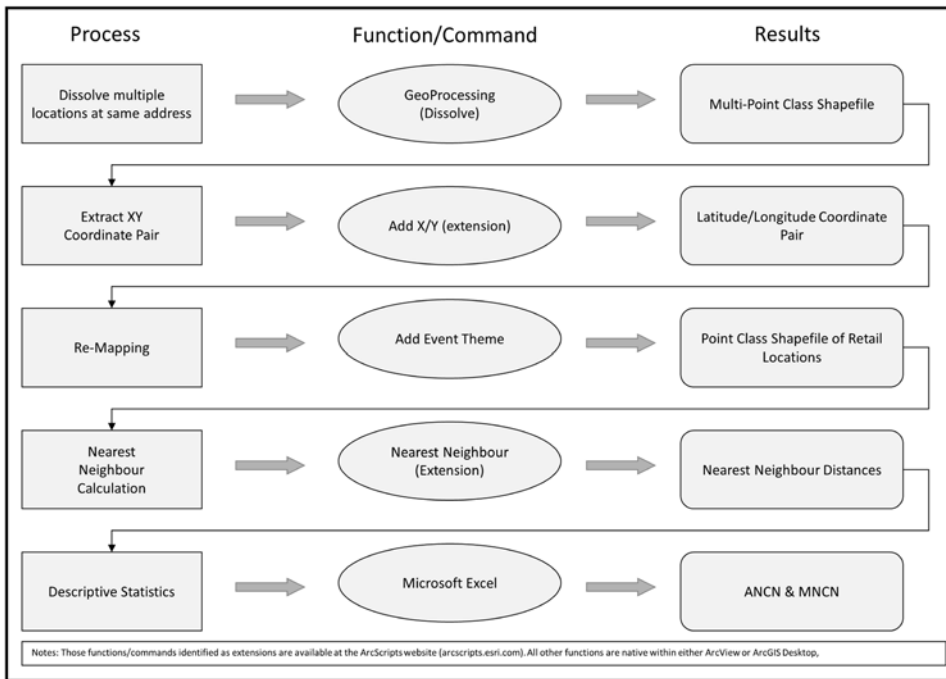
These two descriptive values provide information about the proximity of commercial locations across the entire study area. Another problem, which is easily addressed, is that of multiple stores at a single location - a mall for example. Given that each mall location’s nearest neighbour would be other mall locations (same address), this produces a zero (0) value. These must be removed to avoid skewing the data to the lower end of the distance measure. The calculation of the nearest neighbour values is a straightforward process. A nearest neighbour ArcGIS extension (Sawada, 2002) was used which required point class data (Figure 2).

Cluster Detection Techniques

The cluster detection technique is a two-stage process. In the first stage the commercial locations at unique addresses were analyzed, whereas in the second stage only those locations outside of the CBD were analyzed with the spatial extent of these planned entities being added after the analysis. To empirically detect and then quantify the spatial organization and morphology of the commercial areas using just the commercial location, the following steps were undertaken:

1. Using the derived *NCN* values (*aNCN* and *mNCN*) to cluster the data points using a simple point buffer. The buffer size is defined as the *aNCN* or *mNCN* value or a multiple thereof (2x, 3x, etc.). When two

Figure 2. Nearest commercial neighbour analysis



- or more buffers coalesce they are dissolved into a single unit. This is an iterative process until all overlapping buffer polygons have been dissolved. The result is a series of polygons that generalize commercial zones (Figure 3).
- Once the clusters have been generated, the number of commercial points within each cluster is determined. Those clusters that have less than 5 stores are removed based on the threshold criteria posited by Simmons et al. (2000) who suggest that a cluster of four or fewer stores are not “commercial areas.”
- The remaining clusters are then constrained using a buffered road network (20m buffer). This differs from Simmons et al. (2000) where they have represented commercial zones using a point symbol located at their centroid along a street network. It is felt that this represents a truer cartographic representation of the actual commercial zone.
- Successive iterations will involve a multiple of the $mNCN$ value as the buffer value. The $aNCN$ is not used in multiples as it is typically a larger value than the $mNCN$.
- The value that is determined to have produced the more accurate result, based on field observations and direct comparisons to the aerial imagery is then deemed to be the most effective (Figure 4).

