Project Participants

- Towson University Applied Mathematics Laboratory
  - Undergraduate research projects in applied mathematics.
  - Founded in 1980
- National Institute of Justice
- Special thanks to Stanley Erickson (NIJ), Ron Wilson (NIJ) and Andrew Engel (SAS)
Collaborators

Dr. Coy L. May (Towson University) (2005-2006, 2006-2007)

2005-2006 Students:
- Paul Corbitt
- Brooke Belcher
- Brandie Biddy

2006-2007 Students:
- Chris Castillo
- Adam Fojtik
- Ruozhen Yao
- Melissa Zimmerman
- Gregory Emerson
- Laurel Mount
- Grant Warble

2007-2008 Students:
- Lauren Amrhine
- Colleen Carrion
- Chris Castillo
- Yu Fu
- Natasha Gikunju
- Kristopher Seets
Geographic Profiling

The Question:

Given a series of linked crimes committed by the same offender, can we make predictions about the anchor point of the offender?

The anchor point can be a place of residence, a place of work, or some other commonly visited location.
Existing Methods

- Spatial distribution strategies
- Probability distance strategies

Notation:
- Anchor point- $z = (z^{(1)}, z^{(2)})$
- Crime sites- $x_1, x_2, \cdots, x_n$
- Number of crimes- $n$
Distance

- How do we measure the distance between points?
  - Euclidean: \[ d_2(x, y) = \sqrt{(x^{(1)} - y^{(1)})^2 + (x^{(2)} - y^{(2)})^2} \]
Distance

- How do we measure the distance between points?
- **Manhattan:**

\[
d_1(x, y) = |x^{(1)} - y^{(1)}| + |x^{(2)} - y^{(2)}|
\]

From Google Maps
Distance

- How do we measure the distance between points?
  - Highway distance
Distance

- How do we measure the distance between points?
- Street distance
Spatial Distribution Strategies

- **Centroid:**
  \[
  \hat{z}_{\text{centroid}} = \frac{1}{n} \sum_{i=1}^{n} x_i
  \]
Spatial Distribution Strategies

Center of minimum distance: $\hat{z}_{cmd}$ is the value of $y$ that minimizes

$$D(y) = \sum_{i=1}^{n} d(x_i, y)$$
Spatial Distribution Strategies

- Circle Method (*Canter & Larkin, 1993*):
  - Anchor point contained in the circle whose diameter are the two crimes that are farthest apart.

![Diagram showing anchor point and crime locations within a circle]
Probability Distribution Strategies

• The anchor point is located in a region with a high “hit score”.

• The hit score $S(y)$ has the form
  \[ S(y) = \sum_{i=1}^{n} f(d(y, x_i)) \]
  \[ = f(d(z, x_1)) + f(d(z, x_2)) + \cdots + f(d(z, x_n)) \]

where $x_i$ are the crime locations and $f$ is a decay function and $d$ is a distance.
Probability Distribution Strategies

- Linear:
  \[ f(d) = A - Bd \]
Rossmo (Rigel)

- Manhattan distance metric.
- Decay function

\[
f(d) = \begin{cases} 
  \frac{k}{d^h} & \text{if } d > B \\
  \frac{kB^{g-h}}{(2B-d)^g} & \text{if } d \leq B
\end{cases}
\]

- The constants \( k, g, h \) and \( B \) are empirically defined.
Canter, Coffey, Huntley & Missen (Dragnet)

- Euclidean distance
- Decay functions
  \[ f(d) = Ae^{-\beta d} \]
  \[ f(d) = \begin{cases} 0 & \text{if } d < A, \\ 1 & \text{if } A \leq d < B, \\ Ce^{-\beta d} & \text{if } d \geq B. \end{cases} \]
- Calibrated against homicide data
Levine (CrimeStat)

- Euclidean distance
- Decay functions
  - Linear \( f(d) = A + Bd \)
  - Negative exponential \( f(d) = Ae^{-\beta d} \)
  - Normal \( f(d) = \frac{A}{\sqrt{2\pi S^2}} \exp\left[\frac{-(d - \bar{d})^2}{2S^2}\right] \)
  - Lognormal \( f(d) = \frac{A}{d\sqrt{2\pi S^2}} \exp\left[\frac{-(\ln d - \bar{d})^2}{2S^2}\right] \)
Levine (CrimeStat)

From Levine (2004)
Probability Distribution Strategies

- Existing methods differ in their choices of
  - The distance measure, and
  - The distance decay function;

but share the common mathematical heritage:

\[ S(y) = \sum_{i=1}^{n} f(d(y, x_i)) \]

- In practice, \( S(y) \) may be evaluated only at discrete values \( y_j \) giving us a hit score matrix

\[ S_{ij} = \sum_{i=1}^{n} f(d(y_j, x_i)) \]
Controversies

- What is the right tool?

