ASSESSING POSITIONAL UNCERTAINTY IN GEOCODED DATA

TONY H. GRUBESIC*
Alan T. Murray**

*Department of Geography, University of Cincinnati
USA

**Department of Geography, The Ohio State University
USA

ABSTRACT

The assignment of geo-referenced coordinates to individual households, known as geocoding, is a fundamental and important task for urban data management. Not surprisingly, geographic information systems (GIS) play a critical role in processing the spatial information used for coordinate assignment. There are a wide variety of applications where geocoded information is relied upon for the efficient and effective delivery of services. For example, emergency planning/response efforts, crime mapping and analysis, public facility location modeling and the calculation of municipal tax exposure utilize geocoded data. However, because the results of the geocoding process are largely dependent on the quality of the TIGER-based street files used for geo-referencing, as well as the assumptions built into them, there is the potential for introducing significant spatial inaccuracies during address conversion. The purpose of this paper is to both explore and document the problems associated with the geographic base files used for coordinate assignment. We provide an extended empirical example that highlights the relative uncertainty of household location in geographic space. This includes an assessment of positional uncertainty through the spatial perturbation of geocoded points within an established bound – allowing for alternative, yet equally accurate, spatial realizations of geocoded data. Sensitivity analysis is then conducted to evaluate the significance of locational uncertainties in geocoded data and spatial analysis

1. INTRODUCTION

The use of geographic information systems (GIS) in both public and private agencies has grown considerably over the past 15 years. In part, this can be attributed to significant increases in computing power (e.g. processor speeds, memory and storage) and modest decreases in hardware and software costs. The growing popularity of GIS can also be linked to an increased level of user friendliness and accessibility, particularly with commercial desktop GIS packages such as ArcView and MapInfo (Thrall and Thrall, 1998). As the use of commercial GIS products increases, there are parallel gains in the use of spatial information and data. For example, in the

public sector, urban and regional planners make extensive use of spatial data. This includes the use of spatial information to better allocate public services (e.g. fire and police) and to locate new public facilities (highways, parks, community health clinics and recreation centers). In all cases, spatial information combined with the use of GIS makes urban planning more efficient, cost effective and accurate (Yeh, 1999).

One of the more powerful capabilities of commercial GIS packages is their ability to integrate and use disaggregate spatial information for analysis. Broadly defined, disaggregate spatial information is geographically referenced data that contain a high level of spatial resolution. For example, a disaggregate database for a supermarket can contain thousands of individual customer records and their shopping preferences, coded by street address. That same information can be aggregated to a more general level (e.g. postal code or county) for statistical analysis if necessary, but individual address records have become the standard level for spatial investigation in a wide variety of socioeconomic and planning applications (Ratcliffe, 2001). The concept of spatially enabling disaggregate data, such as a postal address, is fairly easy to understand. One is simply trying to assign grid references (e.g. UTM coordinates) to point locations described by street addresses. For example, a commercial GIS package such as ArcGIS can take a street address and zip code (e.g. 3230 Durbin Road, Cincinnati, OH 45213) and assign a set of geographic coordinates, such as latitude and longitude (39.180850, -084.426171). Clearly, this ability makes both GIS and geocoding an indispensable tool for visualizing and analyzing a wide variety of spatial information.

As mentioned previously, the use of geocoded data is common in both the public and private sector. A wide variety of research institutes, government agencies and commercial organizations collect address information and maintain large spatially enabled databases. These databases include property tax records, electoral registers, customer databases and emergency 911 locations. However, given the increasing reliance on spatial data for critical municipal and social services, it is extremely important that the spatial accuracy of these databases be considered. For example, failure to make the appropriate connection between an emergency location and the routing of a response team (fire, police or medical) could lead to catastrophic property damage or a loss of life (Grubesic and Murray, 2004).

Several studies have examined the spatial accuracy of geocoded records. For example, Krieger et al. (2001) used a test file of 70 addresses, 50 of which contained some type of error, to evaluate the geocoding performance of four commercial geocoding engines. Match accuracies for these data ranged from 44% - 84%. In other work, Burra et al. (2002) explored the impacts of geocoding error on statistical analysis. Results indicated that small errors in the spatial accuracy of geocodes could significantly impact local statistical tests such as the Getis-Ord Gi and Gi*. Considering that the spatial analysis of geocoded information is such a critical component for so many public and private agencies (municipal planning departments, medical practitioners, county auditors, crime analysis units, etc.), maintaining an acceptable and reliable level of accuracy and precision in a spatial database is a significant concern. More importantly, the ability to assess the locational accuracy of a geocoded database would provide a stronger foundation for both spatial analysis and any related policy formulation and assessment.

