The Mathematics of Geographic Profiling

Towson University
Applied Mathematics Laboratory

Dr. Mike O'Leary

Ninth Crime Mapping Research Conference
Pittsburgh, PA
March 28 - 31, 2007

Supported by the NIJ through grant 2005-IJ-CX-K036

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), Iara Infosino (NIJ), Thomas Sexton (NIJ), Andrew Engel (SAS), and Coy May (Towson University).

Students

- 2005-2006:
  - Paul Corbitt
  - Brooke Belcher
  - Brandie Biddy
  - Gregory Emerson
  - Laurel Mount
  - Ruozhen Yao
  - Melissa Zimmerman
- 2006-2007:
  - Chris Castillo
  - Adam Fojtik
  - Jonathan Vanderkolk
  - Grant Warble

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.

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 been able to implement the algorithm in software, and begun testing it on actual crime series.

Existing Methods

- Spatial distribution strategies
- Probability distance strategies

- Notation:
  - Anchor point- $\boldsymbol{z} = (z^{(1)}, z^{(2)})$
  - Crime sites- $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$
  - Number of crimes- n

Distance

- Euclidean

$$d_2(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(x^{(1)} - y^{(1)})^2 + (x^{(2)} - y^{(2)})^2}$$

- Manhattan

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

- Street grid

Spatial Distribution Strategies

- Centroid:

$$\zeta_{centroid} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i = \frac{\boldsymbol{x}_1 + \boldsymbol{x}_2 + \cdots + \boldsymbol{x}_n}{n}$$

- Center of minimum distance: ζ_{cmd} is the value of y that minimizes

$$D(y) = \sum_{i=1}^{n} d(\boldsymbol{x}_i, y)$$

- Circle Method: The anchor point is contained in the circle whose diameter are the two crimes that are farthest apart.

Probability Distribution Strategies

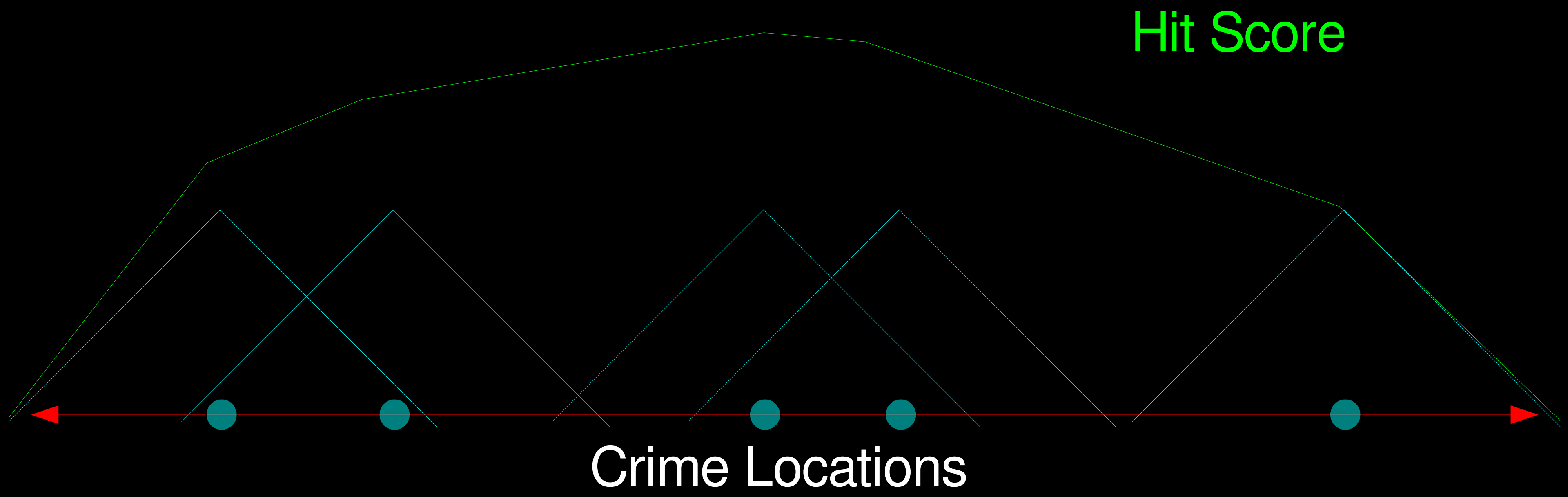
- The anchor point is located in a region with a high *hit score*.
- The hit score $S(\boldsymbol{y})$ has the form

