Applying Mathematics....

... to catch criminals

Mike O'Leary

Department of Mathematics
Towson University

Stevenson University
Kappa Mu Epsion 2008

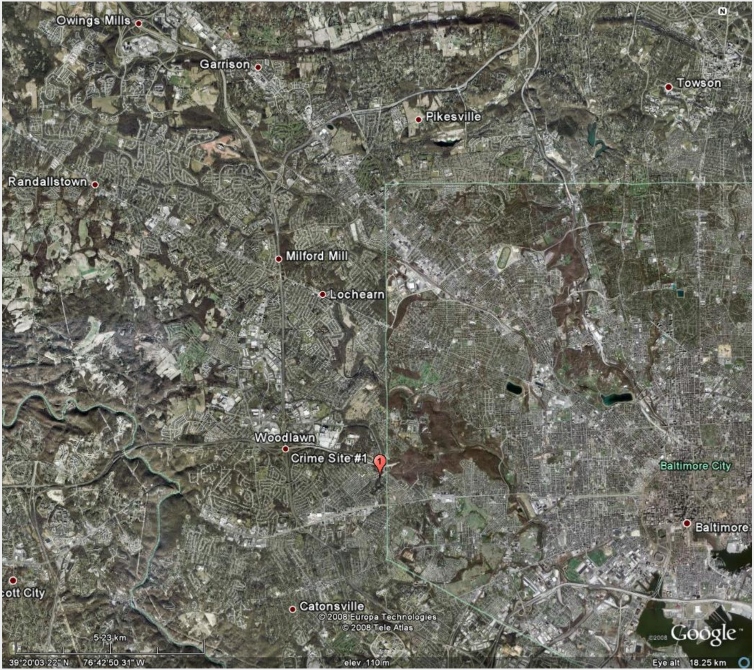
Acknowledgements

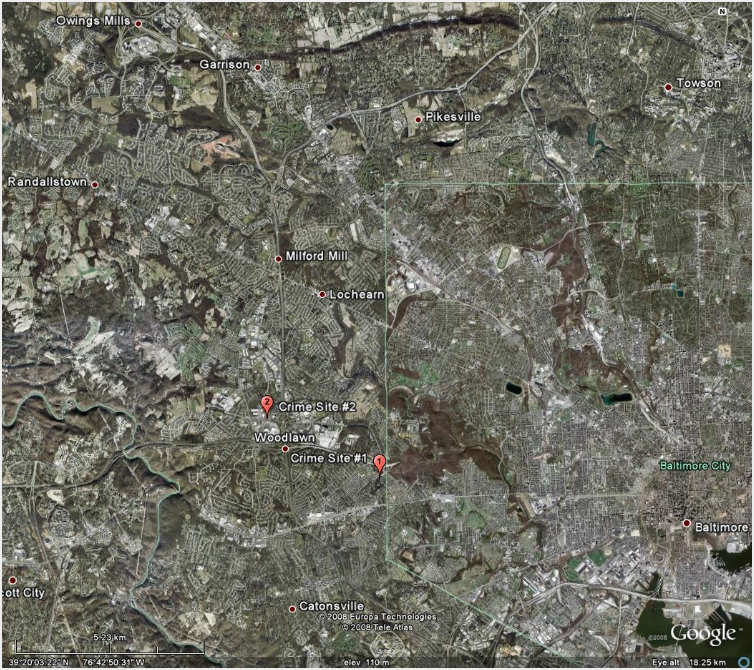
- Towson University Applied Mathematics Laboratory
  - Undergraduate research projects in mathematics sponsored by companies and government agencies
  - Student participants in 2005-06, 2006-07, and 2007-08:
    - Lauren Amrhine, Brooke Belcher, Brandie Biddy, Colleen Carrion, Chris Castillo, Paul Corbitt, Gregory Emerson, Adam Fojtik, Yu Fu, Natasha Gikunju, Laurel Mount, Kristopher Seets, Jonathan Vanderkolk, Grant Warble, Ruozhen Yao, Melissa Zimmerman
  - Faculty participants:
    - Coy L. May (Towson University) (2005-2007)
    - Andrew Engel (SAS) (2005)
- National Institute of Justice
  - Supported by 2005-IJ-CX-K036 and 2007-DE-BX-K005
  - Special thanks to Ron Wilson for his support
  - Thanks to Stanley Erickson, Iara Infosino, and Tommy Sexton
- Phil Canter, Baltimore County Police Department

