INT. J. GEOGRAPHICAL INFORMATION SCIENCE
VOL. 18, NO. 1, JANUARY–FEBRUARY 2004, 61–72

Research Article

# Geocoding crime and a first estimate of a minimum acceptable hit rate

JERRY H. RATCLIFFE
Department of Criminal Justice, Temple University, 1115 W Berks Street, Philadelphia, PA 19122, USA; e-mail: jhr@temple.edu

(*Received 2 September 2002; accepted 2 April 2003*)

**Abstract.** Spatial crime analysis relies not only on accurate geocoding but also the achievement of a high level of geocoding success. Geocoding is the task of converting locations, such as the addresses of burglary victims, into grid coordinates and is a task performed regularly by many crime analysts. Data sources include police offence and incident databases where the quality of geographical references can vary. The reality of dealing with this real world data means that achieving a completely successful geocoding process is rare and few crime analysts can get a hit rate (the percentage measure of success) of 100%. This paper seeks the answer to a seemingly simple question: what is an 'acceptable' minimum geocoding hit rate for crime data? This paper uses a number of different crime patterns and Monte Carlo simulation to replicate a declining geocoding hit rate to answer this question. Reduced crime rates of mapped points, aggregated to census boundaries, are compared for a statistically significant difference. The result indicates 85% as a first estimate of a minimum reliable geocoding rate, and this result is applicable to many address-based, point pattern datasets beyond the crime arena.

## 1. Introduction

One of the rapidly growing application areas of Geographical Information Systems (GIS) is in the analysis of crime. Crime mapping has made rapid advances in recent years with regard to data availability and analytical techniques. Many of the papers presented to the annual US National Institute of Justice (NIJ) Crime Mapping Research Center (now the Mapping and Analysis for Public Safety program) conferences (www.ojp.usdoj.gov/nij/maps/) regularly discuss analytical techniques and analyses performed on high volumes of individual crime locations. The mapping of high volume crime data is made possible through the automated geocoding of address level data extracted from police recorded crime databases. As a result point level analytical tools are now coming to the fore. New aggregation techniques for point pattern data, based on Local Indicators of Statistical Association (LISA), have been developed (Chakravorty 1995, Ratcliffe and McCullagh 1999, Unwin 1996) and spatial crime analysis by law enforcement is now a substantial market for GIS companies.

*International Journal of Geographical Information Science*
ISSN 1365-8816 print/ISSN 1362-3087 online 
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810310001596076

Crime is an inherently spatial phenomenon and crime mapping tends to be point-specific. While some crimes are more difficult to map (internet fraud, tax evasion and some motoring offences such as driving without a licence), the majority of criminal activity and day-to-day incidents that police are required to respond to can be analysed spatially. The location of incidents which have to be mapped are usually well known: businesses have thefts at specific sites, residential burglaries occur at houses, and street crimes (assaults and vehicle crimes) often occur outside premises with known addresses. The process of geocoding—turning an address into a point on a map—is therefore of vital importance in crime mapping. Any error in the initial geocoding process will translate into compounding errors as the analytical and dissemination stages of police intelligence work are undertaken. Moreover some crime sites are not geocodable in that the address information presented to the crime analyst contains insufficient information to determine the incident location. The reality of modern crime analysis is that while crime mapping is an enlightening and practical intelligence tool at many levels, the analyst rarely has time to track down the location of ungeocoded incidents and completely successful geocoding is not the norm. Crime maps, while they may not say as such on any output, are rarely created from 100% of the original data.

This paper statistically tests the accuracy of thematic crime maps generated from data sets with incomplete geocoding in order to arrive at a first estimate of a reliable minimum geocoding level. A Monte Carlo simulation of a declining geocoding hit rate (the percentage of unit records in a crime database that are successfully geocoded) is combined with a statistical analysis of aggregated outcomes to determine a point where the output is significantly different from that generated by maps created with 100% geocoded records. While the discussion and data sets employed have a crime focus, there are technical and policy implications for the spatial analysis of any address-based data, from hospital records and insurance claims to newspaper subscription and voter registers. The paper starts with a brief overview of the use of spatial data within law enforcement.