The purpose of this paper is to both explore and document the problems associated with the geographic base files (GBF) used for coordinate assignment. We provide an extended empirical

example that highlights the relative uncertainty of household location in geographic space. This includes an assessment of positional uncertainty through the spatial perturbation of geocoded points within an established bound – allowing for alternative, yet equally accurate, spatial realizations of geocoded data. Sensitivity analysis is then conducted to evaluate the significance of locational uncertainties in geocoded data and spatial analysis.

2. COORDINATE ASSIGNMENT

Geocoding refers to the assignment of georeferenced coordinates to an entity for the purpose of identifying its location on the Earth's surface (Clark, 2003). The most precise method for accomplishing georeferencing is through the assignment of latitude and longitude (lat/long) coordinates (Goodchild, 1984). A latitude and longitude reference consists of a pair of angles, given in degrees, minutes and seconds. Latitude coordinates are specified for either north or south of the equator and longitude coordinates are specified for east or west of the prime meridian. In geographic information systems, lat/long coordinates are typically expressed in decimal degrees, north and east being positive and south and west being negative. For example, the latitude and longitude of Chioggia, Italy are (45.13, 12.17), meaning that it is 45.13° north of the equator and 12.17° east of the prime meridian (or Greenwich).

As mentioned previously, although the *concept* of spatially enabling disaggregate data such as a postal address is fairly easy to understand, the actual *process* is somewhat complicated. The most critical component needed to assign geographic coordinates to an address is a GBF (geographic base file). GBFs consist of georeferenced, thematic data. In the United States, the primary GBF used for geocoding are the TIGER (Topologically Integrated Geographic Encoding and Referencing) files produced by the U.S. Census Bureau. In Canada, the Area Master Files (AMF) from Statistics Canada are the most commonly used, while in the United Kingdom it is either the *Address-Point* of the Ordinance Survey or the *Postcode Address File* of the Royal Mail. Although these systems are totally independent, all of them utilize a street network and address ranges assigned to each street segment for assigning geographic coordinates. For example, the TIGER line files are based on a combination of digital line graphs from the U.S. Geological Survey and the dual independent map encoding (DIME) scheme developed by the U.S. Census Bureau in the early 1970s (Grubesic and Murray, 2004). Can (1993) notes that the basic observational unit in the TIGER database is a line segment that can represent either a physical or non-physical feature. This might include a highway, street or river or non-visible segments representing political boundaries and statistical units.

Figure 1 illustrates both TIGER data and its file structure. Each line segment has both a "to" and "from" node, providing a sense of direction in the TIGER system. The direction of a line segment always starts with the from-node and ends at the to-node. That said, both left and right can be determined by starting at the from-node and moving toward the end-node. For example, in Figure 1, it is possible to start at N1 and walk toward N2 on L1. In this case, P2 is the left-polygon and P1 is the right polygon.

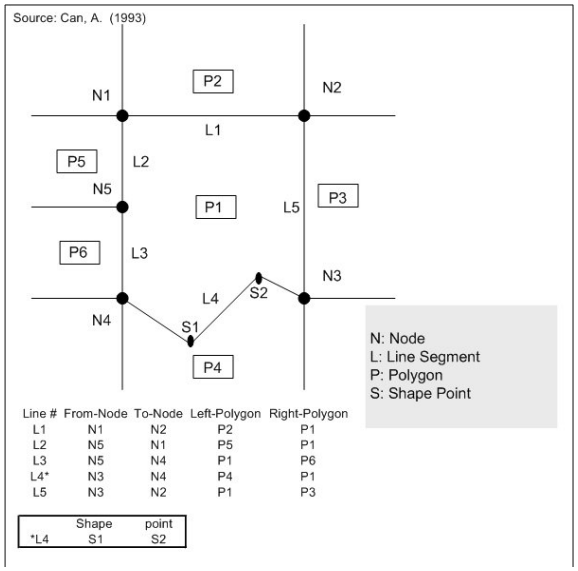

Figure 1: Vector Topology

In cases where there is a curved segment, additional shape points are included for the proper cartographic representation, as illustrated by L4. Once the vector topology is established in the geographic base file, it is possible to add supplementary data. This includes the basic components of the geographic base files that most users are familiar with, such as address ranges and street names. As a result, each line segment is associated with a specific, named street. More importantly, address ranges are assigned to each line segment, as illustrated in Figure 2.