$$S(\boldsymbol{y}) = \sum_{i=1}^{n} f(d(\boldsymbol{y}, \boldsymbol{x}_i))$$
$$= f(d(\boldsymbol{z}, \boldsymbol{x}_1)) + f(d(\boldsymbol{z}, \boldsymbol{x}_2)) + \cdots + f(d(\boldsymbol{z}, \boldsymbol{x}_n))$$

where $\boldsymbol{x}_i$ are the crime locations, f is a decay function and d is a distance.

Probability Distribution Strategies

- Linear:
  - $f(d) = A - Bd$



Probability Distribution Strategies

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

but share the common mathematical heritage:

$$S(\boldsymbol{y}) = \sum_{i=1}^{n} f(d(\boldsymbol{y}, \boldsymbol{x}_i))$$

- In practice, $S(\boldsymbol{y})$ may be evaluated only at discrete values $\boldsymbol{y}_j$ giving us a hit score matrix

$$S_{ij} = \sum_{i=1}^{n} f(d(\boldsymbol{y}_j, \boldsymbol{x}_i))$$

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 $\boldsymbol{z}$ commits a crime at the location $\boldsymbol{x}$ according to a probability density function $P(\boldsymbol{x} \mid \boldsymbol{z})$.
  - 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

$P(\boldsymbol{x} \mid \boldsymbol{z})$.

The Mathematics

- Given crimes located at $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ the *maximum likelihood estimate* for the anchor point $\zeta_{\boldsymbol{mle}}$ is the value of $\boldsymbol{y}$ that maximizes

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

$$= P(\boldsymbol{x}_1 \mid \boldsymbol{y}) P(\boldsymbol{x}_2 \mid \boldsymbol{y}) \cdots P(\boldsymbol{x}_n \mid \boldsymbol{y})$$

or equivalently, the value that maximizes

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

$$= \ln P(\boldsymbol{x}_1 \mid \boldsymbol{y}) + \ln P(\boldsymbol{x}_2 \mid \boldsymbol{y}) + \cdots + \ln P(\boldsymbol{x}_n \mid \boldsymbol{y}).$$

Relation to Spatial Distribution Strategies

- If we make the assumption that offenders choose target locations based only on a distance decay function in normal form, then

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

- The maximum likelihood estimate for the anchor point is the centroid.

Relation to Spatial Distribution Strategies

- If we make the assumption that offenders choose target locations based only on a distance decay function in exponentially decaying form, then

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

- The maximum likelihood estimate for the anchor point is the center of minimum distance.

Relation to Probability Distance Strategies

- What is the log likelihood function?

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

- This is the hit score $S(\boldsymbol{y})$ 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 *a priori* estimates for parameters other than the anchor point.

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

The same process that determines the best choice of $\boldsymbol{z}$ also determines the best choice of σ.

Better Models

- We have recaptured the many results from existing techniques by choosing $P(\boldsymbol{x} \mid \boldsymbol{z})$ appropriately.
- These choices of $P(\boldsymbol{x} \mid \boldsymbol{z})$ are not very realistic.
  - Space is homogeneous and potential crime locations are equi-distributed.
- We want to incorporate the effects of the local geography.

Better Models

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

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

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.

