The Mathematics of Geographic Profiling

Dr. Mike O'Leary

Towson University

Supported by the NIJ through grant 2005-IJ-CX-K036 and 2007-DE-BX-K005

Project Participants

- Towson University Applied Mathematics Laboratory
  - Undergraduate research projects in applied mathematics.
  - Founded in 1980
- National Institute of Justice
- Special thanks to Iara Infosino (CAA), 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
  - Gregory Emerson
  - Laurel Mount
  - Ruozhen Yao
  - Melissa Zimmerman
- 2006-2007 Students:
  - Chris Castillo
  - Adam Fojtik
  - Jonathan Vanderkolk
  - 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.

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.

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

Spatial Distribution Strategies

- Centroid

$$\hat{z}_{centroid} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$$

- Center of minimum distance; $\hat{z}_{cmd}$ minimizes

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

  - We can use different choices for the metric- Euclidean, Manhattan, Travel distance, Travel time.

Spatial Distribution Strategies

- Circle Method (Canter & Larkin, 1993):
  - Anchor point contained in the circle whose diameter are the two crimes that are farthest apart.
  - Offenders who live within the circle are called *marauders*; those who love outside are called *commuters*.

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

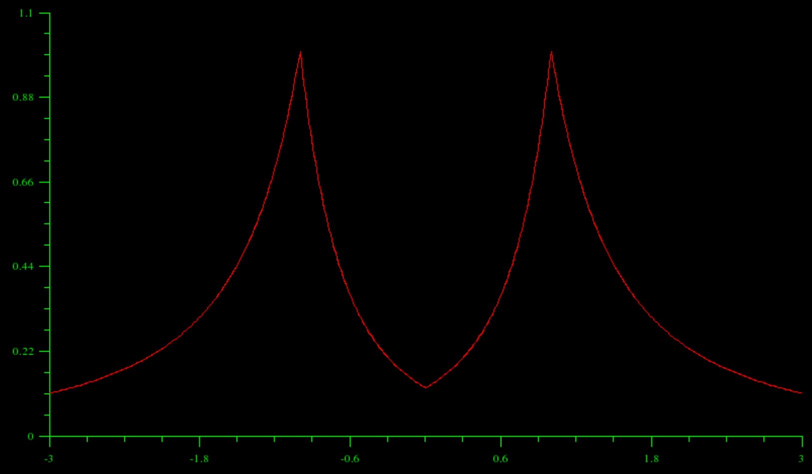
$$S(\boldsymbol{y}) = \sum_{i=1}^{n} f(d(\boldsymbol{y}, \boldsymbol{x}_i))$$
$$= f(d(\boldsymbol{z}, \boldsymbol{x}_1)) + f(d(\boldsymbol{z}, \boldsymbol{x}_2)) + \cdots + f(d(\boldsymbol{z}, \boldsymbol{x}_n))$$

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

Rossmo (Rigel)

- Manhattan distance metric.
- Decay function

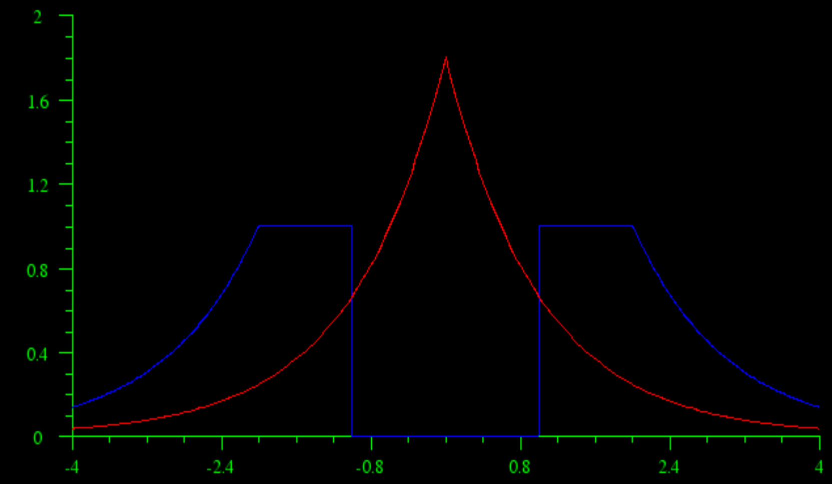
$$f(d)=\begin{cases}\dfrac{k}{d^h} & \text{if } d > B\\[2ex] \dfrac{k\,B^{g-h}}{(2B-d)^g} & \text{if } d \leqslant B\end{cases}$$



- The constants k, g, h and B are empirically defined.

Canter, Coffey, Huntley & Missen (Dragnet)

- Euclidean distance
- Decay functions
  - $f(d) = A e^{-\beta d}$
  - $$f(d) = \begin{cases} 0 & \text{if } d < A, \\ 1 & \text{if } A \le d < B, \\ Ce^{-\beta d} & \text{if } d \ge B. \end{cases}$$