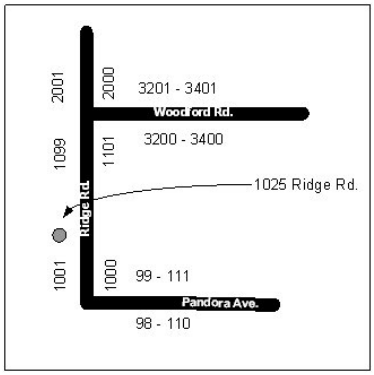

Figure 2: Address Ranges

The combination of an established vector topology and the supplementary data associated with each segment, node and polygon allows for commercial geocoding engines to assign lat/long coordinates. In this example, 1025 Ridge Road is the address of a household that is assigned coordinates using a commercial geocoding engine (Figure 2). This process is completed as follows: 1) the geocoding engine matches the street name in both the event table (e.g. crime location, customer record, etc.) and the geographic base file; 2) a determination is made regarding whether a particular address is located on the odd- or even-numbered side of the street; 3) using spatial interpolation, the geocoding engine estimates the location of the address along the street segment relative to the address range; 4) the geocoding engine applies a hard-coded offset distance; and, 5) assignment of the georeferenced coordinates (lat/long) is made.

Not surprisingly, there are several subroutines in this five step process that can have a dramatic impact on the spatial accuracy of geocoded points. First, the procedures which match street names in the GBF with records in an event database are quite complex. In fact, Grubesic and Murray (2004) note that both deterministic and probabilistic routines can be applied.[1] Deterministic routines are the most common, utilizing the logical components of both the GBF and event database in conjunction with lookup tables to compare input information in order to assign a match score to each record. This can be fraught with errors and the spatial accuracy of this routine is highly dependent on the quality of both the event table information and the GBF. Both Ratcliffe (2001) and Harries (1999) note that outdated or poor quality street files, data entry errors, incorrect directional identifiers (north v. south) or street suffixes (street v. road) can have significant impacts on the accuracy of a geocoded database. This, of course, is reflected in the match scores – where a rooftop hit (the most accurate geocode) gets a perfect score while a geocode placed at an intersection or zip code centroid receives a reduced match score (Grubesic and Murray, 2004). It is important to note, however, that if lat/long coordinates are assigned to a record, the geocoding engine reports this as a successful "hit", regardless of the match quality. As a result, the hit-rate of a geocoding application (e.g. 95%) says *absolutely nothing* about spatial accuracy. Another problem with the five-step routine outlined above is the use of spatial interpolation, based on the address range and the length of a street segment, for the assignment of a geocode. As illustrated in Figure 3, although address ranges apply to the entire length of a street segment, there may or may not be an equivalent number of built structures corresponding to the addresses. For example, the range of the left side of the Park Ave. segment encompasses 101 – 115. However, there are only three built structures on the left side. This means that several potential structures (105, 107, 109, 113, 115) do not exist. Obviously, the geocoding engine has no way of calculating this during the interpolation process. As a result, the geocoded lat/long coordinates for 103 Park Ave. might be located more closely to 107 or 109.

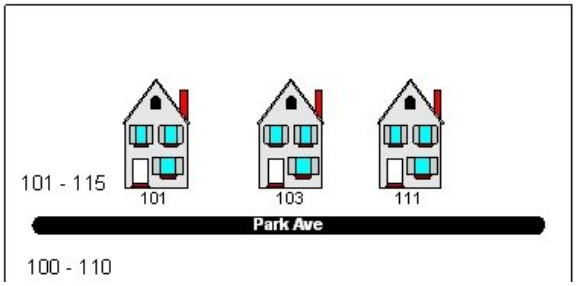

Figure 3: Address Ranges and Structures

A final concern in the five-step process deals with the hard coded offset distance used by geocoding engines. Simply put, the offset distance is an attempt to accommodate the distance between the street centerline (the segment that address ranges are assigned to) and the actual position of a household in geographic space. For example, Figure 4 illustrates parcel data, built structures and a street segment for a residential area in Cincinnati, Ohio.

[1] For a more thorough discussion on deterministic and probabilistic geocoding routines see Grubesic and Murray (2004).

