A Report on Several CrimeStat Spatial Distribution Analyses: A Ottawa-Nepean Area Case Study, January 2005 to March 2006

Alexander Coster

Table of Contents

Introduction

The use of geographic information systems (GIS) in crime analysis helps shape first response placement as well as more broad law enforcement initiatives and planning. Spatial and temporal patterns of crime can be analyzed to pinpoint areas and times where crimes are more likely to occur, allowing for the efficient allocation of law enforcement resources. The following analyses were conducted on residential and commercial break-and-entry events (B&Es), auto thefts, and population (above the age of 15) statistics in the Ottawa-Nepean area, using information gathered from Statistics Canada, DMTI, and the Ottawa Police Department. The analyses were conducted using CrimeStat 4.01 and displayed using ArcMap 10.5.1. The analyses concluded that crime events in Ottawa-Nepean are in fact spatial and temporally clustered (not random) and remain independent of population densities.

1. Nearest Neighbour Analysis

The nearest neighbor index (NNI) compares the distances between nearest points and distances that would be expected on the basis of chance. The NNI is the ratio of the observed nearest neighbor distance to the mean random distance. If the observed average distance is about the same as the mean random distance, then the ratio will be about 1.0. If points are closer together than expected, it will be less than 1.0 (clustering), and if they are further apart than it will be greater than 1.0 (dispersion) (Levine, 2013b).

All three crime statistics first-order indices (residential, commercial, and car theft) are very close to zero (.29803, .188967, and .264159 respectively). This suggests high levels of spatial aggregation. Index does change as a function of crime: commercial B&Es are the most spatially aggregated, while commercial B&Es are the least. The index values increase for all three as the order increases: the average distance between neighbours increases with higher-order nearest neighbors (kth nearest neighbour is further away than first nearest neighbour), so it is expected that spatial aggregation values would decrease (see Graph 1). As the order increases, Auto Theft and Commercial B&Es become similar in their spatial aggregation (see Graph 1). The index values for Commercial B&Es increase the least and the first order index is the smallest: this suggests that commercial buildings are more spatially aggregated, and remain at the relatively same distance as neighbour order increases (commercial zones are not found at all corners of the city due to zoning laws, usually found in clusters in plazas). Residential zones are closer together in some areas, though they appear all over the city (resulting in the furthest away neighbours being quite far away). The spatial aggregation statistics are supported upon inspectio of city zones: commercial zones are found in the central region, while residentai zones appear all over the city (see Map 1).

2. Moran's I Analysis & Moran's Correlogram

Moran`s I is a global summary statistic for measuring spatial autocorrelation. Spatial autocorrelation is a measure of the degree of influence that the presence of a certain phenomenon has on the clustering or dispersion of other phenomenon. The weighted Moran`s I is similar to a correlation coefficient in that it compares the sum of the cross-products of values at different locations, two at a time, weighted by the inverse of the distance between the locations and with the variance of the variable (Levine, 2013a). High values mean nearby points have similar values. Since most crime incidents are represented as a single point, they do not have associated intensities, so we need to assign the crime incidents to geographical zones and count the number of incidents per zone (dissemination areas).

The Moran Correlogram calculates the I value by different distance intervals (or bins) in order to identify clusters. When graphed, the plot indicates how concentrated or distributed is the spatial autocorrelation. A series of concentric circles is overlaid on the points and the Moran`s I statistic is calculated for only those points falling within each circle; as the circle size increases, the I value approaches the global value (2013a). It provides information about the scale of autocorrelation, whether it is more concentrated or diffuse. This is useful for gauging the extent to which `hot spots`are truly isolated concentrations of incidents or are by-products of spatial clustering over a larger area. The Correlogram examines the spatial intensity of crime (number of times a particular crime occurs at a given location) and the effects of population density on the distribution of crime events by looking at points of crime grouped by DA. A high value would signal a DA with a high number of crimes surrounded by DAs with similar high values, or a DA with low crime values surrounded by DAs with low crime values. A low value close to zero indicates no spatial autocorrelation (a DA with high crimes surrounded by a random mix of low and high crime DAs).

Residential B&Es Moran's I maximize at Bin 1 (446 meters) at .325146 (see Graph 2). Spatial autocorrelation is somewhat strong at this distance for Residential B&Es: DAs with high values are surrounded by other DAs with similar high values for the most part (or low values with low values). Car theft and Commercial B&Es maximize at Bin 2 (892 meters) at .125732 and .066318, respectively. Their best distances do not show very strong spatial autocorrelation, though it is not completely random.

All three crimes have higher Moran's I at close distances (under 10 km) before approaching their global I values, and all three are greater than population values. Crime intensities are not distributed by population; crime in Ottawa is somewhat independent of population. Population represents the baseline, or population dependent, situation: if crime was simply a factor of population, it would be very similar to this line.

This analysis differs from the Nearest Neighbour because it suggests that per Dissemination Area, there is spatial autocorrelation of higher density of crimes for Residential B&Es than

Commercial or Car Thefts. Unlike the Moran correlogram, the NNI which found that spatial aggregation is strongest amongst Commercial B&Es. They both conclude that spatial clustering exists in the crime statistics, however.

3. Fuzzy Membership & Standard Nearest Neighbour Hierarchical Spatial Clustering

Both the Fuzzy membership analysis and Standard Nearest neighbour Hierarchical Spatial Clustering method were used to analyze Residential B&E events. Fuzzy membership mode allows the user to define a small search radius around each location to include events that occur at, around or near that location (Levine, 2013c). The radius is decided by the user. The aim of the statistic is to allow the identification of locations where a number of incidents may occur, but where there may not be precision in measurement. It can identify high incident locations more precisely. The dbf file has four output variables:

i. Location with the most incidents, then next location, then next.
ii. Frequency of events at the location
iii. X coordinate
iv. Y coordinate

Areas with the highest frequency of crime are displayed as points. The search radius around each point (which includes other incidents) was set to 1000 meters. Blue points represent points of the lowest frequency of incidents, while orange and red points represent higher rates of crime incidents (see Map 2).

Nearest Neighbor Hierarchical Clustering (Nnh) identifies hot spots, or groups of incidents that are spatially close (opposed to individual points that are clustered or are the center of a cluster), based on several criteria. It repeats the routine until either all points are grouped into a single cluster or else the clustering criteria fail, has a defined threshold distance, and compares this to distances of all pairs of points (2013c). Only points that are closer to one or more points than the threshold distance and belong to a group having the minimum number of points are selected for clustering at the first level. Subsequent clustering produces a hierarchy of clusters, which continues until either all clusters converge into a single cluster or, more likely, the clustering criteria fails.

Instead of a closeness value, NNH determines hot spots in the data through statistical criteria, using ellipses of 1000m (same radius as Fuzzy membership), however the minimum number of points is ten (need at least ten points in each cluster). The NNH illustrates a similar pattern to the Fuzzy membership: the majority of the NNH ellipses are clustered in the downtown area of Ottawa (Lower Town, Chinatown, Little Italy, New Edinburgh, Vanier, Glebe neighbourhoods) (see Image 1). This tells us that Residential B&Es are clustered downtown. Overall, the two analyses line up fairly well (see Map 3).