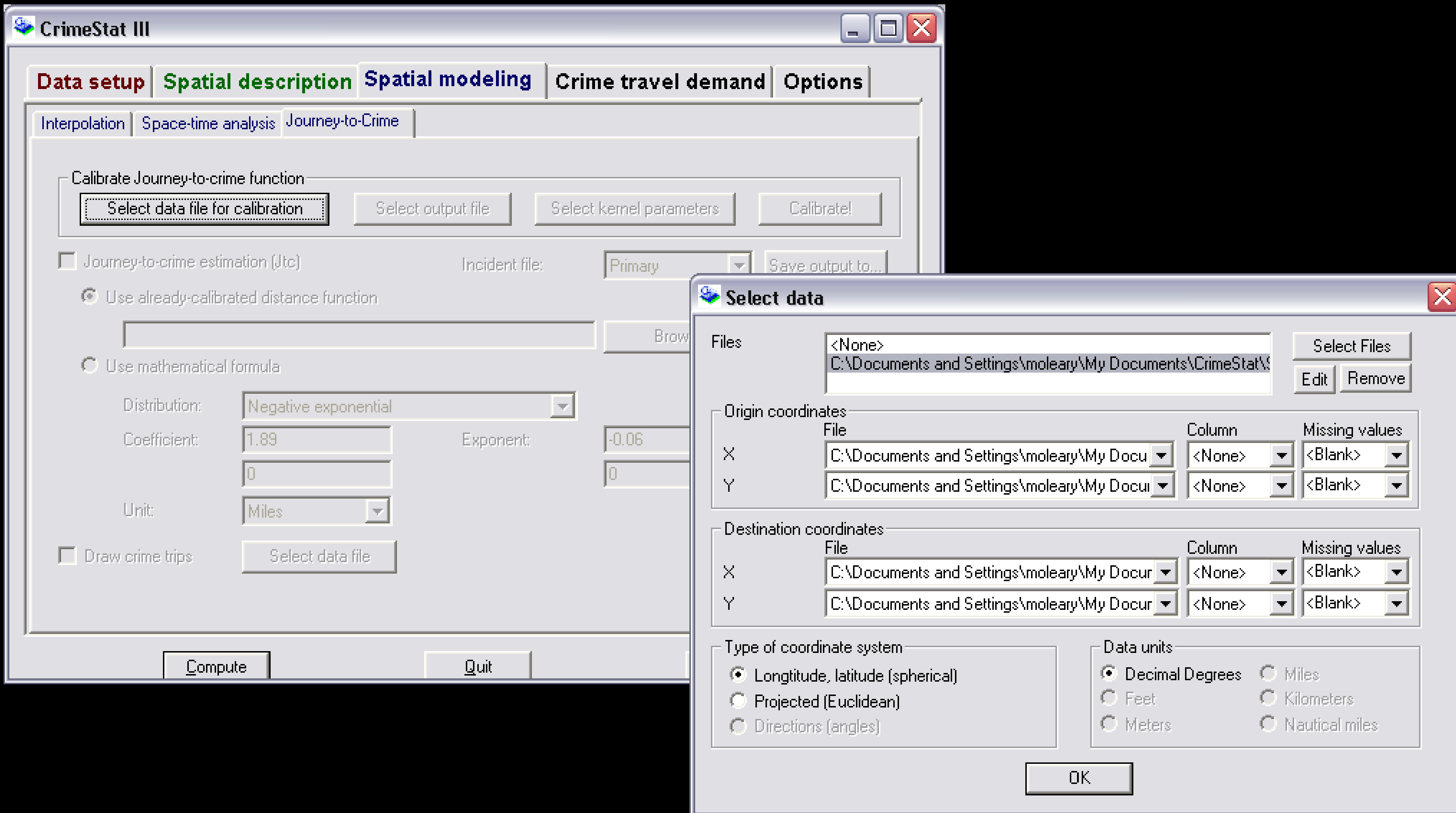


- Calibrated against homicide data

Levine (CrimeStat)

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

CrimeStat


CrimeStat III
Data setup
Spatial description
Spatial modeling
Crime travel demand
Options
Interpolation
Space-time analysis
Journey-to-Crime
Calibrate Journey-to-crime function
Select data file for calibration
Select output file
Select kernel parameters
Calibrate!
Journey-to-crime estimation (Jtc)
Incident file:
Primary
Save output to...
Use already-calibrated distance function
Use mathematical formula
Distribution:
Negative exponential
Coefficient:
1.89
0
Exponent:
-0.06
0
Unit:
Miles
Draw crime trips
Select data file
Compute
Quit
Select data
Files
<None>
C:\Documents and Settings\moleary\My Documents\CrimeStat\
Select Files
Edit
Remove
Origin coordinates
File
Column
Missing values
X
Y
C:\Documents and Settings\moleary\My Docu
<None>
<Blank>
Destination coordinates
Type of coordinate system
Longitude, latitude (spherical)
Projected (Euclidean)
Directions (angles)
Data units
Decimal Degrees
Miles
Feet
Kilometers
Meters
Nautical miles
OK

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

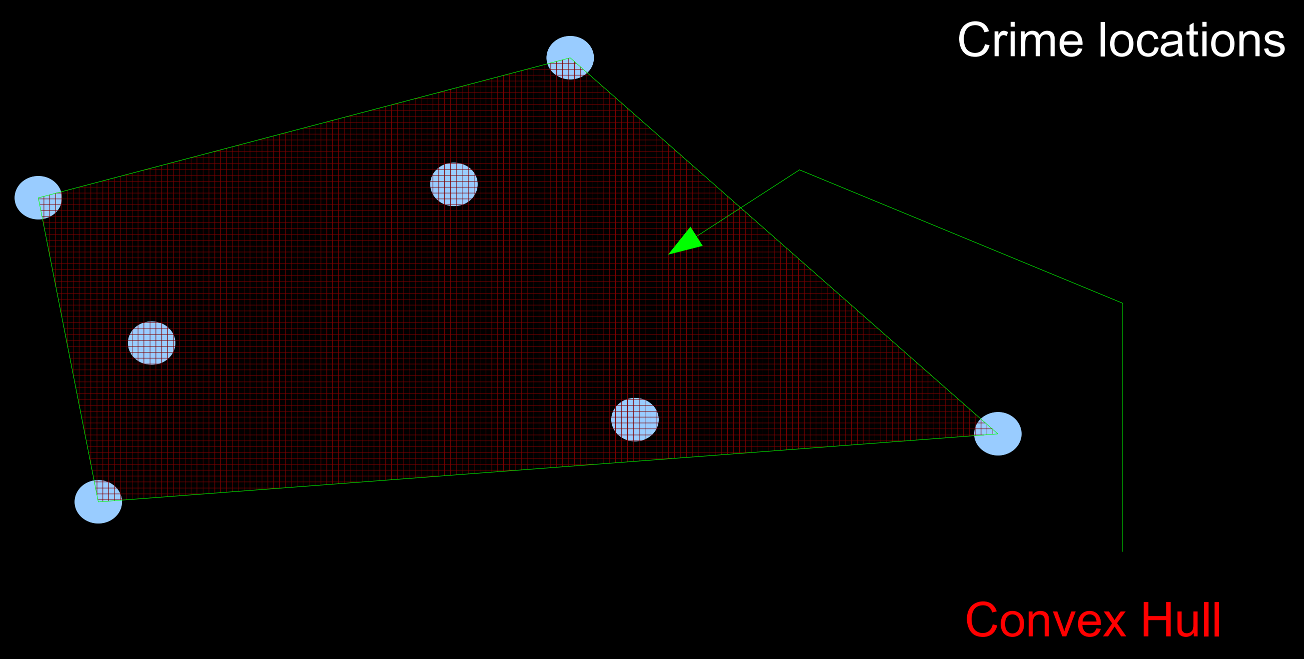
$$S_{ij} = \sum_{i=1}^{n} f(d(\boldsymbol{y}_j, \boldsymbol{x}_i))$$

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?
- How do you choose one method over another?

Shortcomings

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



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*

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.

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.