## 2. Crime mapping

Law enforcement has become increasingly sophisticated over the last few decades, partly due to the realities of increasing fiscal constraint (anti-terrorism is an exceptional area with a seemingly bottomless budget). The absence of increased numbers to swell the ranks means that the existing pool of officers have to work smarter and make better use of limited resources (Morgan and Newburn 1997). The use of GIS has coincided with a general increase in the use of technology and computing to assist in this effort. At the same time the calls on police to reduce crime have not abated and one of the developing areas of current law enforcement strategy is termed intelligence-led policing (Heaton 2000, Maguire 2000, Sheptycki 2000, Ratcliffe 2002). GIS is one of a number of technologies that police are using to achieve more effective intelligence-based operations.

An example of the development of GIS within law enforcement can be seen from the New South Wales (NSW) Police Service in Australia. The NSW Police Service is the state law enforcement agency for the most populous state in Australia, headquartered in the state capital, Sydney. The NSW Police Service has been developing its mapping capability over the last few years. A primary focus of senior management has been Operations and Crime Review (OCR) panels. These

OCR panels are modelled on the CompStat (short for Computer Statistics) planning forums of the New York City Police Department, where maps of crime distributions are projected onto a wall for management and senior police executives to determine policing strategy. The change in management style across the NSW Police Service as a result of the Operations and Crime Review (OCR) panels has been significant, and had for some time a measurable impact on the level of crime in the state (BoCSaR 2001, Chilvers and Weatherburn 2001). The use of mapping technology to chart crime hotspots in an environment where the whole room can see the effectiveness or limitations of crime reduction strategies leaves a powerful impression on all. The use of maps has gone some way to making the OCR a more dynamic and visual environment where management can quickly understand a complex crime distribution. In this situation a picture truly is worth a thousand words. Not surprisingly, with the impetus given to mapping in the OCR, a number of local area commanders have been enthusiastic about developing their local mapping potential. Local intelligence officers are now seen as the hub of the intelligence analysis and dissemination practices that are the cornerstone of intelligence-led policing. Crime mapping can provide a valuable analytical and briefing tool and the use of MapInfo, the chosen mapping software of the NSW Police Service, is growing across the State in local and regional intelligence offices.

Further afield, crime mapping has steadily grown in the United States (Rich 2001) driven to a degree by the NIJ Mapping and Analysis for Public Safety Program (formerly the Crime Mapping Research Center). The level of law enforcement uptake of GIS is difficult to assess due to the uncoordinated nature of policing in the US. Given that federal authorities are unclear as to the actual number of law enforcement agencies in the country (Walker and Katz 2001) a measure of GIS usage is likely to be even harder to estimate! A 1997 nationwide survey by Crime Mapping Research Center staff found that 13% of agencies that responded ($n = 261$) used crime mapping in some form (Mamalian and LaVigne 1999). Two factors are worth bearing in mind. Firstly, this number is likely to have risen in recent years (Rich 2001), in the same way that GIS use has grown in other sectors. Secondly, about half of all police agencies in the US have 10 sworn officers or less and have very limited budgets. With GIS moving into the same cost bracket as mainstream office software, uptake is likely to have increased within these smaller agencies.

In the UK, the statutory obligation placed on every police service and local authority to produce a crime and disorder audit by the 1998 Crime and Disorder Act (Home Office 1998) has had a significant role in bringing crime mapping to the fore in the crime and disorder arena. The UK government, in discussing best practice in analysing crime and disorder problems, advocated the use of GIS to map crime hotspots (HOCD 1998, §3.27). They did however add the *caveat* that while GIS can georeference locations, the database must contain accurate addresses (§3.28). The question of accurate address information is addressed later in this paper. A recent review of audits noted that staff training in GIS was lacking and that less than half of the audit agencies undertook mapping as part of the crime and disorder audit (Bowers *et al.* 2002). Indeed while uptake of GIS is one factor, proficiency is another matter altogether. GIS has yet to reach the general level of acceptance that word processing has reached and GIS skilling in law enforcement has to compete for limited training budgets alongside proficiency in firearms, public order, control and restraint, law, first aid and a plethora of other training agendas. This impacts onto the level of proficiency that intelligence officers and crime analysts have with the tools at their disposal. While crime events can be visualised as point patterns, law enforcement GIS training cannot be limited to point mapping techniques, as polygons also have their place at the analysis and dissemination stage. Notwithstanding the problems of the Modifiable Areal Unit Problem (MAUP) (Openshaw 1984, Bailey and Gatrell 1995), boundary files are commonly used in crime mapping to aggregate crime counts and make comparisons to census data. Law enforcement is an inherently practical business and intelligence officers are often required to perform a myriad of tasks: dedicated mapping officers are rare. It is therefore common to find that while a few intelligence officers are aware of the MAUP, few have the level of training or time to discover solutions.