The Geographic Profiling Problem

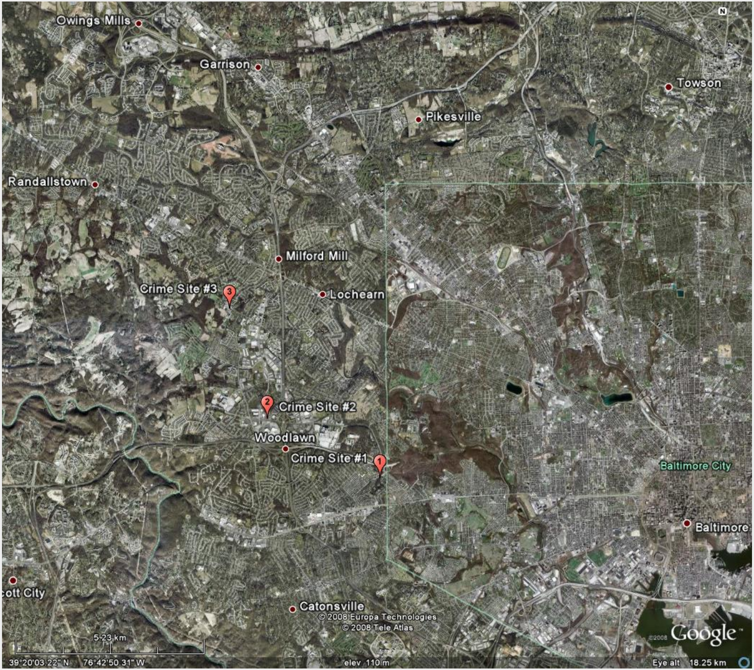
How can we estimate for the location of the anchor point of a serial offender from knowledge of the locations of the offender's crime sites?

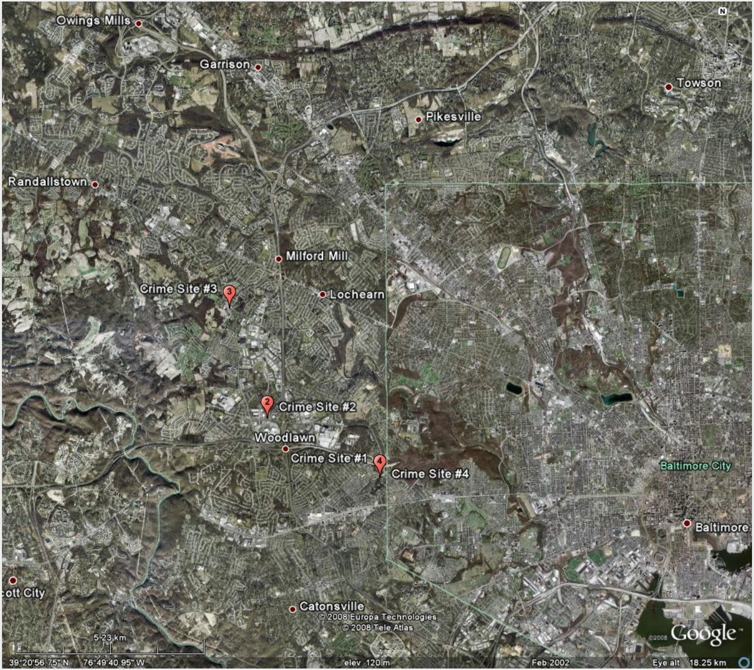
- The anchor point can be the offender's place of residence, place of work, or some other location important to the offender.

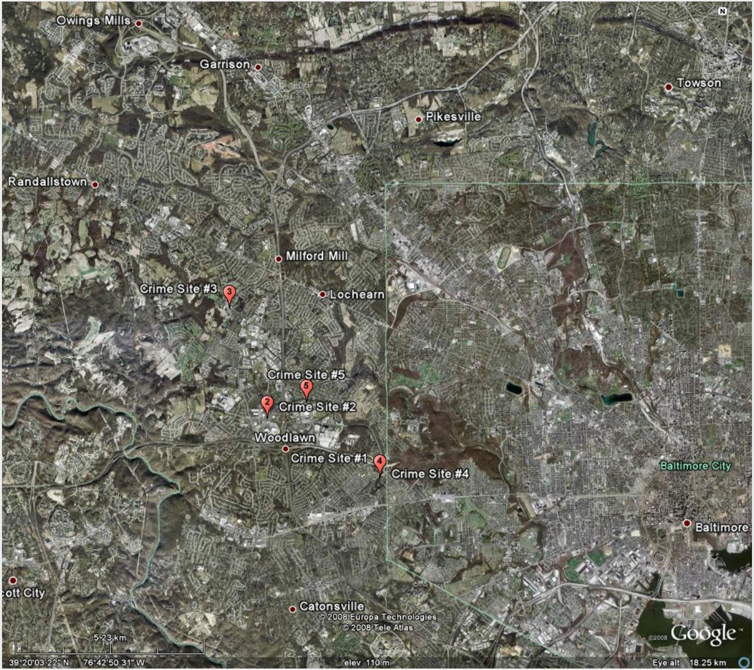

Owings Mills
Garrison
Towson
Pikesville
Randallstown
Milford Mill
Lochearn
Woodlawn
Crime Site #1
Baltimore City
Baltimore
ott City
Catonsville
© 2008 Europa Technologies
© 2008 Tele Atlas
5.23 km
39°20'03.22" N 76°42'50.31" W
elev 110 m
Eye alt 18.25 km
Google

