The Mathematics of Geographic Profiling

Towson University
Applied Mathematics Laboratory

Dr. Mike O'Leary

Crime Hot Spots: Behavioral, Computational and Mathematical Models
Institute for Pure and Applied Mathematics
January 29 - February 2, 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) and Andrew Engel (SAS)

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:**

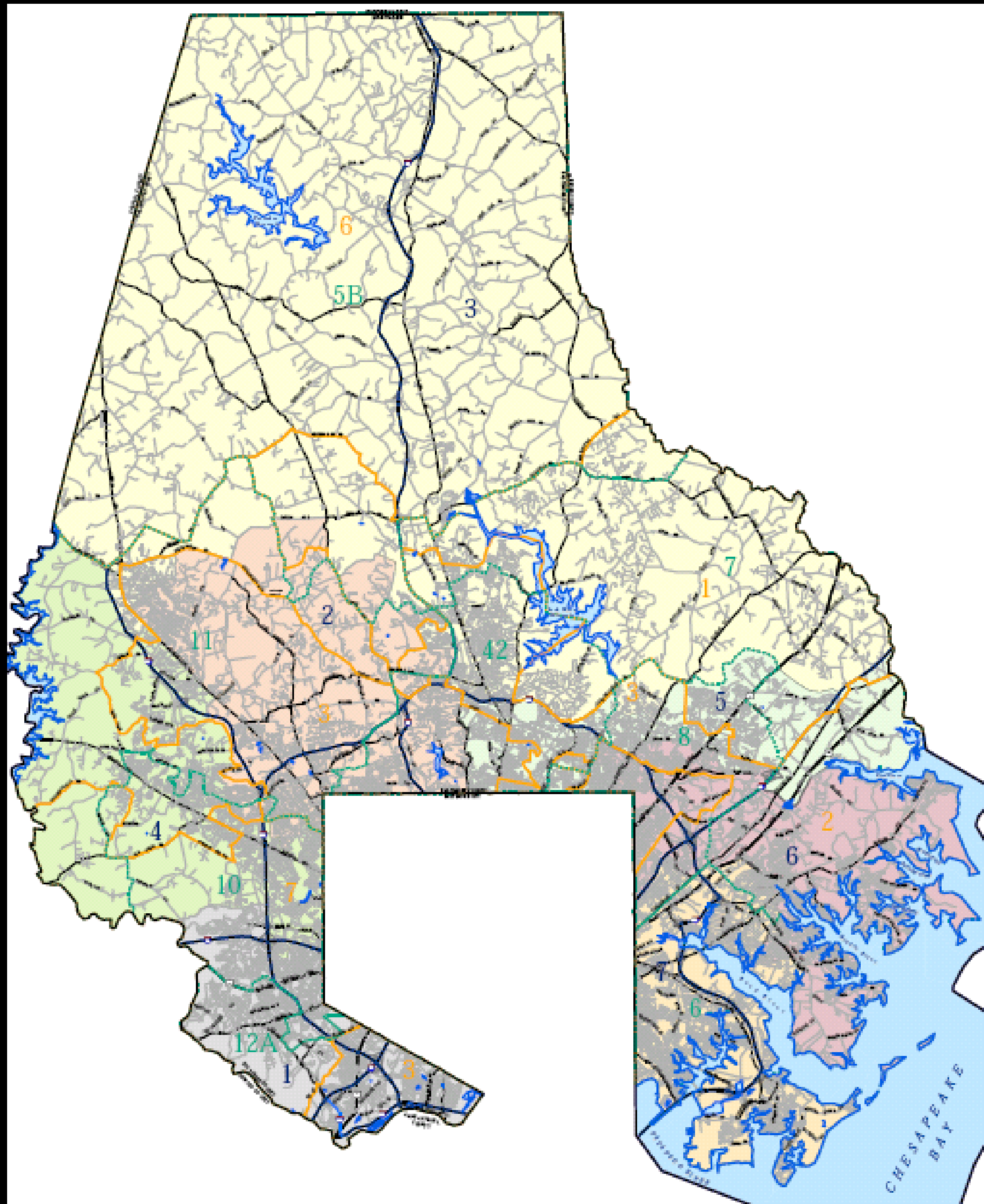
  Given a series of linked crimes committed by the same offender, can we make predictions about the anchor point of the offender?

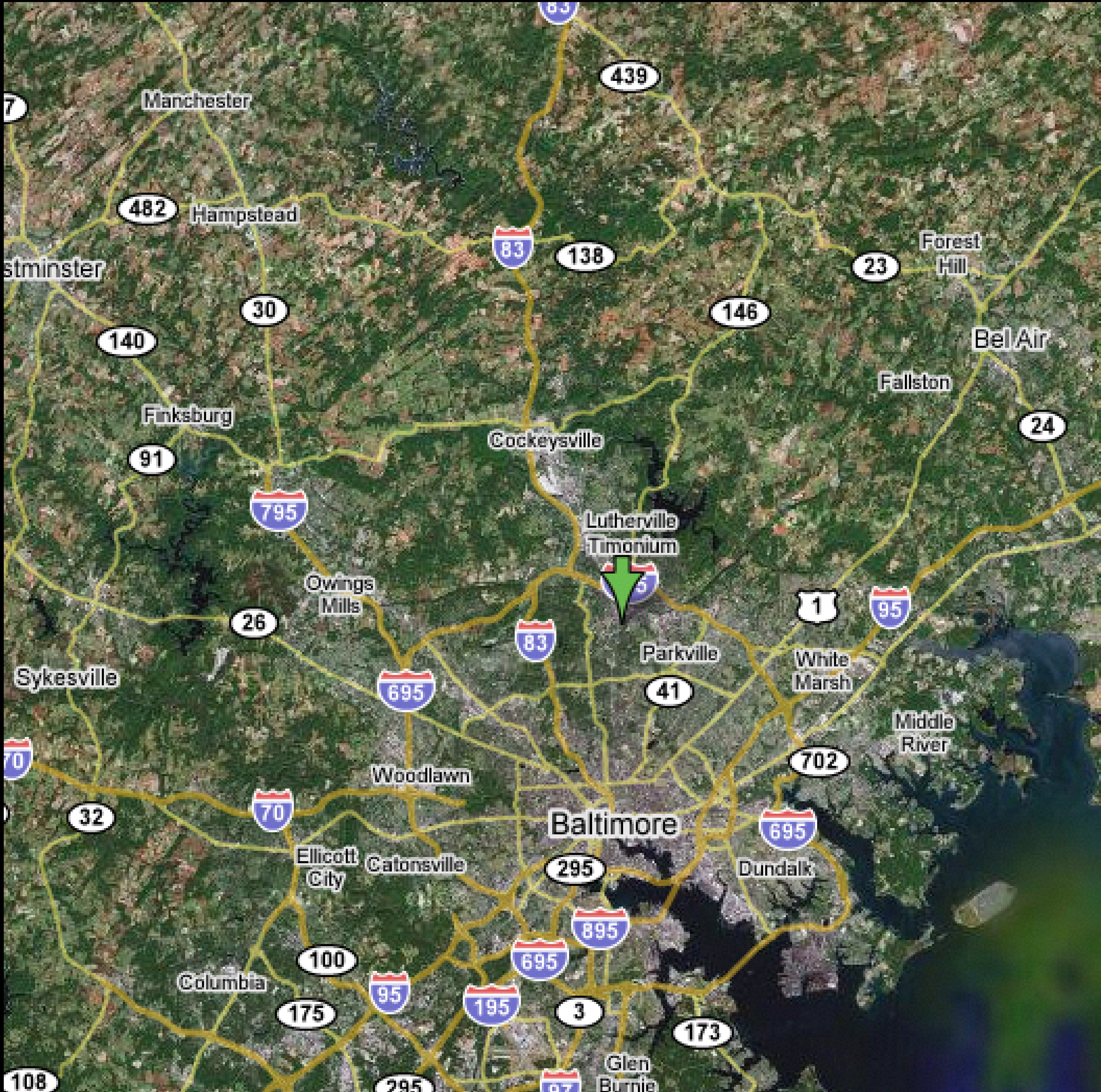
- The anchor point can be a place of residence, a place of work, or some other commonly visited location.

Geographic Profiling

- Our question is *operational*.
  - This places limitations on available data.

- Example
  - A series of 9 linked vehicle thefts in Baltimore County

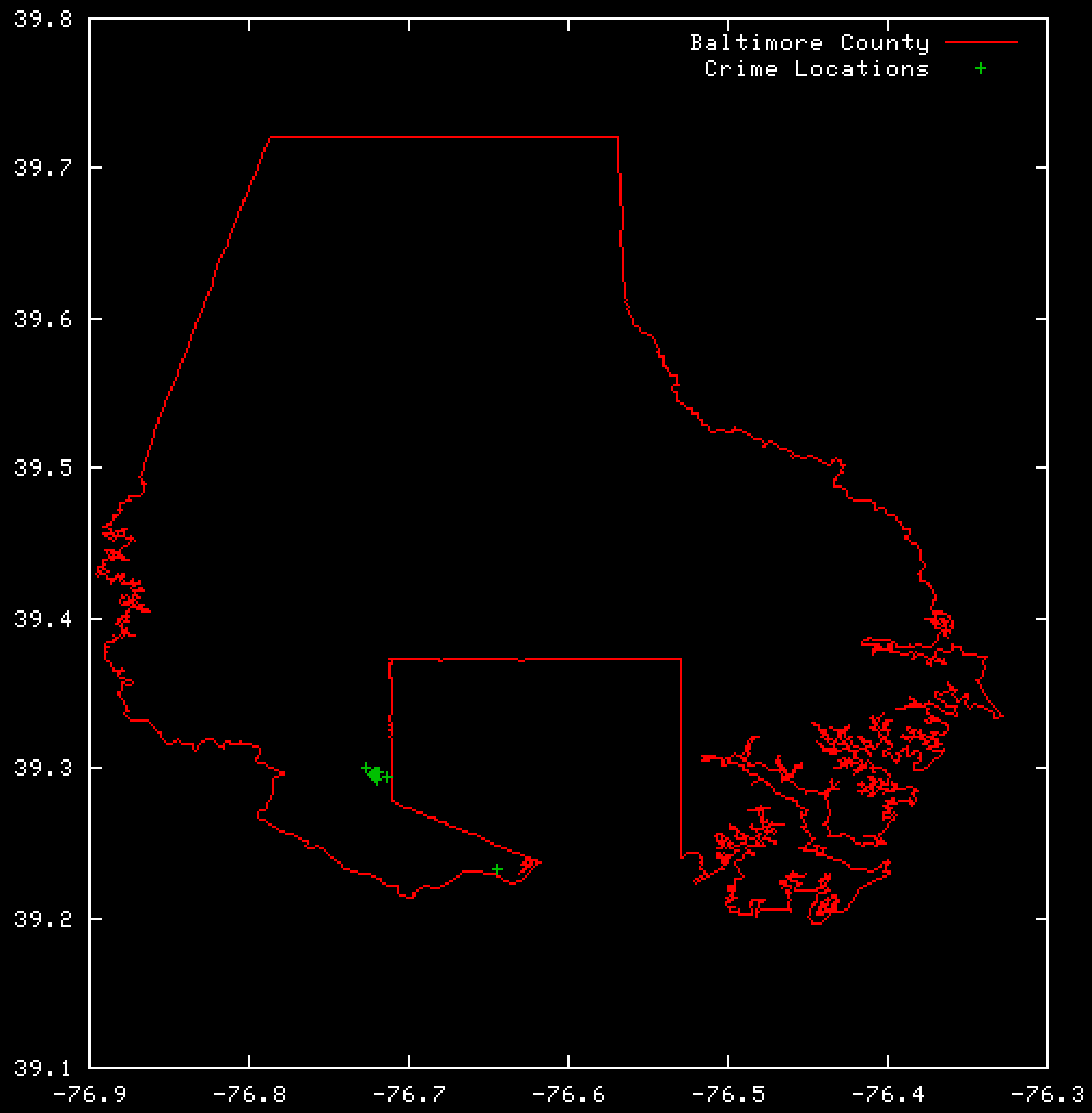

6
5B
3
7
1
2
11
42
3
5
8
2
6
4
10
7
6
12A
1
3
CHESAPEAKE
BAY



439
Manchester
482
Hampstead
83
138
23
Forest Hill
30
146
140
Bel Air
Fallston
Finksburg
Cockeysville
24
91
795
Lutherville Timonium
Owings Mills
1
95
26
83
Parkville
Sykesville
695
White Marsh
41
Middle River
702
Woodlawn
32
70
Baltimore
695
Ellicott City
Catonsville
295
Dundalk
895
695
100
Columbia
95
195
3
175
173
Glen Burnie
108

