The value of retrospective spatial trends of crime for predictive crime mapping

Spencer Chainey  
@SpencerChainey  
Director of Geographic Information Science  
University College London
Introduction

• The paradigm of predictive policing
• Retrospective crime data analysis
  – Hotspot analysis
  – Repeat and near repeat victimisation
• From research to operational police strategy
The paradigm for predictive policing

Los Angeles Police Department:
- Predictive modelling algorithm: analysing crime from last 3 years to identify future hotspots
- Police officers asked to give additional attention to these areas (500 ft x 500 ft area (one square block))

Santa Clara, California
- Applying earthquake models (prediction of aftershocks) to predict crime

IBM predictive analytics
- “apply statistical data exploration and machine-learning techniques to historical information in order to uncover hidden patterns, associations, correlations and trends ... includes vast amounts of textual or unstructured data [alongside recorded crime] such as emails, videos and chat room interactions”

Memphis: claim 30% fall in serious crimes and 15% fall in violent crimes
Hotspot analysis

What is a hotspot?

• Geographical area of higher than average crime or disorder
  – Area of crime or disorder concentration, relative to the distribution of crime and disorder across the whole region of interest
  – Hotspots are areas of clusters of crime or disorder that can exist at different scales of interest

• Basic form of crime prediction
  – Takes data from the past to predict the future
Hotspot analysis
Common hotspot mapping techniques

**Point map**
- Best for location, size, shape and orientation of hotspot
- 9 out of 10 intelligence professionals prefer it

**Grid thematic map**

**Thematic map of geographic administrative units**

**Kernel density estimation map**
- Best for location, size, shape and orientation of hotspot
- 9 out of 10 intelligence professionals prefer it
Hotspot analysis – kernel density estimation
Examples from the UK

Birmingham

Leeds

Middlesbrough

Oxford

Serious Acquisitive Offenders

Birmingham

Liverpool
Hotspot analysis – comparison of techniques
London Metropolitan Police: Camden and Islington BCUs
Hotspot analysis: prediction comparisons

- Results from research
  - Prediction Accuracy Index
    - The higher the Index value, the better the prediction

<table>
<thead>
<tr>
<th>Hotspot mapping technique</th>
<th>Average PAI (01/01/2003)</th>
<th>Average PAI (13/03/2003)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial ellipses 250m</td>
<td>1.74</td>
<td>2.25</td>
</tr>
<tr>
<td>Spatial ellipses 500m</td>
<td>1.24</td>
<td>1.52</td>
</tr>
<tr>
<td>Spatial ellipses HSD</td>
<td>1.69</td>
<td>2.03</td>
</tr>
<tr>
<td>Thematic mapping of output areas</td>
<td>1.91</td>
<td>2.38</td>
</tr>
<tr>
<td>Thematic mapping of grids 250m</td>
<td>2.00</td>
<td>2.34</td>
</tr>
<tr>
<td>Thematic mapping of grids HSD</td>
<td>2.06</td>
<td>2.63</td>
</tr>
<tr>
<td>Kernel density estimation</td>
<td>2.90</td>
<td>3.41</td>
</tr>
</tbody>
</table>

Values in bold indicate the highest values and values in italics indicate the lowest PAI values. Results are presented for each of the dates when hotspot maps were generated. These results show that KDE consistently produced the best hotspot maps for predicting future events.
### Hotspot analysis: prediction comparisons