Given this growing enthusiasm for mapping, an inquiry into the impact of geocoding rate may be timely. After a literature search it became clear that as yet there is a paucity of reliable statistical information regarding the limitations of the geocoding procedure commonly used in crime mapping. It is doubtful that many users in law enforcement are consciously aware of the problems inherent in the geocoding process. Occasionally a point will obviously be in the wrong place and the less experienced user may write this off as a quirk of the software geocoding engine, unaware that every point on the map will be inaccurate to some degree. After data capture, geocoding is the most important part of the crime mapping process, and the value of sophisticated analytical tools such as HotSpot Detective® or Vertical Mapper® is limited by the accuracy of geocoded locations.

Two main questions present themselves:

1. How accurate are geocoded points?
2. Given that the most police services have data quality issues in the recorded crime database, what geocoding hit rate must be achieved to produce an accurate map?

The first question was tackled by Ratcliffe (2001) in a study that compared the accuracy of geocoded points to the building location determined from the cadastral file. In a study of 20 000 addresses in the Eastern Suburbs of Sydney, it was found that the mean error could be minimised by the judicious use of the offset facilities available in both MapInfo and ArcView. A road offset of 25 m, and an end offset (inset in MapInfo-speak) of 15 m reduced the 5% trimmed mean distance between geocoded locations and target buildings to 20.5 m. Combining results with the census Collection Districts (the smallest areal unit of the Australian Census) it was further found that with the same settings, a geocoded point was located in a different polygon in 5% of the cases. These findings give an indication of the level of error for geocoded points in a densely populated urban setting, but do not address the problem of less than perfect geocoding hit rates.

### 2.1. *Why we have incomplete geocoding*

Rules are meant to be broken, and the practical context of crime mapping can be different from other application areas. While an epidemiologist may have to examine the location of every outbreak of a disease in order to track the original source and the spread of the contagion, geocoding every crime or incident location is often impractical for a crime analyst. Usually the sheer volume of records swamps any attempt at perfection. Traditional law enforcement was rarely an issue of overwork, but the introduction of the police car and the telephone changed the nature of policing forever. With these two devices the public both had an expectation of police attendance to every incident and they also had the means to summon assistance with ease. Police services now receive so many requests for service that they must triage the calls. Some crimes (for example theft from a motor vehicle) will now rarely result in a police officer visiting the scene. With so much data available electronically, analysts rarely have the time to check geocoding results, instead relying on a high, but less-than-perfect, geocoding hit rate to get a general picture of crime in a geographical area.

Records are usually geocoded from the address fields in the police crime or incident database. The process of recording an incident (call for service) or a crime is similar. Most requests for police attendance originate in a call from a member of the public. For example, they may call the police station to report a burglary at their house. The dispatcher will record the address for the 'call for service' in a dedicated field on a computer terminal at the station. It is rare that any address verification takes place at this stage. If an officer is dispatched and is unable to find the crime victim, the dispatcher will call the victim on their phone and ask for a better location. In this way, calls for service (incident) databases can have a wide variety of incident locations with an even wider diversity of spelling combinations. The attending officer may confirm that there has been a burglary and will at some point return to the police station. It will either fall to the reporting officer, or to another staff member reading the officer's written report, to enter the crime details onto a crime database. This differs from the incident database in that calls such as false calls to burglar alarms, traffic accidents, and of course the traditional 'person locked out of their car', will appear on the incident database but not the crime database. From an analytical perspective, both databases have value. Unfortunately few agencies have any address verification on either the crime or the incident database, and the crime database can be even more vague in address accuracy if a person has to interpret the handwriting of the reporting officer.