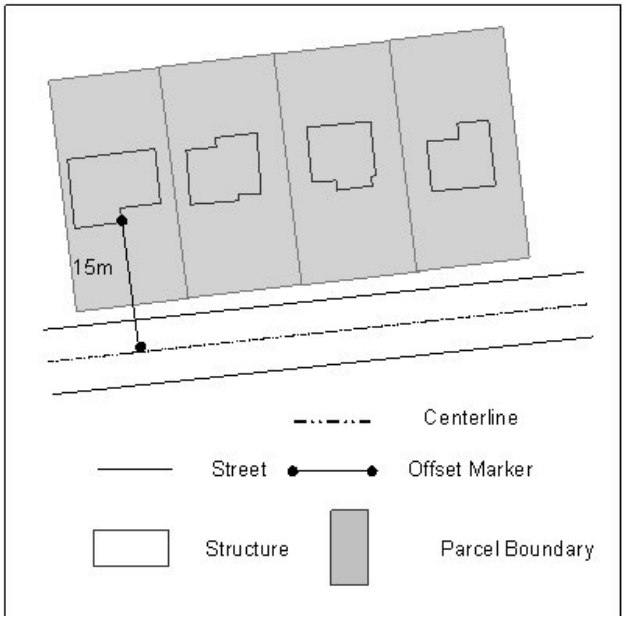

Figure 4: Offset Distance: Cincinnati, Ohio

In addition, Figure 4 shows an offset distance marker for the first built structure, which represents the distance between the street centerline and the household structure (15 meters). Clearly, the appropriate offset distance can vary substantially by neighborhood, city and region. For example, an appropriate offset distance for more densely settled inner city locations might be 2-5 meters, while homes located in less dense suburban areas might be 10-20 meters away from the street. In both cases, a poorly selected offset distance can impact the spatial accuracy of geocoded coordinates.

3. ASSESSING POSITIONAL UNCERTAINTY

The previous sections have illustrated that geocoding introduces considerable spatial uncertainty in produced geo-referenced information. Further, there is no attempt to quantify the degree of spatial uncertainty resident in any geocoded information layer. This is a formidable task no doubt given that a data layer, and its associated positional quality, is dependent on the entities being represented. Nevertheless, considerable research has been devoted to examining data quality issues (see Goodchild and Gopal 1989, Zhang and Goodchild 2002). The reason there has been, and continues to be, so much interest in data quality is that easy-to-use commercial GIS packages have made spatial information readily accessible for planning, management and decision making, either directly or as a component of further quantitative analysis. Thus, work in GIScience recognizes that there are many facets of uncertainty or error in spatial information (Heuvelink, 1999), and any techniques or processes utilizing spatial information must attempt to address the interpretive impacts of uncertainty and/or error.

For point-based objects, recent work by Murray (2003) has examined modeling impacts of locational vagueness through intentional manipulation of object position. In the context of geocoding, the idea is to take into account probable uncertainty associated with geo-referenced

objects by evaluating the analytical impacts of bounded, random positional movement. The rationale is that the identified location of the point objects being geocoded is uncertain, so why not consider this variability/error in subsequent analysis. By carrying out analysis many times under conditions where the spatial data is moved or perturbed, one can assess uncertainty in a controlled way.

An empirical evaluation of potential analysis sensitivity using data with positional uncertainty may be structured by randomly varying the location of objects and subsequently carry out analysis to examine resulting sensitivities. Let the geocoded information be denoted as $layer_0$, and Ω is the set of point objects in this layer. The geographic location of object *i* is represented by (x_i, y_i). Given that we want to perturb an identified geographic reference location, movement may be related to its current position with respect to distance and direction using a polar coordinate referencing approach. Perturbing an identified location may be accomplished by randomly selecting a distance from the current location that is at most a distance *S* away. This provides a bound, and can be empirically evaluated for a given geocoded data layer. Mathematically, we select a random variable, λ, such that $\lambda \in [0, S]$. In addition, a randomly selected direction of movement from the current location is also needed. This is accomplished by selecting a random variable, θ, such that $\theta \in [0,360]$. In this case 0 represents a move directly north of the current location. Perturbing a positional location in this manner gives the change depicted in Figure 5. As a result of perturbations to each data entity, a transformation of sorts is realized that can be defined mathematically as follows for any object *i* with random variables λ_i and θ_i:

$$\hat{x}_i = x_i + \lambda_i \cos\theta_i$$
$$\hat{y}_i = y_i + \lambda_i \sin\theta_i$$

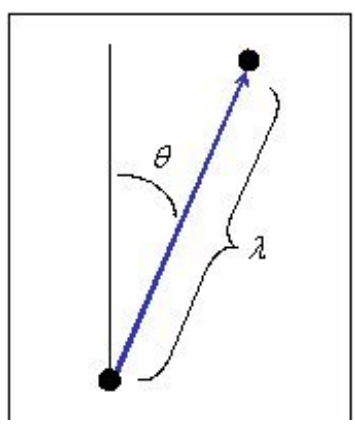

Figure 5: Perturbation of Geocoded Location

Perturbing each object in $layer_0$ gives a new data instance, $layer_1$, representing an equally probable positional reference of our geocoded objects. Clearly this can continue until some *K* perturbations of $layer_0$ are produced. As such, a sensitivity analysis could be structured where a

replicated analysis is performed for each information instance $layer_k$, and some process is necessary for summarizing variability in observed results, if there are any.