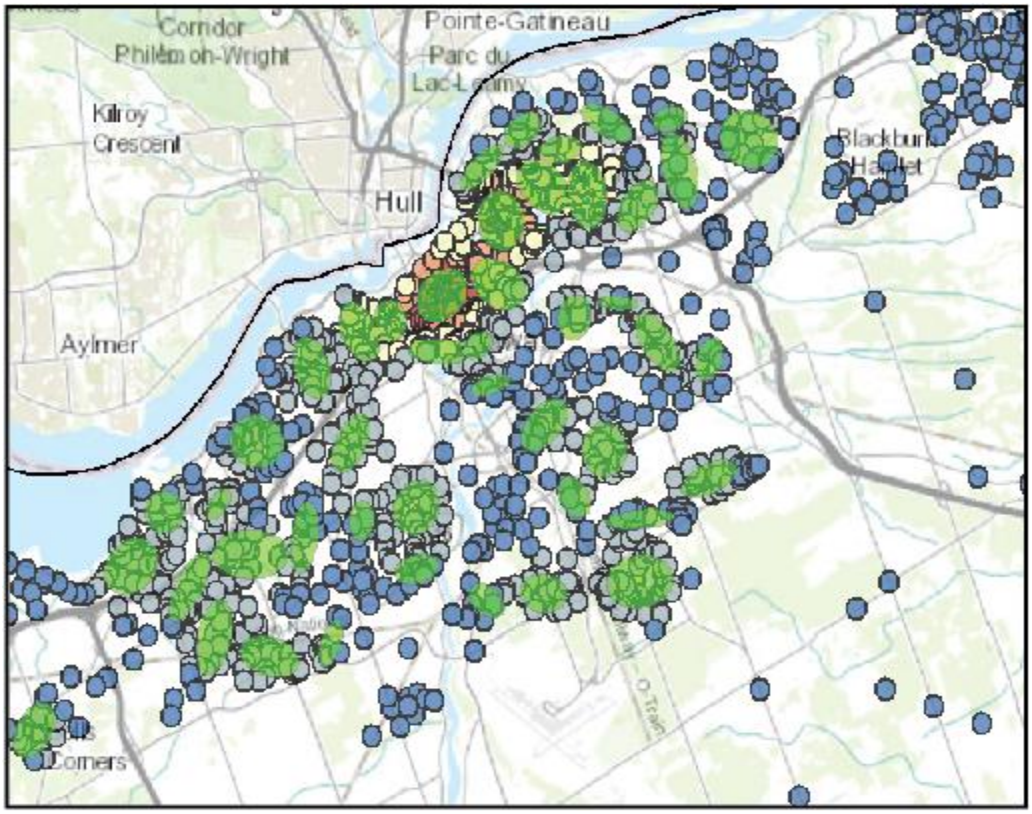


Image 1: Fuzzy membership points & Nearest Neighbour Hierarchical (NNH) Spatial Clustering ellipses results in the Lower Town, Chinatown, Little Italy, New Edinburgh, Rockcliffe Park, Vanier, and Glebe neighbourhoods of Ottawa.

There is some disagreement in areas around Orleans and Manor Park; the Fuzzy membership method detected many hot spot points of residential B&Es, though they seem to be too spatially dispersed to create NNH ellipses (the hot spots are more dispersed) (see Image 2). The majority of the rest of the city's crime hot spots identified by fuzzy membership are much too dispersed to be considered by the NNH analysis. Residential B&Es are mostly dispersed outside of Ottawa's downtown area.

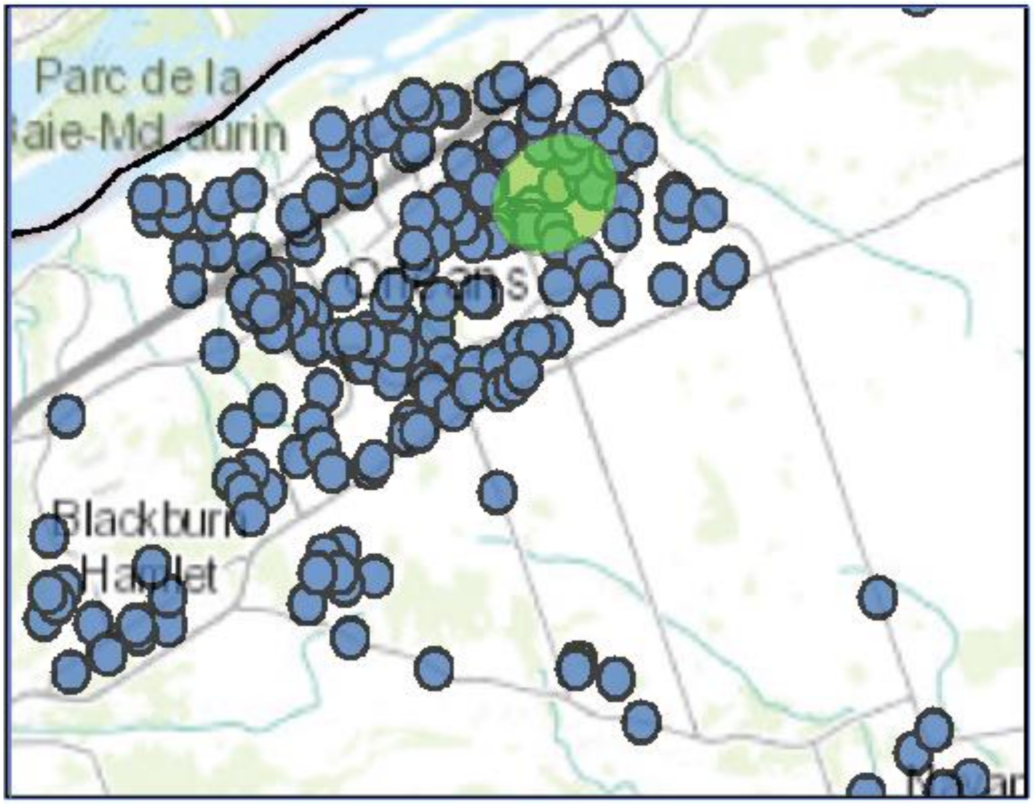
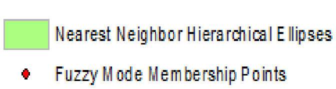


Image 2: Fuzzy membership points & NNH ellipses do not line up, near Manor Park.

The pattern of hot spots corresponds to the distribution of residential zones in Ottawa, which are densely clustered in the same neighborhoods where the fuzzy points and NNH ellipses are located (see Map 1). The neighbourhoods that comprise Orleans and Manor Park are also covered disproportionally by residential zones; it is interesting that the NNH analysis did not include this area more thoroughly (see Image 2). There seems to be hot spots of crime here, but less patterns of clustering. There are clearly small geographical environments where there are concentrated incidents; police can use this for specific targeting.

4. **Standard & Risk-Adjusted Nearest Neighbour Hierarchical Spatial Clustering**

Risk-adjusted nearest neighbor hierarchical clustering (Rnnh) combines the hierarchical clustering capabilities with a kernel density interpolation technique (Levine, 2013c). Groups of points closer than the threshold with a greater than minimum number of points are identified by the Nnh routine and combined with the baseline population. It defines clusters of points that are closer than what would be expected on the basis of a baseline population by dynamically adjusting the threshold distance in the Nnh routine according to the distribution of the baseline variable (2013c). Threshold distance is not constant throughout the study area; the threshold distance changes inversely proportional to the population density of the location (high density DAs will have shorter threshold distances). Rnnh is a risk measure rather than a volume measure (2013c). With the 2nd and 3rd order clusters, each previous order's clusters' centers are treated as the new points, repeated until no further clustering can be conducted or all are



converged. The higher orders of clusters show linkages; there are hot spots of hot spots. Police may be interested in either variable: they may want to simply know where the most crimes are being committed, or where crimes are proportionally higher compared to population density. In this case, the NNH and RNNH were used to analyze Residential B&Es as a function of population.

The relative risk of a residential B&E may be less in a densely populated area than sparsely populated one. However, the basic assumption is that as population increases, crime will increase also. This basic assumption is proven correct, due to the high number of 1st and 2nd order clusters in the downtown area, as well as the only 3rd order cluster which encompasses downtown Ottawa (Residential B&E events cluster relative to the number of people living in DAs) (see Map 4). This also correlates with DAs that are mostly occupied by residential zones (see Map 1). Downtown neighbourhoods directly south of Ottawa River have crime rates that are significantly clustering given their high population densities.