Example

ADDRESS	DATE_FROM	TIME	DATE_TO	TIME	REMARKS
918 M	01/18/2003	0800	01/18/2003	0810	VEHICLE IS 01 TOYT CAMRY, LEFT VEH RUNNING
1518 L	01/22/2003	0700	01/22/2003	0724	VEHICLE IS 99 HOND ACCORD STL-REC, ...B/M PAIR,DRIVING MAROON ACCORD.
731 CC	01/22/2003	0744	01/22/2003	0746	VEHICLE IS 02 CHEV MALIBU STL-REC
1527 K	01/27/2003	1140	01/27/2003	1140	VEHICLE IS 97 MERC COUGAR, LEFT VEH RUNNING
1514 G	01/29/2003	0901	01/29/2003	0901	VEHICLE IS 99 MITS DIAMONTE, LEFT VEH RUNNING
1415 K	01/29/2003	1155	01/29/2003	1156	VEHICLE IS 00 TOYT 4RUNNER STL-REC, (4) ARREST NFI
5943 R	12/31/2003	0632	12/31/2003	0632	VEHICLE IS 92 BMW 525, WARMING UP VEH
1427 G	02/17/2004	0820	02/17/2004	0830	VEHICLE IS 00 HOND ACCORD, WARMING VEH
4449 S	05/15/2004	0210	05/15/2004	0600	VEHICLE IS 04 SUZI ENDORO



Baltimore County
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

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

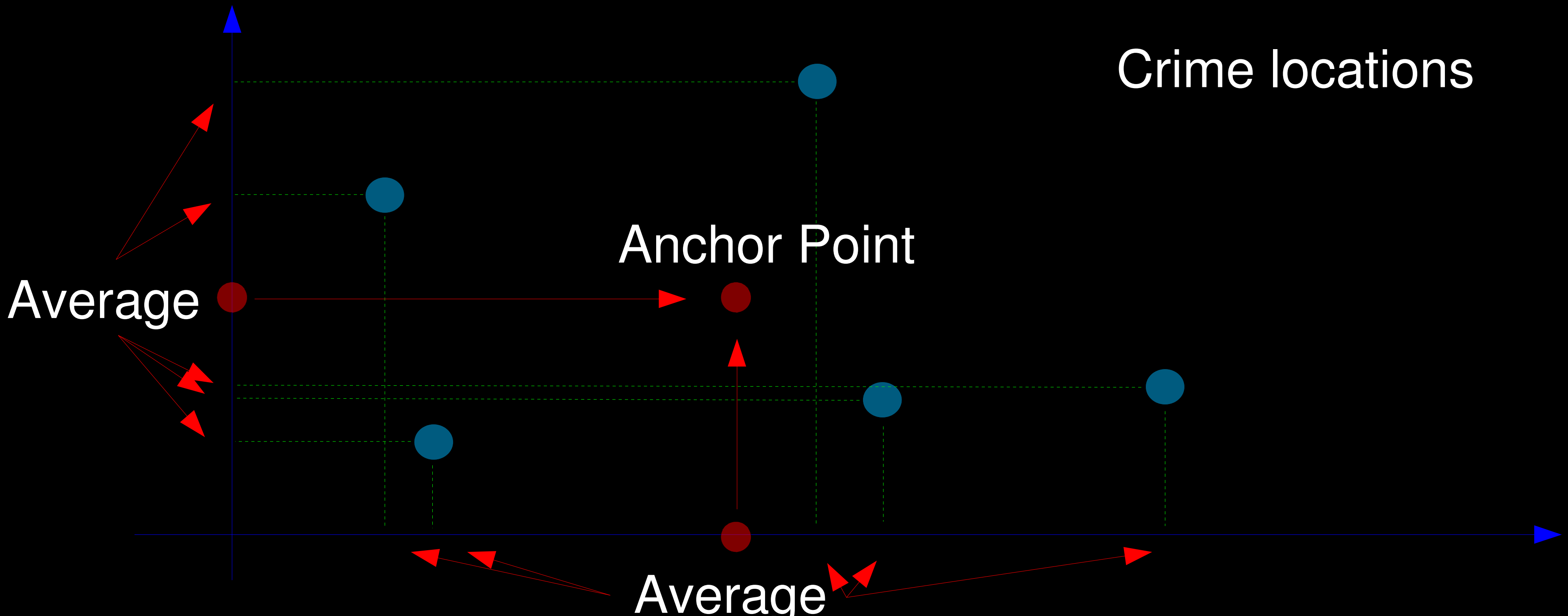
$$d_1(\boldsymbol{x},\boldsymbol{y})=|x^{(1)}-y^{(1)}|+|x^{(2)}-y^{(2)}|$$

- Street grid

Spatial Distribution Strategies

- Centroid:

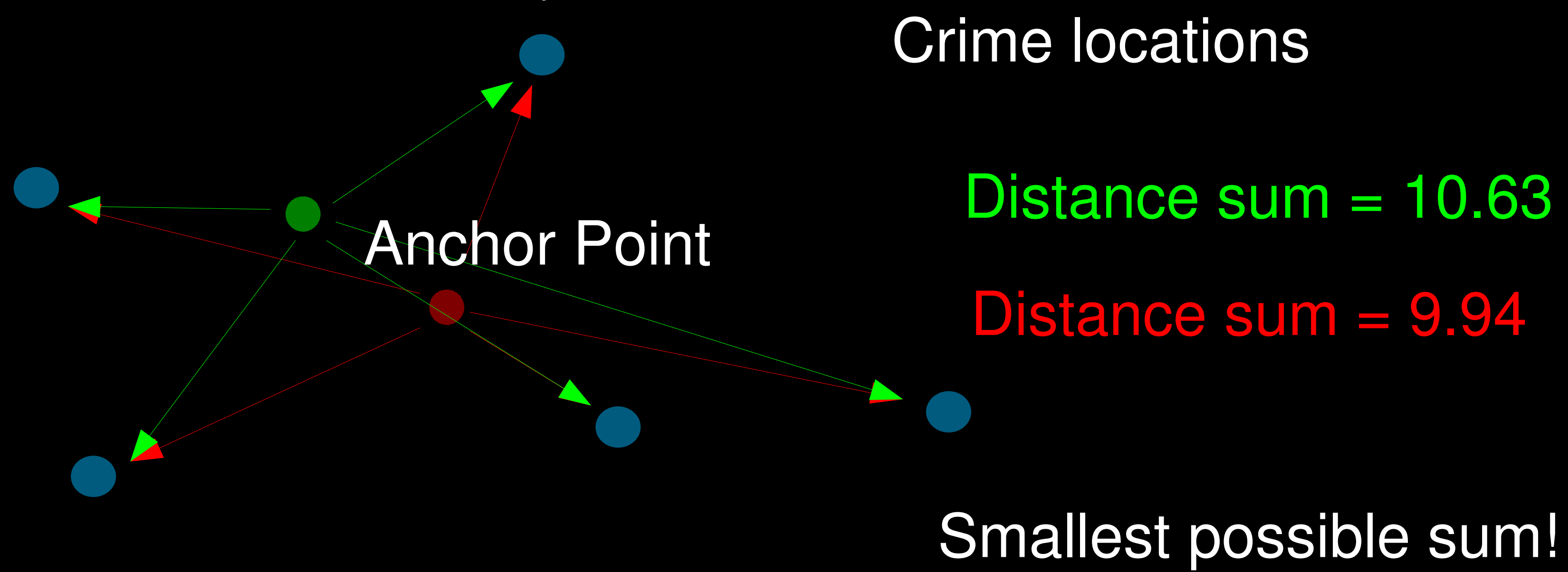
$$\zeta_{centroid} = \frac{1}{n} \sum_{i=1}^{n} x_i$$



Spatial Distribution Strategies

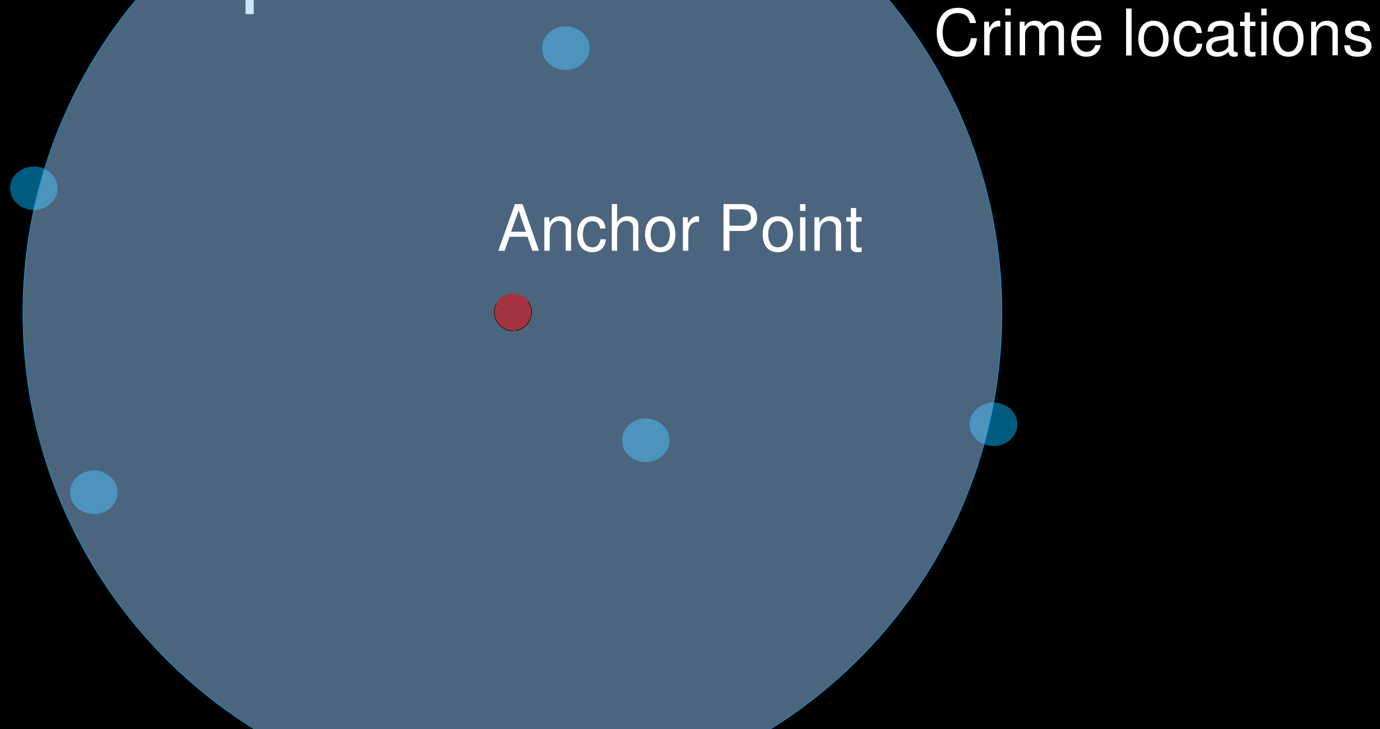
- Center of minimum distance: ζ_{cmd} is the value of y that minimizes