Something to note is that the random variables used to perturb object location may in fact be spatially bounded. S is no doubt a function of and bounded by scale. Consider for example the TIGER geographic base files commonly used for geocoding applications in the United States that rely on USGS digital line graphs. Although the USGS maintains relatively strict standards on the spatial accuracy of these data, the standards vary by map scale. For example, at 1:1,200 feet depicted objects are plus or minus 3.33 feet. At 1:24,000 feet depicted objects are plus or minus 40 feet (USGS, 1999). In addition, if commercial or residential property is being geocoded, the size of the parcels also represents a scale based factor influencing S. It is easy to conceive that θ can also be bounded spatially. Consider for example that positional movement might likely be restricted by the orientation of the street network. Alternatively, if it is know which side of the road a location resides, then clearly θ should be restricted in orientation. For either S or θ, implementing such bounds is readily accomplished using GIS.

4. CASE STUDY: CRIME IN LIMA, OHIO

Crime in cities is a major urban issue. Often crime data is processed by policing agencies using address information, where geocoding is then applied to obtain a spatial layer of crime occurrence. A natural question with such information is what is useful for deterring crime. This is why policing agencies focus on detecting hot spots, those areas with elevated crime occurrence. Identifying significant geographic relationships in the occurrence of criminal activity may be approached using numerous techniques. For spatially distributed point data (i.e. geocoded data), such as crime events, it is often necessary to describe patterns. Patterns are the result of spatial processes which reveal both physical and socioeconomic factors and their influences. For our purposes, we are primarily concerned with the location characteristics of the geocoded points rather than their attributes.

Our database includes all violent crime recorded in Lima, Ohio for 1999. Each incident was recorded as occurring at a specific address in the city. In other words, we have a full postal address (number, street, city and zip code) for each event. We make the basic assumption that there were no errors in the address recording process. Addresses from this database are then geocoded for subsequent analysis using a commercial desktop geocoding engine produced by Group One Software. A standard offset distance of 10 feet was utilized for this application. The results of this process are displayed in Figure 6.

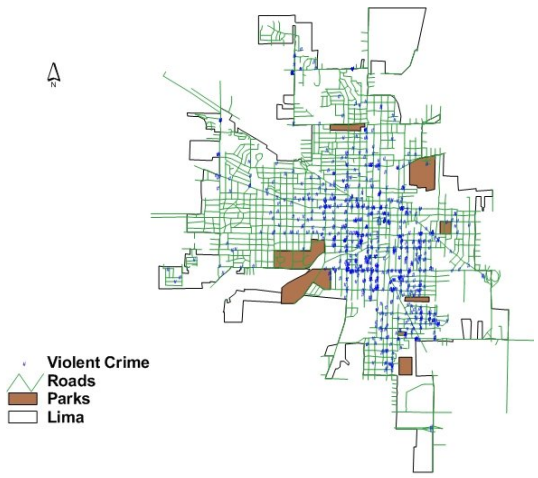

Figure 6: Geocoded Crime Events: Lima, Ohio

Geocoder diagnostics reveal a hit rate of 96%, indicating that 96% of the crime events were assigned a set of georeferenced coordinates. More importantly, the match scores from these "hits" suggest that 96% were assigned a street-level geocode, or supposedly a rooftop hit. This is the best match score possible and it indicates that some level of spatial accuracy (at least according to the TIGER files used for processing) was achieved. Once again, it is important to keep in mind that the actual coordinates assigned in this process are based on a spatial interpolation process using the address ranges assigned to each line segment in the GBF (TIGER). That said, the actual spatial accuracy of these geocodes remains unclear. Utilizing a post-processing routine outlined in Grubesic and Murray (2004), it was determined that the spatial accuracy of the geocoded crime events for Lima, Ohio had an average spatial error of 230 feet.[2] This is an important determination for several reasons. First, it suggests that some level of spatial error exists in the database and that the georeferenced coordinates assigned to each crime event do not necessarily correspond to the appropriate location. Second, it provides the opportunity to explore and assess the positional uncertainties of these geocoded data through basic statistical methods and sensitivity analysis.

