APPENDIX I

HOT SPOT ANALYSIS

Hot spots are concentrations of incidents within a limited geographical area that appear over time. ‘Hot spots’ may not exist in reality, but could be areas where there is sufficient concentration of certain activities (in this case, murder incidents) such that they get labeled as being an area of high concentration. There are no boundaries around these incidents, but a gradient where people draw an imaginary line to indicate the location at which the hot spot starts. There are literally dozens of different statistical techniques designed to identify ‘hot spots’ (Levine, 2004). For this study, one of the two Hierarchical techniques named Nearest Neighbor Hierarchical Clustering was employed to identify and understand hot spots of murder.

NEAREST NEIGHBOR HIERARCHICAL CLUSTERING (Nnh)

The nearest neighbor hierarchical clustering (Nnh) routine in CrimeStat identifies groups of incidents that are spatially close. It is a hierarchical clustering routine that clusters points together on the basis of a criteria. The clustering is repeated until either all points are grouped into a single cluster or else the clustering criteria fails. The CrimeStat Nnh routine uses a method that defines a threshold distance and compares the threshold to the distances for all pairs of points. Only points that are closer to one or more other points than the threshold distance are selected for clustering. In addition, the user has specify a minimum number of points to be included in a cluster (in this study, 10 minimum points is specified which is also the default number of points). Only points that fit both criteria - closer than the threshold and belonging to a group having the minimum number of points, are clustered at the first level (first-order clusters). The routine then conducts subsequent clustering to produce a hierarchy.
of clusters. The first-order clusters are themselves clustered into second-order clusters. Again, only clusters that are spatially closer than a threshold distance (calculated anew for the second level) are included. The second-order clusters, in turn, are clustered into third-order clusters, and this re-clustering process is continued until either all clusters converge into a single cluster or, more likely, the clustering criteria fails (Levine, 2004).

**Criteria 1: Threshold Distance**

The first criteria in identifying clusters is whether points are closer than a specified threshold distance. Random nearest neighbor distance (the default) was chosen for analysis. The mean random distance was defined as:

$$\text{Mean Random Distance} = d(\text{ran}) = 0.5 \sqrt{\frac{A}{N}}$$  \hspace{1cm} (A. 1)

*Source: (Levine, 2004).*

where A is the area of the region and N is the number of incidents. The confidence interval around that distance is defined as:

$$\text{Confidence Interval for Mean Random Distance} = \text{Mean Random Distance} \pm t^* \text{SE}(d(\text{ran}))$$

$$= 0.5 \sqrt{\frac{A}{N}} \pm 0.26136 \sqrt{\frac{N^2}{A}}$$  \hspace{1cm} (A. 2)

*Source: (Levine, 2004).*

where A is the area of the region, N is the number of incidents, t is the t-value associated with a probability level in the Student’s t-distribution (Levine, 2004).
The lower limit of this confidence interval is:

\[
\text{Lower Limit of Confidence Interval for Mean Random Distance} = 0.5 \sqrt{\frac{A}{N}} - t \sqrt{\frac{N^2}{A}} \quad (A. 3)
\]

*Source: (Levine, 2004)*.

and the upper limit of this confidence interval is:

\[
\text{Upper Limit of Confidence Interval for Mean Random Distance} = 0.5 \sqrt{\frac{A}{N}} + t \sqrt{\frac{N^2}{A}} \quad (A. 4)
\]

*Source: (Levine, 2004).*

The confidence interval defines a probability for the distance between any pair of points. In other words, the threshold distance is a probability level for selecting any two points (a pair) on the basis of a chance distribution (Levine, 2004).

**Criteria 2: Minimum Number of Points**

Whatever method is used for selecting a threshold distance, a second criteria is the minimum number of points that are required for each cluster. This criteria is used to reduce the number of very small clusters. To minimize numerous very small clusters as well as reduce the likelihood that clusters could be found by chance, a minimum number was set and for this study it was 10 (default). By decreasing this number, more clusters are produced; conversely, by increasing this number, fewer clusters are produced. The routine will only include points in the final clustering that are part of groups (or clusters) in which the minimum number is found (Levine, 2004). Figure A. 1 shows the Hot Spot Analysis carried out using CrimeStat III software.
Visualizing the Cluster Output

To identify the approximate cluster location, *Convex hull* output was chosen. A convex hull is calculated for each cluster. The convex hull draws a polygon around the points in the cluster. It is a literal definition of the cluster, as opposed to the ellipse which is an abstraction. The convex hull is saved in ArcView ‘.shp’ format (*Levine, 2004*) and generated as maps.

**Fig. A. 1:** Hot Spot Analysis in CrimeStat III