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



Spatial Distribution Strategies

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



Probability Distribution Strategies

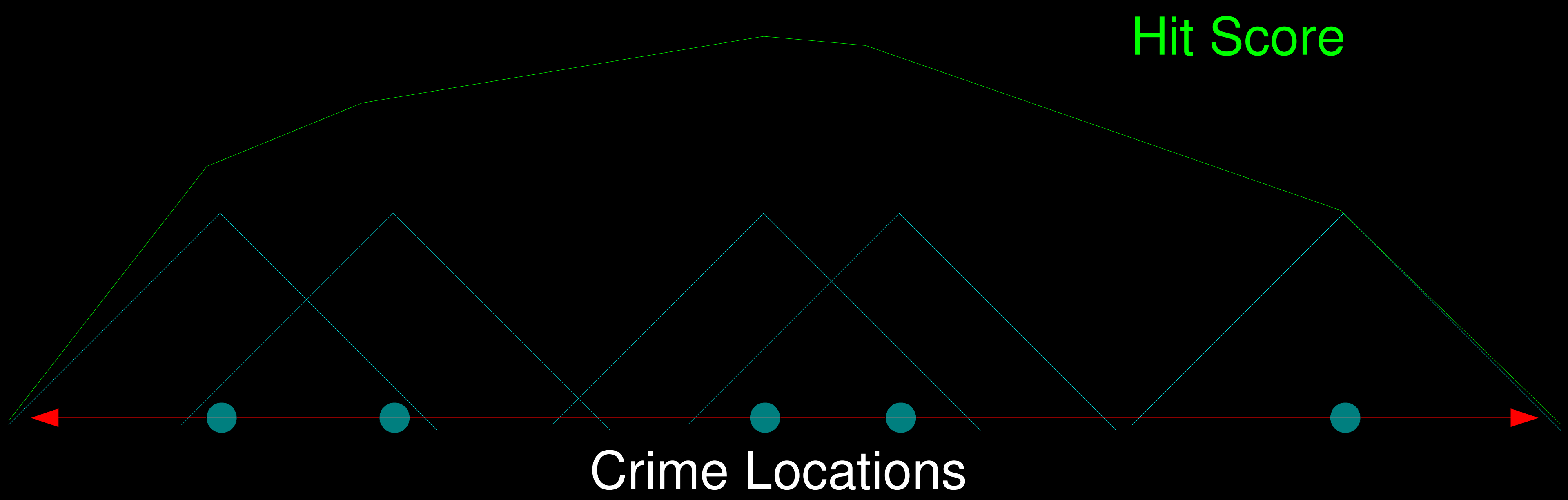
- The anchor point is located in a region with a high "hit score".
- The hit score $S(\boldsymbol{y})$ has the form