5. RESULTS

An exploratory analysis of geocoded point incidents was conducted using the database from Lima, Ohio. As mentioned previously, we are primarily concerned with the locational characteristics of the geocoded points rather than their attributes. Nearest-neighbor analysis (NNA) is commonly applied to assess the spatial clustering of points (O'Sullivan and Unwin, 2003), and can be utilized to evaluate the potential analysis sensitivity with positional uncertainty. Nearest-neighbor analysis measures the distances between each point and their nearest neighboring points. The mean of the nearest-neighbor distance is then compared to the expected mean distance (gleaned from a point pattern generated using a Poisson process) and the

[2] This process compares the locations of geocoded events with the cadastre polygons that correspond to residential structures. In this particular case, the *distance* method was used, where the distance of a geocoded point and the nearest edge of the feature it corresponds to (e.g. residential structure) is calculated and used for determining the overall accuracy of the geocoded database.

size of the study area in order to generate the *R* statistic. Interpretation of the *R* statistic is straightforward. An *R* statistic that is less than 1 suggests a clustered pattern. An *R* statistic that is greater than 1 suggests a uniform pattern. Finally, an *R* statistic that is approximately equal to 1 suggests a random pattern.

The geocoded locations, Ω, in $layer_0$ are represented by the small points displayed in Figure 7. The buffers for $layer_1 \ldots layer_k$ represent the perturbation bounds for modeling positional uncertainties.

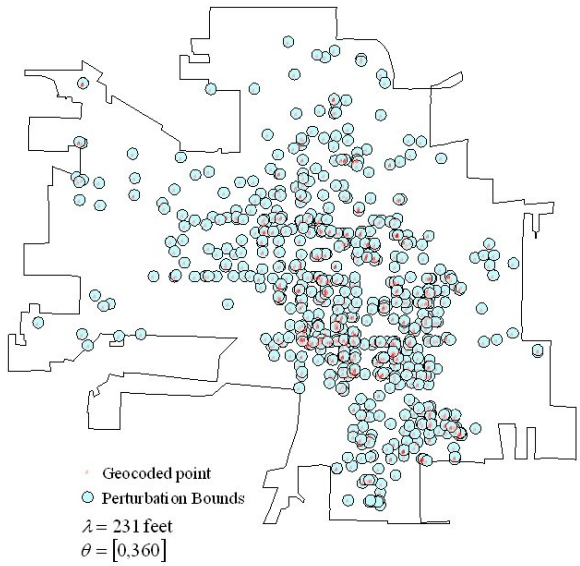

Figure 7: Spatial Bounds for Perturbed Geocoded Locations

The results of the nearest neighbor sensitivity analyses are quite informative (Table 1).

Data	Mean	*R*
Original	159.09	0.3705
Perturbed 1	195.63	0.4580
Perturbed 2	195.30	0.4537
Perturbed 3	194.01	0.4520
Perturbed 4	196.59	0.4564
Perturbed 5	200.99	0.4657

Table 1: Nearest Neighbor Results

In addition to displaying a clustered pattern, the *R* statistic is, in fact, remarkably stable to the spatial perturbation of the geocoded points, with only a moderate jump between the original and perturbed data. In all five of the cases where the geocoded points for Lima were perturbed, the mean distance for Ω changed very little. This is echoed in the *R* statistic, which displayed only a minor increase for all five perturbations – suggesting that the degree to which the crime events are clustered is somewhat less than the original set.

A second approach for evaluating the sensitivity of geocoded data to positional uncertainty is a basic kernel density test. A discussion of kernel density may be found in O'Sullivan and Unwin (2003). In evaluating kernel density we are seeking to interpolate a crime event density surface for the city of Lima using the original geocoded data and the perturbations. Each observation (crime event) is given a weight of 1. In this case, the interpretation of the resulting surfaces is quite intuitive, with darker shading indicating higher densities (Figure 8). For obvious reasons, this particular approach did not yield any statistical or graphical sensitivities between the original data and the perturbed data. With a λ value of 231 feet, and original geocoded points acting as anchors for the perturbations, the graphical sensitivity of the kernel surface is sure to be limited, particularly given the spatial distribution and count of violent crimes in Lima.