#### Table 7: PAI values for different hotspot mapping techniques, by crime type

<table>
<thead>
<tr>
<th>Hotspot mapping technique</th>
<th>Residential burglary</th>
<th>Street crime</th>
<th>Theft from vehicle</th>
<th>Theft of vehicle</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(a) PAI values calculated from the 1 January 2003 measurement date</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial ellipses 250 m</td>
<td>1.38</td>
<td>2.36</td>
<td>2.18</td>
<td>1.65</td>
</tr>
<tr>
<td>Spatial ellipses 500 m</td>
<td>1.34</td>
<td>1.46</td>
<td>1.54</td>
<td>0.82</td>
</tr>
<tr>
<td>Spatial ellipses HSD</td>
<td>1.43</td>
<td>2.45</td>
<td>2.12</td>
<td>1.29</td>
</tr>
<tr>
<td>Thematic mapping of output areas</td>
<td>1.10</td>
<td>4.20</td>
<td>1.17</td>
<td>1.18</td>
</tr>
<tr>
<td>Thematic mapping of grids 250 m</td>
<td>1.70</td>
<td>4.04</td>
<td>1.82</td>
<td>1.37</td>
</tr>
<tr>
<td>Thematic mapping of grids HSD</td>
<td>1.68</td>
<td>3.46</td>
<td>2.12</td>
<td>2.06</td>
</tr>
<tr>
<td><strong>Kernel density estimation</strong></td>
<td><strong>2.31</strong></td>
<td><strong>4.68</strong></td>
<td><strong>2.29</strong></td>
<td><strong>2.32</strong></td>
</tr>
<tr>
<td><em>(b) PAI values calculated from the 13 March 2003 measurement date</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial ellipses 250 m</td>
<td>1.32</td>
<td>2.59</td>
<td>2.15</td>
<td>2.93</td>
</tr>
<tr>
<td>Spatial ellipses 500 m</td>
<td>1.31</td>
<td>1.40</td>
<td>1.55</td>
<td>1.82</td>
</tr>
<tr>
<td>Spatial ellipses HSD</td>
<td>1.29</td>
<td>2.63</td>
<td>2.63</td>
<td>1.59</td>
</tr>
<tr>
<td>Thematic mapping of output areas</td>
<td>1.25</td>
<td>3.32</td>
<td>2.93</td>
<td>2.01</td>
</tr>
<tr>
<td>Thematic mapping of grids 250 m</td>
<td>1.67</td>
<td>3.58</td>
<td>2.43</td>
<td>1.66</td>
</tr>
<tr>
<td>Thematic mapping of grids HSD</td>
<td>1.95</td>
<td>4.14</td>
<td>2.55</td>
<td>1.89</td>
</tr>
<tr>
<td><strong>Kernel density estimation</strong></td>
<td><strong>2.33</strong></td>
<td><strong>4.59</strong></td>
<td><strong>3.66</strong></td>
<td><strong>3.05</strong></td>
</tr>
</tbody>
</table>

Values in bold indicate the highest values and values in italics indicate the lowest PAI values. These results show that KDE consistently produced the best hotspot maps for predicting spatial patterns of crime for all crime types, and that in some cases STAC was not the worst performer. Instead, thematic mapping of output areas generated the lowest PAI values for residential burglary, and in one case for theft from vehicles.
Figure 4. Hotspot maps generated from 3 months of residential burglary input data (measurement date of the 1 January 2003) using (a) STAC, (b) thematic mapping of output areas, (c) grid thematic mapping and (d) KDE. Each map is shown with its PAI value, based on 1 month of measurement data.
Hotspot analysis

How many crimes does KDE predict?

- Controlling for 3% of area as hotspot

<table>
<thead>
<tr>
<th>Crime type</th>
<th>PAI</th>
<th>Crimes committed in January 2010</th>
<th>Number of crimes in hotspots</th>
<th>Percentage of crimes in hotspots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential Burglary</td>
<td>2.77</td>
<td>470</td>
<td>39</td>
<td>8%</td>
</tr>
<tr>
<td>Street Crime</td>
<td>6.59</td>
<td>460</td>
<td>91</td>
<td>20%</td>
</tr>
<tr>
<td>Theft from Vehicle</td>
<td>3.98</td>
<td>962</td>
<td>115</td>
<td>12%</td>
</tr>
<tr>
<td>Theft of Vehicle</td>
<td>3.26</td>
<td>307</td>
<td>30</td>
<td>10%</td>
</tr>
</tbody>
</table>

Number of crimes predicted using kernel density estimation. Hotspot map generated using three months of retrospective crime data to determine where crimes in the next month may occur (i.e. today being 1st January 2010). The area determined as ‘hot’ was controlled to cover only 3% of the study area’s total area.
Hotspot analysis
KDE weaknesses: how many hotspots?!

- Thematic thresholds to apply?
- Left to the whims and fancies of the map producer
- Trial and error, experimentation, experience, whatever suits your circumstance

One main hotspot

Lots of hotspots!

Which hotspots are statistically significant? i.e. is the spatial concentration of these crime incidents really unusual?
Hotspot analysis

Spatial significance mapping

• Statistical significance: 90%, 95%, 99%, 99.9%
  i.e. the spatial clustering of crime is extremely unusual

• LISA statistics (Local Indicators of Spatial Association)
  – Identify the local association between an observation and its neighbours, up to a specified distance from the observation
  – Local Moran’s I, Local Geary’s C, Getis and Ord Gi and Gi* statistics
Hotspot analysis
Gi* statistic
• Gi* results are Z scores
• Z score used extensively to determine statistical significance/confidence thresholds
  – 90% significant: >= 1.645
  – 95% significant: >= 1.960
  – 99% significant: >= 2.576
  – 99.9% significant: >= 3.291 i.e. something exceptionally unusual has happened at this location in terms of the spatial concentration of crime
• Universal Z score values
  – the same values apply, regardless of crime type, the location of your study area, the size of your study area ...
Hotspot analysis

Gi* statistic

Thematic class values:

- 90% significant: $\geq 1.645$
- 95% significant: $\geq 1.960$
- 99% significant: $\geq 2.576$
- 99.9% significant: $\geq 3.291$
Hotspot analysis – KDE versus Gi*