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

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

Probability Distribution Strategies

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



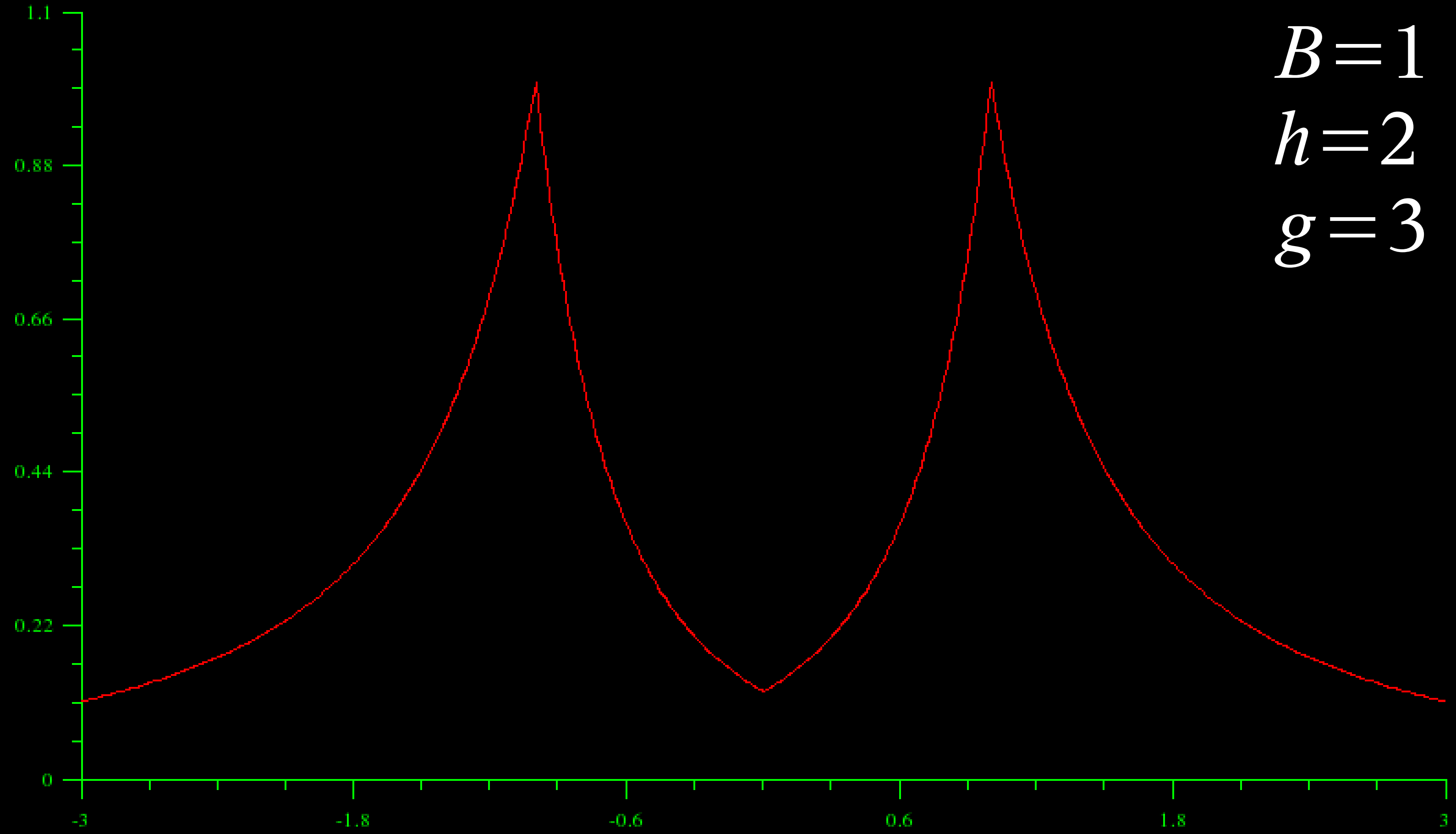
Rossmo

- Manhattan distance metric.
- Decay function

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

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

Rossmo



$B = 1$
$h = 2$
$g = 3$

Canter, Coffey, Huntley & Missen

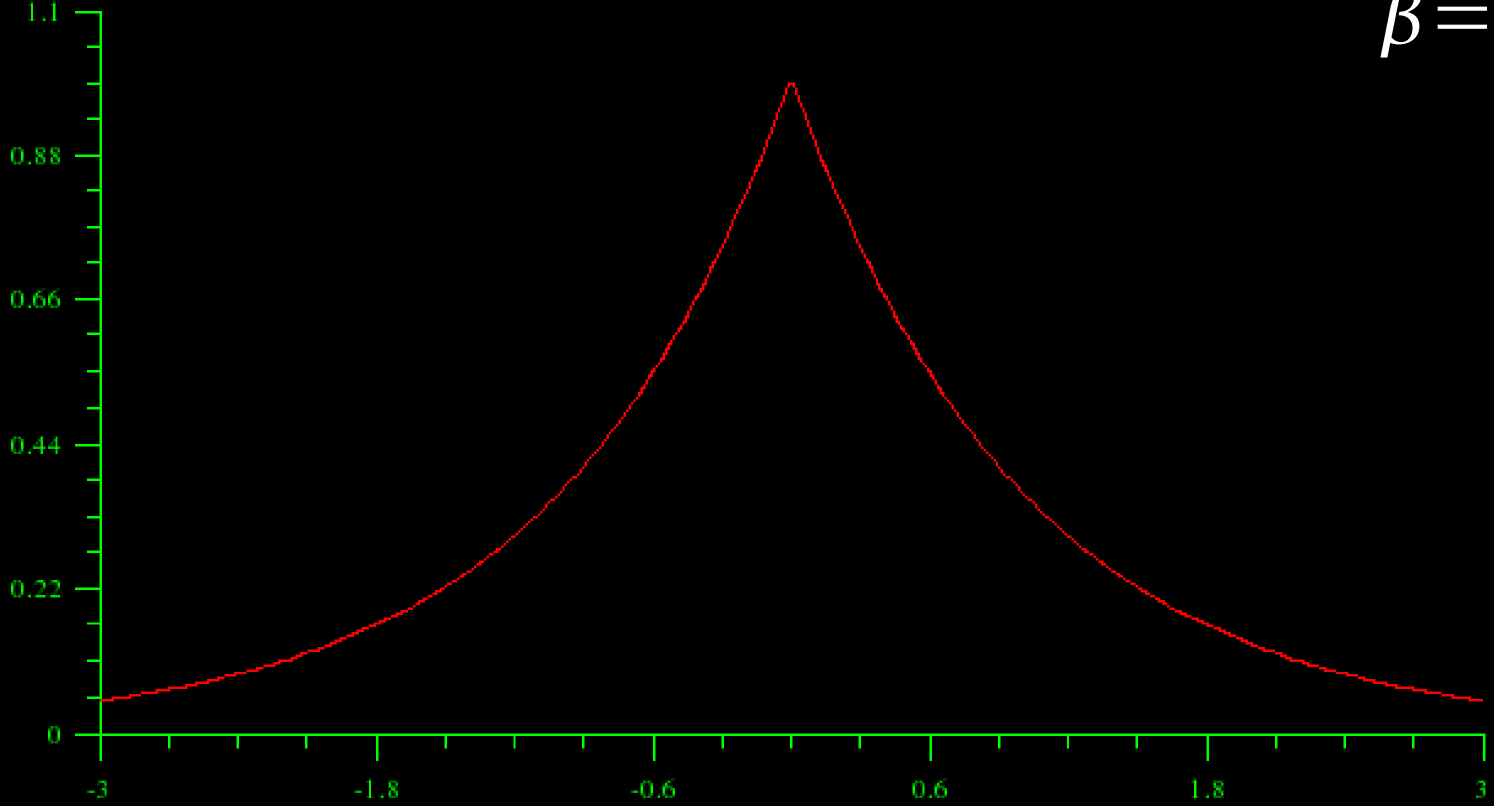
- Euclidean distance
- Decay functions

$$f(d) = A e^{-\beta d}$$

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

Dragnet

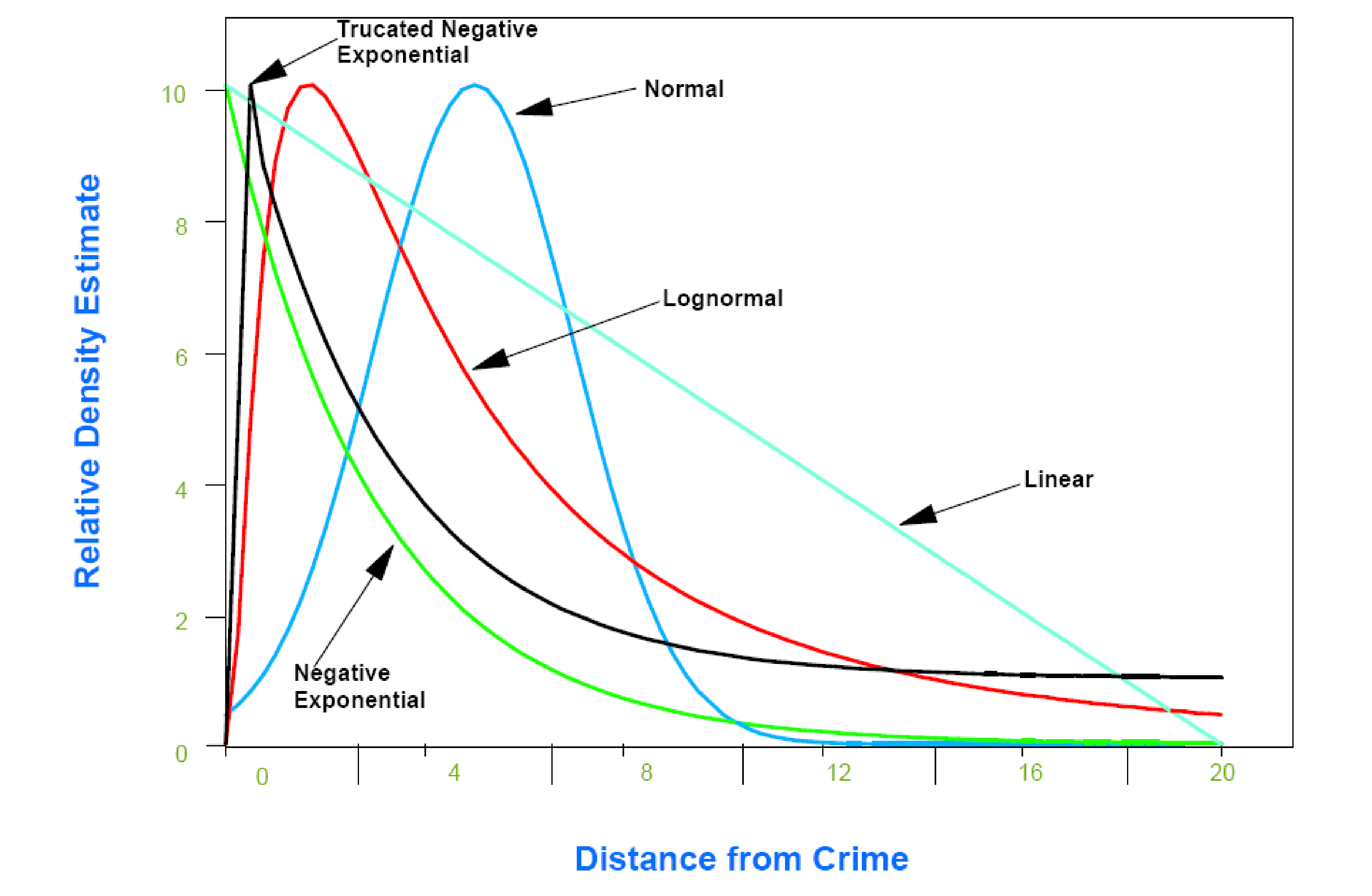
$A = 1$
$\beta = 1$



Levine

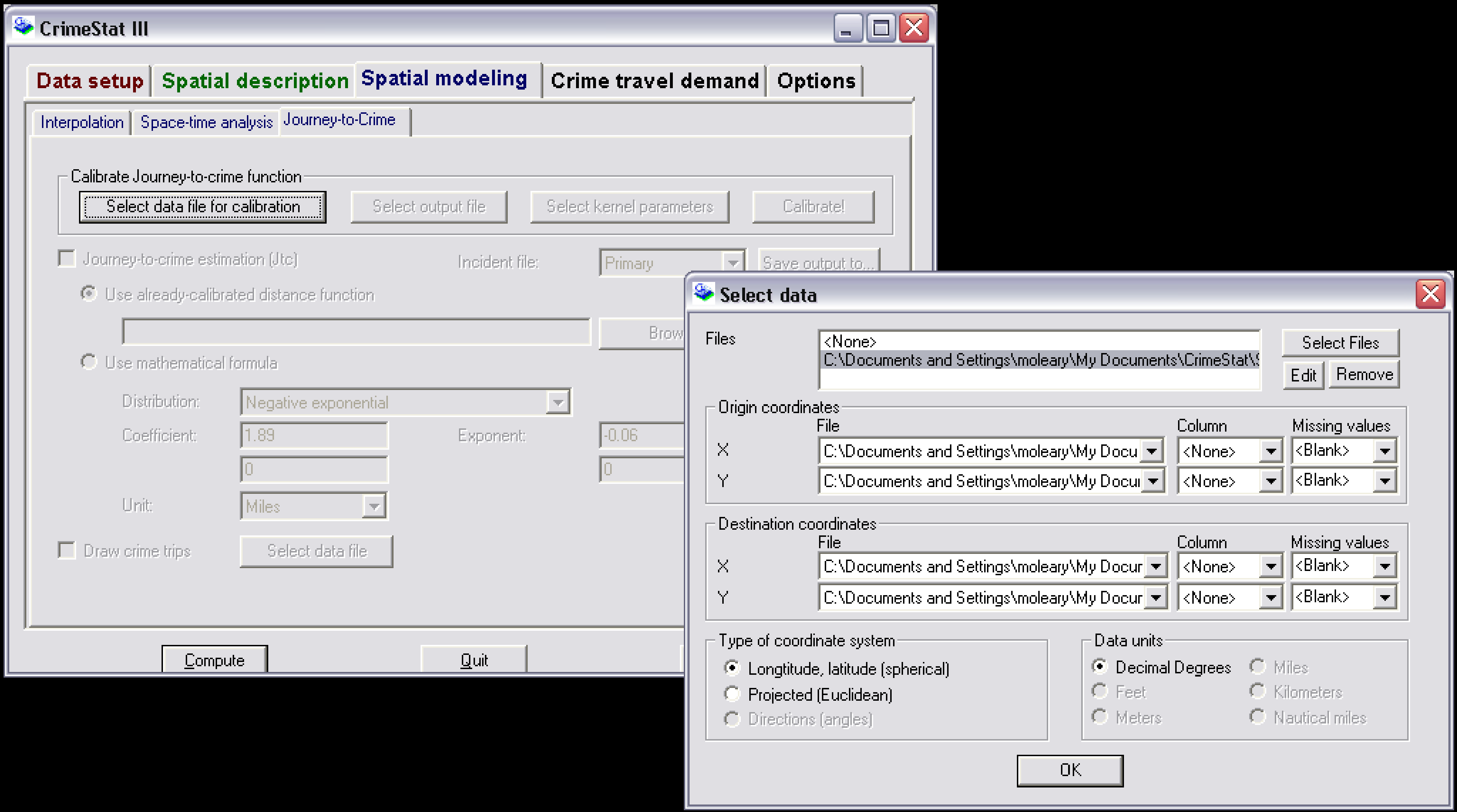
- Euclidean distance
- Decay functions
  - Linear $f(d) = A + Bd$
  - Negative exponential $f(d) = A e^{-\beta d}$
  - Normal $f(d) = \frac{A}{\sqrt{2\pi S^2}} \exp[\frac{-(d-\bar{d})^2}{2S^2}]$
  - Lognormal $f(d) = \frac{A}{d\sqrt{2\pi S^2}} \exp[\frac{-(\ln d-\bar{d})^2}{2S^2}]$

CrimeStat



From Levine (2004)

CrimeStat

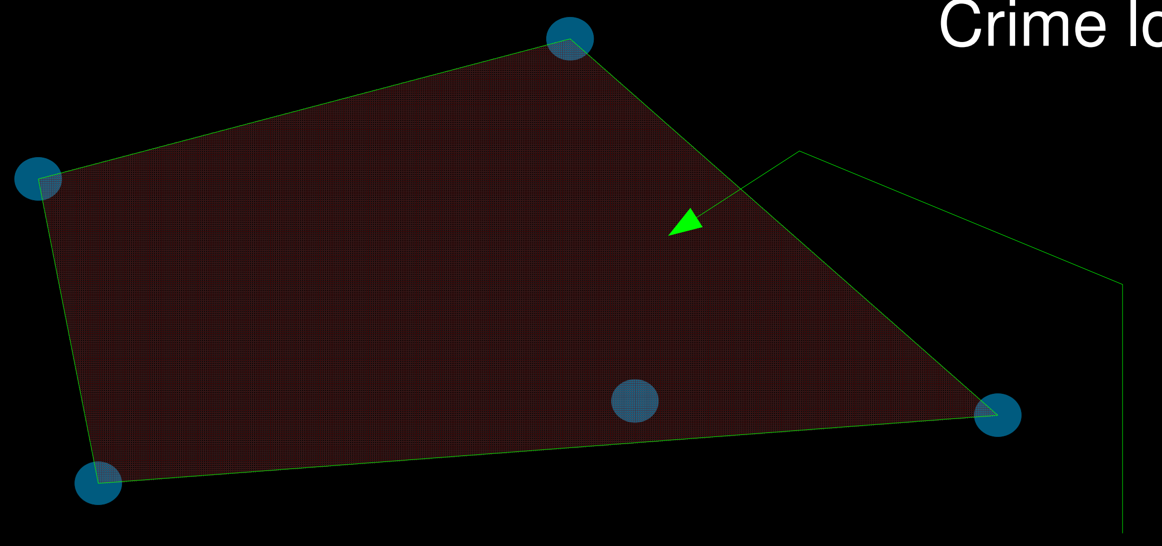

CrimeStat III
Data setup
Spatial description
Spatial modeling
Crime travel demand
Options
Interpolation
Space-time analysis
Journey-to-Crime
Calibrate Journey-to-crime function
Select data file for calibration
Select output file
Select kernel parameters
Calibrate!
Journey-to-crime estimation (Jtc)
Incident file:
Primary
Save output to...
Use already-calibrated distance function
Use mathematical formula
Distribution:
Negative exponential
Coefficient:
1.89
Exponent:
-0.06
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
Longtitude, latitude (spherical)
Projected (Euclidean)
Directions (angles)
Data units
Decimal Degrees
Miles
Feet
Kilometers
Meters
Nautical miles
OK

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*

A New Approach

- In previous methods, the unknown quantity was:
  - The anchor point

    (spatial distribution strategies)
  - The hit score

    (probability distance strategies)
- We use a different unknown quantity.