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 $\hat{\boldsymbol{z}}_{mle}$ is the value of $\boldsymbol{y}$ that maximizes

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

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

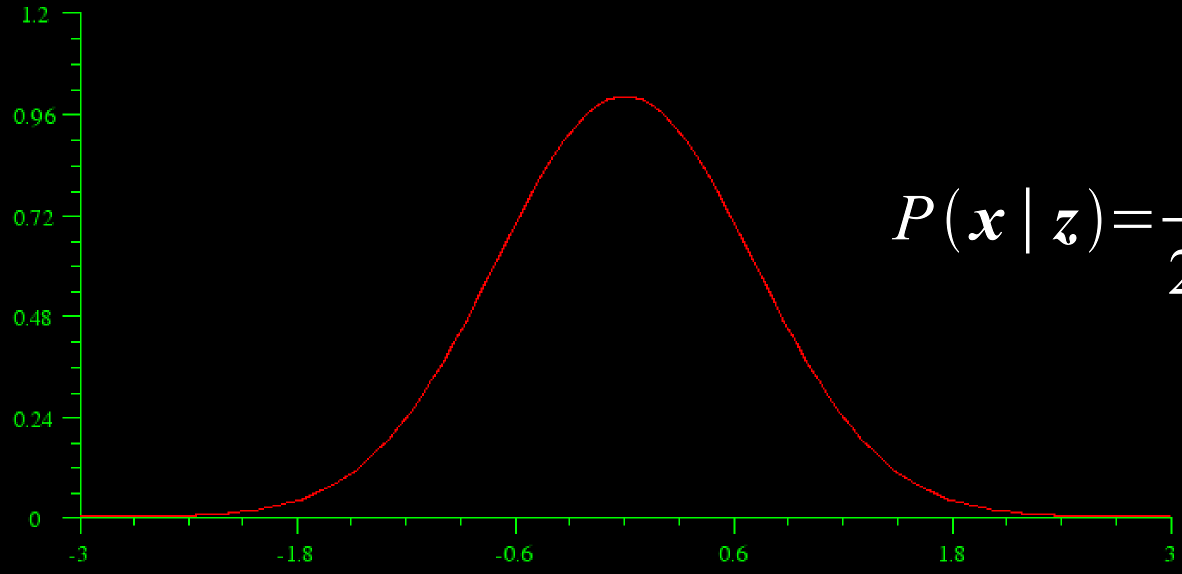
or equivalently, the value that maximizes

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

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

Relation to Spatial Distribution Strategies

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

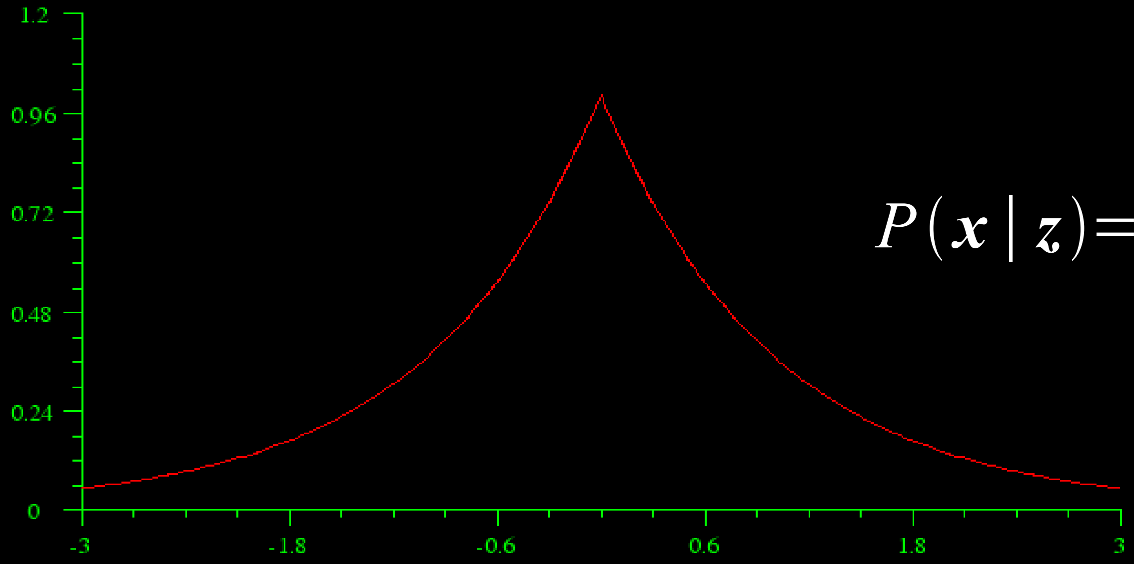


$$P(\mathbf{x} \mid \mathbf{z}) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{|\mathbf{x}-\mathbf{z}|^2}{2\sigma^2} \right]$$

- 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:



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

- Then the maximum likelihood estimate is the center of minimum distance.

Relation to Probability Distance Strategies

- What is the log likelihood function?

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

- 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 σ also determines the best choice of $\boldsymbol{z}$.

Better Models

- We have recaptured the results of 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 crimes are equi-distributed.
  - Space is infinite.
  - Decay functions were chosen arbitrarily.

Better Models

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

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

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.