The Risk-Adjusted NNH also recognizes a 2nd order cluster of residential B&Es in the Orleans area, which the NNH did not. Crime is higher here than what is expected based on the population. However, although there are many fuzzy membership method points in the neighbourhoods directly south of downtown Ottawa (Barrhaven, Honeygables, Hearts Desire), the RNNH doesn't display any clusters there.

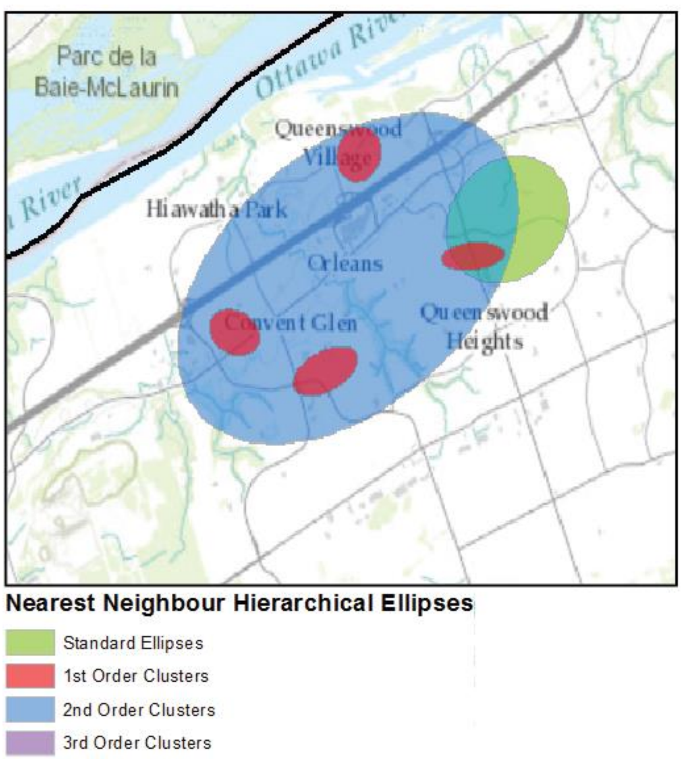


Image 3: Risk-Adjusted NNH depicts clusters of residential B&Es in Orleans, where the other NNH did not. There is risk-adjusted spatial autocorrelation..

While the 1st order areas (red ellipses) may warrant direct police intervention more routinely, the higher orders may signal different responses. 2nd order areas (blue ellipses) could be elevated patrol areas, while 3rd order areas (purple ellipses) should be thought of as integrated management strategy zones (crime prevention, community involvement, long-range planning) (2013c).

5. **Space-time Analysis: The Knox Index**

The Knox test evaluates spatial-temporal interaction using specified distance and time thresholds. The ‘close’ time (or clustering time) was set to 6 hours, and the ‘close’ distance was set to 5km. There can be spatial clustering within a time period. There can also be space-time clustering: a number of events could occur within a short time period within a concentrated area, which is very common among auto thefts (car thief gang may binge attack one neighbourhood for an hour before moving on; the cluster moves, and there is an interaction between space and time) (Levine, 2013e). There can also be space-time interaction (relationship between space and time is more complex) (see Appendix A). Auto Thefts were examined using the Knox Index to determine spatial and temporal clustering of these events by comparing pairs of points in terms of distance and time to identify spatial-temporal hot spots. The analysis identifies the number of pairs of points that fall into each of the listed categories, and sorted into a ‘close time, close distance table’ (see Appendix A).

According to the Knox Index, there are more Auto Thefts that are spatially and temporally interacting than expected; about 3,147 more incidents than expected inside six hours and five kilometers (see Appendix A). This could be the result of clustering. The Chi-square is quite high (94.016): this indicates spatial-temporal interaction, and a marked difference between the observed number of pairs in each cell and the expected number. The P value is .0001, meaning that the Chi Square is significant.

6. **Comparison of Single and Dual Kernel Density Interpolation Spatial Clustering Analyses**

The single kernel density routine in CrimeStat is applied to a distribution of point locations, such as crime incidents. The primary file is the location of Residential B&Es. Absolute density was used for the calculation, which means estimates at each reference cell are re-scaled so that the sum of the densities over all reference grids equals the total number of incidents. That is, the estimate is the number of incidents/points that occurred in each grid cell (2013d). Kernel estimates are a hot spot identifier as well, but it is a continuous surface: the densities are calculated at all locations. For both single and dual, a triangular (conical) distribution was used: it emphasizes ‘peaks’ and ‘valleys’. The output files are polygon grid shapefiles. The kernel estimates are good methods of visually representing areas that have a higher risk of residential B&Es.

The single kernel matches up very well with the fuzzy membership points and NNH ellipses: downtown, directly south of Ottawa river, where the points are dense, is matched by the denser kernel values (.047 - .101) (see Map 5). To the west (Constance Bay, Fitzroy Harbour), kernel density values are lower and fuzzy points are dispersed. Kanata, Eaglesons Corners, Bridlewood, Hazeldean, Lynwood Village, Bellwood, and Westcliffe Park all line up well. This is due to the single kernel estimate and fuzzy method points being independent of other factors (population) and just absolute counts of residential B&Es. RNNH ellipses also line up with single kernel estimates in the downtown areas and Orleans area neighbourhoods. However, neighbourhoods to the south and west are underrepresented by ellipses, though are still considered hot spots by the kernel estimate (see Image 4). Old Stittsville and Stittsville, which has many fuzzy membership points and high kernel density estimates, lacks any RNNH ellipses. Overall, the highest densities of crime estimated by the single kernel occur in the downtown area directly south of Ottawa River.

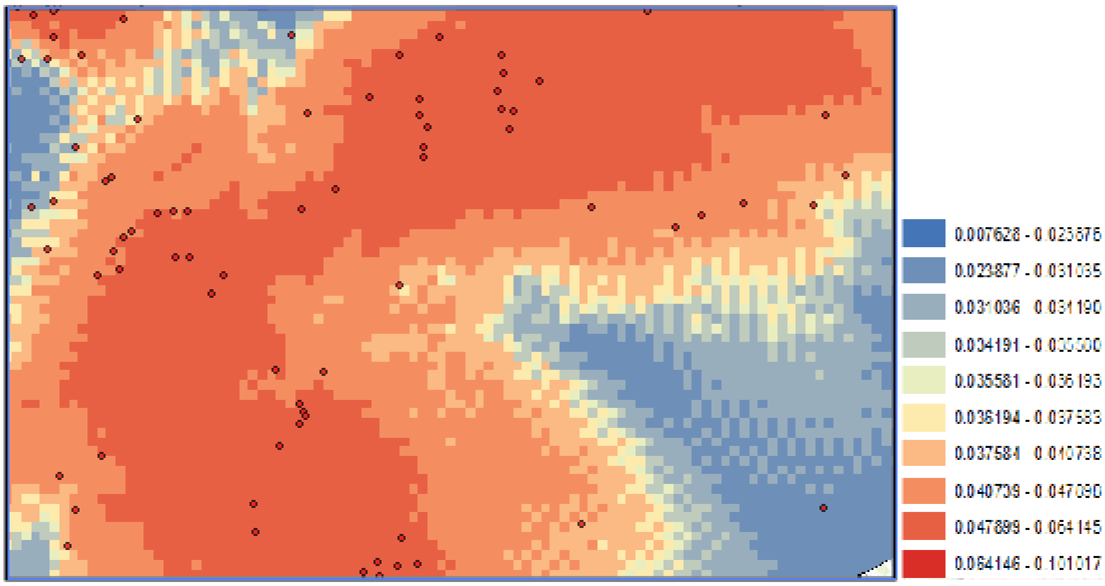


Image 4: Higher density estimates by single kernel function in neighbourhoods south of downtown Ottawa (Greely, Herbert Corners, Waterson Corners, Osgoode) where RNNH ellipses did not detect clustering.