The next level of detection introduces a second data layer. Using the existing point location dataset, those points which are located within malls, power centres, and the central business district are removed from the analysis. The result is that only freestanding locations (unplanned) outside of clearly defined (planned) areas are analyzed. The resulting polygons are then combined with the planned commercial zones to produce a final commercial structure. The results of this level are compared against the aerial-imagery and field observation to identify its level of accuracy.

Figure 3. Clustering process: 3a. Point locations of individual commercial operations; 3b. Using of NCN value as a clustering threshold; 3c. Final merged polygons

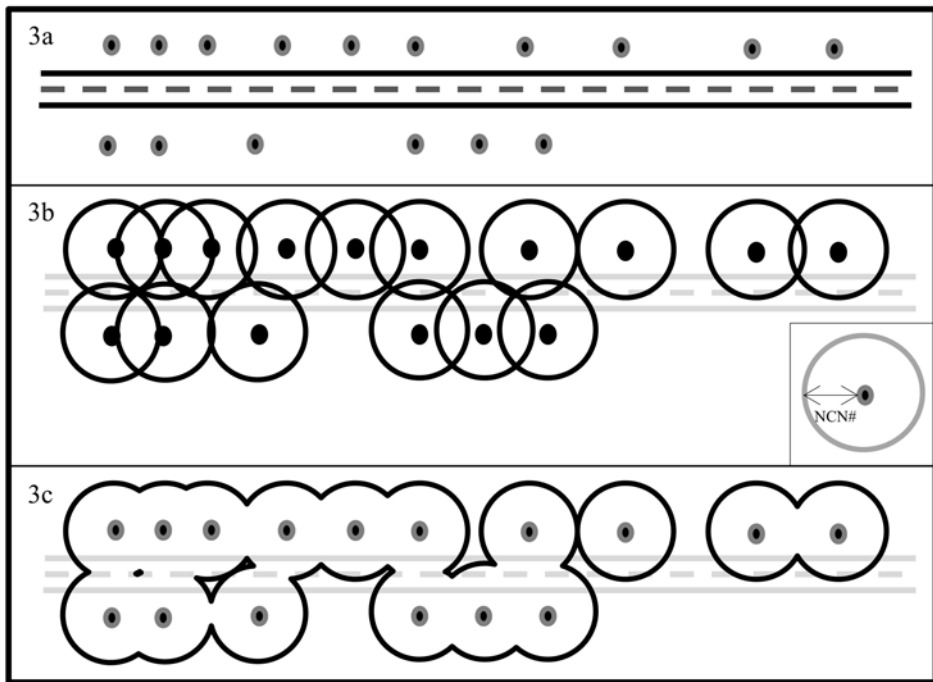
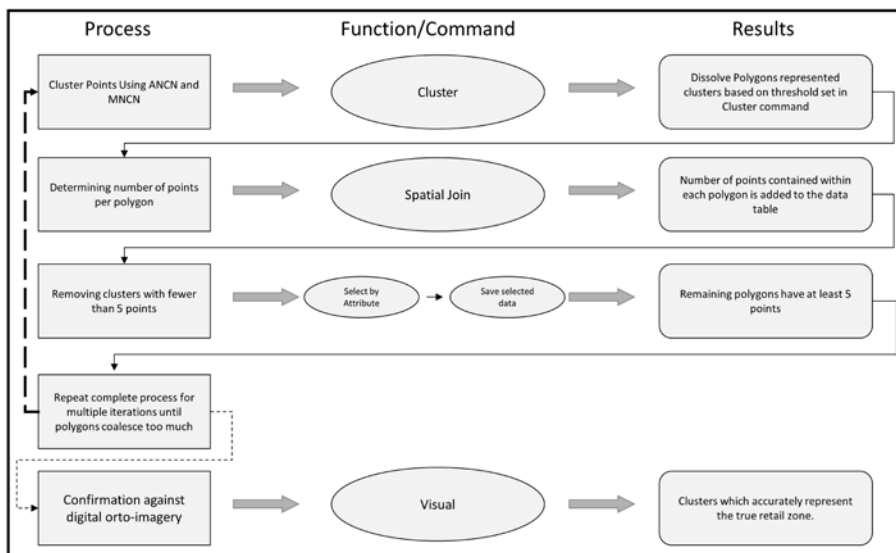