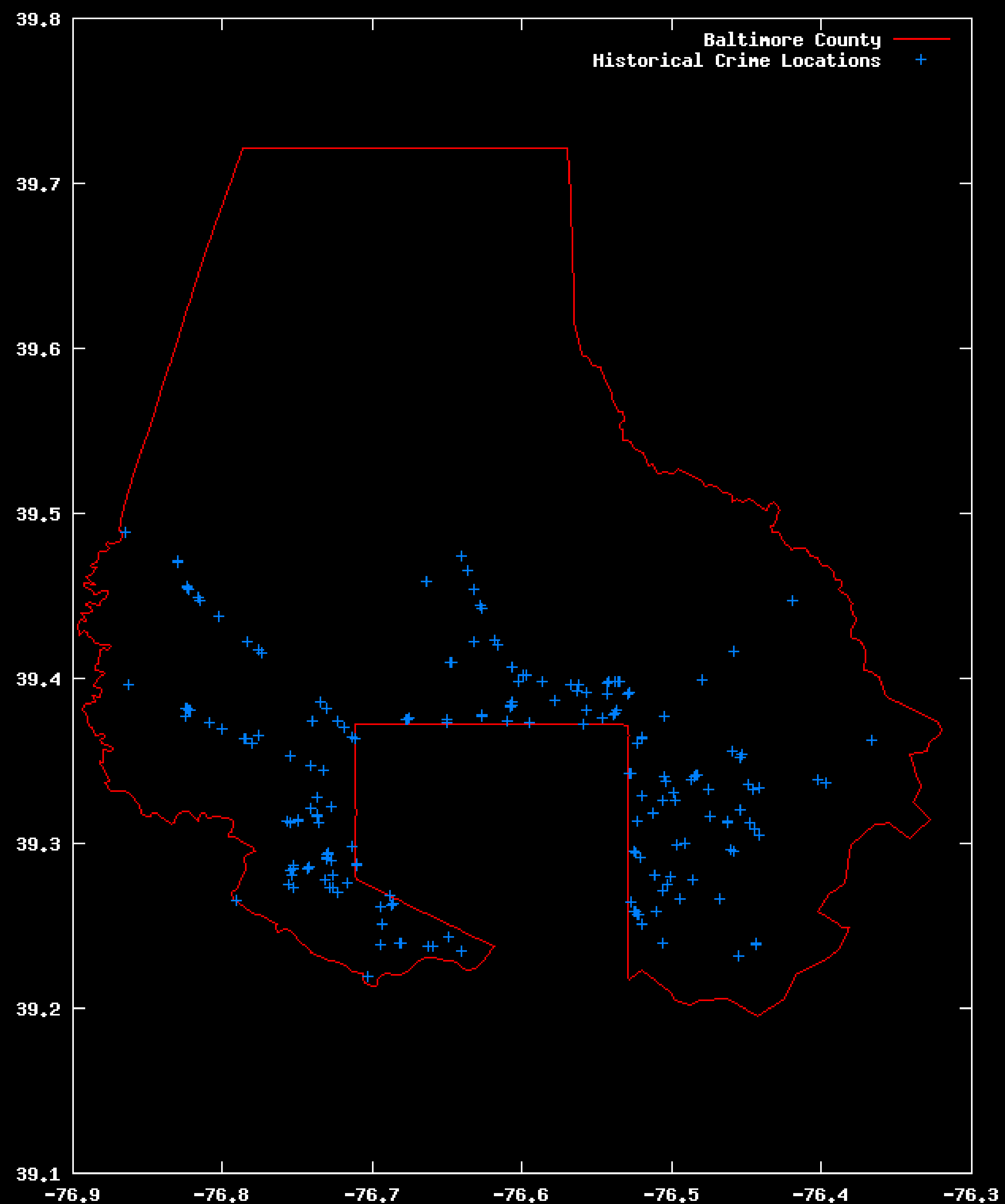
Geography

- 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.

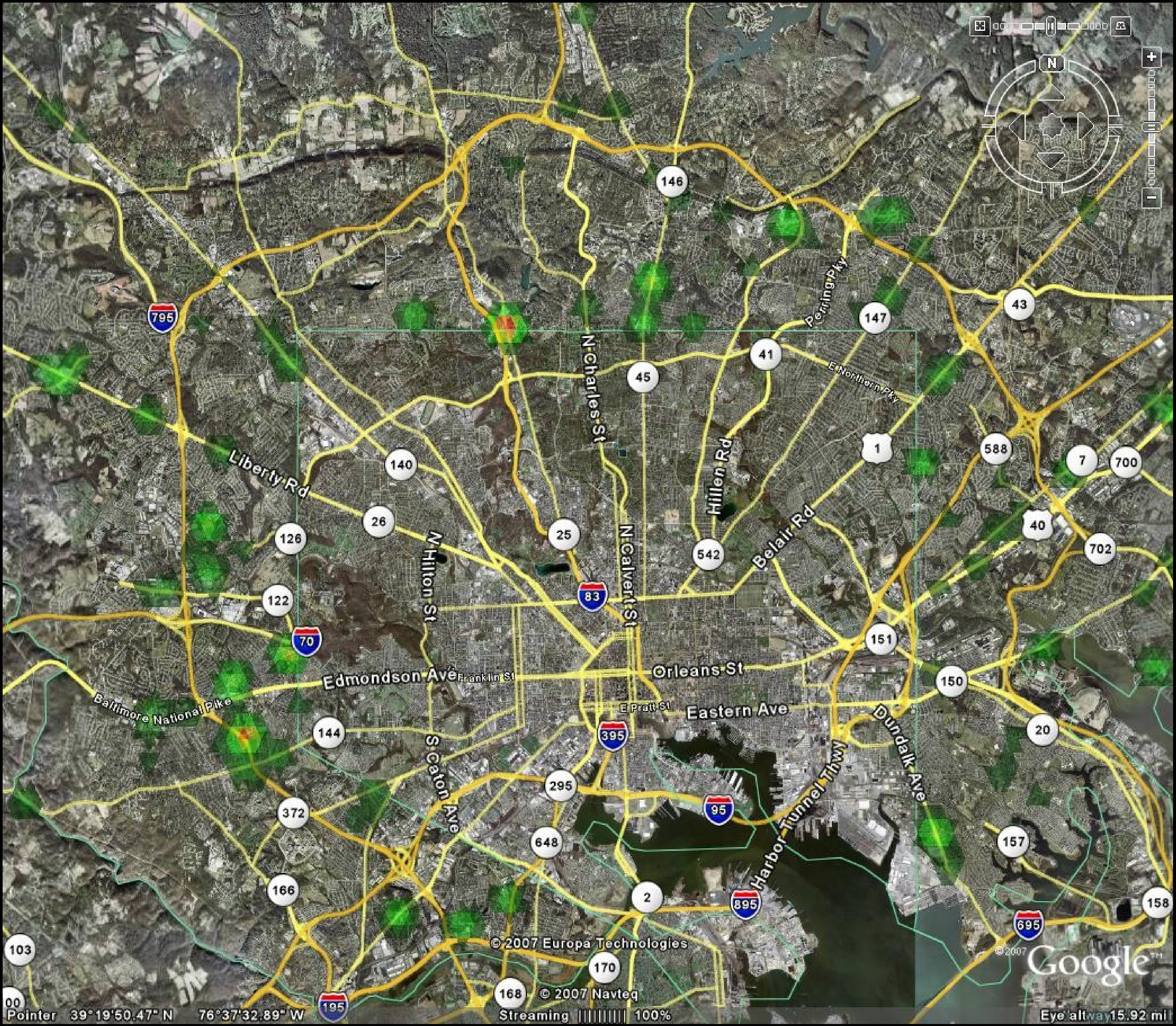
Geography

- 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 by kernel density parameter estimation.

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



Baltimore County
Historical Crime Locations
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



N Charles St
Perring Pky
E Northern Pky
Hillen Rd
Liberty Rd
N Hilton St
N Calvert St
Belair Rd
Edmondson Ave
Franklin St
Orleans St
Baltimore National Pike
E Pratt St
Eastern Ave
Dundalk Ave
S Caton Ave
Harbor Tunnel Thwy
© 2007 Europa Technologies
© 2007 Navteq
Google
Pointer 39°19'50.47" N 76°37'32.89" W
Streaming 100%
Eye alt 15.92 mi