A New Approach

- Let $P(\boldsymbol{x};\boldsymbol{z})$ be the density function for the probability that an offender with anchor point $\boldsymbol{z}$ commits a crime at location $\boldsymbol{x}$.
  - This distribution is our new unknown.
  - This has criminological significance.
    - In particular, assumptions about the form of $P(\boldsymbol{x};\boldsymbol{z})$ are equivalent to assumptions about the offender's behavior.

The Mathematics

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

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

$$= P(\mathbf{x}_1, \mathbf{y}) P(\mathbf{x}_2, \mathbf{y}) \cdots P(\mathbf{x}_n, \mathbf{y})$$

or equivalently, the value that maximizes

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

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

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

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

- 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

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

- 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.

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

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

Better Models

- We have recaptured the results of existing techniques by choosing $P(\boldsymbol{x}\,;\boldsymbol{z})$ appropriately.
- These choices of $P(\boldsymbol{x}\,;\boldsymbol{z})$ 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 $P(\boldsymbol{x};\boldsymbol{z})$.
- Consider

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

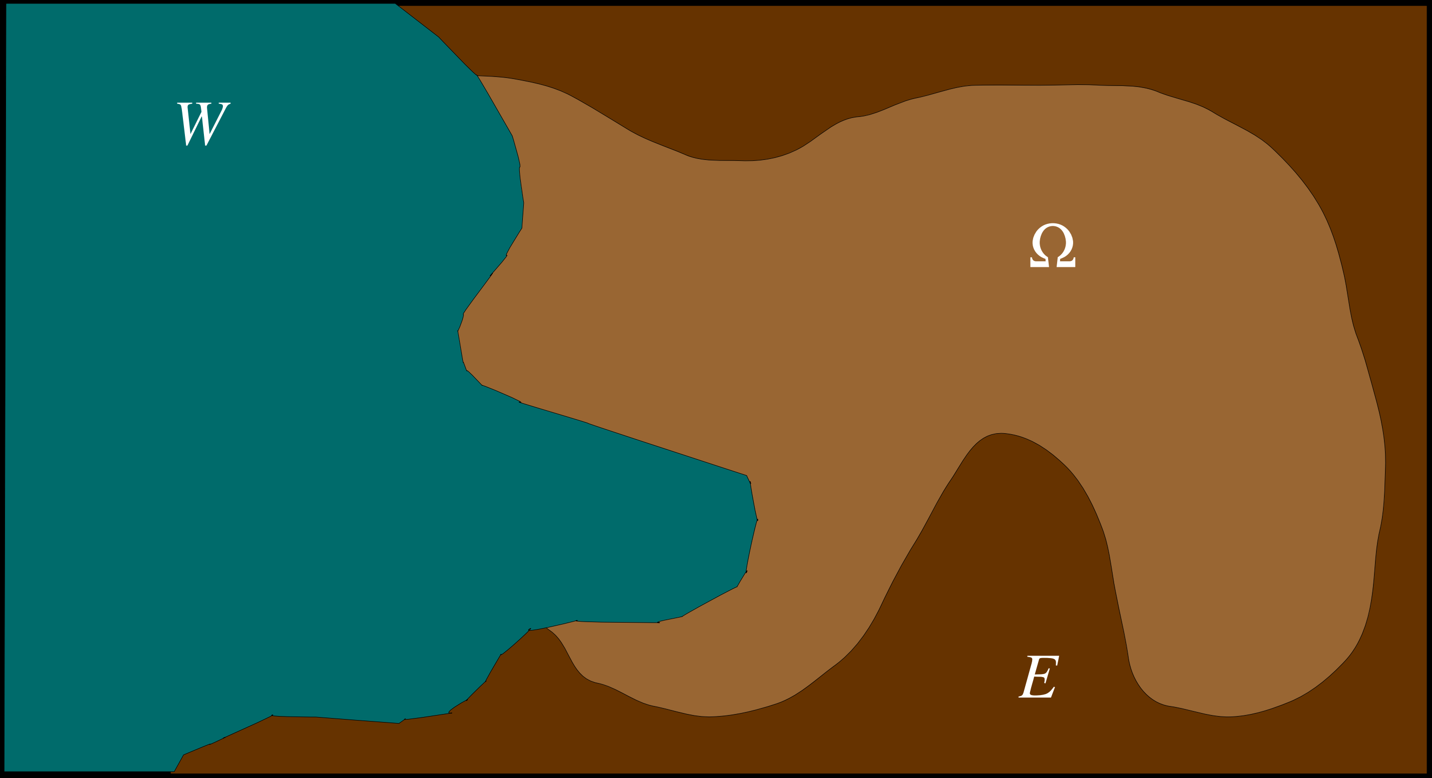
Distance Decay (Dispersion Kernel)

Geographic factors

Normalization

The Simplest Case

- Suppose we have information about crimes committed by the offender only for a portion of the region.



The Simplest Case

- Regions
  - Ω: Jurisdiction(s). Crimes and anchor points may be located here.
  - E: “elsewhere”. Anchor points may lie here, but we have no data on crimes here.
  - W: “water”. Neither anchor points nor crimes may be located here.
- In all other respects, we assume the geography is *homogeneous*.

The Simplest Case

- We set

$$G(x)=\begin{cases}1 & x\in\Omega\\ 0 & x\notin\Omega\end{cases}$$

  We choose an appropriate decay function

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

- The required normalization function is

$$N(\boldsymbol{x};z)=\left[\iint_{\Omega}\exp\left(-\frac{|y-z|^2}{2\sigma^2}\right)dy^{(1)}dy^{(2)}\right]^{-1}$$

The Simplest Case

- Our estimate ζ_{mle} of the anchor point is the choice of y that maximizes

$$\frac{\exp\left(-\sum_{i=1}^{n} \frac{|\boldsymbol{x}_i - \boldsymbol{y}|^2}{2\sigma^2}\right)}{\left[\iint_{\Omega} \exp\left(-\frac{|\boldsymbol{\eta} - \boldsymbol{y}|^2}{2\sigma^2}\right) d\eta^{(1)} d\eta^{(2)}\right]^n}$$

The Simplest Case

- Our students wrote code to implement this method last year, and tested it on real crime data from Baltimore County.
- We used Green's theorem to convert the double integral to a line integral.

$$\iint_{\Omega} \exp\left(\frac{-|\eta-y|^2}{2\sigma^2}\right) d\eta^{(1)}\, d\eta^{(2)} = \oint_{\partial\Omega} \frac{-\sigma^2}{|\eta-y|} \exp\left(\frac{-|\eta-y|^2}{\beta}\right) (\boldsymbol{e_r} \cdot \boldsymbol{n})\, ds + \begin{cases} \beta\pi & z \in \Omega \\ 0 & z \notin \Omega \end{cases}$$

  - Baltimore county was simply a polygon with 2908 vertices.

The Simplest Case

- To calculate the maximum, we used the BFGS method.
  - Search in the direction $D_n \nabla f(\boldsymbol{y_n})$ where

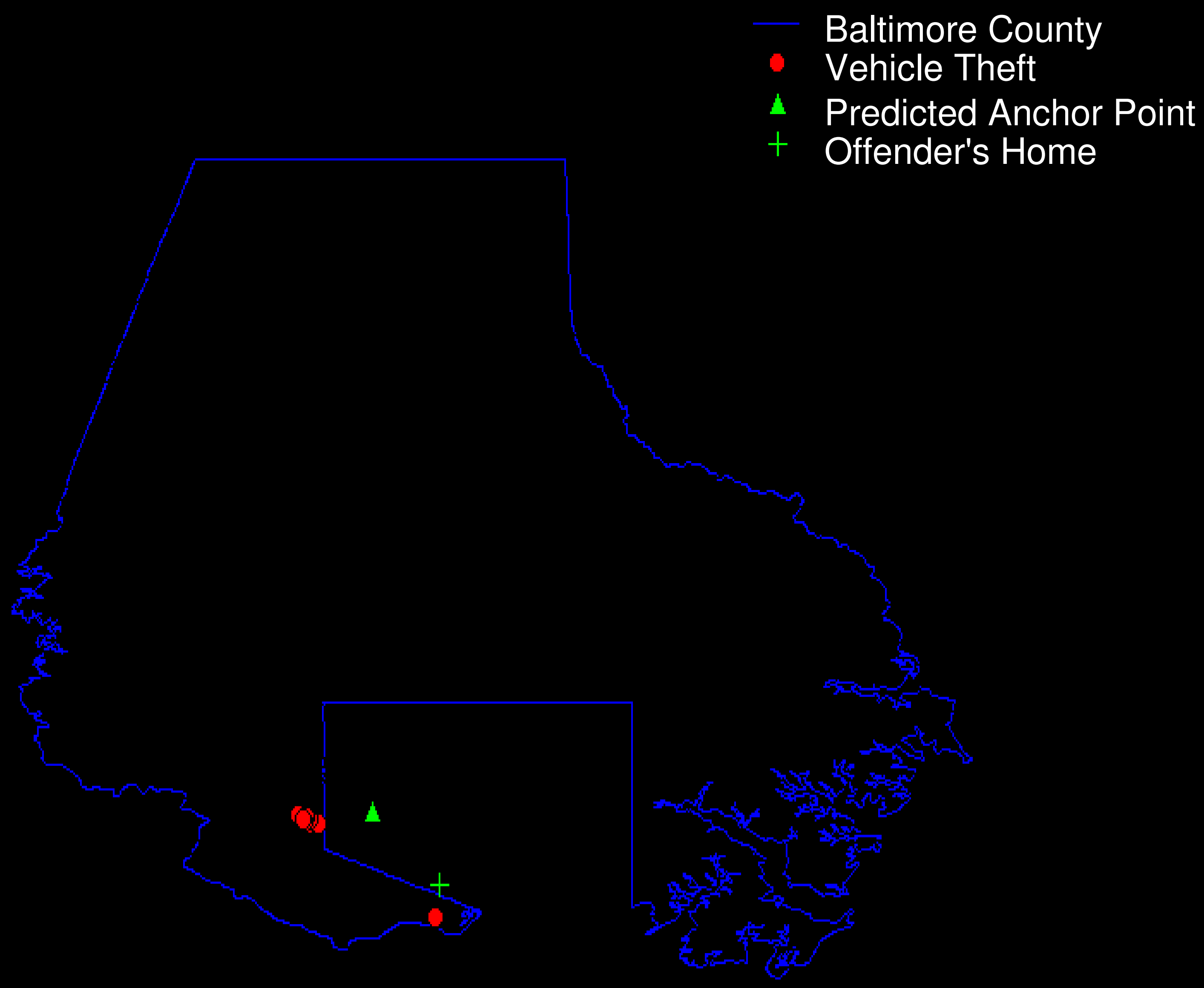
$$D_{n+1} = D_n + \left(1 + \frac{\mathbf{g}^T D_n \mathbf{g}}{\mathbf{d}^T \mathbf{g}}\right) \frac{\mathbf{d}\mathbf{d}^T}{\mathbf{d}^T \mathbf{g}} - \frac{D_n \mathbf{g}\mathbf{d}^T + \mathbf{g}\mathbf{d}^T D_n}{\mathbf{d}^T \mathbf{g}}$$

$$\boldsymbol{d} = \boldsymbol{y}_{n+1} - \boldsymbol{y}_n$$

$$\boldsymbol{g} = \nabla f(\boldsymbol{y}_{n+1}) - \nabla f(\boldsymbol{y}_n)$$

  - For the 1-D optimization we used the bisection method.