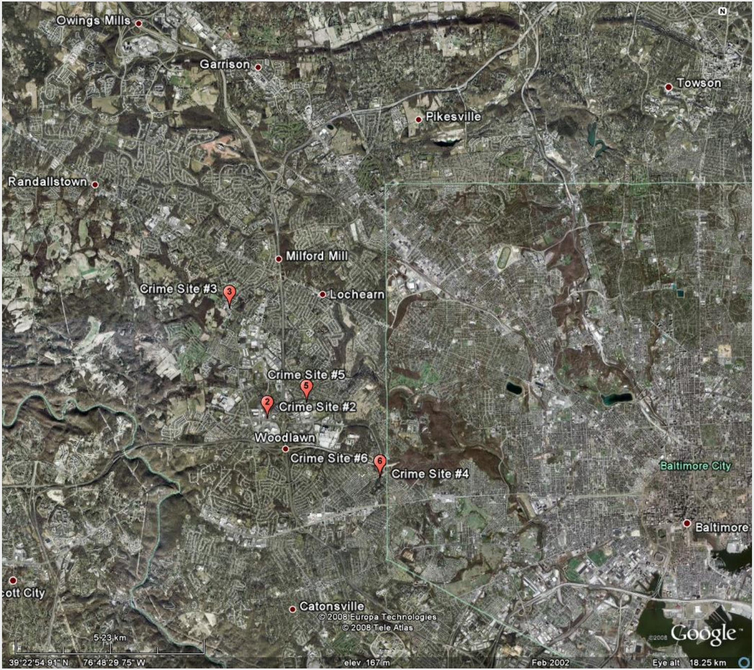


Owings Mills
Garrison
Towson
Pikesville
Randallstown
Milford Mill
Lochearn
Crime Site #2
Woodlawn
Crime Site #1
Baltimore City
Baltimore
ott City
Catonsville
© 2008 Europa Technologies
© 2008 Tele Atlas
5.23 km
39°20'03.22" N 76°42'50.31" W
elev 110 m
Eye alt 18.25 km
Google


Owings Mills
Garrison
Towson
Pikesville
Randallstown
Milford Mill
Crime Site #3
Lochearn
Crime Site #2
Woodlawn
Crime Site #1
Baltimore City
Baltimore
ott City
Catonsville
© 2008 Europa Technologies
© 2008 Tele Atlas
5.23 km
39°20'03.22" N 76°42'50.31" W
elev 110 m
Eye alt 18.25 km
Google


Owings Mills
Garrison
Towson
Pikesville
Randallstown
Milford Mill
Crime Site #3
Lochearn
Crime Site #2
Woodlawn
Crime Site #1
Crime Site #4
Baltimore City
Baltimore
ott City
Catonsville
© 2008 Europa Technologies
© 2008 Tele Atlas
5.23 km
39°20'56.75" N 76°49'40.95" W
elev 120 m
Feb 2002
Eye alt 18.25 km


Owings Mills
Garrison
Towson
Pikesville
Randallstown
Milford Mill
Crime Site #3
Lochearn
Crime Site #5
Crime Site #2
Woodlawn
Crime Site #1
Crime Site #4
Baltimore City
Baltimore
ott City
Catonsville
© 2008 Europa Technologies
© 2008 Tele Atlas
5.23 km
39°20'03.22" N 76°42'50.31" W
elev 110 m
Eye alt 18.25 km


Owings Mills
Garrison
Towson
Pikesville
Randallstown
Milford Mill
Crime Site #3
Lochearn
Crime Site #5
Crime Site #2
Woodlawn
Crime Site #6
Crime Site #4
Baltimore City
Baltimore
ott City
Catonsville
© 2008 Europa Technologies
© 2008 Tele Atlas
5.23 km
39°22'54.91" N 76°48'29.75" W
elev 167 m
Feb 2002
Eye alt 18.25 km
Google

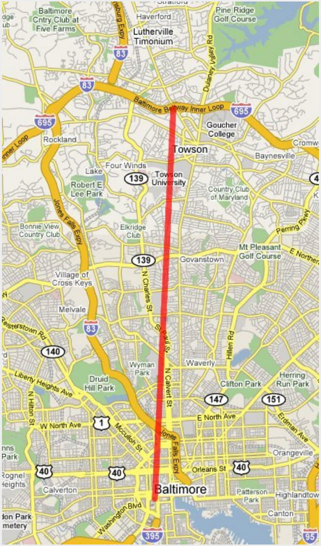
Developing a Model

To understand how we might proceed let us begin by adopting some common notation

- A point $\mathbf{x}$ will have two components $\mathbf{x} = (x^{(1)}, x^{(2)})$.
    - These can be latitude and longitude
    - These can be the distances from a pair of reference axes
- The series consists of n crimes at the locations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$
- The offender's anchor point will be denoted by $\mathbf{z}$.
- Distance between the points $\mathbf{x}$ and $\mathbf{y}$ will be $d(\mathbf{x}, \mathbf{y})$.

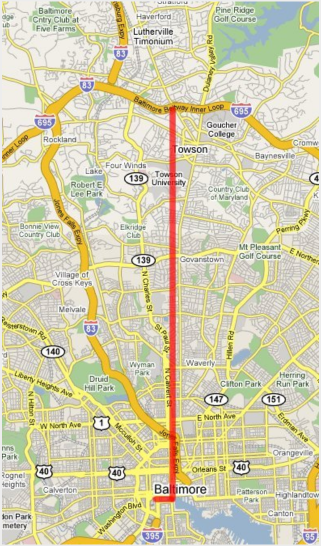
How should we measure distances?

- The Euclidean distance $d_2(\mathbf{x}, \mathbf{y}) = \sqrt{(x^{(1)} - y^{(1)})^2 + (x^{(2)} - y^{(2)})^2}$



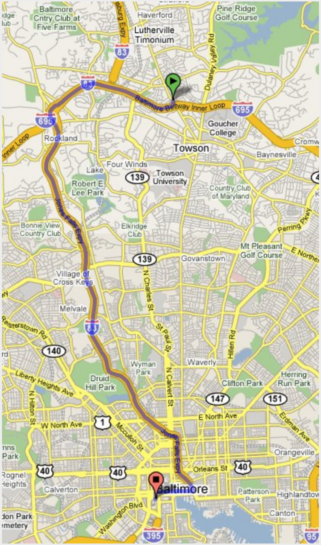
How should we measure distances?