Figure 4. Point only cluster detection of commercial locations



The final step involves comparing the results generated in the two detection methods to accurate parcel data for the Kitchener CMA. This is to function as a measure of accuracy of the detection techniques and involves extracting and grouping land use parcels for the Kitchener CMA that have been defined as commercial and will be used to represent the known commercial structure of the CMA. The generated clusters from the two detection techniques are then overlaid onto this layer and visually compared to the actual land use parcels for their level of visual agreement.

RESULTS

The first stage was to reduce each of the commercial datasets into a subset that contained a single point location for each unique address. The reduction of data points for the Kitchener CMA was approximately 52% indicating that the majority of commercial locations had more than one unique business at that address.

Once the subset was generated the nearest neighbour point distances were calculated. Based on the inter-point distances (measured in meters) the mean and median values were extracted. Kitchener produced an average value of 79.58m between commercial locations. The median value was determined to be 31.5m. Lastly, threshold values for each of the study areas need to be calculated for each of the clustering iterations. Five threshold values representing multiples of the median value were calculated. By the fifth multiple the amount of cluster coalescence was so high that in many cases independent commercial clusters were joined together and identified as being of the same cluster which they should not have been. However, it was prudent to ensure that the model “went past” the accurate values to ensure strong comparison between each of the thresholds.

Based on the threshold values calculated a series of cluster maps for the study regions were generated. Each of the commercial clusters was generated by using a buffer routine with the buffer size set to the threshold value.

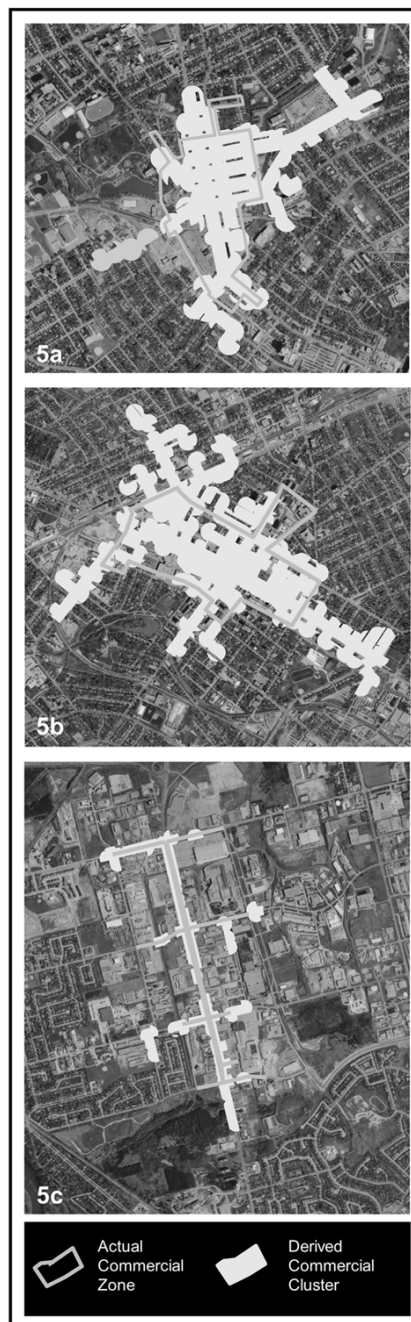
Overlapping polygons were dissolved creating a larger commercial cluster. Once the clusters were generated the number of commercial points was added, and those clusters that contained four or less points were removed resulting in commercial polygons that had five commercial locations or more contained within them. These remaining polygons were then clipped using a road network with a 40m wide buffer. This was done in order to constrain the polygons to the existing road network and to make them more visually accurate in relation to the roads.

The Kitchener CMA has well developed and geographically varied commercial structure. Of the six iterations conducted, a maximum number of clusters (298) was generated using the 1x iteration and a minimum of seventy-three was generated using the 5x iteration. These polygons were compared against the ortho-imagery and the land use planning polygons to determine which iteration was the most accurate. The comparison was done using three readily identifiable commercial areas; the Kitchener CBD, the Waterloo CBD and the Highway 24 linear commercial complex in Cambridge (Figure 5).