Sample Results



Better Models

- This is just a modification of the centroid method that accounts for possibly missing crimes outside the jurisdiction.
- Clearly, better models are needed.

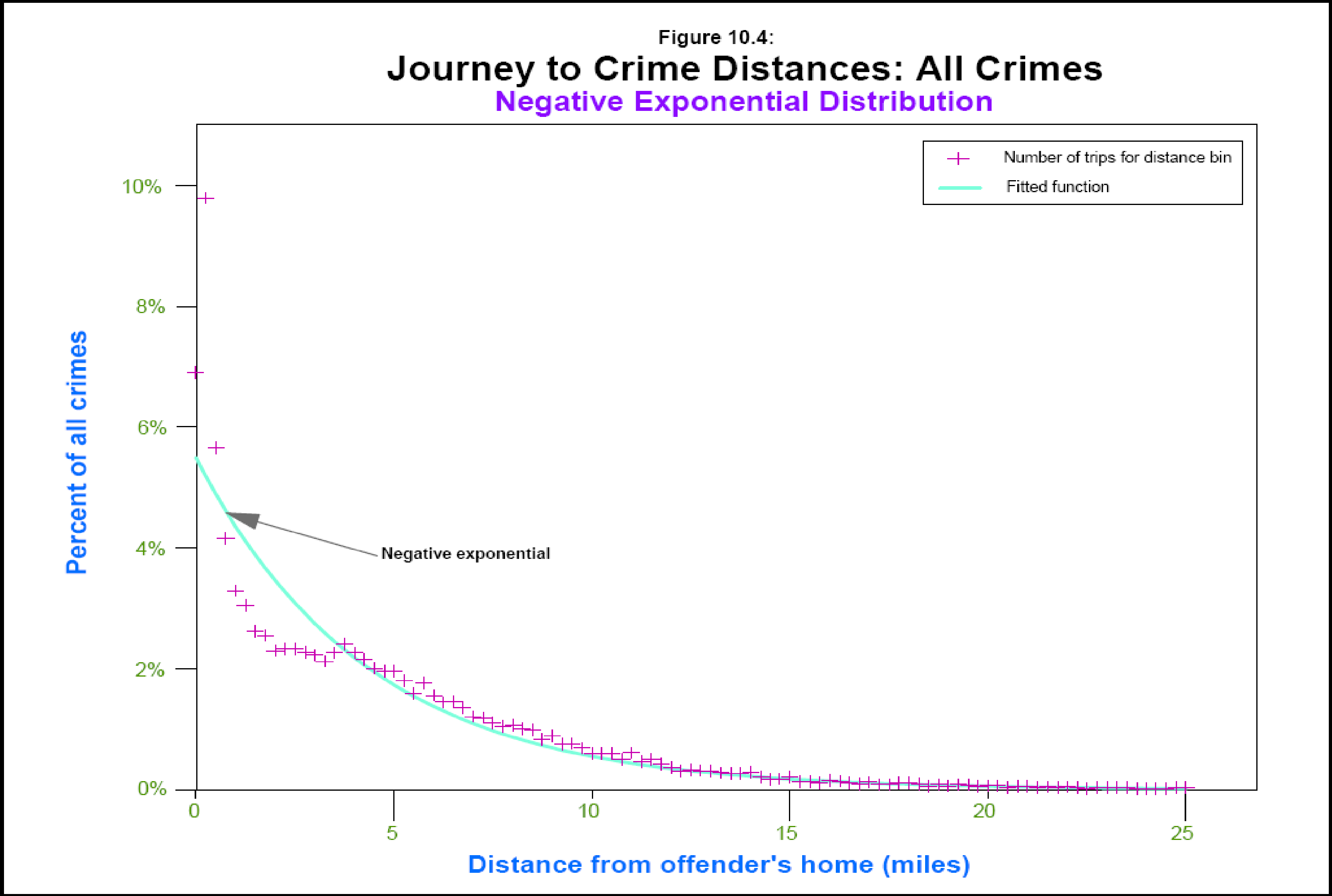
Better Models

- Recall our ansatz

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

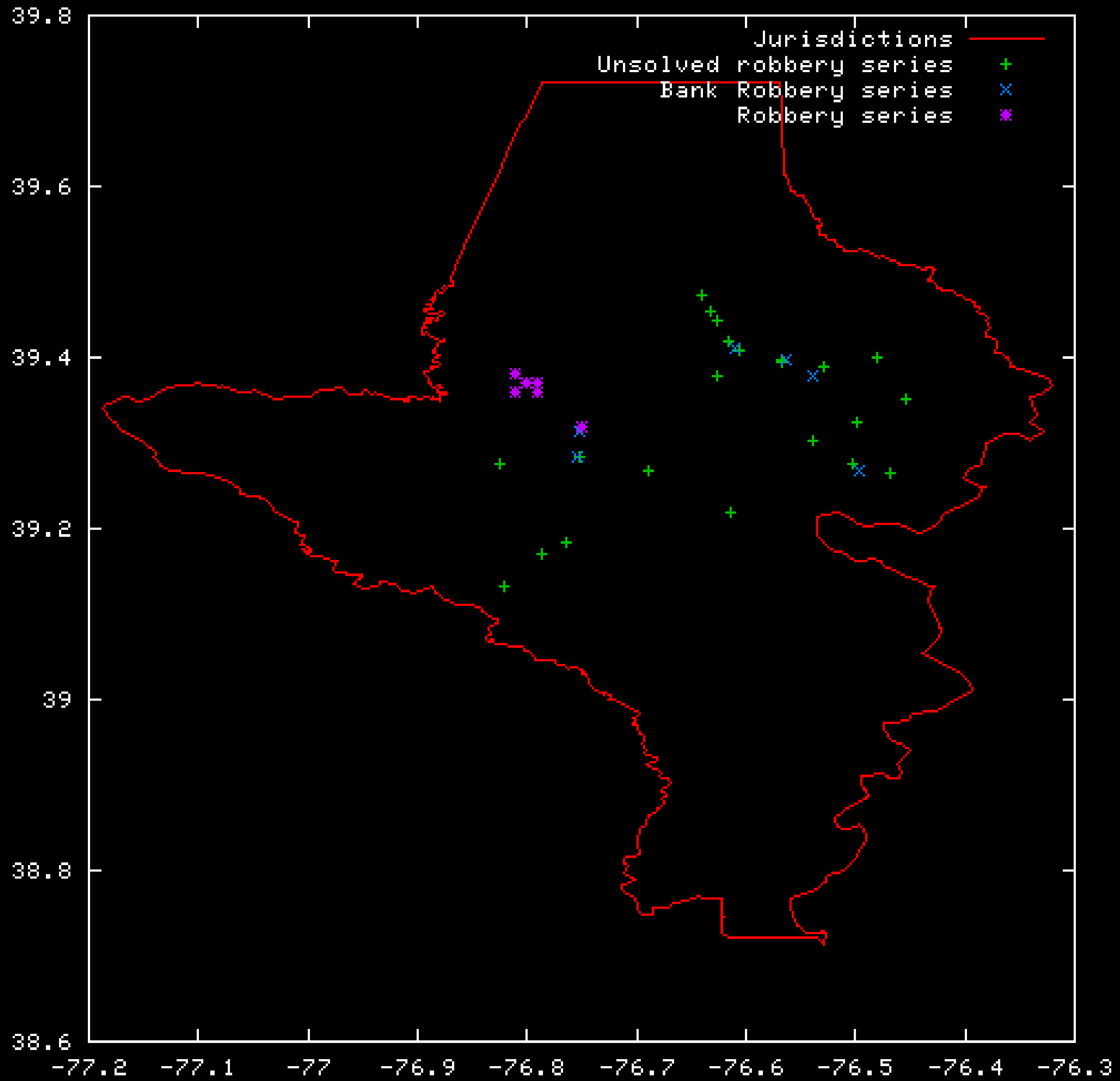
  - What would be a better choice of D?
  - What would be a better choice of G?

Distance Decay



From Levine (2004)