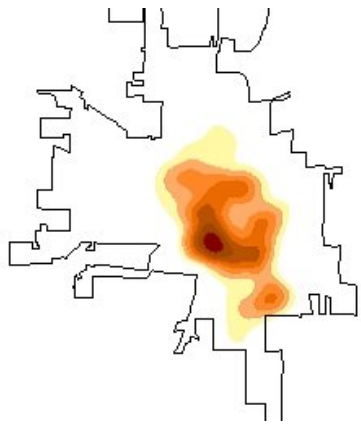

Figure 8: Kernel Density Surface

A final approach for evaluating potential analysis sensitivity for geocoded point data is cluster analysis. There are two families of clustering models, hierarchical and non-hierarchical or partitioning (Murray and Grubesic 2002). Hierarchical clustering techniques begin with all observations in separate groups and proceed to join the most similar observations (or groups of observations) according to some pre-specified criteria.[3] In hierarchical clustering, nearest neighbor distance is frequently used as the dissimilarity measure (Bailey and Gatrell, 1995). The nearest neighbor measure is a comparison of the distances between two points (or groups of points) with the average distance between all points. If the distance meets the *a priori* criterion (usually the calculated probabilities of a threshold distance between observations occurring by chance), observations are linked to form a new cluster (Bailey and Gatrell, 1995). Partitioning or non-hierarchical approaches are somewhat different. This type of cluster analysis attempts to split observations into a pre-specified number of groups, *p*, where the specified criterion is optimized globally over all possible splits. The *k*-means heuristic is a common non-hierarchical approach (Murray and Grubesic, 2002) for solving this type of problem.

For this case study, the hierarchical approach is utilized for evaluating analysis sensitivity for the Lima data. Figure 9 displays the results of hierarchical, second-order nearest neighbor cluster

[3] Alternatively, one can begin with all observations in a single group and break them up into separate clusters.

analysis generated using CrimeStat 2.0.[4] The original, non-perturbed geocoded locations are used in this example, with each ellipse representing two standard deviations. The standard deviational ellipse is frequently used to summarize the approximate location of a cluster and its members, with two standard deviations only covering about 99% of the cluster membership (Levine, 2001). Clearly, this is highly motivated by the actual spatial distribution of the points used for analysis, making it an excellent example for evaluating positional uncertainty. It is interesting to note that there are three identified second-order cluster groups in the original, non-perturbed data. However, there is only one cluster group identified in the perturbed distribution. To a certain extent, this mimics the results displayed in Table 1, suggesting that the degree to which the crime events are clustered (using NNA) were moderately sensitive between the original data and the spatial perturbations. However, unlike the standard NNA, the outcome of the hierarchical cluster analysis suggests that analytical results generated by using geocoded data are sensitive to positional change under conditions of uncertainty.

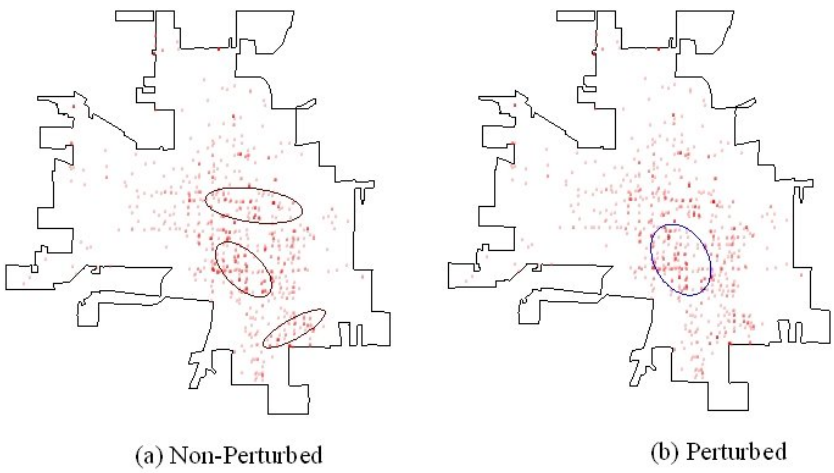

Figure 9: Hierarchical Clustering Results

6. Discussion and Conclusion

Previous research suggests that common statistical measures and spatial analytic modeling results vary significantly merely by altering geographic representation (O'Sullivan and Unwin, 2003). This problem is commonly known as the modifiable areal unit problem, or MAUP. The implications of MAUP for assessing the locational uncertainties in the geocoding process are significant. By explicitly testing for MAUP, one can better understand the behavior of an analytical technique or model with respect to changes in scale, units or spatial representations. In this case, the use of geocoded data and their perturbed variations reveal certain statistical sensitivities. In the case of NNA, the results remained fairly stable, with the most significant shift in mean and *R* occurring between the original data and the five perturbations. Where the kernel estimation was concerned, there was little observed variation. As mentioned previously,