This vagueness in address recording results in a number of common errors, including:

- Misspelling the street name (e.g. 12 W Braod Street instead of 12 W Broad Street),
- Recording streets with the incorrect directional prefix or suffix (e.g. 12 E Broad Street instead of 12 W Broad Street),
- Using an abbreviation not recognised by the geocoding engine (e.g. Strt. in '12 E Broad Strt.'),
- An incorrect street type (e.g. 12 W Broad Avenue, instead of Street),
- Entering an impossible address (e.g. 120000 W Broad Street),
- Entering a location not known to the geocoding database (e.g. 'the Crown Cinema'),
- Omitting to enter any address at all,
- Confusing an address with unit numbers (e.g. Appt 4/12 W Broad Street),
- Baffling a geocoding engine with preliminary text (e.g. 50 yards E of 12 W Broad Street)

(adapted from Harries 1999 p. 98). These common errors do not include any problems associated with the street database that the GIS has to use for geocoding. This could be an additional factor if the area to be examined has experienced rapid development. New streets and new housing developments can spring up rapidly leaving a lag time before these streets are available on a street database for geocoding purposes. In these circumstances, a dispatcher or police officer may correctly record the address, but the analyst is still unable to geocode the location. This can occur in new developments where theft of building supplies or builders equipment is common.

A few police agencies have more rigid data entry systems for their databases. These do have the advantage of checking address spelling and house number ranges, and this increases geocoding hit rates to near 100%. However they do require considerable maintenance to keep the address database current, including a system to verify and enter new addresses.

Given therefore that quality geocodable data are only available to a limited number of law enforcement agencies, we return to the second question posed in the last section: what geocoding hit rate must be achieved to produce an accurate map? To answer this question, the remainder of the paper reports on the use of a Monte Carlo simulation technique to estimate a minimum acceptable geocoding hit rate.

## 3. Monte Carlo simulation of the geocoding hit rate problem

The problem was tackled from a practical standpoint. As stated earlier, many analysts in law enforcement are taught to aggregate crime counts to census boundaries—usually census blocks (called enumeration districts in many countries). Although this approach to mapping has the potential to cause interpretative issues due to the MAUP, it is an easy technique to teach non-GIS specialists. If we could therefore generate such a map based on a hit rate of 100 percent, how many points would have to be removed (to simulate points not geocoded) before a generated map differed statistically from the notional 100% map? For example, let us assume that an intelligence analyst has a data set of 300 incidents to map. The reality is that questionable data quality in some cases will mean that some addresses are not geocodable automatically. Therefore what percentage of the 300 points does the analyst have to geocode such that the final thematic map, aggregated to census blocks, would accurately reflect the map that would have been generated by all of the data had the analyst been able to map every location?

The following process was employed on a number of different data sets, and is shown graphically in figure 1.

1. Generate a speculative 100% hit rate with notional crime distribution aggregated to census boundaries.
2. Remove 1% of randomly selected points.
3. Generate a thematic map of the incomplete data and statistically test the distribution of points against the 100% map.
4. If the distribution is not statistically different from the notional 100% map then continue with the removal of points (back to 2). If the distribution is statistically different, record the percentage of points removed.
5. Return to (1). Complete this process a number of times so that any chance of an unusual random point selection does not unduly influence the results (Monte Carlo simulation).