- The Manhattan distance $d_1(\mathbf{x}, \mathbf{y}) = |x^{(1)} - y^{(1)}| + |x^{(2)} - y^{(2)}|$



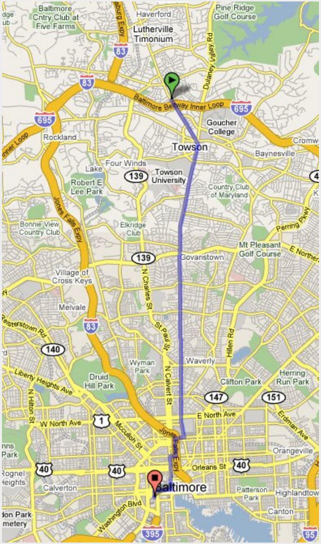
How should we measure distances?

- The highway distance?



How should we measure distances?

- The street distance?



A Mathematical Model

- Suppose that we know nothing about the offender, only that the offender chooses to offend at the location x with probability density $P(\mathbf{x})$.
  - The probability density does not mean that the offender chooses randomly (though he may), rather we are modeling our lack of complete information about the offender.
  - Probabilistic models are common in modeling deterministic phenomena, including
    - The stock market
    - Population dynamics
    - Genetics
    - Epidemiology
    - Heat flow

A Mathematical Model

- On what variables should the probability density $\mathrm{P}(\mathbf{x})$ depend?
  - The anchor point $\mathbf{z}$ of the offender
    - Each offender needs to have a unique anchor point
    - The anchor point must have a well-defined meaning- *e.g.* the offender's place of residence
    - The anchor point needs to be stable during the crime series
  - The average distance α the offender is willing to travel from their anchor point
    - Different offender's have different levels of mobility- an offender will need to travel farther to commit some types of crimes (*e.g.* liquor store robberies, bank robberies) than others (*e.g.* residential burglaries)
    - This varies between offenders
    - This varies between crime types
  - Other variables can be included
- We are left with the assumption that an offender with anchor point $\mathbf{z}$ and mean offense distance α commits an offense at the location $\mathbf{x}$ with probability density $\mathrm{P}(\mathbf{x} \,|\, \mathbf{z}, \alpha)$

A Mathematical Model

- Our mathematical problem then becomes the following:
  - Given a sample $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ (the crime sites) from a probability distribution $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$, estimate the parameter $\mathbf{z}$ (the anchor point).
- This is a well-studied mathematical problem
- One approach is the theory of *maximum likelihood*.
  - Construct the likelihood function

$$L(\mathbf{y}, a) = \prod_{i=1}^{n} P(\mathbf{x}_i \,|\, \mathbf{y}, a) = P(\mathbf{x}_1 \,|\, \mathbf{y}, a) \cdots P(\mathbf{x}_n \,|\, \mathbf{y}, a)$$

  - Then the best choice of $\mathbf{z}$ is the choice of $\mathbf{y}$ that makes the likelihood as large as possible.
  - This is equivalent to maximizing the log-likelihood

$$\lambda(\mathbf{y}, a) = \sum_{i=1}^{n} \ln P(\mathbf{x}_i \,|\, \mathbf{y}, a) = \ln P(\mathbf{x}_1 \,|\, \mathbf{y}, a) + \cdots + \ln P(\mathbf{x}_n \,|\, \mathbf{y}, a)$$

Bayesian Analysis

- Suppose that there is only one crime site $\mathbf{x}$. Then Bayes' Theorem implies that

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}) = \frac{P(\mathbf{x} \,|\, \mathbf{z}, \alpha)\pi(\mathbf{z}, \alpha)}{P(\mathbf{x})}$$

  - $P(\mathbf{z}, \alpha \,|\, \mathbf{x})$ is the *posterior* distribution
    - It gives the probability density that the offender has anchor point $\mathbf{z}$ and the average offense distance α, given that the offender has committed a crime at $\mathbf{x}$
  - $\pi(\mathbf{z}, \alpha)$ is the *prior* distribution.
    - It represents our knowledge of the probability density for the anchor point $\mathbf{z}$ and the average offense distance α before we incorporate information about the crime
    - If we assume that the choice of anchor point is independent of the average offense distance, we can write

$$\pi(\mathbf{z}, \alpha) = H(\mathbf{z})\pi(\alpha)$$

      where $H(\mathbf{z})$ is the prior distribution of anchor points, and $\pi(\alpha)$ is the prior distribution of average offense distances
  - $P(\mathbf{x})$ is the *marginal* distribution