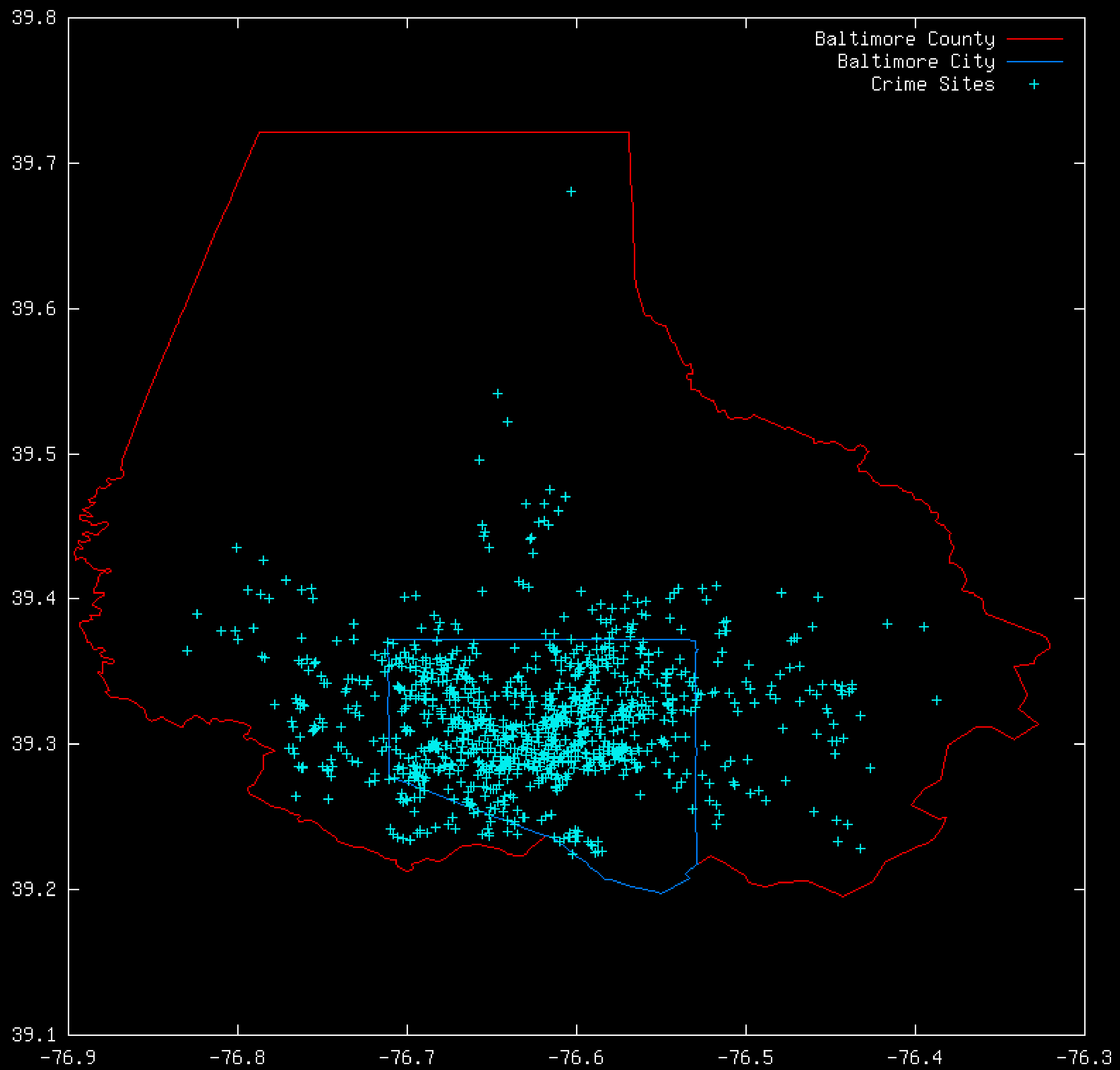
Geography

- 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 density of potential targets.