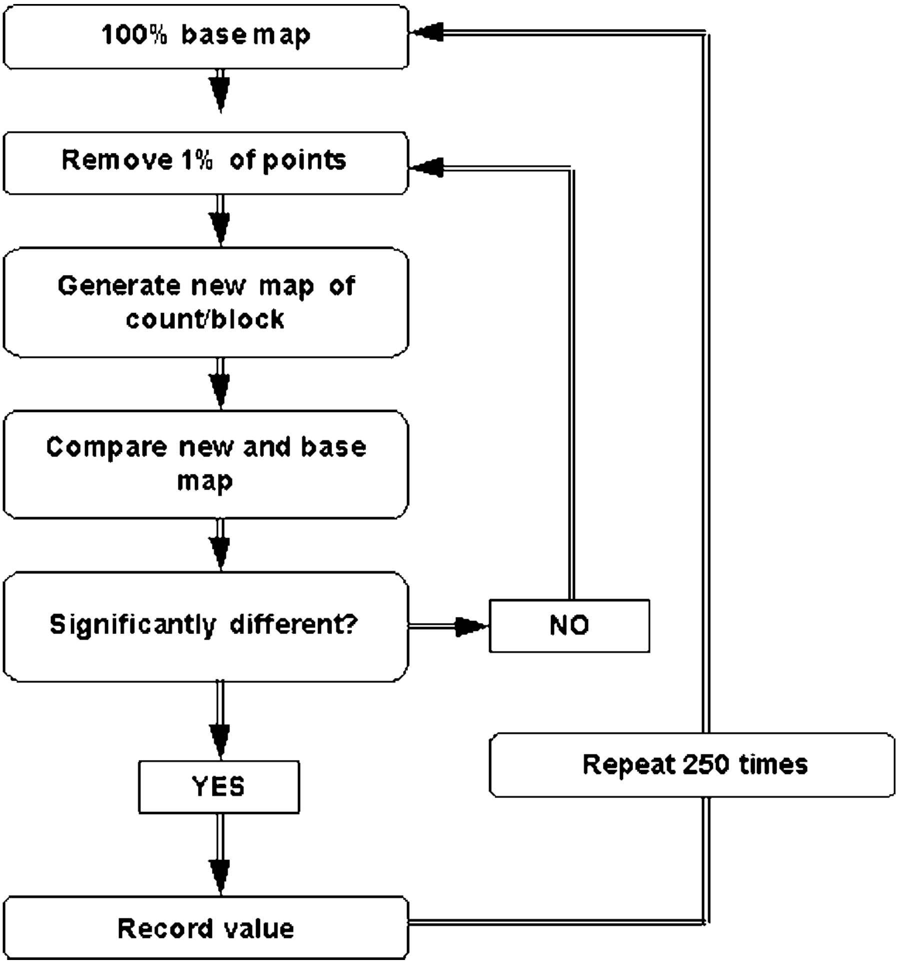


Figure 1. Flow chart of the geocoding Monte Carlo simulation process.

6. Construct a frequency distribution of the results, and interpret the Monte Carlo estimate in line with the parameters of the simulation.

The principle of Monte Carlo simulation is based on the notion that there may be situations where it is impossible to realistically model sample data many times such that the behaviour of the system can be evaluated (Mooney 1997). Extrapolation of results to predict wider outcomes therefore becomes difficult. Monte Carlo simulation can solve this problem by using random samples with a distribution similar or identical to the population, to resemble the real world problem as closely as necessary—a real world that can include spatial problems (Fisher and Langford 1995, Zhang and Murayama 2000). These samples are tested and observed using an empirical process to model the problem. We can therefore formulate the geocoding hit rate problem such that a Monte Carlo simulation can generate a set of observed values. By following the process outlined above, each run of the simulation can generate a percentage of removed points after which the map fails to resemble the 100% map in either magnitude of points or distribution of values. The necessity to repeat the process a number of times is caused by the principle that a single 'realisation' of the random process (here the selection of points for removal from consideration) could be a statistical quirk because it yields only a single returned value (the minimum hit rate for the current run). Repeated testing removes the impact of any quirks due to unusual random selection and conclusions are drawn from the aggregated results of all realisations of the Monte Carlo process.

In practical terms, the aim of this research was to model the impact of inaccuracies in law enforcement geocoding. Given that the population of all law enforcement geocoding situations is too vast to investigate, a pseudo-population (Mooney 1997) of five geocoded data sets were used to model the larger population. These complete sets represent the 100% geocoded data sets that make up the base maps for the Monte Carlo simulation. These were drawn from different crime types and from different locations in New South Wales (Australia) and they represent a range of data sets that would be commonly analysed spatially. Table 1 describes some basic characteristics of each data set.