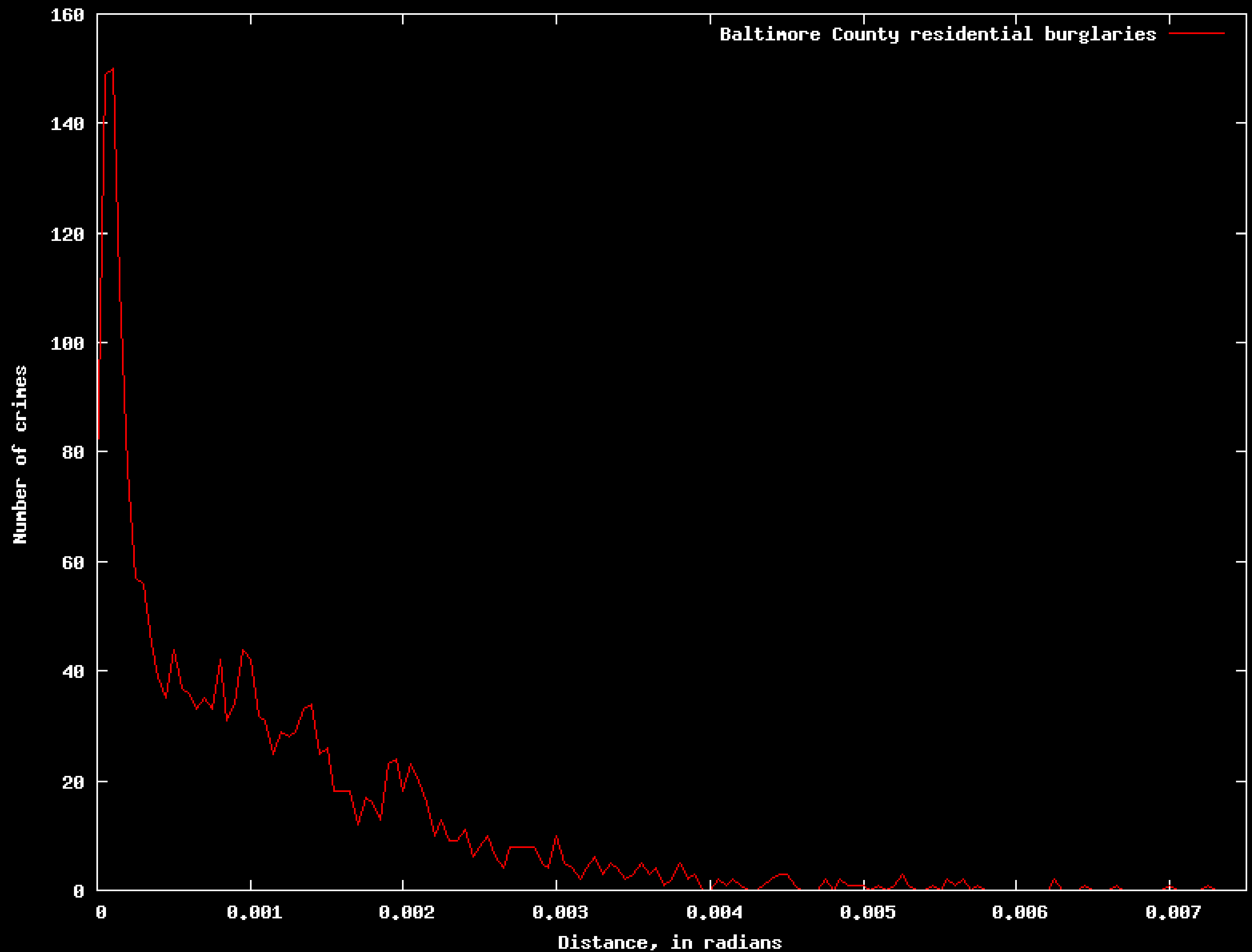
Geography

- The target attractiveness 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

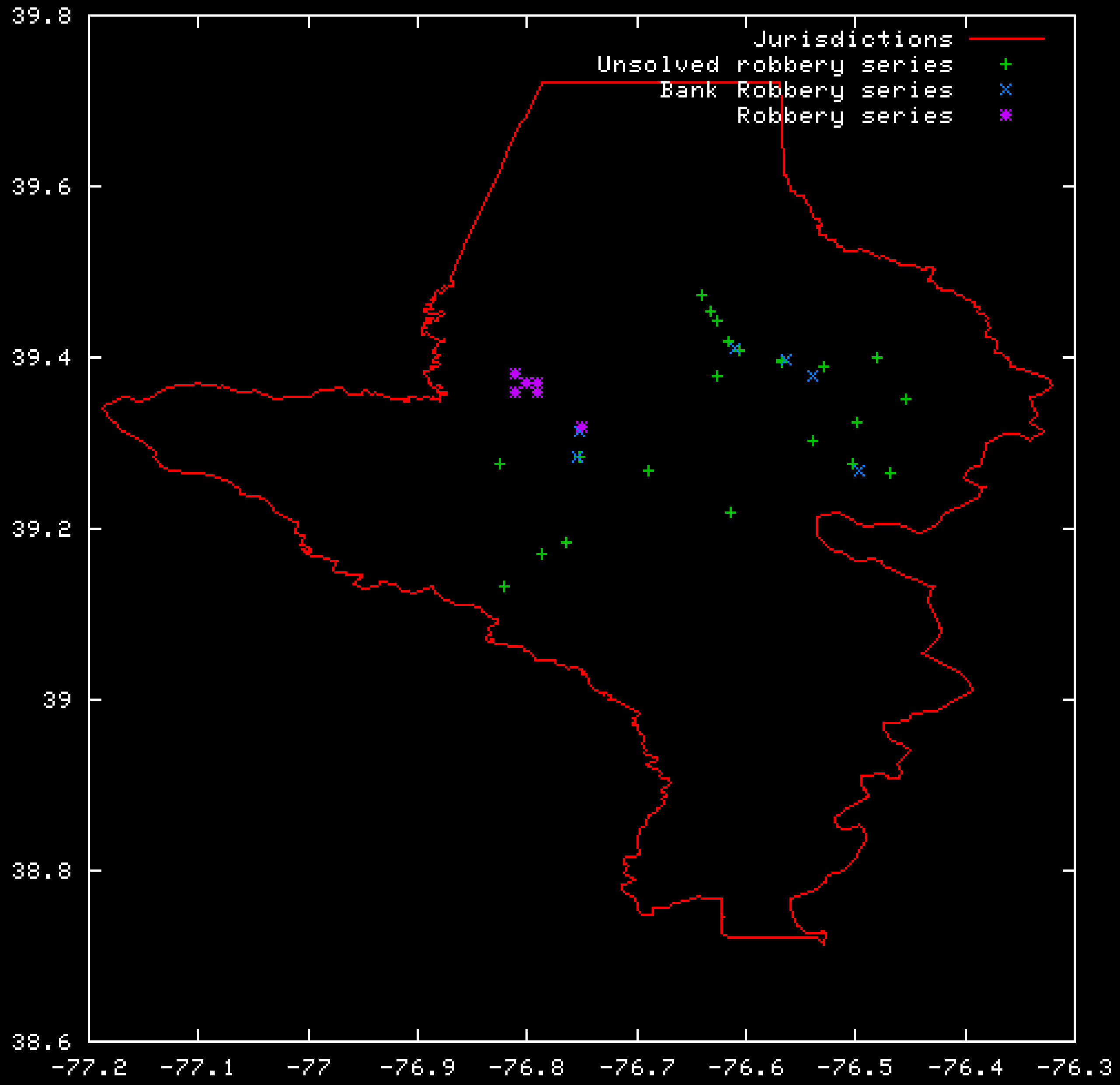
$$G(x) = 0 \quad \text{for all } x \notin J$$