90% significant: Gi* z score > 1.645; 95% significant: Gi* z score > 1.960;
99% significant: Gi* z score > 2.576; 99.9% significant: Gi* z score > 3.291
Hotspot analysis

Gi* statistic

- **Results from research** - higher Prediction Accuracy Index (PAI), better it is at predicting where crime will happen
  - Gi* gives best results (shown for 95% significance level)

**PAI results for Gi* and common hotspot mapping techniques**

- Thematic mapping of geographic units
- Thematic mapping of grids
- Kernel density estimation
- Gi*
Hotspot analysis

How many crimes does Gi* predict?
• Controlling for 3% of area as hotspot

<table>
<thead>
<tr>
<th>Crime type</th>
<th>Crimes committed in January 2010</th>
<th>KDE Percentage of crimes in hotspots</th>
<th>Gi* statistic Number of crimes in hotspots</th>
<th>Gi* statistic Percentage of crimes in hotspots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential Burglary</td>
<td>470</td>
<td>8%</td>
<td>54</td>
<td>12%</td>
</tr>
<tr>
<td>Street Crime</td>
<td>460</td>
<td>20%</td>
<td>106</td>
<td>23%</td>
</tr>
<tr>
<td>Theft from Vehicle</td>
<td>962</td>
<td>12%</td>
<td>125</td>
<td>13%</td>
</tr>
<tr>
<td>Theft of Vehicle</td>
<td>307</td>
<td>10%</td>
<td>58</td>
<td>19%</td>
</tr>
</tbody>
</table>

Number of crimes predicted using Gi* (in comparison to KDE). Hotspot map generated using three months of retrospective crime data to determine where crimes in the next month may occur (i.e. ‘today’ being 1st January 2010). The area determined as ‘hot’ was controlled to cover only 3% of the study area’s total area.
Repeat victimisation

- Hotspot analysis is good for identifying persistent trends
- Very recent victimisation is an excellent predictor of immediate future risk
  - Risk doubles following a victimisation
  - Burglary, domestic violence, bike theft, vehicle crime ...
  - Repeats occur swiftly offering a limited opportunity for intervention
- 7-15% of all burglaries are RVs

<table>
<thead>
<tr>
<th>Offense</th>
<th>Repeat Offences</th>
<th>Repeat Victims</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic violence</td>
<td>62%</td>
<td>28%</td>
<td>Merseyside, England(^9)</td>
</tr>
<tr>
<td></td>
<td>42%</td>
<td>31%</td>
<td>West Yorkshire, England(^9)</td>
</tr>
<tr>
<td>Commercial robbery</td>
<td>65%</td>
<td>32%</td>
<td>Indianapolis, Indiana(^11)</td>
</tr>
<tr>
<td>Gas station robbery</td>
<td>62%</td>
<td>37%</td>
<td>Australia(^12)</td>
</tr>
<tr>
<td>Bank robbery</td>
<td>58%</td>
<td>36%</td>
<td>England(^13)</td>
</tr>
<tr>
<td>Residential burglary</td>
<td>32%</td>
<td>15%</td>
<td>Nottinghamshire, England(^14)</td>
</tr>
<tr>
<td></td>
<td>13%</td>
<td>7%</td>
<td>Merseyside, England(^15)</td>
</tr>
<tr>
<td></td>
<td>32%</td>
<td>16%</td>
<td>Beenleigh, Australia(^15)</td>
</tr>
<tr>
<td></td>
<td>25%</td>
<td>9%</td>
<td>Enschede, Netherlands(^11)</td>
</tr>
<tr>
<td>Commercial burglary</td>
<td>66%</td>
<td>36%</td>
<td>Austin, Texas(^18)</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>14%</td>
<td>Merseyside, England(^19)</td>
</tr>
<tr>
<td>Residential and</td>
<td>89%</td>
<td>18%</td>
<td>Charlotte, North Carolina(^20)</td>
</tr>
<tr>
<td>commercial burglary</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Near Repeat Victimisation

- Near repeat victimisation
  - Neighbours are at heightened risk
  - Decays rapidly in space and time

- Near repeat patterns have been found for:
  - Theft from motor vehicle
  - Gun crime
  - Street robbery
  - Bicycle theft
  - Road-side bombs in Iraq

- Near repeats account for 10-50% of all crime
- Near repeat patterns found in many areas in UK, USA, Canada, Australia, Netherlands, New Zealand
Hotspot analysis, repeats and near repeats

• Hotspot analysis is good at identifying persistent problem areas
  – And at predicting where crime is likely to persist in the future
  – Spatial significance mapping using Gi* illustrates the effective role of hotspot analysis for predicting future events: almost a quarter of all crime in 3% of area
  – Straightforward to apply: ArcGIS