Waterloo CBD

In the case of the Waterloo CBD, the 1x iteration was the only result that generally fell within the official CBD boundaries with little spill over. However, it can be argued that though visually this would be the most accurate there is a discrepancy between the land use definition of the CBD and the location of commercial and commercial operations within this inner city area. Like many larger urban environments, arterial roads connect the suburban and outlying areas with the core. Waterloo is no different. To the north-east edge of the CBD is Bridgeport Road which connects the Conestoga Parkway to the east with the downtown area to the west. This arterial strip is mostly comprised of commercial operations starting at the Bridgeport Mall and continuing into the core. In addition, this is a one-way road, west-bound into the core. As a result of road connectivity and continuous commercial development those two areas

Figure 5. Comparison of derived clusters to aerial photography for (a) Waterloo CBD, (b) Kitchener CBD, (c) Highway 24 Linear Commercial Complex



may function as a single commercial area. Therefore, this research suggests that the 2x iteration more accurately captures this commercial area. Furthermore, the 2x also most accurately captures the CBD and captures the commercial development that borders the CBD as a result of externality effects.

Kitchener CBD

The Kitchener CBD produced results that were similar to the Waterloo CBD. The 1x iteration accurately captures the CBD as well as the arterial development radiating outside of the core area. The 2x iteration produces a similar, albeit a more area encompassing result. Once again the differences between formally planned and identified areas (the CBD) and the more functional commercial zone that developments within and around the CBD emerges. Similar to the Waterloo CBD, the Kitchener CBD has several major arterial roads the permit access to and from the core. These arterials are almost fully developed for commercial service provision and in most cases are directly “connected” to the core – in terms of continuous stores, with little to no discernable gaps (houses, green space, etc.). Therefore, if one considers only the CBD then the 1x is more accurate, whereas if one considers the commercial area as a functional zone then the 2x iteration is most accurate.

Highway 24 Cambridge

The final area studied within the Kitchener CMA is the Highway 24 (Cambridge) linear commercial complex (LCC). This is unique area consisting of a primary arterial strip (Highway 24) with a super-regional new-format center anchoring the north and a super-regional mall anchoring the southern extent. Radiating off this main core are a series of smaller arteries that have developed because of the externality effects resulting from the proximity to this large commercial area. When looking at the results it becomes clear that the 2x and 3x iterations most accurately capture this area. However, the 3x iteration also over-captures surrounding areas and includes them within

the LCC. The challenge being faced in this context is the difference between the dense CBD and the “sprawling,” suburban orientated arterial developments. Businesses along the linear commercial complex tend to be located in free-standing buildings, strip malls, and even larger regional and super-regional malls.

SUMMARY

Based on the results it is apparent that the multiple iterations were effective in capturing existing commercial agglomerations. The 1x and 3x iterations captured commercial agglomerations that had density variations that were unique to that particular area. Furthermore, the 2x proved to be the most effective overall at defining existing commercial agglomerations. The technique for determining commercial clusters, especially those associated with heavily developed arterial strips also provided the ability to detect externality effects as a result of commercial spill over along roads perpendicular to the main arterial. For example, Hespeler Road Highway 24 area perpendicular development is very apparent. Moreover, the issue of defined commercial areas (i.e., CBD) versus a broader area (i.e., inner city) also arises. While there is a planned CBD, the density of development within the inner city precipitates a situation where smaller community orientated commercial clusters have become linked with larger arterial clusters and to the CBD. This is reflective of inner city areas that have high degrees of residential and commercial development next door to one another. The second stage of the analysis involved all points that existed within malls (as identified by planning documents) and the central business district(s) were removed from the nearest neighbour generation. The result was a decrease in the overall database size by 37% for Kitchener. In addition, the reduced dataset was then reduced again so that there was one point per address resulting in a further reduction of 58% for the study area.