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

Distance Decay



Distance Decay



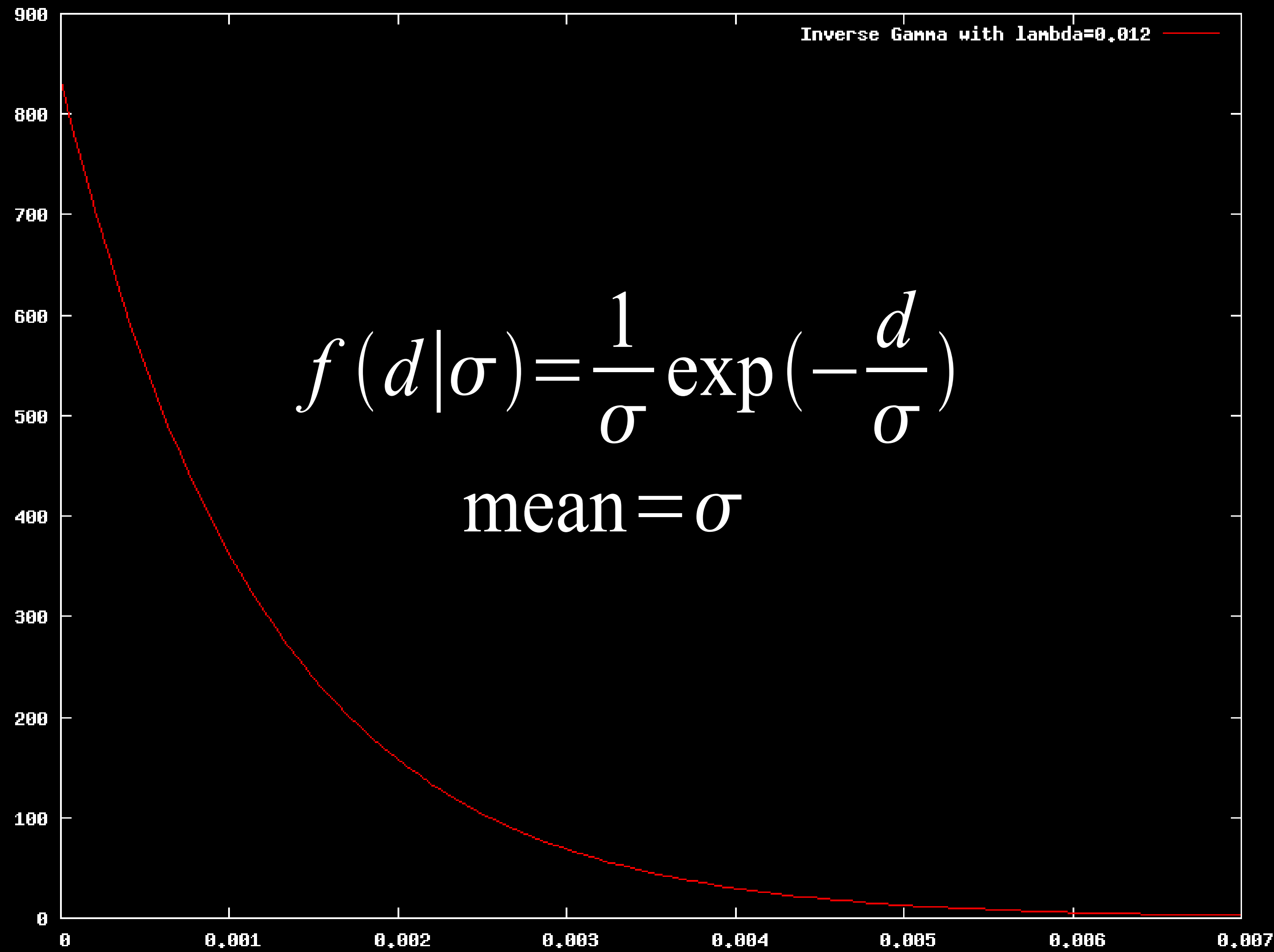
Distance Decay

- Suppose that each offender has a decay function $f(d|\sigma)$ where $\sigma \in \Sigma$ varies among offenders according to the distribution $\pi(\sigma)$.
- Then if we look at the decay function for all offenders, we obtain the aggregate distribution

$$F(d) = \int_{\Sigma} f(d|\sigma) \cdot \pi(\sigma)\, d\sigma$$

Distance Decay

- Suppose that the distance decay behavior of an individual offender is exponentially decaying, so that