Bayesian Analysis

- A similar analysis holds when there is a series of n crimes; in this case
$$P(\mathbf{z},\alpha\,|\,\mathbf{x}_1,\ldots,\mathbf{x}_n) = \frac{P(\mathbf{x}_1,\ldots,\mathbf{x}_n\,|\,\mathbf{z},\alpha)\pi(\mathbf{z},\alpha)}{P(\mathbf{x}_1,\ldots,\mathbf{x}_n)}.$$
- If we assume that the offender's choice of crime sites are mutually independent, then
$$P(\mathbf{x}_1,\ldots,\mathbf{x}_n\,|\,\mathbf{z},\alpha) = P(\mathbf{x}_1\,|\,\mathbf{z},\alpha)\cdots P(\mathbf{x}_n\,|\,\mathbf{z},\alpha)$$
giving us the relationship
$$P(\mathbf{z},\alpha\,|\,\mathbf{x}_1,\ldots,\mathbf{x}_n) \propto P(\mathbf{x}_1\,|\,\mathbf{z},\alpha)\cdots P(\mathbf{x}_n\,|\,\mathbf{z},\alpha)H(\mathbf{z})\pi(\alpha).$$
- Because we are only interested in the location of the anchor point, we take the conditional distribution with respect to α to obtain the following

Fundamental Theorem of Geographic Profiling

Suppose that an unknown offender has committed crimes at $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, and that

- The offender has a unique stable anchor point $\mathbf{z}$
- The offender chooses targets to offend according to the probability density $P(\mathbf{x} \mid \mathbf{z}, \alpha)$ where α is the average distance the offender is willing to travel
- The target locations in the series are chosen independently
- The prior distribution of anchor points is $H(\mathbf{z})$, the prior distribution of the average offense distance is $\pi(\alpha)$ and these are independent of one another.

Then the probability density that the offender has anchor point at the location $\mathbf{z}$ satisfies

$$P(\mathbf{z} \mid \mathbf{x}_1, \dots, \mathbf{x}_n) \propto \int_0^\infty P(\mathbf{x}_1 \mid \mathbf{z}, \alpha) \cdots P(\mathbf{x}_n \mid \mathbf{z}, \alpha) H(\mathbf{z}) \pi(\alpha) \, d\alpha$$

Remarks

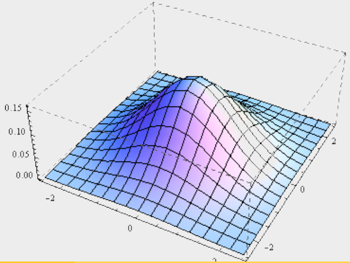
1. The framework is independent of the significance of the anchor point $\mathbf{z}$
2. This framework holds for any model of offender behavior $P(\mathbf{x} \mid \mathbf{z}, \alpha)$
3. This framework holds for any choice of prior distributions $H(\mathbf{z})$ and $\pi(\alpha)$
4. The framework is independent of the choice of distance metric
5. Geographic features that affect crime selection can be incorporated into the form of $P(\mathbf{x} \mid \mathbf{z}, \alpha)$
6. Geographic features that affect the selection of anchor points are incorporated into the form of $H(\mathbf{z})$
7. The framework provides a prioritized search area; the framework estimates $P(\mathbf{z} \mid \mathbf{x}_1, \ldots, \mathbf{x}_n)$ which is the probability density for the offender's anchor point; by definition locations where $P(\mathbf{z} \mid \mathbf{x}_1, \ldots, \mathbf{x}_n)$ are larger are more likely to contain the anchor point than regions where it is smaller.

Using the Fundamental Theorem

- For the mathematics to be useful, we need to be able to:
  - Make some reasonable choice for our model for offender behavior
  - Make some reasonable choice for the prior distribution of anchor points
  - Make some reasonable choice for the prior distribution of the average offense distance, and
  - Be able to evaluate the mathematical terms that appear

Models of Offender Behavior

- Suppose that we assume that offenders choose offense sites according to a normal distribution, so that
$$P(\mathbf{x} \mid \mathbf{z}, \alpha) = \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4\alpha^2}|\mathbf{x}-\mathbf{z}|^2\right).$$
- If we also assume that all offenders have the same average offense distance α, and that all anchor points are equally likely, then
$$P(\mathbf{z} \mid \mathbf{x}_1, \ldots, \mathbf{x}_n) = \left(\frac{1}{4\alpha^2}\right)^n \exp\left(-\frac{\pi}{4\alpha^2}\sum_{i=1}^{n}|\mathbf{x}_i-\mathbf{z}|^2\right).$$
  - The mode of this distribution- the point most likely to be the offender's anchor point- is the mean center of the crime site locations.



Models of Offender Behavior

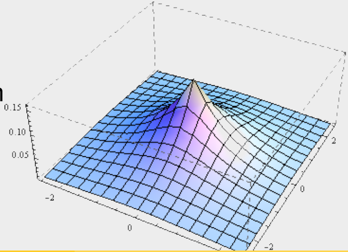
- Suppose that we assume that offenders choose offense sites according to a negative exponential distribution, so that

$$P(\mathbf{x} \mid \mathbf{z}, \alpha) = \frac{2}{\pi\alpha^2} \exp\left(-\frac{2}{\alpha}|\mathbf{x} - \mathbf{z}|\right).$$

- If we also assume that all offenders have the same average offense distance α, and that all anchor points are equally likely, then

$$P(\mathbf{z} \mid \mathbf{x}_1, \dots, \mathbf{x}_n) = \left(\frac{2}{\pi\alpha^2}\right)^n \exp\left(-\frac{2}{\alpha}\sum_{i=1}^{n} |\mathbf{x}_i - \mathbf{z}|\right)$$

  - The mode of this distribution- the point most likely to be the offender's anchor point- is the center of minimum distance of the crime site locations.