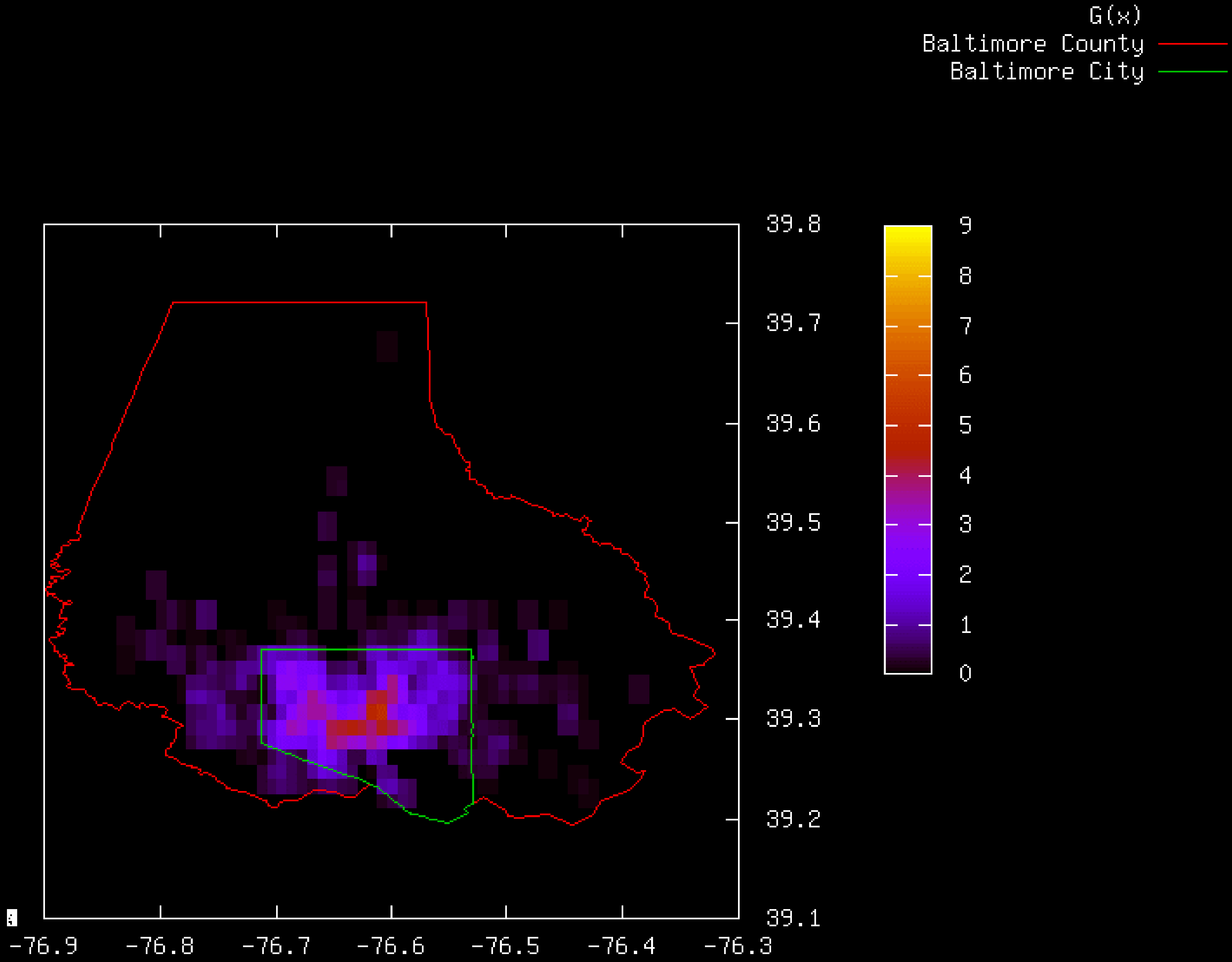
Geography

- 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 $G(x)$.
- $G(x)$ could then be calculated in the same fashion as hot spots; *e.g.* by kernel density parameter estimation.