[4] CrimeStat is a statistical software package distributed by the National Institute of Justice. http://www.icpsr.umich.edu/NACJD/crimestat.html

the sheer number of crime events and the relatively low lambda value certainly influenced the stability of these results. Finally, there were significant statistical sensitivities for hierarchical cluster analysis. While the original geocoded data yielded three clusters, the perturbed data yielded only one. This, of course, suggests that locational uncertainty of geocoded data and their impact on analytical tests are worth considering. However, this finding is tempered by the knowledge that the significance of these sensitivities is dependent on the analysis being carried out. In this case, certain approaches (e.g. nearest-neighbor analysis and kernel density) look stable, while the hierarchical clustering routine is not. Using this approach we can be confident in how we use such analysis. If we find sensitivities in the results, then we can temper our substantive conclusions. Alternatively, if there is consistency in our analysis, then we can be confident in making strong summary statements.

7. REFERENCES

Bailey, T.C. and A.C. Gatrell. (1995): Interactive Spatial Data Analysis. London, Longman Scientific and Technical.

Burra, T., M. Jerrett, R.T. Burnett, and M. Anderson (2002): Conceptual and practical issues in the detection of local disease clusters: a study of mortality in Hamilton, Ontario. Canadian Geographer 46: pp. 160-71.

Can, A. (1993): TIGER/Line files in teaching GIS. International Journal of Geographical Information Systems. 7(6), 561-572

Clarke, K.C. (2003): Getting Started with Geographic Information Systems. New York: Prentice Hall.

Goodchild, M. (1984): Geocoding and Geosampling. In: Spatial statistics and models. Gaile, G.L. and C.J. Willmott (Eds.). Boston: D. Reidel.

Goodchild, M. and S. Gopal (1989). Accuracy of Spatial Databases. London: Taylor and Francis

Grubesic, T.H. and A.T. Murray (2004): Geocoding Accuracy: Implications for Urban and Regional Planning. Draft Manuscript, University of Cincinnati.

Harries, K. (1999): *Mapping Crime: Principle and Practice*. Washington DC: National Institute of Justice.

Heuvelink, G.B.M. (1999): "Propagation of error in spatial modelling with GIS." In: Geographical Information Systems: Management Issues and Applications, ed. Longley, P.A., Goodchild, M.F., Maguire, D.J. and D.W. Rhind. New York: John Wiley and Sons.

Krieger, N., Waterman, P., Lemieux, K., Zierler, S. and J.W. Hogan (2001): On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. American Journal of Public Healt*h*. 91(7), pp. 1114 – 1116.

Murray, A.T. (2003): "Site placement uncertainty in location analysis." Computers, Environment and Urban Systems 27, 205-221.

Murray, A.T. and T.H. Grubesic. (2002): Identifying non-hierarchical spatial clusters. International Journal of Industrial engineering. 9, pp. 86-95.

O'Sullivan, D. and D. Unwin (2003). Geographical Information Analysis. New York: Wiley.

Ratcliffe, J.H (2001): On the accuracy of TIGER-type geocoded address data in relation to cadastral and census areal units. International Journal of Geographical Information Science. 15(5), pp. 473-485.

Thrall, G. and S. Thrall (1998): "Desktop GIS". In P A Longley , M F Goodchild, D J Maguire, D W Rhind (Eds). Geographical Information Systems: Principles, Techniques, Management and Applications. New York: Wiley.

Yeh, A.G-O (1999): Urban Planning and GIS. In: Geographical Information Systems: Management Issues and Applications, ed. Longley, P.A., Goodchild, M.F., Maguire, D.J. and D.W. Rhind. New York: John Wiley and Sons.

Zhang, J. and M. Goodchild (2002). Uncertainty in Geographical Information. London: Taylor and Francis.

CVS

TONY H. GRUBESIC

ASSISTANT PROFESSOR, UNIVERSITY OF CINCINNATI, DEPARTMENT OF GEOGRAPHY.

ALAN T. MURRAY

ASSOCIATE PROFESSOR, OHIO STATE UNIVERSITY, DEPARTMENT OF GEOGRAPHY

CO-ORDINATES

Tony H. Grubesic

Department of Geography, University of Cincinnati

P.O. Box 210131

Cincinnati, Ohio 45221-0131

USA

(Tel) +1-513-556-3357

(Fax) +1-513-556-3370

tony.grubesic@uc.edu

http://www.geography.uc.edu/grubesic/

Alan T. Murray

Department of Geography, The Ohio State University
1110 Derby Hall

154 North Oval Mall
Columbus, Ohio 43210
USA

(Tel) +1 614-688-5441

(Fax) + 1-614-292-6213

murray.308@osu.edu

http://geog-www.sbs.ohio-state.edu/faculty/murray/personal/murray.htm