The dual kernel density routine is applied to two distributions: Residential B&Es and populations (over 15) for Dissemination Areas. The kernel estimate for B&Es (primary file) is divided by the kernel estimate for population (secondary file) for each grid cell. So in theory, an areas with a dense population will probably receive a lower relative estimate of B&Es than an area of equal B&E statistics with lower population density. Where the single kernal is a volume estimation, the dual kernel is more of a risk estimation.

There is greater disagreement between the dual kernel estimate and the NNH and fuzzy membership in downtown Ottawa directly south of Ottawa River, but greater agreement with the RNNH analysis (see Map 6). Due to high population density, the dual kernel estimates a lower relative risk of B&Es in the downtown area: there is a smaller ratio of crimes to population. There is an even greater difference of statistical analysis outcomes in the Kanata, Hazel Dean and Eaglesons Corners area: a high number of fuzzy membership points and NNH ellipses coincides with very low dual kernel estimates (very low B&E to population ratio) (see image 5). Bellwood, Lynwood Village and Barrhaven, however, line up with all three functions.

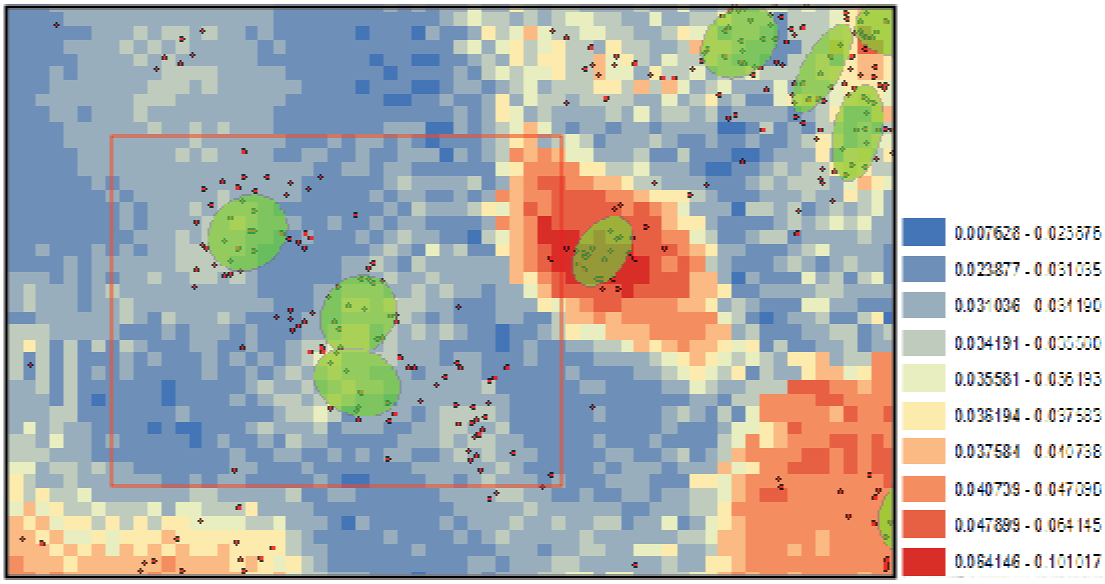


Image 5:

Image 5: Neighbourhoods of Kanata, Hazel Dean and Eaglesons. Dual kernel estimates, fuzzy membership and NNH ellipses showing different statistical outcomes

The biggest difference between the single and dual kernel estimates is that the dual kernel's concentrations of crime extend outward from the city centre whereas the single kernel stays mostly concentrated in the centre. There are large swaths of land to the south and east of downtown which are designated at higher risk of Residential B&E activity that the other analyses do not show. Simply put, there is a greater risk of crime outside of the city centre.

7. Conclusion

There is spatial aggregation, autocorrelation, and significant clustering of crime events. The Nearest Neighbour Analysis illustrates how the distribution of crimes are related to the placement of different zones. The Moran's I Correlogram indicates higher levels of spatial autocorrelation independent of population distance. The Fuzzy membership analysis, NNH, and RNNH reinforce the hypothesis of spatial clustering of Residential B&Es. The single and dual kernel density estimates show the distribution of Residential B&Es as continuous surfaces, allowing for further visual interpretation by the user. CrimeStat performed the analyses which allowed us to conclude that all three crime types in the Ottawa-Nepean area (Residential B&Es, Commercial B&Es and Auto Thefts) are neither randomly occurring or dependent upon population densities.

Appendix A: Results of Knox Index analysis on Auto Theft Incidents in Ottawa-Nepean, January 2005 - March 2006

```
Knox Index: Interaction of Space and Time
-----------

    Sample size ..........:  2152
    Measurement type ......:  Direct
    Input units .... ......:  Meters
    Time units ............:  Hours
    Simulation runs .......:  19
    Start time ............:  12:47:45 PM, 02/27/2018

    "Close" time ........:          6.00000 hours
    "Close" distance ....:       5000.00000 m

                         | Close in space(1)  | Not close in space(0) |
    ---------------------+--------------------+-----------------------+----------------
    Close in time(1)     |             325473 |                986929 |       1312402
    Not close in time(0) |             242964 |                759110 |       1002074
    ---------------------+--------------------+-----------------------+----------------
                         |             568437 |               1746039 |       2314476

    Expected:
                         | Close in space(1)  | Not close in space(0) |
    ---------------------+--------------------+-----------------------+----------------
    Close in time(1)     |       322326.89199 |          990075.10801 | 1312402.00000
    Not close in time(0) |       246110.10801 |          755963.89199 | 1002074.00000
                         |                    |                       |
                         |       568437.00000 |         1746039.00000 | 2314476.00000

    Chi-square ..........:          94.01612
    P value of Chi-square:           0.00010

    End time .............:  12:47:45 PM, 02/27/2018

    Distribution of simulated index (percentile):
    Percentile      Chi-square
    ---------- ---------------
           min         0.09891
           0.5         0.09891
           1.0         0.09891
           2.5         0.09891
           5.0         0.09891
          10.0         0.17390
          90.0         6.84384
          95.0         7.66375
          97.5         7.66375
          99.0         7.66375
          99.5         7.66375
           max         7.66375

    Simulation ended ......:  12:47:48 PM, 02/27/2018
```

References

Klinkenberg, Brian (2017). Lab 4: Crime analysis using CrimeStat [Web blog post]. Retrieved from http://ibis.geog.ubc.ca/courses/geob479/labs/lab4.htm

Levine, Ned (2013a). Chapter 5: spatial autocorrelation statistics. *Crimestat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations,* 4. Retrieved from http://ibis.geog.ubc.ca/courses/geob479/notes/crime/CrimeStat%20IV%20Chapter%205.pdf