$$G(x) = \sum_{i=1}^{N} K(x - y_i)$$



Baltimore County
Baltimore City
Crime Sites
39.8
39.7
39.6
39.5
39.4
39.3
39.2
39.1
-76.9
-76.8
-76.7
-76.6
-76.5
-76.4
-76.3


G(x)
Baltimore County
Baltimore City
39.8
39.7
39.6
39.5
39.4
39.3
39.2
39.1
-76.9
-76.8
-76.7
-76.6
-76.5
-76.4
-76.3
9
8
7
6
5
4
3
2
1
0

Geography

- The target density function $G(x)$ 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

$$G(x) = 0 \quad \text{for all } x \notin J$$

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

Distance Decay

- The mathematical method does not depend upon any particular choice of the distance decay function, or a particular distance measure.
- We begin with the simple choice

$$D(d(\boldsymbol{x}, \boldsymbol{z})) = \exp(-\sigma|x - z|)$$

where the parameter σ is determined by the crime series data along with the anchor point z.

Normalization

- The expression

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

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

$$N(z) = \frac{1}{\iint_J G(y)\, D(d(y, z))\, dy^{(1)} dy^{(2)}}$$

Mathematics

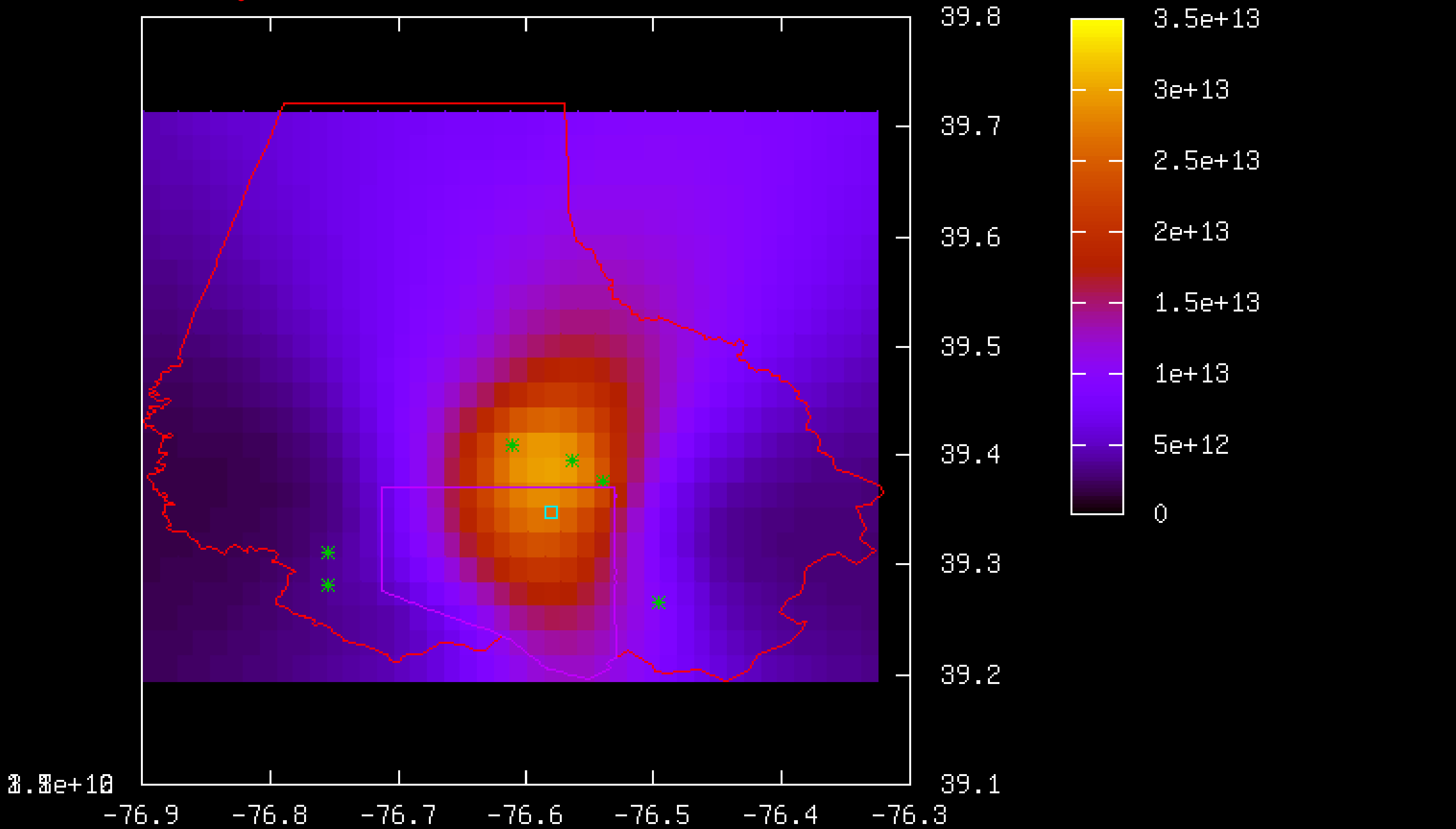
- We are then left with the mathematical problem of finding the maximum value of the likelihood function