An Evaluation of NIJ’s Evaluation Methodology for Geographic Profiling Software

D. Kim Rossmo
Research Professor and Director
Center for Geospatial Intelligence and Information
Department of Criminal Justice
Texas State University

May 8, 2005

The Evaluation of Geographic Profiling Software:
Response to Kim Rossmo’s Critique of the NIJ Methodology

The recent discussion of geographical profiling (GP) methodology has been a testament to the contribution that Kim Rossmo made in critiquing the GP roundtable discussion that was posted on the MAPS website. I’ve had some time to think about his critique and would like to offer some thoughts.

Let me first start by saying that I had nothing to do with the GP roundtable even though my CrimeStat Journey to Crime (Jtc) was one of the tested methods (what Rossmo calls ‘the NIJ methodology’). Like any developer, there are things that the roundtable evaluation did that I wouldn’t necessarily agree with. But, that’s the prerogative of an evaluator and usually just shows that there is more than one way to do things.
Controversies

- Are they operationally effective?
Shortcomings

- These techniques are all *ad hoc*.
- What is their theoretical justification?
  - What assumptions are being made about criminal behavior?
  - What mathematical assumptions are being made?
Shortcomings

- The convex hull effect:
  - The anchor point always occurs inside the convex hull of the crime locations.
Shortcomings

- These methods require some *a priori* knowledge of the offender's distance decay function.
  - In particular, they require an estimate of the distance that the serial offender is likely to travel before the analysis process begins.
  - Indeed, the constant(s) that appear in the distance decay function must be selected before starting the analysis.
Shortcomings

- How do you add in local information?
- How could you incorporate socio-economic variables into the model?

Snook, *Individual differences in distance travelled by serial burglars*
Malczewski, Poetz & Iannuzzi, *Spatial analysis of residential burglaries in London, Ontario*
Bernasco & Nieuwbeerta, *How do residential burglars select target areas?*
Osborn & Tseloni, *The distribution of household property crimes*
Geographic Profiling

What characteristics should a good geographic profiling method possess?

1. It should be mathematically rigorous.
2. There should be explicit connections between assumptions on offender behavior and components of the mathematical model.
Geographic Profiling

What (other) characteristics should a good geographic profiling technique possess?

3. It should take into account local geographic features that affect:
   a. The selection of a crime site;
   b. The selection of an anchor point.

4. It should rely only on data available to local law enforcement.

5. It should return a prioritized search area.
Main Result

- We have developed a fundamentally new mathematical technique for geographic profiling.
  - We have implemented the algorithm in software, and begun testing it on actual crime series.
A New Approach

Let us start with a model of offender behavior.

In particular, let us begin with the ansatz that an offender with anchor point $z$ commits a crime at the location $x$ according to a probability density function $P(x \mid z)$.

This is an inherently continuous model.
Modeling with Probability

- Probabilistic models are commonly used to model problems that are deterministic.
  - Stock market
  - Population genetics
  - Heat flow
  - Chemical diffusion
A New Approach

- Assumptions about
  - The offender's likely behavior, and
  - The local geography

can then be incorporated into the form of $P(x \mid z)$. 
The Mathematics

Given crimes located at $x_1, x_2, \cdots, x_n$ the maximum likelihood estimate for the anchor point $\hat{z}_{mle}$ is the value of $y$ that maximizes

$$L(y) = \prod_{i=1}^{n} P(x_i \mid y)$$

$$= P(x_1 \mid y) P(x_2 \mid y) \cdots P(x_n \mid y)$$

or equivalently, the value that maximizes

$$\lambda(y) = \sum_{i=1}^{n} \ln P(x_i \mid y)$$

$$= \ln P(x_1 \mid y) + \ln P(x_2 \mid y) + \cdots + \ln P(x_n \mid y)$$
Relation to Spatial Distribution Strategies

If we assume offenders choose target locations based only on a distance decay function in normal form:

\[
P(x | z) = \frac{1}{2\pi \sigma^2} \exp \left[ -\frac{|x - z|^2}{2\sigma^2} \right]
\]

Then the maximum likelihood estimate for the anchor point is the centroid.
Relation to Spatial Distribution Strategies

If we assume offenders choose target locations based only on a distance decay function in exponentially decaying form:

\[ P(x \mid z) = \frac{1}{2\pi \sigma^2} \exp \left[ -\frac{|x - z|}{\sigma} \right] \]

Then the maximum likelihood estimate is the center of minimum distance.
Relation to Probability Distance Strategies

What is the log likelihood function?