Because the Monte Carlo process relies on repeated realisations, or 'trials', of the process to produce sufficient output for a generalisation to be made, the researcher has to determine an appropriate number of trials. A sufficient number of trials is necessary to sufficiently model the stochastic processes in the system, while additional testing beyond the necessary number adds little to the analysis and increases computational effort. Although a large number of trials is recommended, Monte Carlo tests have been successfully conducted with as low as twenty (Hope 1968) to fifty runs (Davis and Keller 1997). Although larger trials are recommended, there is no generally accepted theoretical guideline for a minimum number of trials (Mooney 1997 p. 58). Given that the output will be a frequency distribution, statistical power increases with increased number of trials. This is because increased sample sizes tend to generate smaller, and hence more applicable, standard deviations. As said, the trade-off with increased testing is computational effort.

In this study, the programming engine to generate the Monte Carlo process was MapBasic for MapInfo. MapBasic is a relatively effective programming instrument, easy to teach and learn, and useful for a wide variety of mapping tasks requiring automation within MapInfo. However it does lack the speed of more advanced, higher level programming languages. It was determined here that 250 trials were sufficient to generate a frequency distribution that approximated a normal distribution, with a skew approximating 0 and a kurtosis of close to 3.0.

For each of the data sets, a 100% thematic map was generated and then each subsequent map (with progressively less points in the aggregation) was compared

Table 1. Basic characteristics of data sets employed.

| Data set | Location | Crime type | Records | Census blocks |
|---|---|---|---|---|
| 1 | Regional, coastal | All reported crime | 1,362 | 149 |
| 2 | Urban, coastal | Vehicle crime | 1,278 | 261 |
| 3 | Inner city | Malicious damage | 884 | 144 |
| 4 | Urban, coastal | Burglary | 783 | 177 |
| 5 | Inner city | All reported crime | 908 | 217 |

using a non-parametric test. The Mann-Whitney U test was employed at the 0.01 significance level. When the distribution of census block counts between the two maps was significantly different the simulated hit rate was recorded at the point where they differed, and the test repeated 250 times. For each study data set 250 values were therefore recorded and plotted.

## 4. Results

The results from the analysis conducted are summarised in table 2. This table shows; the reference number for the data set; the mean of the level at which the maps became statistically different; the standard deviation of the 250 realisations; the acceptable minimum hit rate for this data set calculated as the mean plus two standard deviations, rounded up.

The results for the census blocks indicate similarity for both the point at which the reduced maps became statistically divergent and in their distribution of values. The range of values across sets 1 to 5 was between 71% and 85%, with the lowest values being recorded in the regional, coastal area, the least urban of the areas studied. This can probably be explained by the distribution of crime in this area. The concentration of crime in small pockets of census blocks instead of more evenly across the 149 blocks would have the effect of reducing the result values as the Mann-Whitney test would have to have more values removed before there was a significant shift in the relative ranks of a number of census polygons.

The decision to determine an 'acceptable' minimum geocoding hit rate is solely based on the interpretation of the normally distributed frequency plots of the analyses. It does not take into consideration other operational factors, some of which are discussed in the final section. By choosing a level of the mean plus two standard deviations, rounded upwards, we can say that for each of the study areas (1–5) generated maps will not differ statistically from a notional 100% aggregated map in at least 95% of the cases.

As the area with the highest values at which the maps became statistically divergent has an acceptable hit rate of 85% (meaning that to generate a statistically reliable map 85% of the points in a crime table must be geocoded) we can use this value as the benchmark for other areas. By using this value we know that an 85% hit rate is acceptable over 95% of the time for a range of areas.