• Repeat and near repeat victimisation analysis identifies where crime is likely to occur in next few days
Example: predictive policing in practice
Trafford, Greater Manchester Police
Predictive policing in practice
Using RV, NRV and hotspot analysis in Trafford, Manchester

- Trafford, Manchester
- Target: reduce burglary in 2011 by 9.3%
- Four burglaries per day
Predictive policing in practice
Using RV, NRV and hotspot analysis in Trafford, Manchester

• Tactical focus: reduce repeat and near repeat victimisation
  – Produce maps and time charts every four days
  – 100m buffer around burgled property
Predictive policing in practice
Using RV, NRV and hotspot analysis in Trafford, Manchester

We know where there is a heightened risk

Crime prevention 1: Minimise risk of RV
• Crime prevention officer – improve security: visit burgled properties within 24 hours
• Immediate action to address vulnerabilities
Predictive policing in practice
Using RV, NRV and hotspot analysis in Trafford, Manchester

We know where there is a heightened risk

Crime prevention 2: Minimise risk of NRV

- Visit neighbouring properties; as much face-to-face contact with residents as possible:
  - Inform – Reassure – Advise
  (start with those within 100m)

I’m not sure whether you are aware, but there was a burglary a few doors up yesterday.

The chances of you being burgled are very low.

There are a couple of things you can do to help us out
- Report suspicious behaviour
- Offer seasonally-sensitive, tailored crime prevention advice
Predictive policing in practice
Using RV, NRV and hotspot analysis in Trafford, Manchester

• Persistent and emerging problem areas
  – 20% of burglary (in addition to RVs and NRVs) in 4% of area
  – Working with residents in hotspots
    • Security improvements
  – Crime prevention advice
    • Targeted, tailored and seasonally-sensitive
    • E.g. Targeting of car keys in summer
Predictive policing in practice
Using RV, NRV and hotspot analysis in Trafford, Manchester

Results: 27% reduction in burglary

- Targeted areas: 42% reduction
- Non-targeted area: Very little change

Repeat victimisation
- Expect 4 per month
- 1 every two months

Near repeat victimisation
- Expect 4 per week
- 2 per week

Public confidence significantly higher than GMP average
- 3000+ visits since Oct 2011
Summary

• Predictive analytics
  – Much is being promised from what sounds like sophisticated (meaning expensive!) solutions
  – We need to avoid being seduced by the technology

• Hotspot analysis
  – Basic form of prediction: use of retrospective patterns of crime to predict the future
  – Identifies small proportion of area where lots of crime happens!

• Repeat and near repeat victimisation
  – More sensitive in predicting what may happen in next few days from the very recent events

• Predictive mapping that uses a combination RV, NRV and hotspot analysis already shown good promise ...

• And is easy to do ...
Thank you

Spencer Chainey
s.chainey@ucl.ac.uk
@SpencerChainey

• Our research publications
  – Hotspot analysis: me
  – RV and NRV: Bowers and Johnson

www.ucl.ac.uk/scs
Hotspot analysis
KDE parameter selection

• Cell size; resolution
• Bandwidth; search radius
  – Fixed interval (size of bandwidth); adaptive (number of points)
Hotspot analysis
KDE parameters – cell size

- Very little influence on hotspot prediction accuracy

![Maps showing hotspot analysis](image)

- Cell size: 75m
  - Bandwidth: 440m
  - PAI: 6.72

- Cell size: 15m
  - Bandwidth: 440m
  - PAI: 6.75

![Influence of cell size on the accuracy of hotspot maps](image)
Hotspot analysis
KDE parameters – bandwidth

- The smaller the bandwidth, the higher the hotspot prediction accuracy
  - Affected street crime (robbery and theft from person) more than burglary and vehicle crime

Cell size: 50m
Bandwidth: 125m
PAI: 9.84

Cell size: 50m
Bandwidth: 225m
PAI: 6.51
Hotspot analysis

Gi* statistic

- Requires data to be aggregated to some form of geographic unit (e.g. count of crime per beat, grid cell)
  - Define neighbourhood: units within a specified radius
Hotspot analysis
Shelf-life of the hotspot map

• Does a hotspot map’s accuracy change over its shelf-life?
  – Street crime hotspot map’s accuracy improved as the map aged
  – Residential burglary, theft from vehicles and theft of vehicle hotspot maps showed no significant change over time
Hotspot analysis

Influence of ‘recency’ of crime data to hotspot map accuracy

• E.g. residential burglary
  – Very recent events (crime in last 2 and 3 days) were most accurate at showing what would happen in the next few days
  – Stabilise thereafter

![Comparison between input data volumes and 'shelf-life' hotspot accuracy for residential burglary (using KDE)]