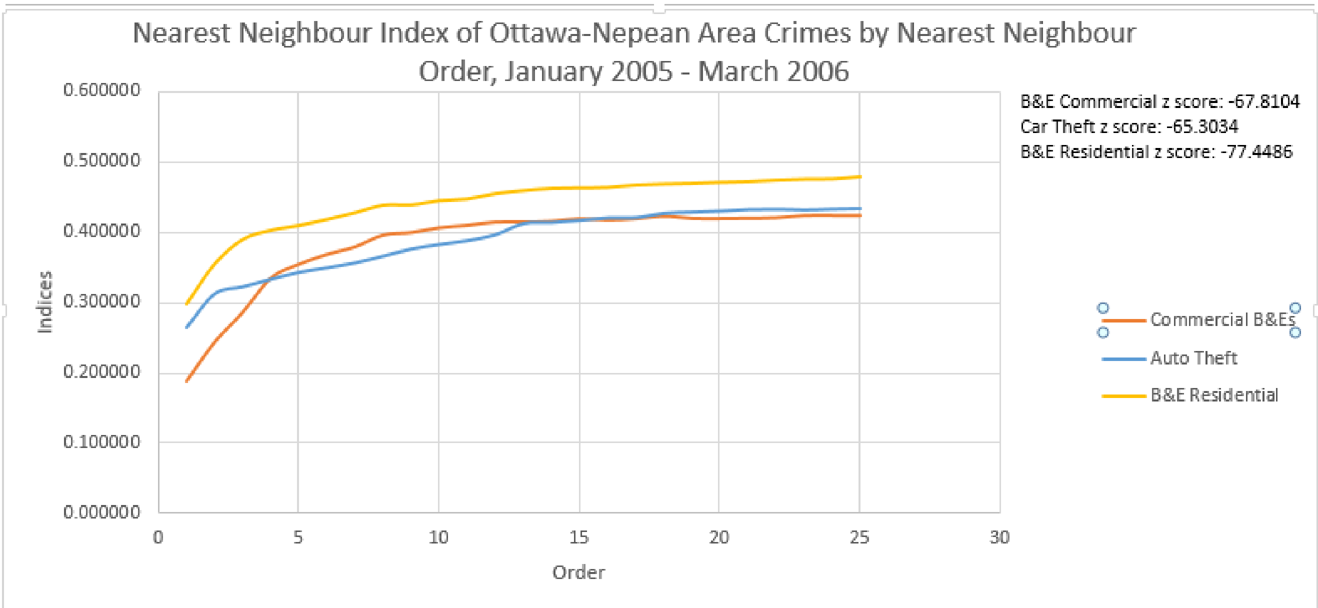
Levine, Ned (2013b). Chapter 6: distance analysis I and II. *Crimestat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations,* 4. Retrieved from http://ibis.geog.ubc.ca/courses/geob479/notes/crime/CrimeStat%20IV%20Chapter%206.pdf

Levine, Ned (2013c). Chapter 7: hot spot analysis of points: I. *Crimestat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations,* 4. Retrieved from http://ibis.geog.ubc.ca/courses/geob479/notes/crime/CrimeStat%20IV%20Chapter%207.pdf

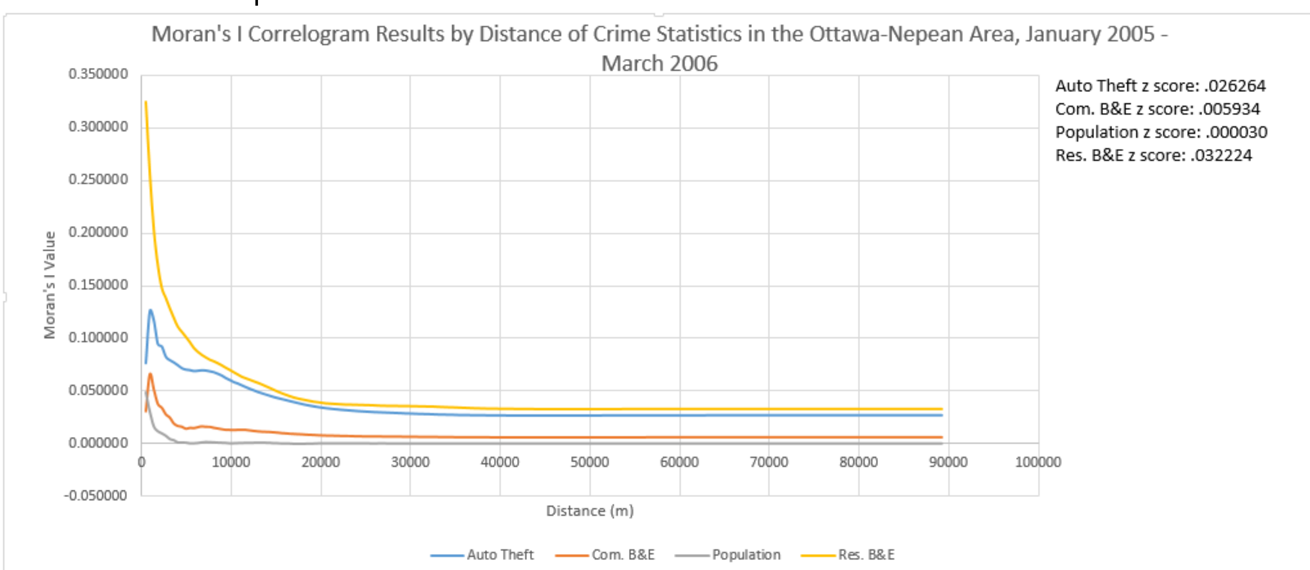
Levine, Ned (2013d). Chapter 10: kernel density interpolation. *Crimestat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations,* 4. Retrieved from https://www.nij.gov/topics/technology/maps/documents/crimestat-files/CrimeStat%20IV%20Chapter%2010.pdf

Levine, Ned (2013e). Chapter 12: space-time analysis. *Crimestat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations,* 4. Retrieved from https://www.nij.gov/topi)cs/technology/maps/documents/crimestat-files/CrimeStat%20IV%20Chapter%2012.pdf

Graph 1: Nearest Neighbour Analysis Results Of Crimes in Ottawa-Nepean Area

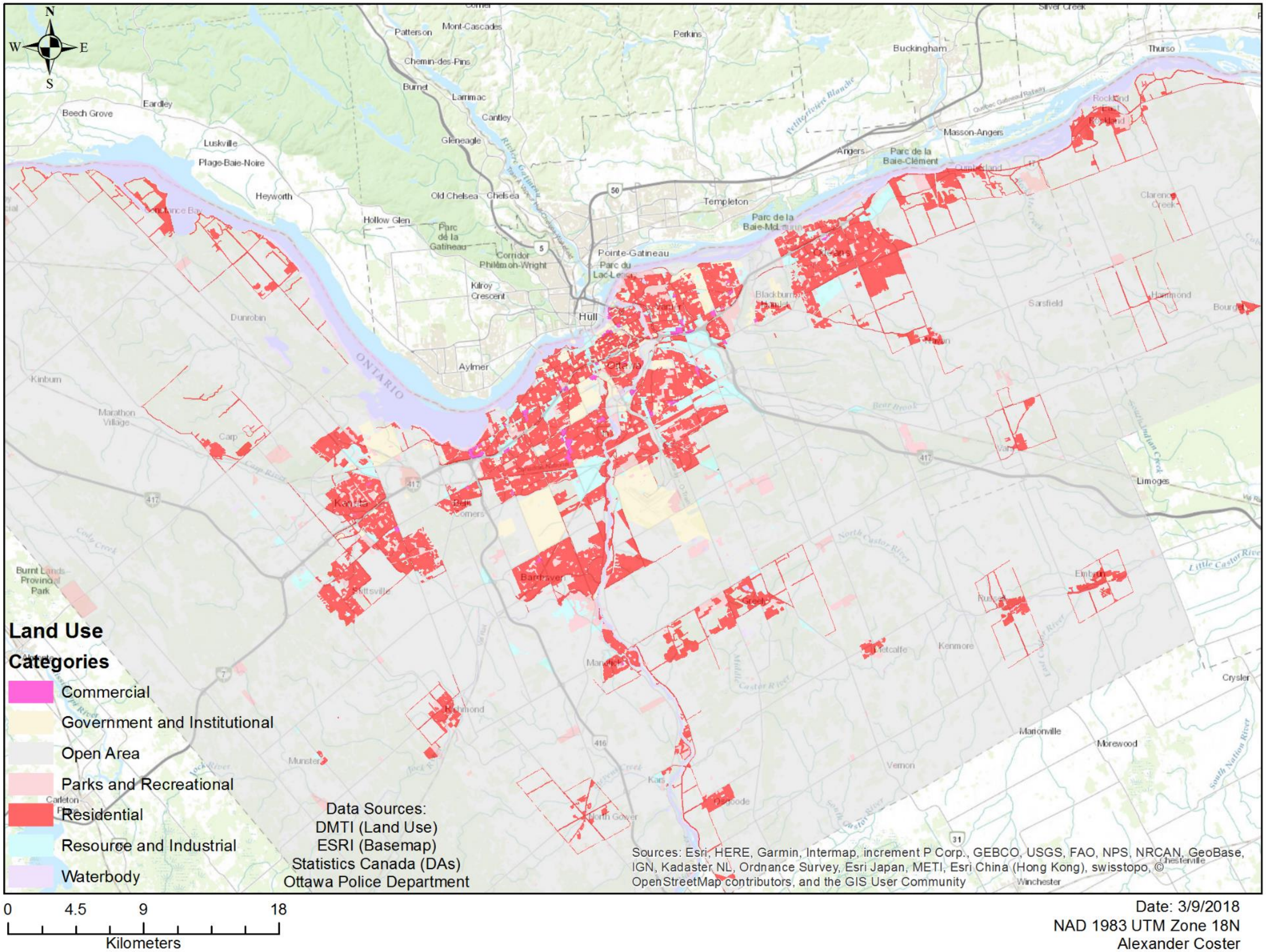


Graph 2: Moran's I Correlogram of Auto Theft, Commercial B&Es, Population, and Residential B&Es in Ottawa-Nepean Area