\[
\lambda(y) = \sum_{i=1}^{n} \left[ -\ln(2\pi\sigma^2) - \frac{|x_i - y|}{\sigma} \right]
\]

This is the hit score \(S(y)\) provided we use Euclidean distance and the linear decay \(f(d) = A + Bd\) for

\[
A = -\ln(2\pi\sigma^2)
\]

\[
B = -1/\sigma
\]
Parameters

- The maximum likelihood technique does not require \textit{a priori} estimates for parameters other than the anchor point.

\[
P(x \mid z, \sigma) = \frac{1}{2\pi \sigma^2} \exp \left[ -\frac{|x - z|^2}{2\sigma^2} \right]
\]

The same process that determines the best choice of \( z \) also determines the best choice of \( \sigma \).
Better Models

- We have recaptured the results of existing techniques by choosing $P(x \mid z)$ appropriately.
- These choices of $P(x \mid z)$ are not very realistic.
  - Space is homogeneous and crimes are equi-distributed.
  - Space is infinite.
  - Decay functions were chosen arbitrarily.
Better Models

- Our framework allows for better choices of $P(x \mid z)$.
- Consider

$$P(x \mid z) = D(d(x, z)) \cdot G(x) \cdot N(z)$$

- Distance Decay (Dispersion Kernel)
- Geographic factors
- Normalization
Geography

• What geographic factors should be included in the model?

Snook, *Individual differences in distance travelled by serial burglars*
Malczewski, Poetz & Iannuzzi, *Spatial analysis of residential burglaries in London, Ontario*
Bernasco & Nieuwbeerta, *How do residential burglars select target areas?*
Osborn & Tseloni, *The distribution of household property crimes*
Geography

- This approach has some problems.
  - Different crimes have different etiologies.
    - We would need to study each different crime type.
  - There are regional differences.
    - Tseloni, Wittebrood, Farrell and Pease (2004) noted that increased household affluence indicated higher burglary rates in Britain, and indicated lower burglary rates in the U.S.
Geography

- Instead, we assume that historical crime rates are reasonable predictors of the likelihood that a particular region will be the site of an offense.
  - Rather than explain crime rates in terms of underlying geographic variables, we simply measure the resulting geographic variability.
  - Let $G(x)$ represent the local attractiveness of potential targets.
Geography

- An analyst can determine what historical data should be used to generate the geographic target density function.
  
  - Different crime types will necessarily generate different functions $G(x)$.
  
  - $G(x)$ is calculated in the same fashion as hot spots; e.g. by kernel density parameter estimation.

\[
G(x) = \sum_{i=1}^{N} K(x - y_i)
\]
The target attractiveness $G(x)$ must also account for jurisdictional boundaries.

Suppose that a law enforcement agency gets reports for all crimes within the region $J$, and none from outside $J$.

Then we must have

$$G(x) = 0 \quad \text{for all} \quad x \not\in J$$

as no crimes that occur outside $J$ will be known to that agency.
Distance Decay

Figure 10.4:
Journey to Crime Distances: All Crimes
Negative Exponential Distribution

From Levine (2004)
Distance Decay

- Buffer Zones
  - A region around the offenders anchor point where they are less likely to offend.
  - Fear of recognition.
Distance Decay: Buffer Zones
Distance Decay: Buffer Zones

Rayleigh Distribution

[Graph showing a Rayleigh distribution curve with a peak around 0.001 and decreasing values as the distance increases.]
Distance Decay
Distance Decay
Distance Decay

- Different offenders have a different average distance they will travel to their target.
  - Suppose that each offender has a decay function \( f(d|\sigma) \) where \( \sigma \) is the average distance the offender travels.
  - Suppose the average distance an offender travels has distribution \( \phi(\sigma) \).
- Then the aggregate distribution over all offenders is
  \[
  F(d) = \int_0^\infty f(d|\sigma) \cdot \phi(\sigma) \, d\sigma
  \]
Distance Decay

Distance between home and crime site, for 2000 burglaries in Baltimore County.

Warning: Preliminary results!

From Levine (2004)
Distance Decay

Estimate for the average distance an offender will travel from home to the offense site.

Warning: Preliminary results!
Distance Decay

Comparing the actual data (yellow) to the fitted model (purple).
Normalization

The expression

\[ P(x \mid z) = D(d(x, z)) \cdot G(x) \cdot N(z) \]

is to represent a probability density function; as a consequence,

\[ N(z) = \frac{1}{\iint G(y) D(d(y, z)) \, dy^{(1)} \, dy^{(2)}} \]
We can estimate the offender's anchor point by calculating the *maximum likelihood estimate*.