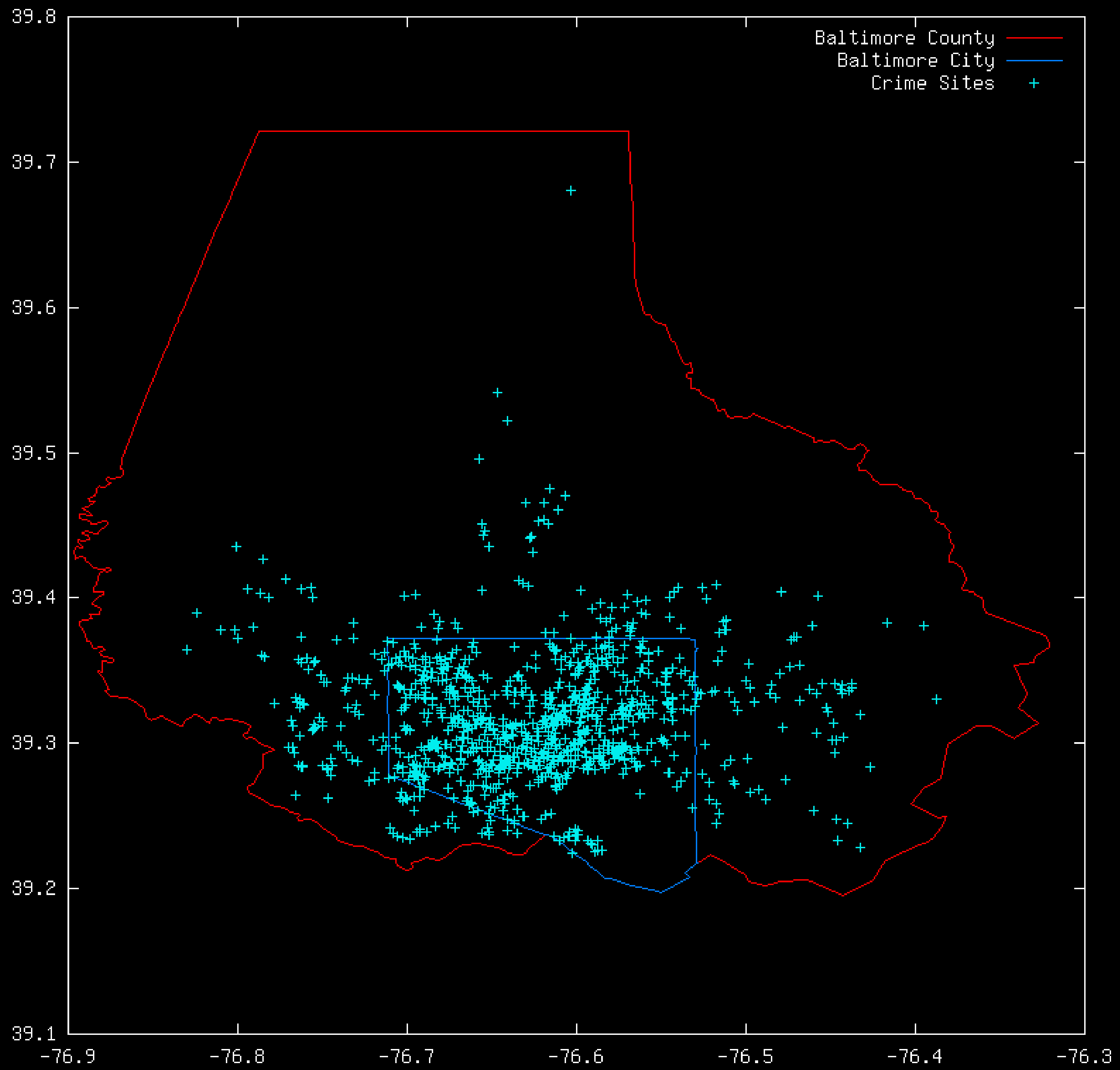
$$L(y) = \frac{\prod_{i=1}^{n} D(d(x_i, y))G(x_i)}{\left[\iint_J D(d(\xi, y))G(\xi)\,d\xi^{(1)}\,d\xi^{(2)}\right]^n}$$