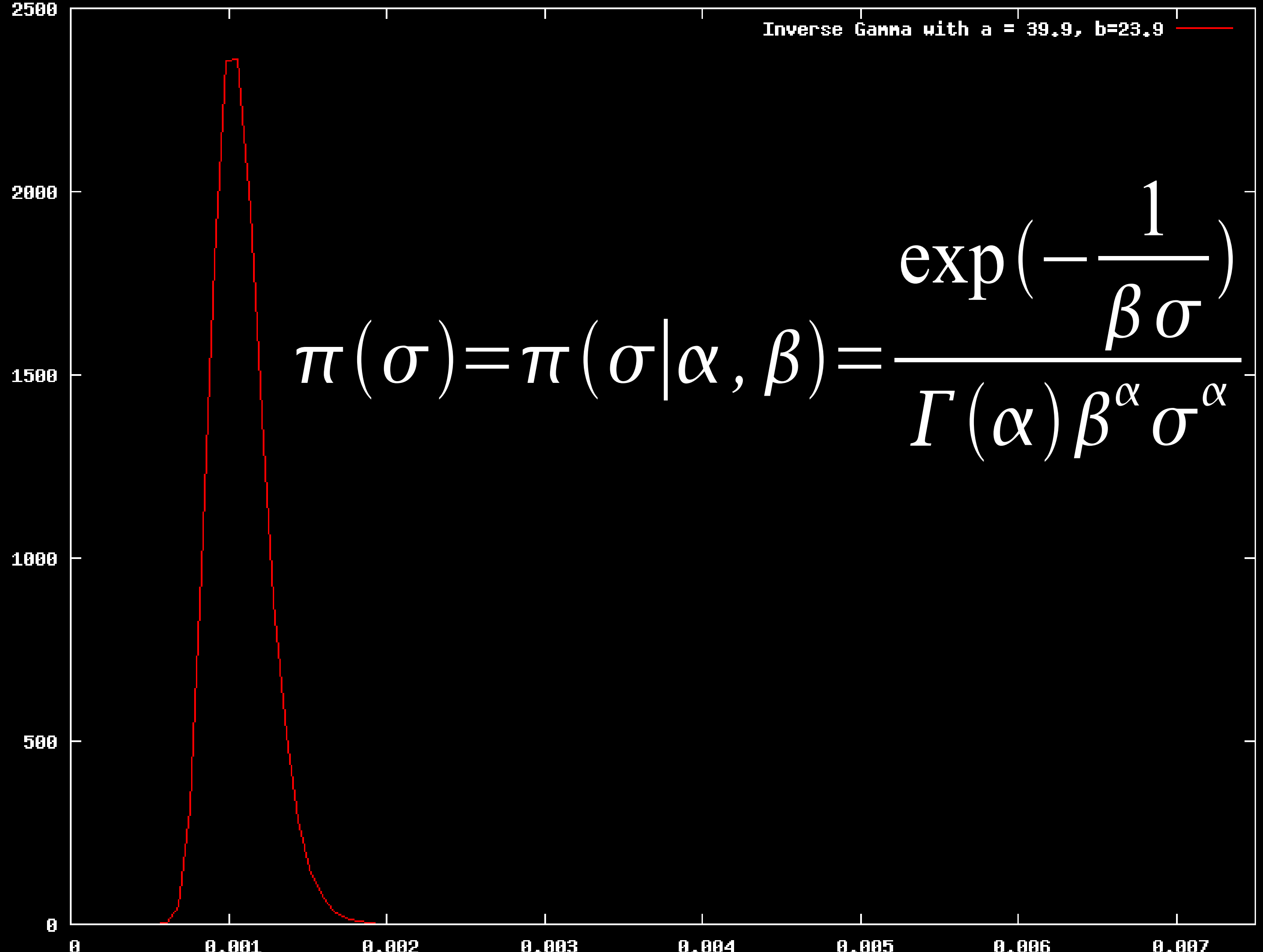


$$f(d|\sigma) = \frac{1}{\sigma}\exp\left(-\frac{d}{\sigma}\right)$$

$$\text{mean} = \sigma$$

Distance Decay

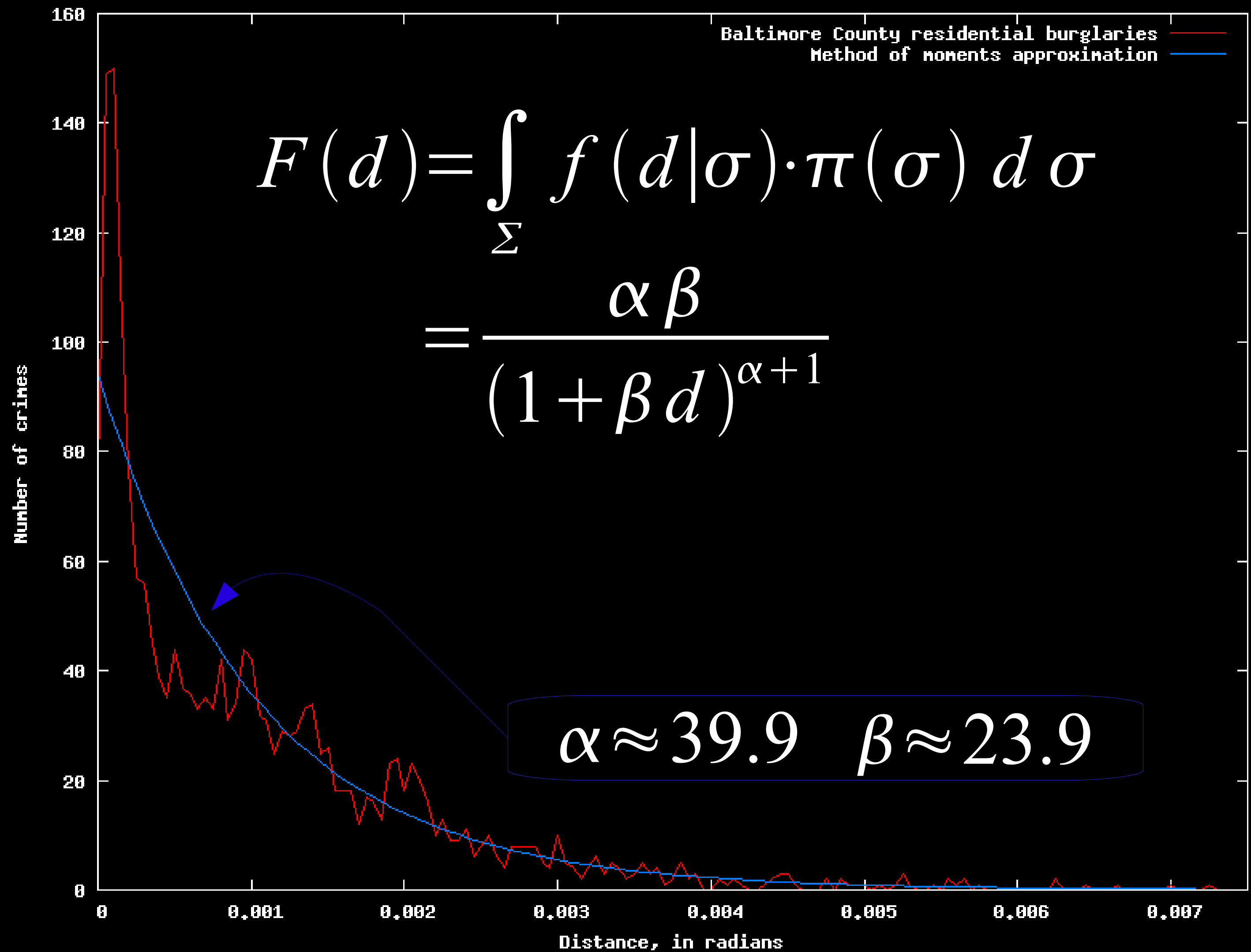
- Suppose that the mean distance an offender travels follows an inverse gamma distribution



$$\pi(\sigma)=\pi(\sigma|\alpha,\beta)=\frac{\exp\left(-\frac{1}{\beta\sigma}\right)}{\Gamma(\alpha)\beta^{\alpha}\sigma^{\alpha}}$$

Distance Decay

- Then we can equate the first and second moments of the mixture distribution with the empirical data.



$$F(d) = \int_{\Sigma} f(d|\sigma) \cdot \pi(\sigma)\, d\sigma$$

$$= \frac{\alpha \beta}{(1+\beta d)^{\alpha+1}}$$

$$\alpha \approx 39.9 \quad \beta \approx 23.9$$

Distance Decay

- This gives us a statistical method to estimate the distribution $\pi(\sigma)$.
  - Note that information about $\pi(\sigma)$ is equivalent to prior information about the offender before the characteristics of the crime series are considered.

Distance Decay

- We can also try to estimate $\pi(\sigma)$ by solving for $\pi(\sigma)$ in

$$F(d)=\int_0^\infty \frac{1}{\sigma} e^{-d/\sigma} \pi(\sigma)\, d\sigma$$

  - If $\pi(\sigma)$ is bounded, then $F(d)$ is differentiable as

$$|F^{(k)}(d)| \leqslant \int_0^\infty \frac{1}{\sigma^{k+1}} e^{-d/\sigma} \|\pi\|_\infty \, d\sigma \leqslant \frac{\Gamma(k)\|\pi\|_\infty}{d^k}$$

    so the map $\pi \rightarrow F$ is regularizing, (in fact, it is essentially a Laplace transform) so the problem of determining $\pi(\sigma)$ from $F(d)$ is unstable.