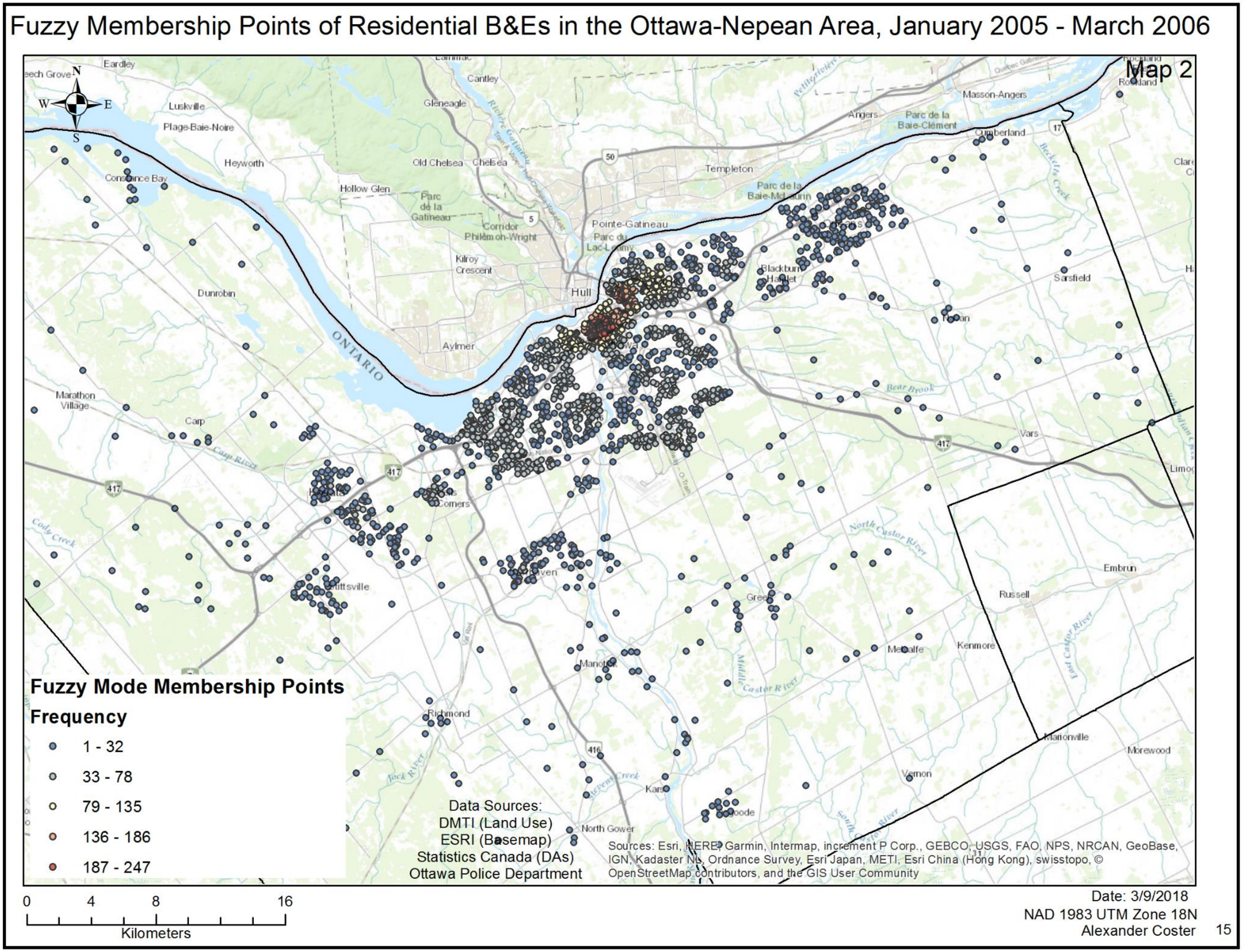


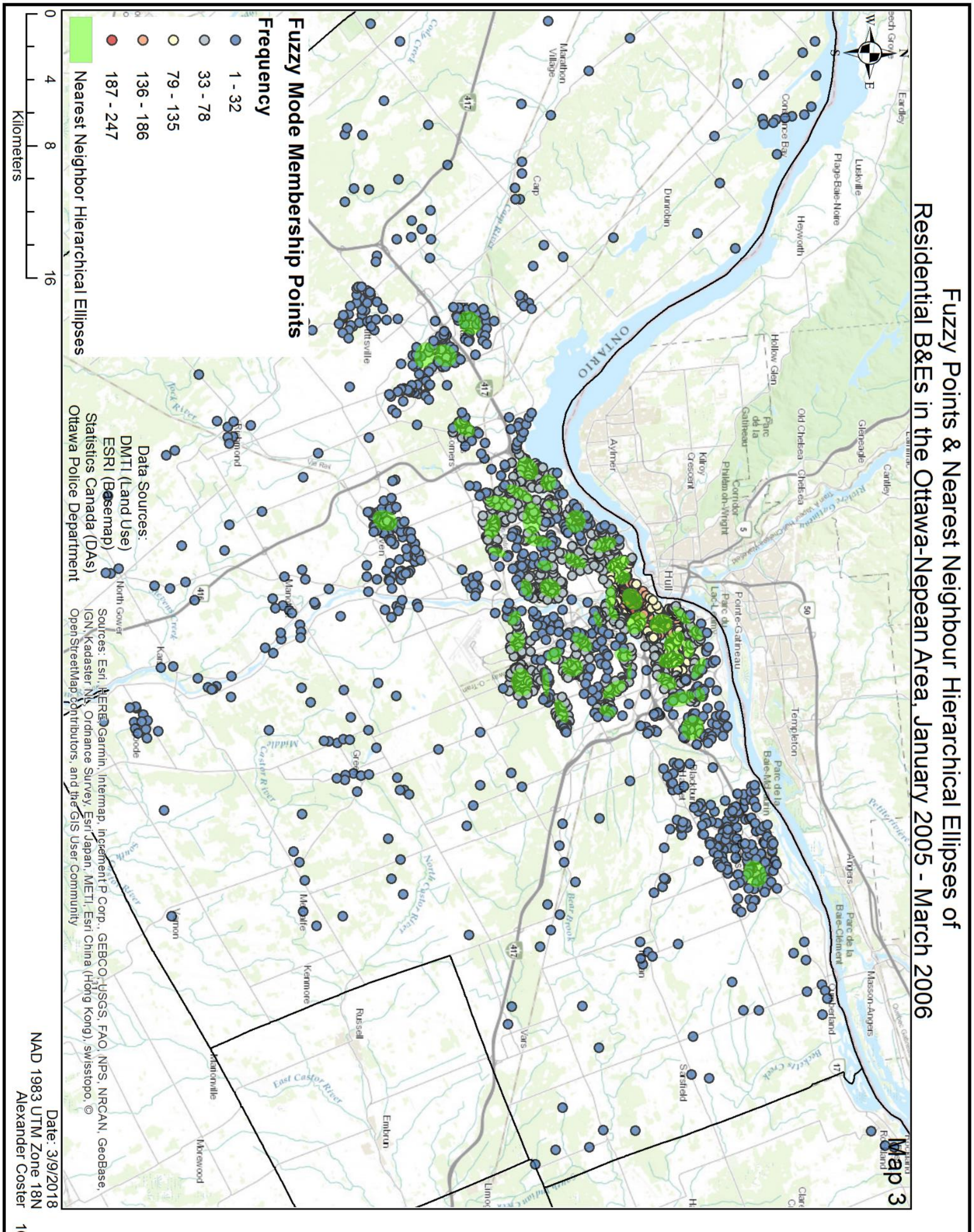
Land Use Categories of Ottawa-Nepean Area