Distance Decay



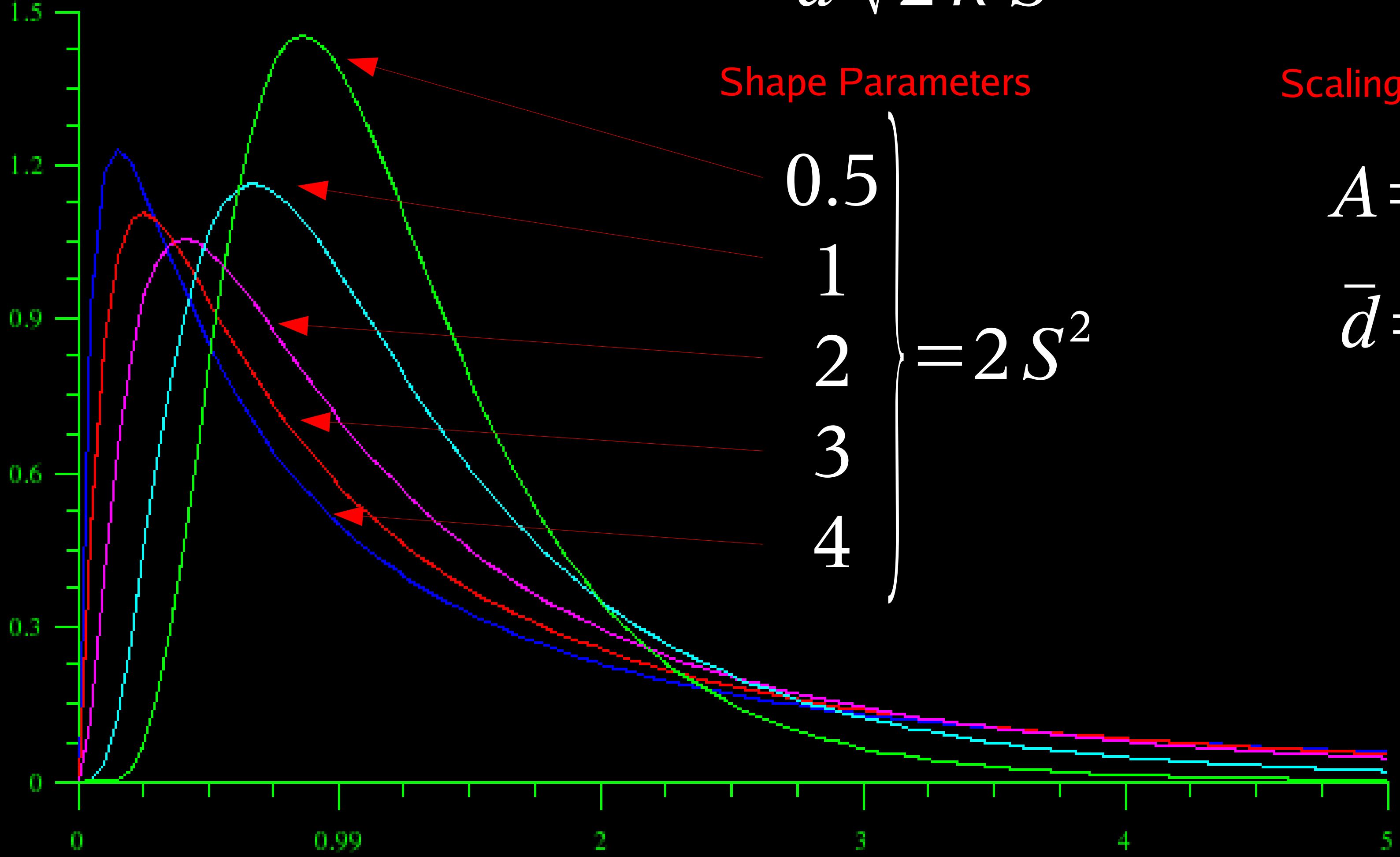
Distance Decay

- Suppose that each offender has a decay function $f(d;\lambda)$ where $\lambda \in (0,\infty)$ varies among offenders according to the distribution $\phi(\lambda)$.
- Then if we look at the decay function for all offenders, we obtain the aggregate distribution

$$F(d) = \int_0^\infty f(d;\lambda) \cdot \phi(\lambda)\, d\lambda$$

Distance Decay

$$f(d)=\frac{A}{d\sqrt{2\pi S^2}}\exp\left[\frac{-(\ln d-\bar{d})^2}{2S^2}\right]$$



Shape Parameters

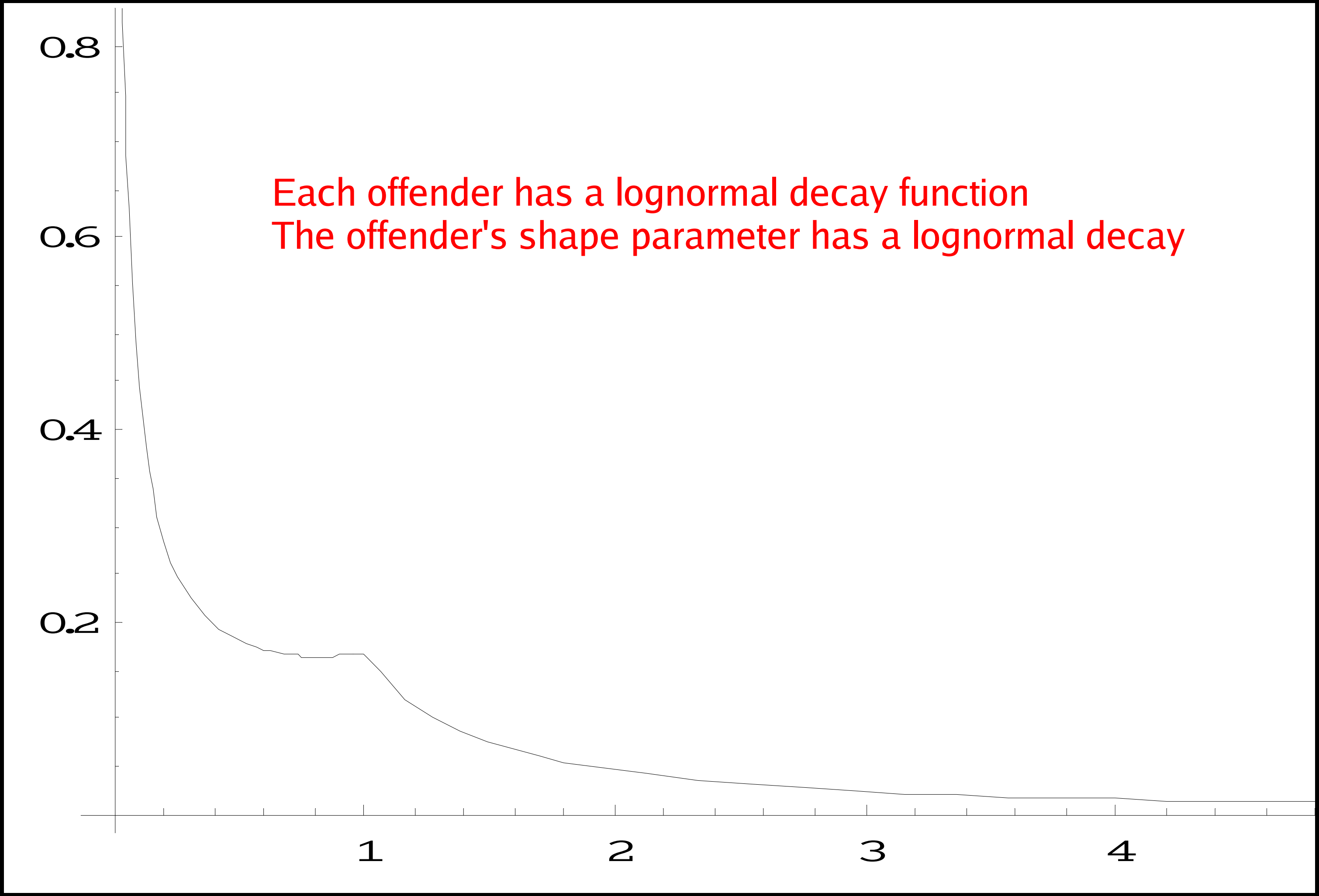
$$\left.\begin{matrix}0.5\\1\\2\\3\\4\end{matrix}\right\}=2S^2$$