Implementation

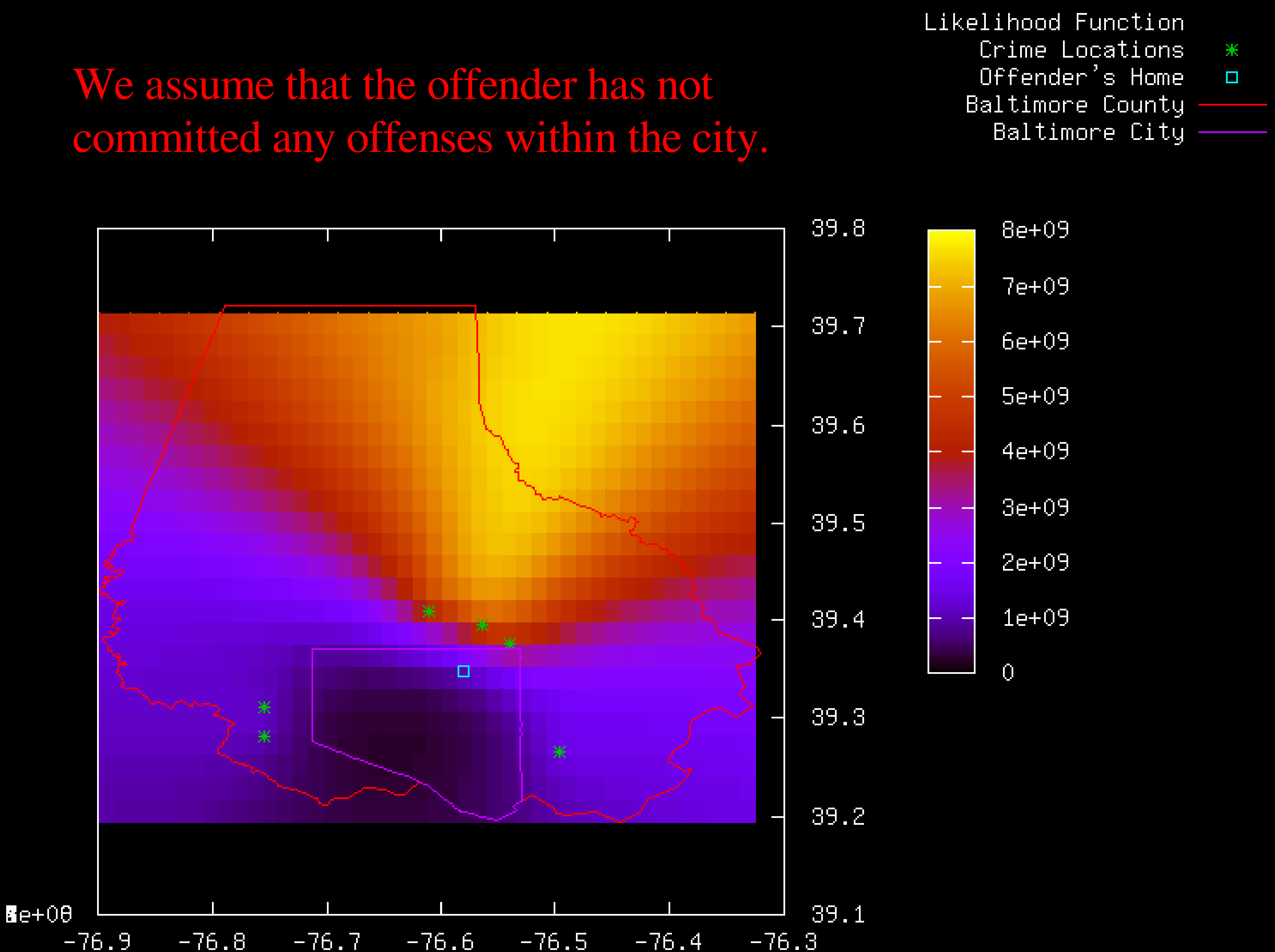
- We have implemented this algorithm in software.
  - Integration was performed using a seven-point fifth-order Gaussian method.
  - Optimization was performed using a cyclic coordinate technique with a Hooke and Jeeves accelerator.
  - Running time with ~650 boundary vertices and ~1000 historical crimes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