Map 1

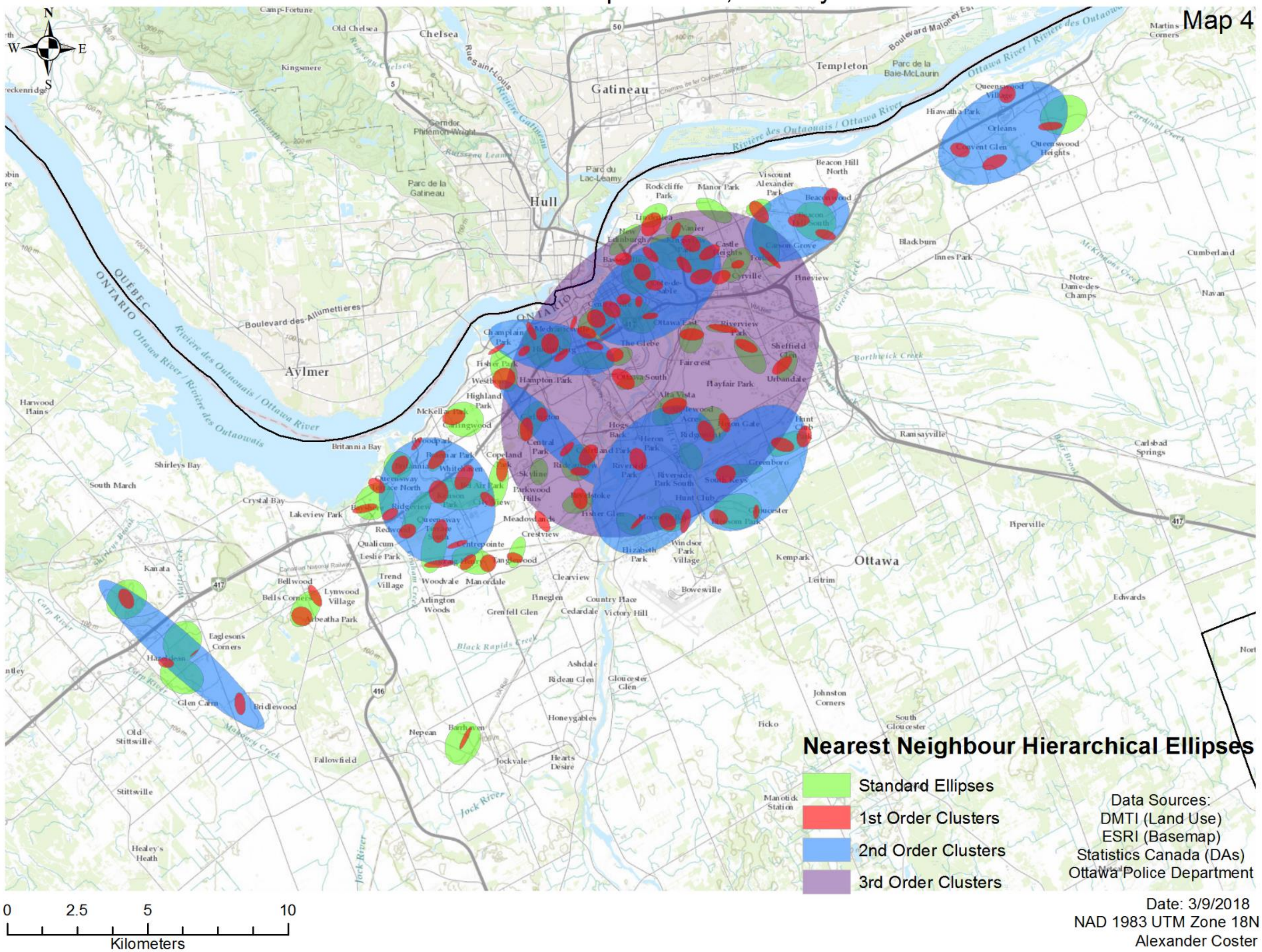


Date: 3/9/2018
NAD 1983 UTM Zone 18N
Alexander Coster

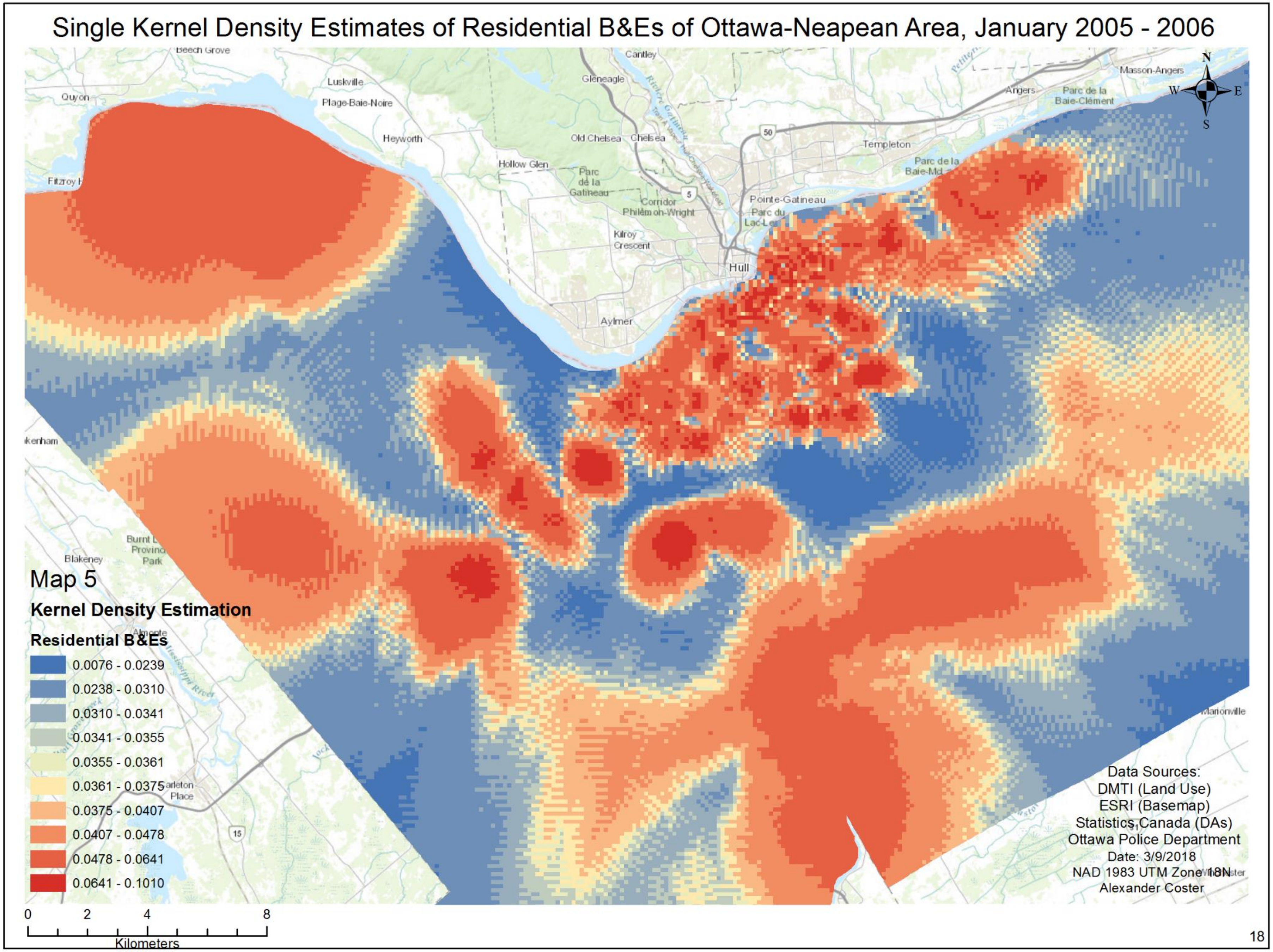

Fuzzy Membership Points of Residential B&Es in the Ottawa-Nepean Area, January 2005 - March 2006
Map 2
Fuzzy Mode Membership Points
Frequency
1 - 32
33 - 78
79 - 135
136 - 186
187 - 247
0 4 8 16
Kilometers
Data Sources:
DMTI (Land Use)
ESRI (Basemap)
Statistics Canada (DAs)
Ottawa Police Department
Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community
Date: 3/9/2018
NAD 1983 UTM Zone 18N
Alexander Coster

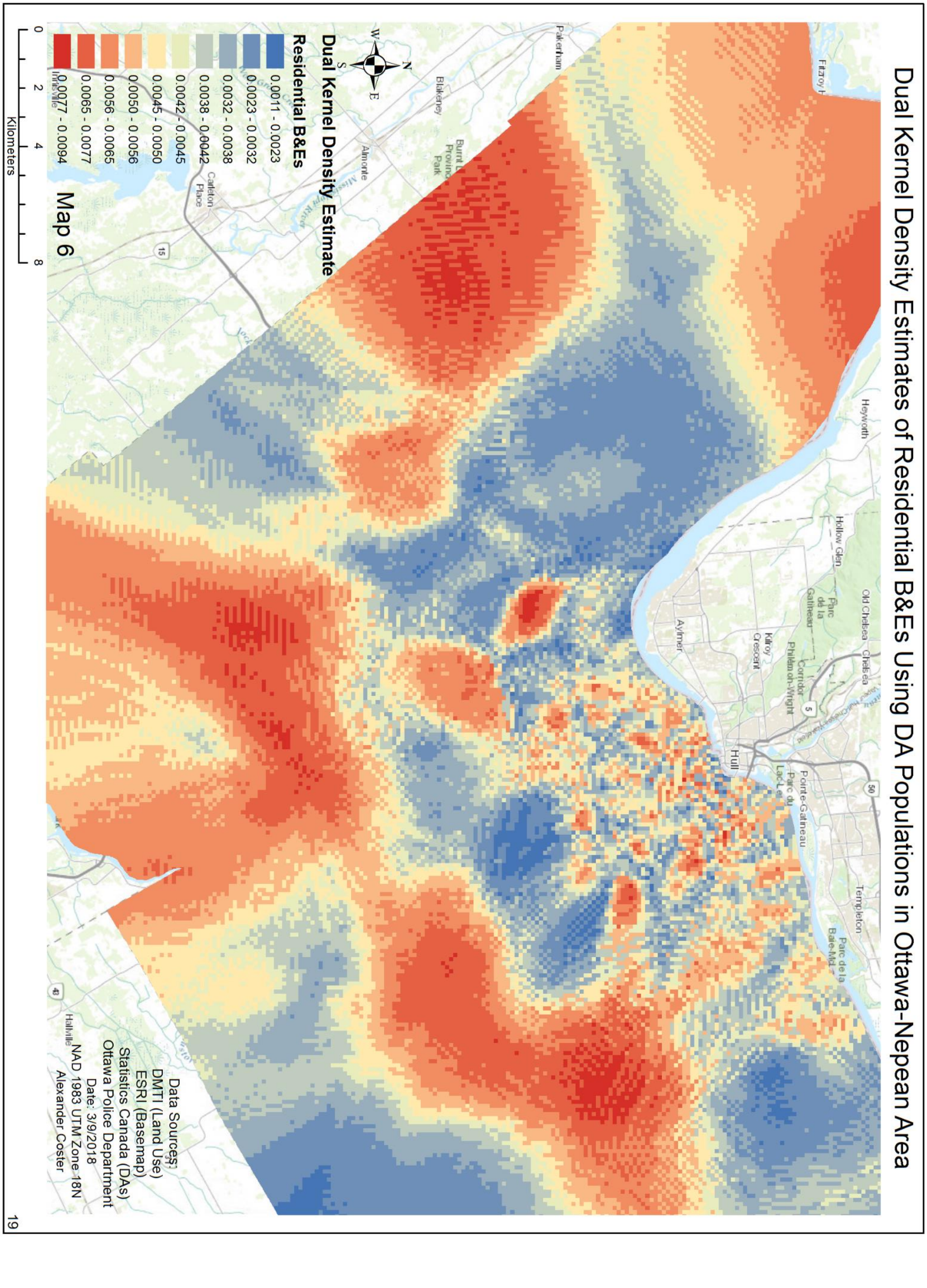

Map 3
Fuzzy Points & Nearest Neighbour Hierarchical Ellipses of
Residential B&Es in the Ottawa-Nepean Area, January 2005 - March 2006
Fuzzy Mode Membership Points
Frequency
1 - 32
33 - 78
79 - 135
136 - 186
187 - 247
Nearest Neighbor Hierarchical Ellipses
0 4 8 16
Kilometers
Data Sources:
DMTI (Land Use)
ESRI (Basemap)
Statistics Canada (DAs)
Ottawa Police Department
Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase,
IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, ©
OpenStreetMap contributors, and the GIS User Community
Date: 3/9/2018
NAD 1983 UTM Zone 18N
Alexander Coster 16
ONTARIO