Scaling Parameters

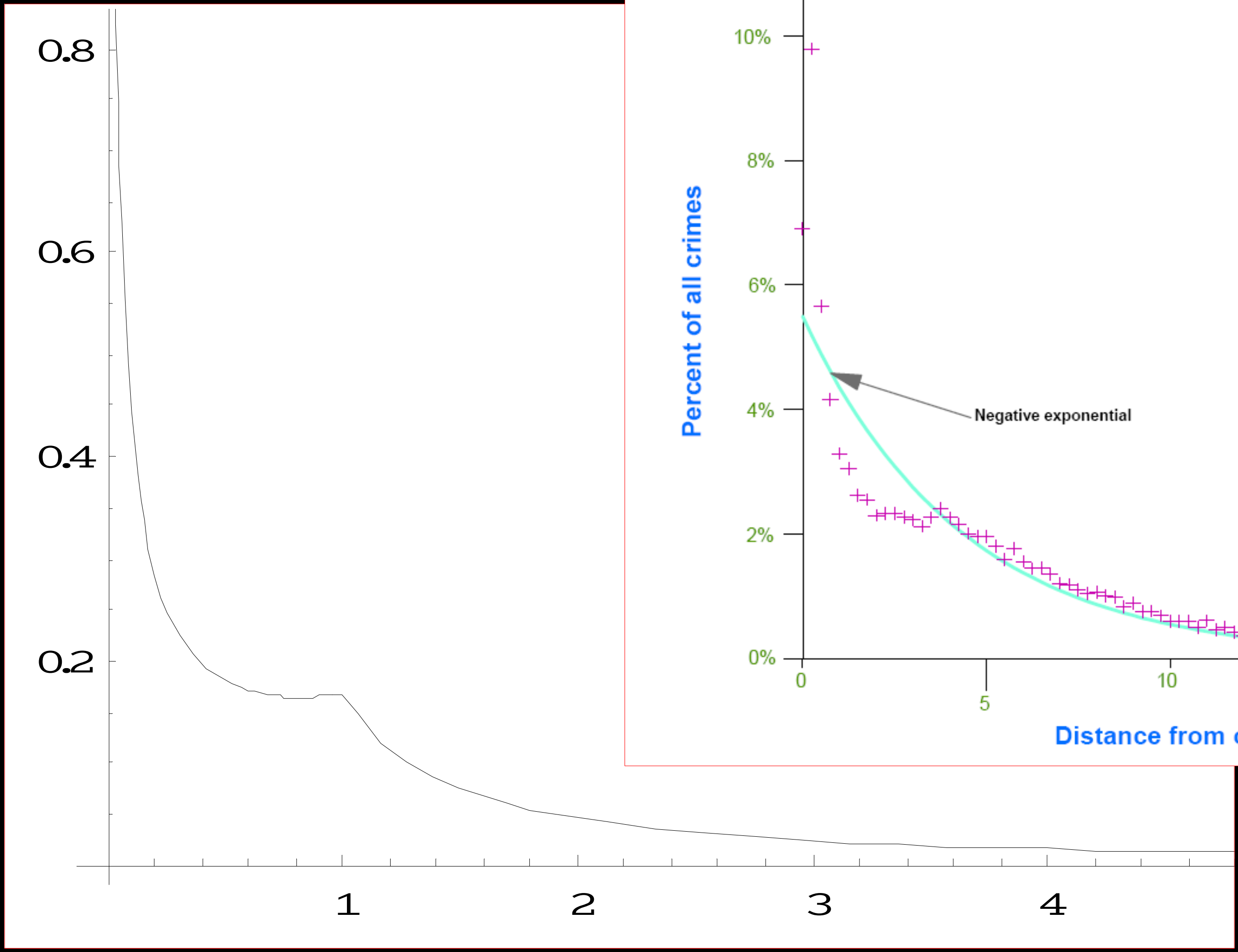
$$A=\sqrt{\pi}$$

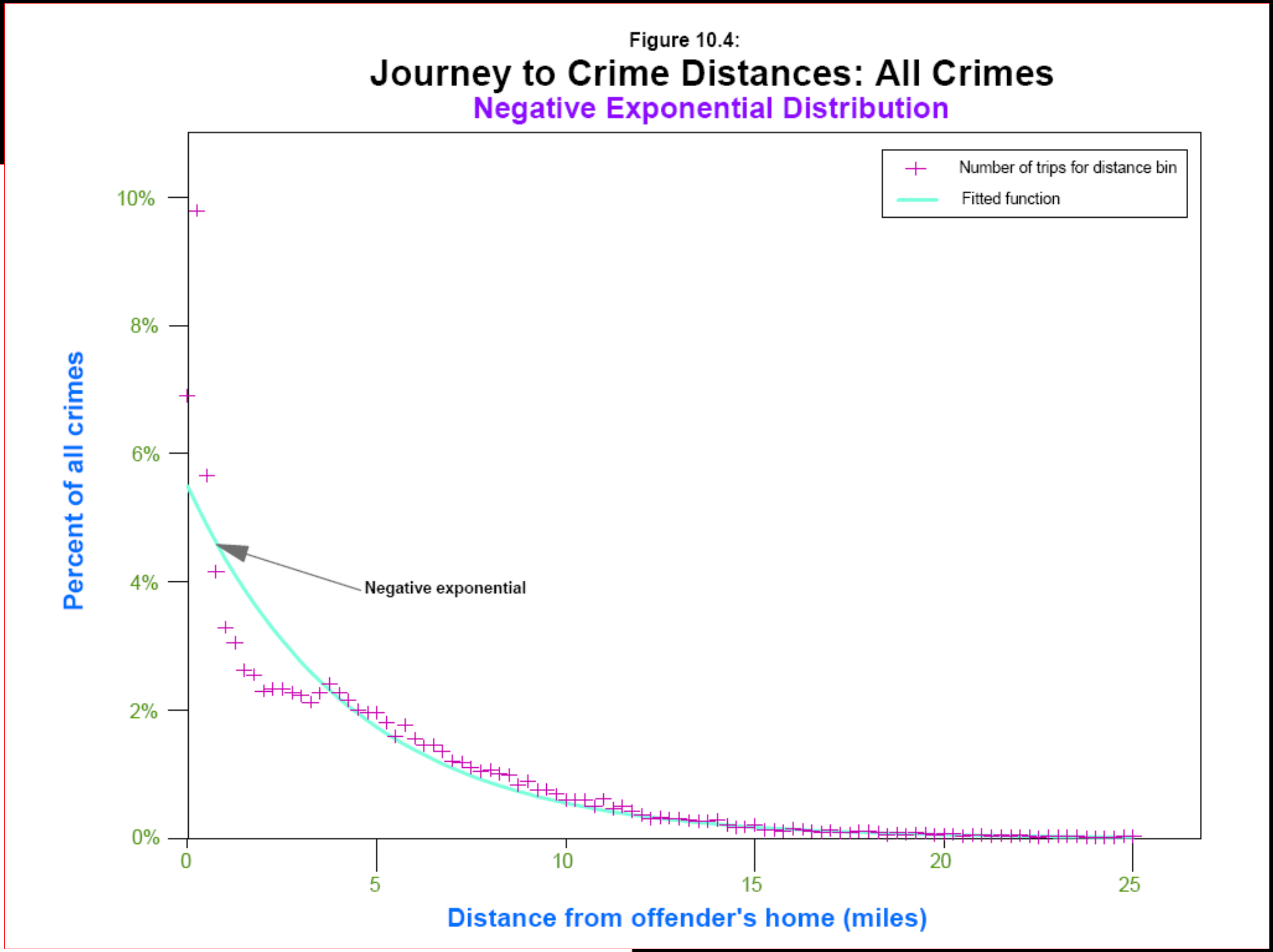
$$\bar{d}=0.1$$

Distance Decay



Distance Decay





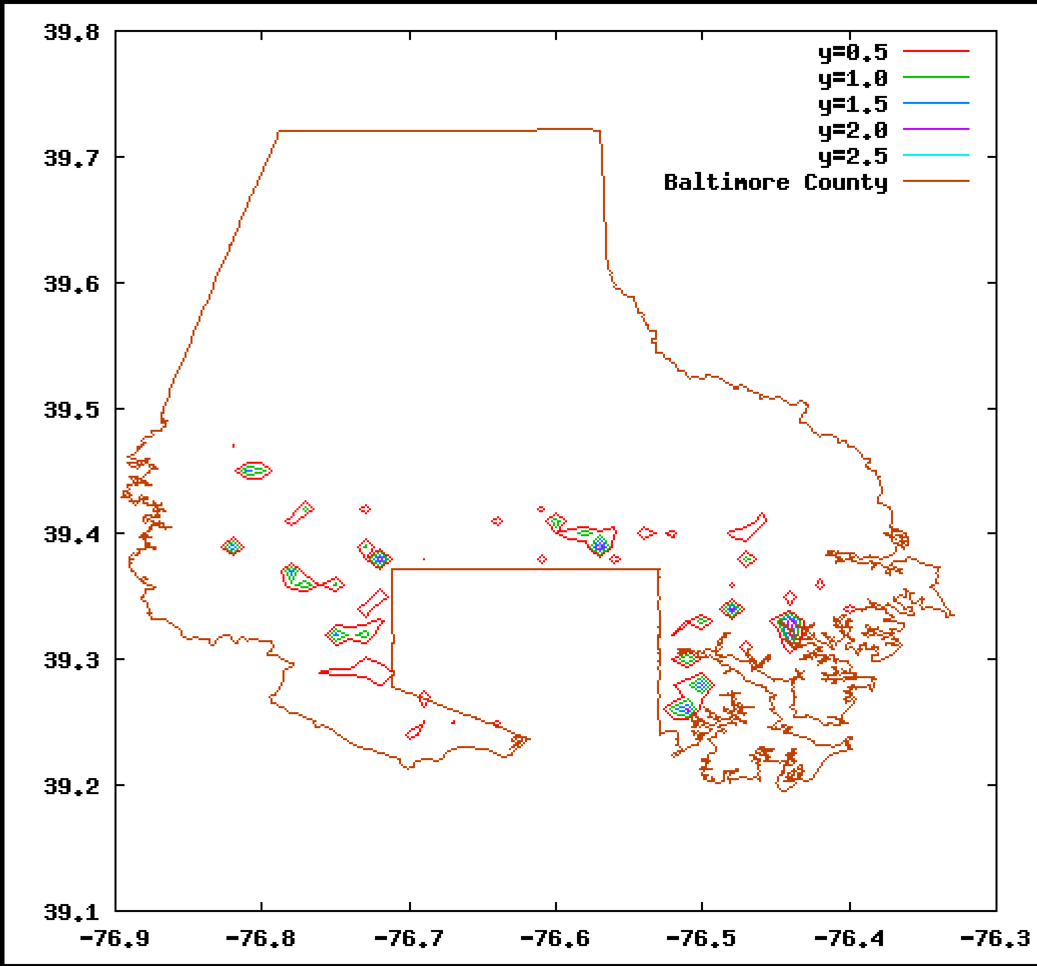
Distance Decay

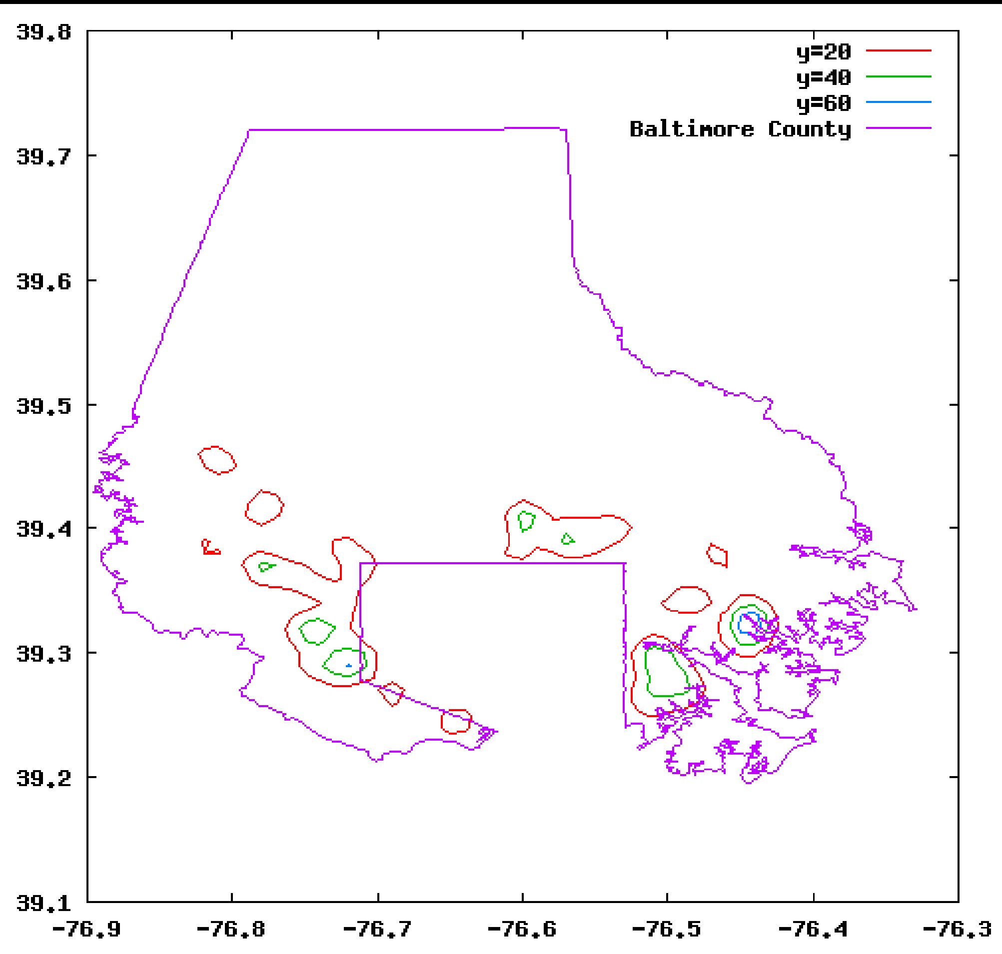
- Is this real, or an artifact?
- How do we determine the “best” choice of decay function?
  - This needs to be determined in advance.
- Will it vary depending on
  - crime type?
  - local geography?

Geography

- Let $G(x)$ represent the local density of potential targets.
  - Rather than look for features (demographic, geographic) to predict it, we can use historical data to measure it.
  - $G(x)$ could then be calculated in the same fashion as hot spots; *e.g.* by kernel density parameter estimation.
    - Issues with boundary conditions

Geography


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
y=0.5
y=1.0
y=1.5
y=2.0
y=2.5
Baltimore County


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
y=20
y=40
y=60
Baltimore County

Geography

- No calibration is required if $G(x)$ is calculated in this fashion.
  - 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 $G(x)$.

Strengths of this Framework

- All of the assumptions on criminal behavior are made in the open.
  - They can be challenged, tested, discussed and compared.

Strengths

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

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 $P(\boldsymbol{x};\boldsymbol{z})$.
- It is unclear what the right choice is for $P(\boldsymbol{x};\boldsymbol{z})$
  - Even with the simplifying assumption that

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

    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 $P(\boldsymbol{x}\,;\boldsymbol{z})$.
- The error analysis for maximum likelihood estimators is delicate when the number of data points is small.

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...

Weaknesses

- We only produce the point estimate of ζ_{mle}.
  - Law enforcement agencies do not want "X Marks the Spot".
  - A search area, rather than a point estimate is far preferable.
- This should be possible with some Bayesian analysis

Questions?

Contact information:

Dr. Mike O'Leary

Director, Applied Mathematics Laboratory

Towson University

Towson, MD 21252

410-704-7457

moleary@towson.edu