Once the subset was generated, the nearest neighbour point distances were calculated as

were the average and median values for each study area. The variation of values for the *aNCN* and *mNCN* followed a similar pattern to the point only cluster detection (*aNCN* = 107 and *mNCN* = 58). Based on these derived values threshold values were generated. However, for this stage of the analysis the maximum number of iterations was capped at three. In the previous stage, it became clear that iterations above 3 were “capturing” more area than was accurate. Therefore the goal of this stage was to integrate the findings of the previous level of analysis with the pre-defined polygons representing the CBDs and malls. Malls and CBDs have clearly defined spatial extent and therefore do not need to be detected in the same manner.

Using the threshold values generated, four cluster maps were generated for each study area. The clusters were generated in the same manner as those in the point only cluster detection to ensure consistency across the various levels of analysis. The number of points contained by each cluster was determined and those which had 4 or fewer points were removed. The remaining polygons were then clipped using a buffered road network with a 40m wide buffer. As with the previous level of analysis this was done to constrain the polygons to the existing road network and to make them more visually appealing. At this point a new step is added whereby the existing CBD and mall polygons are merged with the generated commercial clusters producing a final series of commercial polygons.

Of the four iterations conducted the smallest number of clusters generated was 61 with a high of 199. When the malls and CBD polygons were added a minimum of 132 and maximum of 226 resulted. These differences can be attributed to the reasons previously identified. In addition, the geographic organization of this CMA has several isolated areas whereby a small community mall services the surrounding area; an area which does not have any discernible means of development beyond the mall. The same three areas of interest were examined: The Waterloo CBD, the Kitchener CBD, and the Cambridge LCC. When looking at the results of the Waterloo

CBD it becomes clear that the 1x was the most accurate but with certain caveats. Firstly, the merging of the CBD polygon with the generated polygons produced an overall polygon that not only represented the actual CBD but captured the corresponding growth that is connected, but outside of its formal boundaries. The results for the Kitchener CBD followed a similar pattern to those of the Waterloo CBD. The 1x iteration was the most accurate and captured development occurring along three major arteries leading into and out of the core area. All other iterations over captured the core area. The results for the linear commercial complex in Cambridge again follow a similar pattern. However because of its geographic location and particularly its relative isolation in relation to other commercial areas, one major difference becomes evident. Unlike the previous two CBDs, this area is comprised of freestanding stores, enclosed and open malls, and a power centre. Whereas the CBD can be defined by a single polygon, this area is comprised of at least 7 separate polygons representing the largest of the malls (including the power centre as a single polygon). The connective retailing between these larger entities is apparent especially in the northern half of the LCC where the polygon connects with its neighbour, but in the lower half – the location of a super-regional mall – this connectivity is less apparent. The 1x iteration is accurate but does not connect all the individual polygons together into a single uniform area, whereas the 2x iteration (and the average iteration) does and also captures development perpendicular to the main complex along a secondary artery. The 3x iteration over captures along the horizontal axis but is accurate along the main artery.

Land-Use Validation

The final analysis conducted was to compare the results of the Kitchener CMA cluster detections against the land use parcels for the area for the same time period. All those parcels that were defined as commercial were extracted and then compared against the derived commercial clusters for the point and point + fixed entity

detections. When a regional comparison is conducted between the land use parcels and the six iterations for the point only cluster detection three patterns emerge: (1) all iterations to some extent duplicate the general pattern exhibited by the land use parcels. The general geography and spatial orientation is similar as well as the general orientation of the different commercial areas; (2) when examining each of the iterations in greater detail it becomes clear that the 2x and 3x average most closely match the “original” in both extent and orientation. This parallels the results for the point only analysis when compared against the aerial imagery; and, (3) the 4x and 5x iterations over-capture the commercial structure and begin to blend together as a result of the larger cluster sizes (Figure 6).

When a closer analysis of each of the three points of interest is conducted the following patterns are identifiable: (a) with the point only cluster detection the 2x iteration best matches the general pattern exhibited by the commercial parcels. Though the spatial extent is not exact as a result of the buffering and clipping technique employed it does however produce a reliable fit when compared to the original; and, (b) for the point + fixed entity cluster detection the addition of the malls does add precision in terms of spatial extent as the malls match the parcels precisely. However, the removal of a clustering point results in several individual clusters needed to define the commercial area. The addition of the CBD polygon, like the malls polygons, does add precision but impacts the overall cluster detection in the same method. The 1x iteration represented the commercial area most accurately as it was the only iteration that did not over capture the area.