Models of Offender Behavior

- What would a more realistic model for offender behavior look like?
  - Consider a model in the form

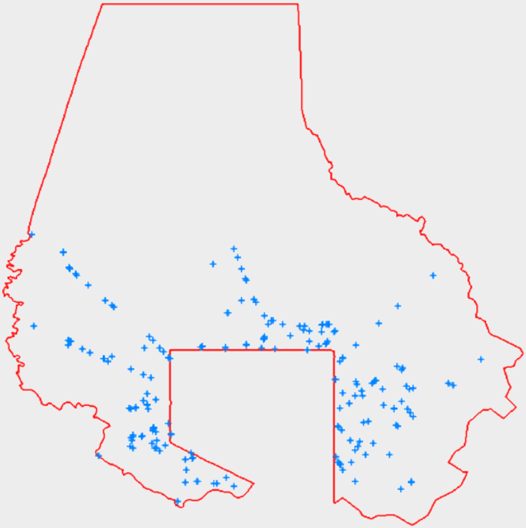
$$P(\mathbf{x} \,|\, \mathbf{z}, \alpha) = D(d(\mathbf{x}, \mathbf{z}), \alpha) \cdot G(\mathbf{x}) \cdot N(\mathbf{z})$$

  - D models the effect of distance decay using the distance metric $d(\mathbf{x}, \mathbf{z})$
    - We can specify a normal decay, so that $D(d, \alpha) = \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4\alpha^2} d^2\right)$
    - We can specify a negative exponential decay, so that $D(d, \alpha) = \frac{2}{\pi\alpha^2} \exp\left(-\frac{2}{\alpha} d\right)$
    - Any choice can be made for the distance metric (Euclidean, Manhattan, *et.al*)
  - G models the geographic features that influence crime site selection
    - High values for $G(\mathbf{x})$ indicate that $\mathbf{x}$ is a likely target for typical offenders;
    - Low values for $G(\mathbf{x})$ indicate that $\mathbf{x}$ is a less likely target
  - N is a normalization factor, required to ensure that P is a probability distribution
    - $N(\mathbf{z}) = \left[\iint D(d(\mathbf{y}, \mathbf{z}), \alpha) G(\mathbf{y}) dy^{(1)} dy^{(2)}\right]^{-1}$
    - N is completely determined by the choices for D and G.

Geographic Features that Influence Crime Selection

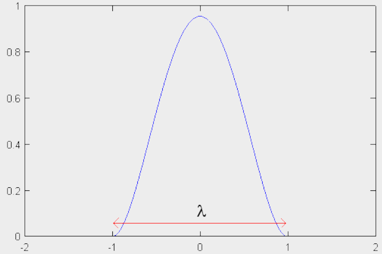
- G models the geographic features that influence crime site selection, with high values indicating the location was more likely to be targeted by an offender.
- How can we calculate G?
  - Use available geographic and demographic data and the correlations between crime rates and these variables that have already been published to construct an appropriate choice for $G(\mathbf{x})$
    - Different crime types have different etiologies; in particular their relationship to the local geographic and demographic backcloth depends strongly on the particular type of crime. This would limit the method to only those crimes where this relationship has been well studied
  - Some crimes can only occur at certain, well-known locations, which are known to law enforcement
    - For example, gas station robberies, ATM robberies, bank robberies, liquor store robberies
    - This does not apply to all crime types- *e.g.* street robberies, vehicle thefts.
  - We can assume that historical crime patterns are good predictors of the likelihood that a particular location will be the site of a crime.

Convenience Store Robberies, Baltimore County



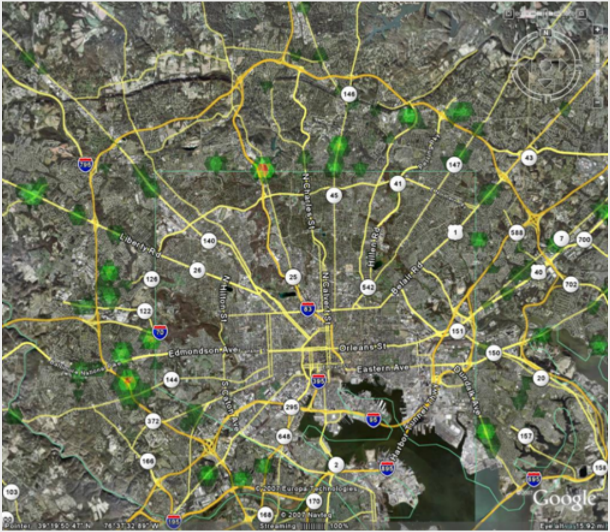
Geographic Features that Influence Crime Selection

- Suppose that historical crimes have occurred at the locations $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N$.
- Choose a kernel density function $K(y \mid \lambda)$
  - λ is the bandwidth of the kernel density function



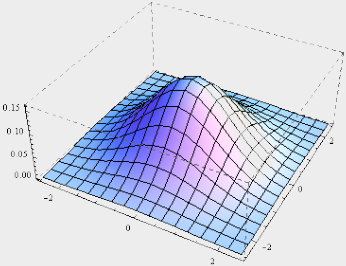
- Calculate $G(\mathbf{x}) = \sum_{i=1}^{N} K(d(\mathbf{x}, \mathbf{c}_i) \mid \lambda)$
  - The bandwidth λ can be *e.g.* the mean nearest neighbor distance
  - Effectively this places a copy of the kernel density function on each crime site and sums