We assume that we do not know if the offender has committed any offenses within the city.




Baltimore County
Baltimore City
Crime Sites
39.8
39.7
39.6
39.5
39.4
39.3
39.2
39.1
-76.9
-76.8
-76.7
-76.6
-76.5
-76.4
-76.3

We assume that the offender has not committed any offenses within the city.



Results

- The software is currently pre-release quality, and is undergoing testing and debugging.
  - I would be glad to share it with interested parties.
- When this completes, we will begin testing it against real data.
  - Volunteers are welcome- please help.

Future Work

- We have met many of our goals for a geographic profiling algorithm, but two issues remain:

  3b. It should take into account local geographic features that affect the selection of an anchor point.

  5. It should return a prioritized search area.

- Work on these areas continues using Bayesian techniques.

Strengths of this framework

- The framework is extensible.
  - Vastly different situations can be modelled by making different choices for the form and structure of $P(\mathbf{x} \mid \mathbf{z})$.
    - *e.g.* angular dependence, barriers.
- The framework is otherwise agnostic about the crime series; all of the relevant information must be encoded in $P(\mathbf{x} \mid \mathbf{z})$.

Strengths

- This framework is mathematically rigorous.
  - There are mathematical and criminological meanings to the maximum likelihood estimate ζ_{mle}.

Weaknesses of this Framework

- GIGO
  - The method is only as accurate as the accuracy of the choice of $P(\boldsymbol{x} \mid \boldsymbol{z})$.
- It is unclear what is the right choice for $P(\boldsymbol{x} \mid \boldsymbol{z})$.
  - Even with the simplifying assumption that

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

    this is difficult.

Weaknesses

- There is no simple closed mathematical form for ζ_{mle}.
  - Relatively complex techniques are required to estimate ζ_{mle} even for simple choices of $P(\boldsymbol{x} \mid \boldsymbol{z})$.
- The error analysis for maximum likelihood estimators is delicate when the number of data points is small.

Weaknesses

- The framework (so far) assumes that crime sites are independent, identically distributed random variables.
  - This is probably false in general!
- The mathematics in the framework can (probably) be adjusted to take this issue into account.

Questions?

Contact information:

Dr. Mike O'Leary

Director, Applied Mathematics Laboratory

Towson University

Towson, MD 21252

410-704-7457

moleary@towson.edu