Effects of Geocoding Quality on Predictive Crime Hotspot Mapping

Paul Zandbergen
Department of Geography
University of New Mexico

Timothy Hart
Department of Criminal Justice
University of Nevada, Las Vegas
Predictive Hotspot Mapping

Time Period 1

Time Period 2

Time periods can be years, seasons, months, weeks, shifts, etc.
NIJ Funded Project

• Project #: 2010-696J-00

• People: Timothy Hart, UNLV
          Paul Zandbergen, UNM
          2 graduate students

• Start date: February 2010
• End date: January 2012
Research Questions

How reliable is predictive hotspot mapping?

• Which factors influence this reliability?
• Which hotspot techniques are most reliable?
• How sensitive are the results to data quality?
Factors Influencing Hotspot Robustness

- Type of crime
- Time period
- Urban morphology
- Geocoding quality
- Type of hotspot technique
- Hotspot parameters
Measures of Hotspot Robustness

- **Predictive Accuracy Index (PAI)**
  - Ratio of hit rate to the area percentage
  - Measures predictive accuracy of hot spot
  - Higher values are better

- **Recapture Rate Index (RRI)**
  - Ratio of hot spot crime densities for periods 2 and 1
  - Standardized for change in total number of crimes
  - Higher values are better

- Source: Chainey et al. (2008), Levine (2008), and Van Patten et al. (2009)
Example Calculation – Assaults in Las Vegas

1,366 crimes in 2007 overlay 1 km grid
1,035 km² total area

653 crimes in 2007 within hotspot of 68 km²

1,531 crimes in 2008, of which 580 within hotspot based on 2007 data

Predictive Accuracy Index = \( \frac{580}{1,531} \) / \( \frac{68}{1,035} \) = 5.77

Recapture Rate Index = \( \frac{580}{663} \) * \( \frac{1,366}{1,531} \) = 0.78
Geocoding Crime Events

<table>
<thead>
<tr>
<th>ADDRESS</th>
<th>CITY</th>
<th>STATE</th>
<th>OCCURRED</th>
<th>CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1840 LYELL CANYON LN</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/24/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>1849 LYELL CANYON LN</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/24/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>6548 HOLLY BLUFF CT</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/23/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>3880 CAMBRIDGE ST</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/24/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>3760 S LAS VEGAS BLVD</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/24/2008</td>
<td>ASSAULT</td>
</tr>
<tr>
<td>955 E TWAIN AVE</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/25/2006</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>S GRAND CANYON DR &amp; W KATIE AV</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/26/2006</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>4374 CY CLIFF VIEW CIR</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/25/2006</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>524 S JONES BLVD</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/26/2006</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>1461 W MICHAEL AVE</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/26/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>1261 W SAHARA AVE</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/27/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>1964 LEONA ST</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/27/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>3403 GREENWOOD SPRINGS DR</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/27/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>3354 LAPSALLE LN</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/27/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>3112 MERRITT AVE</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/28/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>6250 SKY POINTE DR</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/26/2008</td>
<td>ASSAULT</td>
</tr>
<tr>
<td>1901 J ST</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/26/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>5917 GRANADA AVE</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/28/2008</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>9200 W ANN RD</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/28/2008</td>
<td>ASSAULT / VICTIM OVER 60</td>
</tr>
<tr>
<td>9200 EY ANN RD</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/28/2006</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
<tr>
<td>4538 LITTLE FINCH LN</td>
<td>LAS VEGAS</td>
<td>NV</td>
<td>4/28/2006</td>
<td>ASSAULT WITH A DEADLY WEAPON</td>
</tr>
</tbody>
</table>
1. Find the zone (ZIP, City, etc.)
2. Match the street (by Name, Type, Dir, etc.)
3. Match the segment with the proper range
4. Linear interpolation along segment
5. Apply offsets
Geocoding Quality Issues

1. Match rate
   – % of events that geocoded reliably

2. Positional accuracy
   – Distance from the “true location”

3. Repeatability
   – Variability from different geocoding methods
Positional Error

Legend
- Address Points
- Street Geocoded Locations
- Connector Lines
- Street Centerlines
- Parcels
Typical street geocoding: 90th percentile is ~ 100 meters
Research Design

• Crime types:
  – Assault, Auto theft, Burglary, Drugs, Homicide, Robbery

• Jurisdictions
  – Albuquerque, NM
  – Arlington, TX
  – Charlotte, NC
  – Las Vegas, NV
  – Seattle, WA
  – Tampa, FL

• Geocoding methods
  – Local street centerlines
  – TIGER 2009
  – ESRI StreetMap
  – 2 commercial services

• Use 2007 data to predict 2008
# Hotspot Methods

<table>
<thead>
<tr>
<th>Type</th>
<th>Method</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate</td>
<td>Grid-based thematic</td>
<td>Grid cell size, threshold</td>
</tr>
<tr>
<td>Aggregate</td>
<td>Local Moran’s I</td>
<td>Areal units</td>
</tr>
<tr>
<td>Point</td>
<td>Kernel density</td>
<td>Kernel type, bandwidth, threshold</td>
</tr>
<tr>
<td>Point</td>
<td>Nearest neighbor hierarchical clustering</td>
<td>Distance type, minimum events per cluster, ellipse vs. convex hull</td>
</tr>
<tr>
<td>Point</td>
<td>Spatial and Temporal Analysis of Crime</td>
<td>Search radius, minimum events per cluster, scan type</td>
</tr>
</tbody>
</table>
General Approach