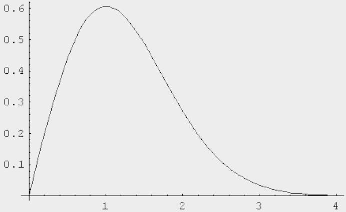
Convenience Store Robberies, Baltimore County



Distance Decay

- To understand distance decay we begin by examining the idea of a buffer zone
  - A buffer zone is a region around the offender's anchor point where they are less likely to offend, presumably due to a fear of being recognized.
  - Consider the following models of offender behavior:

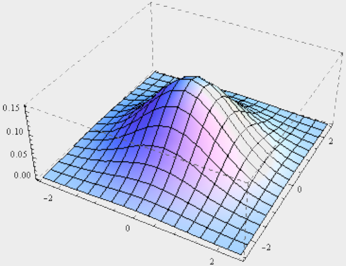


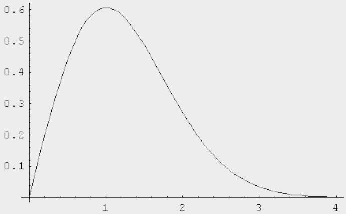


  - Which shows evidence of a buffer zone?

Distance Decay

- To understand distance decay we begin by examining the idea of a buffer zone
  - A buffer zone is a region around the offender's anchor point where they are less likely to offend, presumably due to a fear of being recognized.
  - Consider the following models of offender behavior:

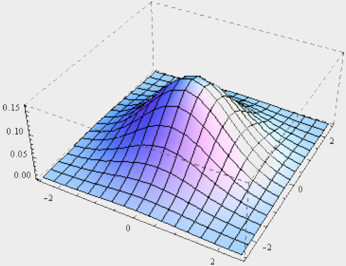


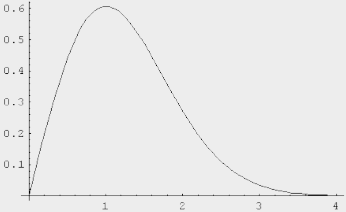


  - Which shows evidence of a buffer zone?
    - These are two views of the same distribution

Distance Decay

- To understand distance decay we begin by examining the idea of a buffer zone
  - A buffer zone is a region around the offender's anchor point where they are less likely to offend, presumably due to a fear of being recognized.
  - Consider the following models of offender behavior:





  - Which shows evidence of a buffer zone?
    - These are two views of the same distribution
    - If the offender is using a two-dimensional normal distribution to select targets, then the appropriate distribution for the offense distance is the *Rayleigh* distribution.

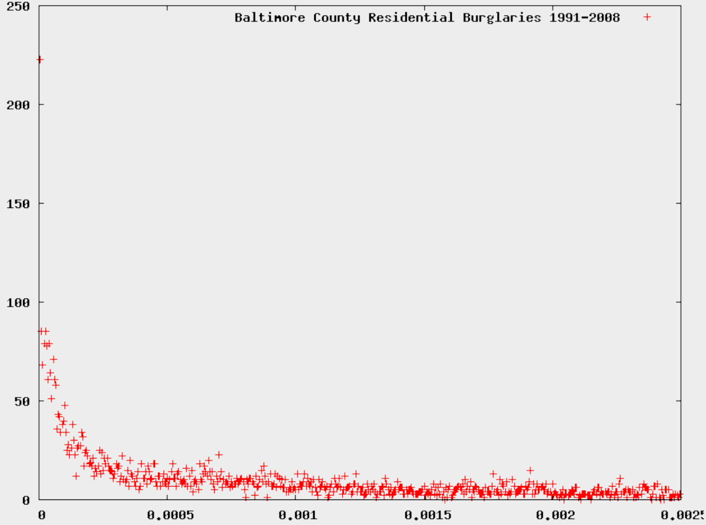
Distance Decay

- Suppose that the (two-dimensional) distance decay component $D(d(\mathbf{x}, \mathbf{z}) \mid \alpha)$ is modeled with a Euclidean distance d
- Then the (one-dimensional) distribution of offense distances $D_{\text{one-dim}}(d \mid \alpha)$ is given by

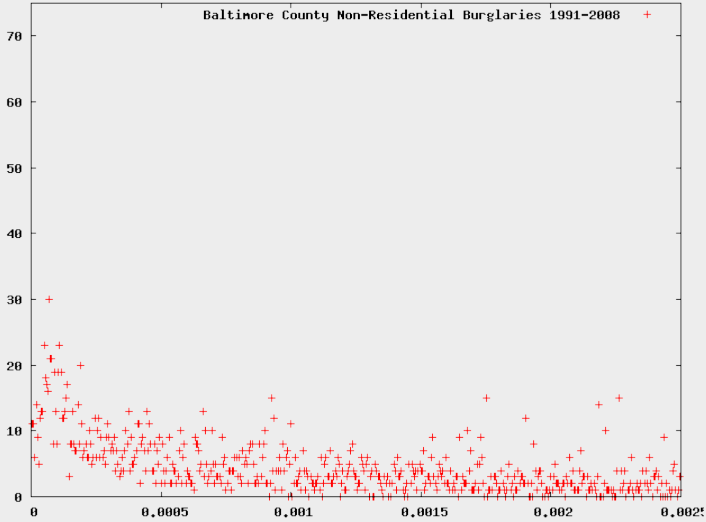
$$D_{\text{one-dim}}(d \mid \alpha) = 2\pi d \cdot D(d \mid \alpha)$$

- In particular, $D_{\text{one-dim}}(d \mid \alpha) \to 0$ as $d \to 0$, regardless of the particular choice of $D(d \mid \alpha)$, provided $D(0 \mid \alpha) < \infty$.