We do this by finding the maximum of

$$L(y) = \prod_{i=1}^{n} D(d(x_i, y)) G(x_i)$$

$$\left[ \iint J \frac{D(d(\xi, y)) G(\xi) d\xi^{(1)} d\xi^{(2)}}{\iint J} \right]^n$$
Implementation

- We have implemented this algorithm in software.
  - Integration was performed using a seven-point fifth-order Gaussian method.
  - Optimization was performed using a cyclic coordinate technique with a Hooke and Jeeves accelerator.
  - Running time with \(~650\) boundary vertices and \(~1000\) historical crimes is \(~10\) minutes.
Profiler
Version 0.12 (Pre-Release)

Using Default Parameter file: \Parameters\Parameters.txt
Using Geography file: \Parameters\baltimore_county.txt
Using Crime Series file: \Parameters\BCData\Crimes.txt
Using Historical data file: \Parameters\BCData\History.txt
Using Output file: \Parameters\BCData\Likelihood.kml

Triangulating region
Setting up target density
Calculating mean nearest neighbor distance
Precomputing target density
Constructing Likelihood Function
Constructing Initial Guess
Initial spatial guess = ( -76.731598 , 39.311223 )
Initial sigma guess = 44.217570

Approximations to anchor point and sigma

<table>
<thead>
<tr>
<th>i</th>
<th>x</th>
<th>y</th>
<th>sigma</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-76.731598</td>
<td>39.311223</td>
<td>44.217570</td>
<td>1.892049e+023</td>
</tr>
<tr>
<td>1</td>
<td>-76.716733</td>
<td>39.312793</td>
<td>71.545719</td>
<td>3.909148e+023</td>
</tr>
<tr>
<td>2</td>
<td>-76.716733</td>
<td>39.314912</td>
<td>73.128008</td>
<td>4.040361e+023</td>
</tr>
<tr>
<td>3</td>
<td>-76.715180</td>
<td>39.314912</td>
<td>72.770779</td>
<td>4.052184e+023</td>
</tr>
<tr>
<td>4</td>
<td>-76.715180</td>
<td>39.314912</td>
<td>72.770779</td>
<td>4.052184e+023</td>
</tr>
</tbody>
</table>

Estimate of anchor point = ( -76.715180 , 39.314912 )
Estimate of sigma = 72.770779

Writing KML file for likelihood function

C:\Documents and Settings\moleary\Desktop\v 0.12 devel\Profiler\release>.
Bayesian Methods

We can get a better estimate of the search area via Bayesian methods.

- We need a *prior* estimate for the distribution of anchor points, before the crime series is considered.
- We also need a *prior* estimate of the average distance an offender is likely to travel.
To calculate the prior distribution of anchor points, we suppose that they are proportional to the local population density, and use block level census data.

- Choose a kernel functions $K(x|\lambda)$ with bandwidth $\lambda$.
- Let block $i$ have center $y_i$, population $P_i$ and area $A_i$, and set $\lambda_i = C\sqrt{A_i}$ for some constant $C$.
- Then $H(z) = \sum_{i \in I} P_i K(z - y_i | \lambda_i)$
Bayesian Methods

- If we have $n$ crimes, and we assume that the crime locations are all independent then

$$P(z|x_1, x_2, \cdots, x_n)$$

$$\propto \int_0^\infty \prod_{i=1}^n P(x_i|z, \sigma) H(z) \pi(\sigma) \, d\sigma$$

$$\propto \int_0^\infty \prod_{i=1}^n D(d(x_i, z), \sigma) G(x_i) \cdot N(z) H(z) \pi(\sigma) \, d\sigma$$
Implementation

- Code to implement these techniques is under development.
  - Estimated completion dates:
    - First draft: late May
    - Complete: late August.
Operational Questions

- A good decision support system should not require that the end user be a content expert.
- Our software is being developed with ease of use uppermost in mind.
Strengths

- This approach meets all of our requirements for a good geographic profiling technique.
  - All of the assumptions on criminal behavior are made in the open.
  - They can be challenged, tested, discussed, and compared.
Weaknesses of this Framework

- GIGO
  - The method is only as accurate as the accuracy of the choice of $P(x \mid z)$
  - It is unclear what the right choice is for $P(x \mid z)$
  - Even with the simplifying assumption that
    $$P(x \mid z) = D(d(x, z)) \cdot G(x) \cdot N(z)$$
    this is difficult.
Weaknesses

- The framework assumes that crime sites are independent, identically distributed random variables.
  - This is probably false in general!
- This should be a solvable problem though...
Next Steps

- Model improvements:
  - What would a better choice for the model of criminal behavior?
  - Model selection and multi-model inference.
Questions?

Contact information:

Dr. Mike O'Leary
Director, Center for Applied Information Technology
Towson University
Towson, MD 21252
410-704-7457
moleary@towson.edu