Aggregation

Point Pattern Analysis

OR

OR
Grid-based Thematic

10 class quintile classification

2 km grid

highest class
Local Moran’s I Clusters

2 km grid

statistically significant high-high cluster

census tracts
Kernel Density
Nearest Neighbor Hierarchical Clustering
Spatial and Temporal Analysis of Crime
Las Vegas Dataset

<table>
<thead>
<tr>
<th>Crime Type</th>
<th>Count 2007</th>
<th>Count 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assault</td>
<td>1,733</td>
<td>1,811</td>
</tr>
<tr>
<td>Auto Burglary</td>
<td>10,985</td>
<td>10,098</td>
</tr>
<tr>
<td>Auto Theft</td>
<td>16,444</td>
<td>11,486</td>
</tr>
<tr>
<td>Burglary</td>
<td>18,064</td>
<td>17,522</td>
</tr>
<tr>
<td>Homicide</td>
<td>124</td>
<td>133</td>
</tr>
<tr>
<td>Robbery</td>
<td>5,300</td>
<td>4,684</td>
</tr>
<tr>
<td>Drugs (calls for service)</td>
<td>10,953</td>
<td>10,954</td>
</tr>
</tbody>
</table>

Source: City of Las Vegas Police Department
Comparison of Robustness

Using auto thefts in 2007 to predict auto thefts in 2008

<table>
<thead>
<tr>
<th>Method</th>
<th>Predictive Accuracy Index</th>
<th>Recapture Rate Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid-based thematic</td>
<td>5.01</td>
<td>0.90</td>
</tr>
<tr>
<td>Local Moran’s I</td>
<td>4.80</td>
<td>0.92</td>
</tr>
<tr>
<td>Kernel density</td>
<td>7.19</td>
<td>0.87</td>
</tr>
<tr>
<td>Nearest neighbor hierarchical clustering</td>
<td>3.90</td>
<td>0.92</td>
</tr>
<tr>
<td>Spatial and Temporal Analysis of Crime</td>
<td>4.04</td>
<td>0.95</td>
</tr>
</tbody>
</table>
## Geocoding Match Rates (%)

<table>
<thead>
<tr>
<th>Crime Type</th>
<th>Local Street Centerlines</th>
<th>ESRI StreetMap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assault</td>
<td>80.0</td>
<td>73.4</td>
</tr>
<tr>
<td>Auto Burglary</td>
<td>82.9</td>
<td>76.8</td>
</tr>
<tr>
<td>Auto Theft</td>
<td>89.2</td>
<td>85.5</td>
</tr>
<tr>
<td>Burglary</td>
<td>90.0</td>
<td>85.7</td>
</tr>
<tr>
<td>Homicide</td>
<td>79.8</td>
<td>75.8</td>
</tr>
<tr>
<td>Robbery</td>
<td>79.9</td>
<td>66.5</td>
</tr>
<tr>
<td>Drugs (calls for service)</td>
<td>63.3</td>
<td>53.3</td>
</tr>
</tbody>
</table>
Geocoding Errors – Out-of-Date Street Network

Las Vegas burglaries 2007

Street Geocoded
- Building/Parcel Geocoded
Geocoding Errors – Positional Errors

Las Vegas burglaries 2007

Street Geocoded

- 1 - 2
- 3 - 6
- 7 - 15
- 16 - 28
- 29 - 61

Building/parcel Geocoded

- 1 - 2
- 3 - 6
- 7 - 15
- 16 - 28
- 29 - 61
Geocoding Errors – Spelling Issues

Las Vegas auto thefts 2007
Effects of Geocoding Quality on Hotspots
Effects of Hotspot Parameters

Las Vegas auto thefts 2007 and 2008

- PAI = 9.67, RRI = 0.82
- PAI = 5.51, RRI = 0.87
- PAI = 5.01, RRI = 0.90
- PAI = 3.79, RRI = 0.94
- PAI = 2.48, RRI = 0.96

Same 1 km grid of auto thefts in 2007, but varying threshold of hotspot boundary
Effects of Hotspot Parameters

![Graphs showing the effects of hotspot parameters.](image)

- Left graph: PAI vs. Kernel Bandwidth (m) for different thresholds.
- Right graph: RRI vs. Kernel Bandwidth (m) for different thresholds.

Legend:
- Blue: > 5' mean
- Green: > 4' mean
- Red: > 3' mean
- Orange: > 2' mean
- Gray: > mean
Conclusions

• Geocoding quality varies strongly by type of crime

• Effects geocoding quality can be substantial for local areas

• Differences between hotspot techniques can be substantial

• Effects of hotspot parameters on robustness is very strong
Next steps

• Additional geocoders

• Expand for all study areas, all crime types

• Consistent comparison of hotspot methods

• Guidelines for geocoding
Contact

Paul Zandbergen
University of New Mexico

zandberg@unm.edu
www.paulzandbergen.com