Standard & Risk-Adjusted Nearest Neighbour Hierarchical Ellipses of Residential B&Es in the Ottawa-Nepean Area, January 2005 - March 2006



Date: 3/9/2018
NAD 1983 UTM Zone 18N
Alexander Coster


Single Kernel Density Estimates of Residential B&Es of Ottawa-Neapean Area, January 2005 - 2006
Map 5
Kernel Density Estimation
Residential B&Es
0.0076 - 0.0239
0.0238 - 0.0310
0.0310 - 0.0341
0.0341 - 0.0355
0.0355 - 0.0361
0.0361 - 0.0375
0.0375 - 0.0407
0.0407 - 0.0478
0.0478 - 0.0641
0.0641 - 0.1010
0 2 4 8
Kilometers
Data Sources:
DMTI (Land Use)
ESRI (Basemap)
Statistics Canada (DAs)
Ottawa Police Department
Date: 3/9/2018
NAD 1983 UTM Zone 18N
Alexander Coster
Quyon
Luskville
Plage-Baie-Noire
Heyworth
Hollow Glen
Old Chelsea
Chelsea
Gleneagle
Cantley
Aylmer
Hull
Pointe-Gatineau
Templeton
Angers
Masson-Angers
Parc de la Baie-Clément
N
S
E
W
18


Dual Kernel Density Estimates of Residential B&Es Using DA Populations in Ottawa-Nepean Area
Dual Kernel Density Estimate
Residential B&Es
0.0011 - 0.0023
0.0023 - 0.0032
0.0032 - 0.0038
0.0038 - 0.0042
0.0042 - 0.0045
0.0045 - 0.0050
0.0050 - 0.0056
0.0056 - 0.0065
0.0065 - 0.0077
0.0077 - 0.0094
Map 6
0 2 4 8
Kilometers
N
W E
S
Data Sources:
DMTI (Land Use)
ESRI (Basemap)
Statistics Canada (DAs)
Ottawa Police Department
Date: 3/9/2018
NAD 1983 UTM Zone 18N
Alexander Coster
19