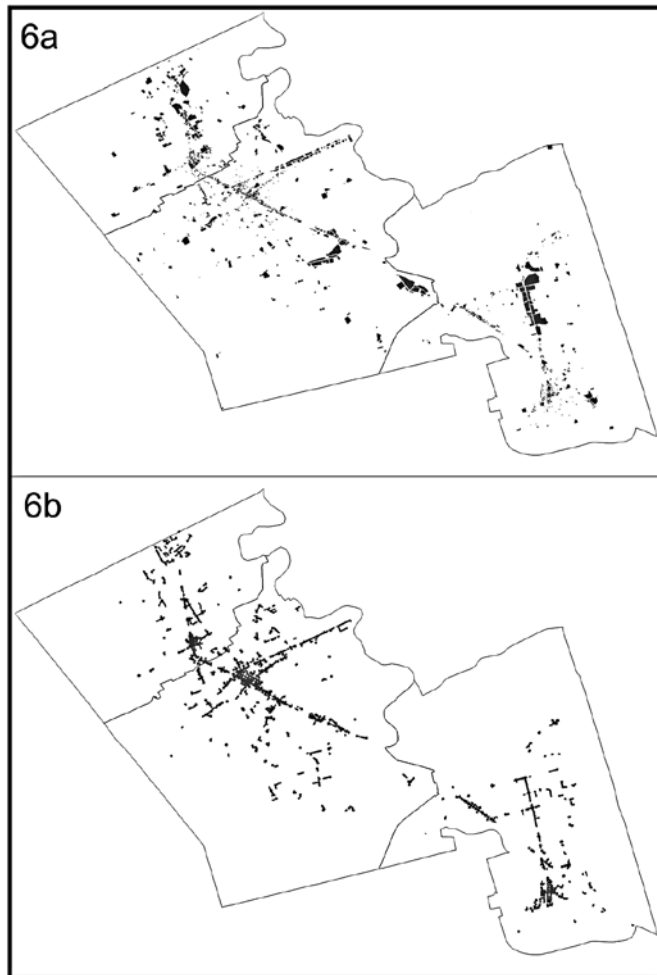
DISCUSSION

Based on the results presented for each of the levels of analyses, three sets of conclusions can be drawn – those which pertain to the general and intermediate levels, those which pertain to the intermediate level of analysis, and those that deal directly with the land use comparison.

Point Only and Point + Fixed Entity Detections

Several commonalities emerged when examining the results of the point only and point + fixed entity cluster detections. First, when dealing with those areas that are clearly defined compared to areas that have grown together, in many instances there is a discrepancy between the actual planned area, such as a downtown and the way the commercial development actually transpired. In most cases the defined commercial areas appear to function as nodes of future growth with development occurring on the boundaries beyond the officially planned area. Second, variations in density between core and suburban areas are a challenge. For example, the difficulties in the definition of the downtown - Is the downtown the area that has been officially identified by an urban planning commission or is it the official area plus the area that has coalesced around it? When considering this from a commercial perspective the manner in which the consumer shops and the boundaries that they adhere to becomes the critical determinant. When someone refers to “going shopping downtown” they are probably referring to all the stores in the general area and not just those that are in the officially defined area. In addition, most consumers do not know what the official planning boundaries actually are. Therefore, when defining a commercial area it is important to inclusively examine the greater surrounding area as well. Variations in density between core and suburban areas were evident visually. The effects could most readily be noticed in over-capturing in the inner city and under-capturing in the suburban areas. However, the areas of focus within the CMA of Kitchener did not present this problem. The final issue identified is over-capturing of commercial areas. Due to the nature of the technique being used, whereby a threshold value is used to define a buffer equally around a commercial point location, inevitably some over-capturing will occur. The amount directly varied with the size of the threshold value. In almost all cases, thresholds above and including 3x were too

Figure 6. Regional comparison of analysis against land use planning information; (a) Areas defined as commercial through land use planning, (b) Derived (2x) commercial clusters



great and resulted in too much over-capturing. Only in more suburban areas did the 3x perform reasonably well.