## 5. Limitations

A number of *caveats* should be stated at this point. Firstly, it is possible that different crime distributions for the same area will have markedly different spatial patterns, and this will influence the number of census blocks that contain values. This in turn will influence the Mann-Whitney test employed. While the non-parametric U-test is both robust and effective to produce a meaningful result, it can

Table 2. Results for 250 Monte Carlo simulations of the hit rate analysis.

| Data set | Mean (%) | Standard deviation | Acceptable minimum (%) |
|---|---|---|---|
| 1 | 75 | 1.59 | 78 |
| 2 | 82 | 1.06 | 85 |
| 3 | 82 | 1.25 | 85 |
| 4 | 81 | 0.94 | 84 |
| 5 | 82 | 1.24 | 85 |

be influenced by an uneven distribution of values, as was noticed in the first data set. This has the effect of increasing the percentage of points removed from the analysis before a map is statistically different from the 100% map. Furthermore, the test also examines the rank order of census blocks without taking into account contiguous areas. It may be that the effect of statistical difference in maps is ameliorated by the use of a small number of classes in any final thematic map. The use of classes was not examined in this paper, as the intention was to examine the error level between areal units prior to the influence of map class aggregation. The use of a few, large polygon boundary sets would solve a number of these problems, but these are rarely used for any operational purpose in policing as they are not detailed enough for any practical purpose.

The Monte Carlo selection procedure for the choice of points to remove is a pseudo-random one (Mooney 1997), but given the limited sizes of the data sets in relation to the millions of iterations necessary before a computer random number program has a return period, this is not deemed to be a problem for this study.

Of a more practical consideration is that the process applies a uniform distribution random point selection process. Each point has an equal chance of selection for removal on each trial. An examination of the common causes of geocoding error mentioned earlier in the paper will suggest that some geocoding errors may not be randomly distributed spatially. For instance, if a base street file has not been updated recently there may be a whole housing development comprising of a number of streets that are not geocodable. Any crime events in these streets will not appear on a map, and their location will not be randomly distributed around the image but will cluster in one location—the new housing development. In the same manner, common usage of an ungeocodable landmark such as a cinema or other civic building will generally cluster in town centres. In these circumstances the minimum hit rate would be raised, as the change in the relative order of some census blocks would change rapidly.

## 6. Concluding remarks

What must be stressed in regard to this analysis is that this is a first estimate of a geocoding hit rate. Common sense dictates that we should attempt to achieve a hit rate of 100% every time. It must not be forgotten that even if an 85% hit rate is achieved, more than 1 in 10 addresses in a crime table are not being geocoded. This means that if a police analyst wanted to map 10 000 crime sites, up to 1500 are not represented in the final map. That is not an insignificant number.

The sensible approach for an analyst is to examine the ungeocoded records and determine if any pattern can be discerned from the geocoding 'misses'. These regular misses may be concentrated in one area, or may be easily resolved using an address scrubbing routine. Address scrubbers work by providing a first pass over a spatial database prior to geocoding. This first pass is designed to correct common spelling mistakes, remove unwanted textual complications, and prepare the address base for maximum geocoding efficiency. Common examples of address scrubbing operations including changing 'Gdns' to 'Gardens', removing unit numbers or apartment numbers, and replacing landmarks with their actual addresses. After the address file has been run through the scrubber, increases in geocoding efficiency and accuracy are usually seen. Improvements are usually ongoing if the analyst always examines the geocoding misses to determine the cause of the problem. This ongoing process of continual improvement is one of the easiest ways to increase geocoding efficiency.

Effort should be measured against reward. Expending significant effort to increase geocoding of a theft from motor vehicle database is unlikely to be worthwhile if the resultant analysis will not be acted upon and law enforcement priorities are elsewhere. Most people would not struggle to determine the appropriate policy objectives between geocoding a theft from motor vehicle database or a serial homicide database. There are lessons to be learned from both databases however, and an understanding of the error characteristics of the former may assist with geocoding of the latter.

An e-mail enquiry distributed on the list server of the Crime Mapping Research Center (now the Mapping and Analysis for Public Safety program) of the US NIJ indicated that in general, law enforcement geocoding hit rates were in the acceptable range. Nearly forty individuals described their geocoding experiences with numerous different agencies. The mean average geocoding hit rate was 87.5%, with a standard deviation of 14.1%. The lowest was 41%, while the highest was 99.7%. Slightly more than two thirds of the responses were 90% or greater.