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,

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

Maximum Likelihood Estimation

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

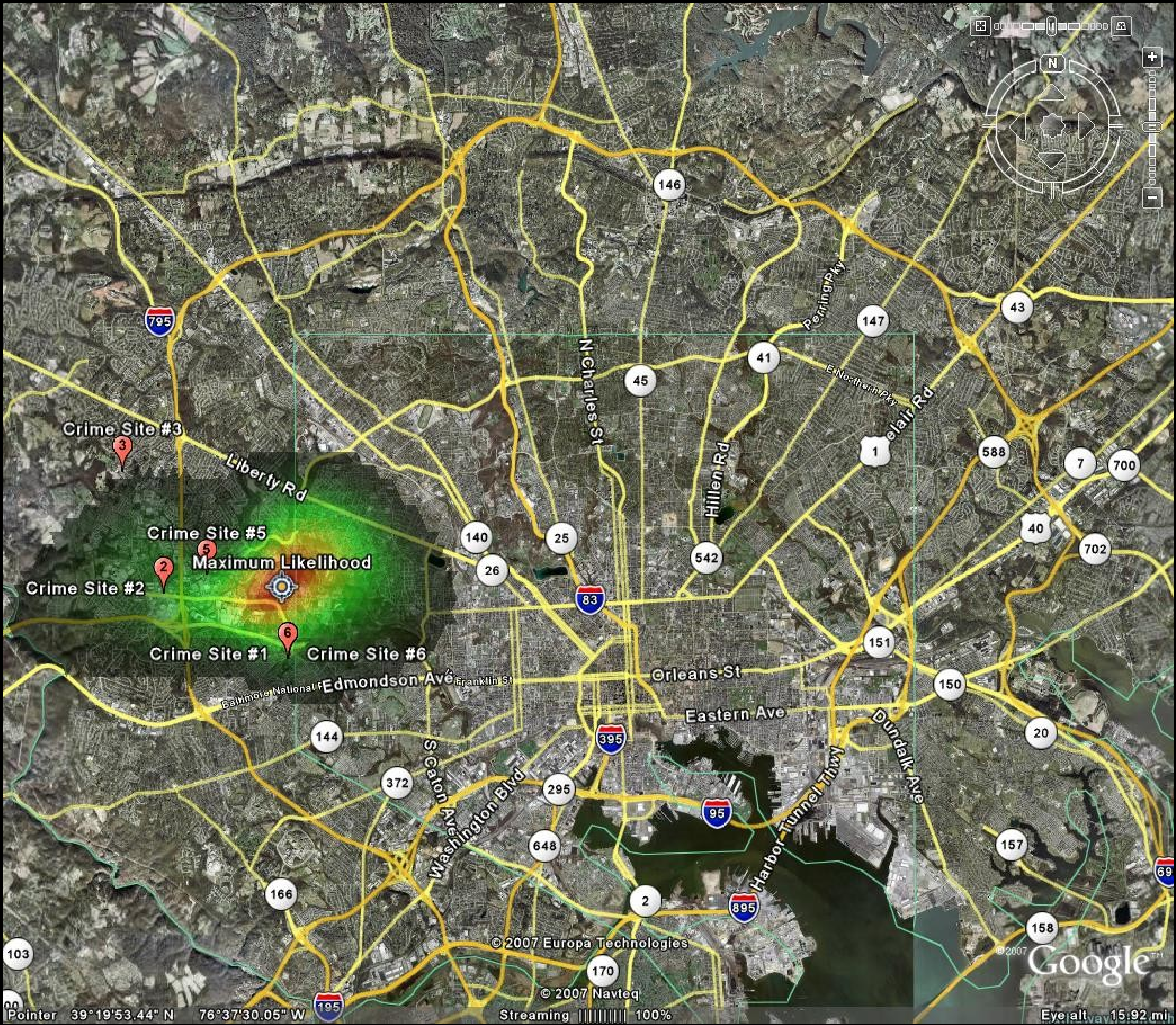
$$L(y)=\frac{\prod_{i=1}^{n} D(d(x_i, y))G(x_i)}{\left[\iint_J D(d(\xi, y))G(\xi)\,d\xi^{(1)}\,d\xi^{(2)}\right]^n}$$

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.

Command Prompt

```
C:\Documents and Settings\moleary\Desktop\v 0.12 devel\Profiler\release>Profiler
.exe

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
-----------------------------------------------
 i      x          y       sigma    Likelihood
-----------------------------------------------
 0 -76.731598 39.311223 44.217570 1.892049e+023
 1 -76.716733 39.312793 71.545719 3.909148e+023
 2 -76.716733 39.314912 73.128008 4.040361e+023
 3 -76.715180 39.314912 72.770779 4.052184e+023
 4 -76.715180 39.314912 72.770779 4.052184e+023

------------------------------------------------------
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>_
```


Crime Site #3
Liberty Rd
Crime Site #5
Maximum Likelihood
Crime Site #2
Crime Site #1
Crime Site #6
Edmondson Ave
Franklin St
Baltimore National Pike
Orleans St
Eastern Ave
N Charles St
Hillen Rd
Perring Pky
E Northern Pky
Belair Rd
S Caton Ave
Washington Blvd
Harbor Tunnel Thwy
Dundalk Ave
© 2007 Europa Technologies
© 2007 Navteq
Google
Pointer 39°19'53.44" N 76°37'30.05" W
Streaming 100%
Eye alt 15.92 mi

Likelihood Functions

- The estimate for the maximum likelihood is mathematically rigorous.
- The contour surface shows the likelihood function for the optimal choice of σ.
  - This gives a probability surface for the offender's anchor point *only* if
    - the estimate for sigma is correct, and
    - all anchor points are equally likely.

Bayesian Methods

- Police agencies would prefer a search area to the point estimate $\hat{z}_{\text{mle}}$.
- We take a Bayesian approach.
- If we have one crime

$$P(z,\sigma|x)=\frac{P(x,z,\sigma)}{P(x)}$$

$$=\frac{P(x|z,\sigma)H(z)\pi(\sigma)}{\int_{\varsigma\in\Sigma}\int_{\zeta\in J}P(x|z,\varsigma)H(\zeta)\pi(\varsigma)\,d\zeta\,d\varsigma}$$

where $H(z)$ is the prior distribution for anchor points.

Bayesian Methods

- 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 λ.
  - 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

$$P(z,\sigma|x_1,x_2,\cdots,x_n) = \frac{\prod_{i=1}^{n} P(x_i|z,\sigma)H(z)\pi(\sigma)}{\int\limits_{\varsigma\in\Sigma}\int\limits_{\zeta\in J}\prod_{i=1}^{n} P(x_i|\zeta,\varsigma)H(\zeta)\pi(\sigma)\,d\zeta\,d\varsigma}$$

Bayesian Methods

- Since the relevant distribution is the marginal distribution for z, we easily see that

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

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

$$\propto \int_{\sigma\in\Sigma} \prod_{i=1}^{n} D(d(x_i, z), \sigma) G(x_i)$$

$$\cdot N(z) H(z) \pi(\sigma)\, d\sigma$$

Strengths of this Framework

- 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 $P(\boldsymbol{x} \mid \boldsymbol{z})$.
- It is unclear what the right choice is 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

- 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