Point + Fixed Entity Cluster Detection

When examining the point +fixed entity cluster detection two major conclusions can be drawn. First, the inclusion of the CBD and mall polygons does increase the accuracy, but only in those areas where these are the main features

of the commercial landscape. In other cases this polygon prevents a potential overlap with a neighbouring location and consequently does not join with that location creating two separate commercial polygons. This affects the ability to “completely” define a single commercial area. The removal of those points which has an immediate and direct effect on the generation of the *NCN* values. In all cases the values increased thereby making the iterations above 1x too large. Second, the geographic arrangement of commercial types is an issue. When a

commercial area is made up of malls exclusively, with little variety between them, the intermediate technique fails to properly combine the neighbouring polygons into a greater single commercial area.

Land Use Comparison

The inclusion of the land use comparison proved to be a useful step in confirming the results of the cluster detection techniques. Several broad conclusions can be made. First, the point only and point + fixed entity approaches accurately generalized the commercial structure at the regional scale. Moreover, the 2x and 3x iterations for the point only and the 1x iteration for the point + fixed entity were the most accurate. Second, the point only cluster detection was the most accurate in defining the general geography of the commercial area. Though spatial extent may not be entirely accurate, the commercial area was clearly delineated and easily detected. The 2x iteration proved to be the most successful in all areas of interest for all study areas. Lastly, the point + fixed entity, while adding accuracy as a result of the inclusion of malls and CBD polygons did not detect the commercial areas as effectively because these included polygons were not incorporated into the nearest commercial neighbour analysis and consequently represented physical gaps in the clustering. In almost all instances the 1x iteration proved to be the most effective, though its limitations were still visible.

CONCLUSION

The exercise of analyzing commercial data within a GIS is a complex one. Couple this with the inherent variety contained within any urban environment and the challenge of trying to achieve a systematized and operational analysis of urban commercial structure can become quite daunting. The main goal of this research was to define a technique for detecting and classifying the spatial organization and morphology of commercial zones within the Canadian urban

environment and to apply that approach on a GIS platform.

The development of the *NCN* proved to be an effective way of analyzing commercial locations. By generating *NCN* values it became apparent that the use of them as clustering thresholds was an excellent way to group locations together. What also became readily apparent is that the median value was far more effective when used in multiples when compared against the average value for the dataset. In addition, it became clear that the *NCN* values also function effectively as a surrogate density measure. When malls and CBDs were removed from the calculations the *NCN* values were larger and more reflective of development outside of heavily planned areas. It is suggested that this value could be used to monitor density within urban environments whereby increases in the value over time represent an ongoing dispersion of commercial locations.

The analysis of commercial data is still a challenging undertaking. Analyses of this nature are constrained by not only the availability of data but the techniques that can be applied. A fully automated analytical technique while desirable remains elusive. However, this paper has presented an approach to the automation of commercial polygons combined with visual ground-truthing. The two main detection techniques explored can be used with a combination of data types to generate commercial clusters. The point only cluster detection is a more automated process and can generate relatively accurate commercial areas. The point and fixed entity cluster detection combines both point and polygon data and as a result achieves some higher levels of accuracy but overall the analysis is cumbersome and does not dramatically improve point only results.

REFERENCES

- Adolphe, L. (2001). A simplified model of urban morphology: An application to an analysis of the environmental performance of cities. *Environment & Planning A*, 28, 183–200. doi:10.1068/b2631