And if an 85% hit rate cannot be achieved? While this study does not suggest that maps created with data that are geocoded at a lower hit rate are necessarily showing an incorrect distribution or significantly lower quantity of points, it does follow that the lower the hit rate the greater the potential for error in spatial patterns, and there certainly exists the potential to underestimate the magnitude of any problem. It is suggested here that that this first estimate of an empirically derived minimum acceptable hit rate should be used as a minimum standard.

## References


BAILEY, T. C., and GATRELL, A. C., 1995, *Interactive Spatial Data Analysis*, Second edition (London: Longman).

BoCSaR-NSW Bureau of Crime Statistics and Research 2001, Media Release: 2000 Recorded crime statistics, pp. 2, Sydney.

BOWERS, K., JENNINGS, J., and HIRSCHFIELD, A., 2002, The crime and disorder audit process: a post-mortem on the first round. *Crime Prevention and Community Safety*, **4**, 19–32.

CHAKRAVORTY, S., 1995, Identifying crime clusters: The spatial principles. *Middle States Geographer*, **28**, 53–58.

CHILVERS, M., and WEATHERBURN, D., 2001, Operation and Crime Review panels: Their impact on break and enter. *Crime and Justice Statistics: Bureau Brief (NSW Bureau of Crime Statistics and Research)* pp. 1–11.

DAVIS, T. J., and KELLER, C. P., 1997, Modelling uncertainty in natural resource analysis using fuzzy sets and Monte Carlo simulation: slope stability prediction. *International Journal of Geographical Information Science*, **11**, 409–434.

FISHER, P. F., and LANGFORD, M., 1995, Modelling the errors in areal interpolation between zonal system by Monte Carlo simulation. *Environment and Planning A*, **27**, 211–244.

HARRIES, K., 1999, *Mapping Crime: Principles and Practice* (Washington DC: US Department of Justice).

HEATON, R., 2000, The prospects for intelligence-led policing: Some historical and quantitative considerations. *Policing and Society*, **9**, 337–356.

HOCD–Home Office Communication Directorate 1998, *Guidance on Statutory Crime and Disorder Partnerships*, Crime and Disorder Act 1998, chapter 3, www.homeoffice.-gov.uk/cdact/actgch3.htm.

Home Office 1998, *Crime and Disorder Bill* (London: House of Commons).

HOPE, A. C. A., 1968, A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society, Series B*, **30**, 583–598.

MAGUIRE, M., 2000, Policing by risks and targets: Some dimensions and implications of intelligence-led crime control. *Policing and Society*, **9**, 315–336.

MAMALIAN, C. A., and LAVIGNE, N. G., 1999, The use of computerized crime mapping by law enforcement: Survey results. NIJ Crime Mapping Research Center, Washington DC.

MOONEY, C. Z., 1997, *Monte Carlo Simulation* (Thousand Oaks, CA: Sage).

MORGAN, R., and NEWBURN, T., 1997, The Future of Policing *(Oxford: Oxford University Press)*.

OPENSHAW, S., 1984, The modifiable areal unit problem. *Concepts and Techniques in Modern Geography*, **38**, 41.

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**, 473–485.

RATCLIFFE, J. H., 2002, Intelligence-led policing and the problems of turning rhetoric into practice. *Policing and Society*, **12**, 53–66.

RATCLIFFE, J. H., and MCCULLAGH, M. J., 1999, Hotbeds of crime and the search for spatial accuracy. *Geographical Systems*, **1**, 385–398.

RICH, T. F., 2001, Crime mapping and analysis in community organization in Hartford, Connecticut. National Institute of Justice, Washington DC.

SHEPTYCKI, J., 2000, Editorial reflections on surveillance and intelligence-led policing. *Policing and Society*, **9**, 311–314.

UNWIN, D. J., 1996, GIS, spatial analysis and spatial statistics. *Progress in Human Geography*, **20**, 540–551.

WALKER, S., and KATZ, C. M., 2001, *The Police In America: An Introduction*, Fourth edition (Boston: McGraw-Hill).

ZHANG, C., and MURAYAMA, Y., 2000, Testing local spatial autocorrelation using k-order neighbours. *International Journal of Geographical Information Science*, **14**, 681–692.