- Arnold, S. J. (2000). Market impacts of large format retailers. *Journal of Retailing and Consumer Services*, 7, iii–v. doi:10.1016/S0969-6989(00)00026-6
- Batty, M. (1999). A research programme for urban morphology. *Environment & Planning A*, 26, 475–476.
- Berry, B. J. L. (1963). *Commercial structure and commercial blight*. Chicago, IL: University of Chicago.
- Bond, D. (2002). *GIS and spatial analysis in urban and regional research*. Retrieved January 21, 2002, from <http://www.car.busmgt.ulst.ac.uk/papers/gis.pdf>
- Bourne, L. S. (1997). Polarities of structure and change in urban systems: A Canadian example. *GeoJournal*, 43(3), 339–349. doi:10.1023/A:1006878605465
- Bunting, T. E., & Filion, P. (1999). Dispersed city form in Canada: A Kitchener CMA case example. *The Canadian Geographer*, 43(3), 268–287. doi:10.1111/j.1541-0064.1999.tb01385.x
- Bunting, T. E., & Millward, H. (1998). A tale of two CBDs I: The decline and revival of downtown retailing in Halifax and Kitchener. *Canadian Journal of Urban Research*, 7(2), 139–167.
- Jones, K. (2000). Dynamics of the Canadian retail environment. In Bunting, T. E., & Filion, P. (Eds.), *Canadian cities in transition: The twenty first century* (2nd ed., pp. 404–422). Toronto, ON, Canada: Oxford University Press.
- Jones, K., & Doucet, M. (2000). Big-box retailing and the urban retail structure: The case of the Toronto Area. *Journal of Retailing and Consumer Services*, 7, 233–247. doi:10.1016/S0969-6989(00)00021-7
- Jones, K., & Simmons, J. (1993). *Location, location, location*. Toronto, ON, Canada: Neilson Canada.
- Longley, P. A., Goodchild, M. F., Maguie, D. J., & Rhind, D. W. (2001). *Geographic information system and science*. Chichester, UK: John Wiley and Sons.
- Martin, D., Longley, P., & Higgs, G. (1994). The use of GIS in the analysis of diverse urban databases. *Computers, Environment and Urban Systems*, 18(1), 55–66. doi:10.1016/0198-9715(94)90035-3
- Millward, H., & Bunting, T. E. (1999). A tale of two CBDs II: The internal retail dynamics of downtown Halifax and downtown Kitchener. *Canadian Journal of Urban Research*, 8(1), 1–25.
- Muncaster, R. W. (1998). *New format retailing and the commercial structure of the region of Waterloo* (CSCA Research Report 1998-04). Toronto, ON, Canada: Centre for the Study of Commercial Activity.
- Proudfoot, M. J. (1937). City retail structure. *Economic Geography*, 13, 425–428. doi:10.2307/141589
- Simmons, J. W. (1964). *The changing pattern of retail location*. Chicago, IL: The University of Chicago.
- Simmons, J. W. (1991). The regional mall in Canada. *The Canadian Geographer*, 35(3), 232–240. doi:10.1111/j.1541-0064.1991.tb01097.x
- Simmons, J. W., Barbiero, P., Bylov, G., Kamikihara, S., & Zsigovics, G. (2000). *Exploring a national database of commercial activity* (CSCA Research Report 2000-12). Toronto, ON, Canada: Centre for the Study of Commercial Activity.
- Statistics Canada. (2006). *Canadian census*. Retrieved February 20, 2010, from <http://www.statcan.ca>
- Storie, C. D., Oakley, C., & Muncaster, R. W. (2001). The emergence of new-format retailing into the commercial structure of Cambridge, Ontario, Canada: A GIS-Based Analysis. *The Canadian Journal of Regional Science*, 24(3), 505–524.
- Tenenbaum, D., & Zsigovics, G. (1999). *Guide to the national chains and franchise database*. Toronto, ON, Canada: Centre for the Study of Commercial Activity.
- Wrigley, N. (2002). 'Food deserts' in British cities: Policy context and research priorities. *Urban Studies (Edinburgh, Scotland)*, 39(11), 2029–2040. doi:10.1080/0042098022000011344
- Wrigley, N., & Lowe, M. S. (2002). *Reading retail: A geographical perspective on retailing and consumption spaces*. New York, NY: Arnold.

Christopher D. Storie earned his PhD in Geography from Wilfrid Laurier University in Waterloo, Ontario, Canada in 2009. From 2005-2010 he was an Instructor of Geography at Winthrop University in Rock Hill, South Carolina. Since 2010, Dr. Storie is an Assistant Professor of Geography at the University of Winnipeg, in Winnipeg, Manitoba, Canada where he specializes in the application of remote sensing and GIS to urban systems with a specific focus on the application of synthetic aperture radar for urban structure mapping and monitoring.