Distance Decay: Residential Burglaries in Baltimore County



Distance Decay: Non-Residential Burglaries in Baltimore County



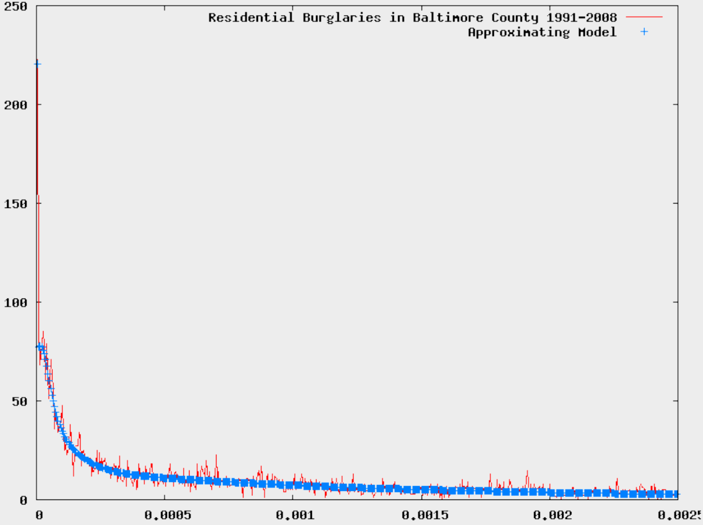
Distance Decay: Data Fitting

- Suppose that we measure the aggregate number of offenders who commit a crime at a distance d from their anchor point; call the relative fraction $A(d)$.
- Different offenders are willing to travel different distances to offend; $\pi(\alpha)$ was defined to be the probability distribution for the average offense distance across offenders.
- Suppose that each offender chooses targets according to $D_{\text{one-dim}}(d \mid \alpha)$
- Then

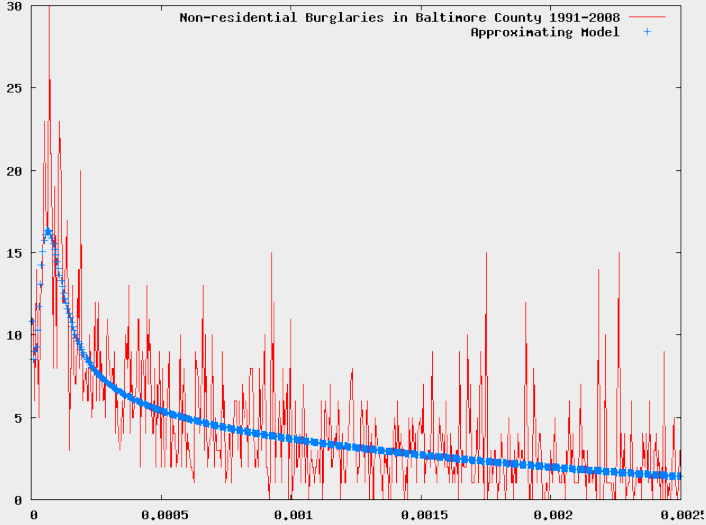
$$A(d) = \int_0^\infty D_{\text{one-dim}}(d \mid \alpha)\pi(\alpha)\,d\alpha$$

- Since $A(d)$ can be measured and $D_{\text{one-dim}}(d \mid \alpha)$ modeled, we can solve this equation for the prior $\pi(\alpha)$
  - This is a mathematically difficult process

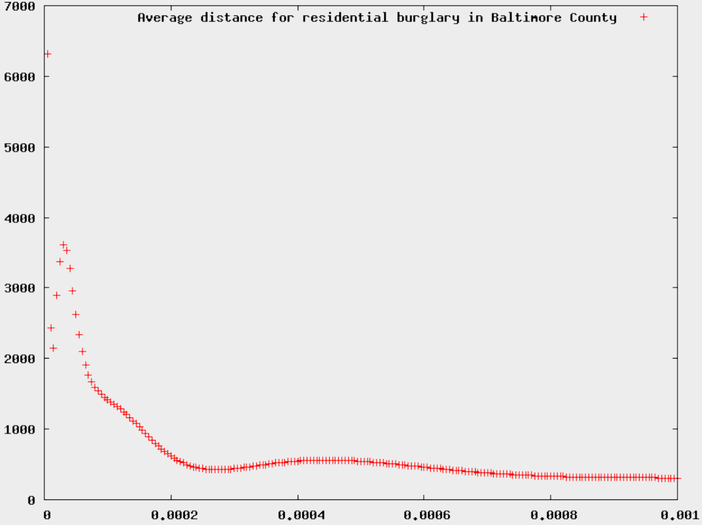
Distance Decay: Residential Burglaries in Baltimore County- Model Fit



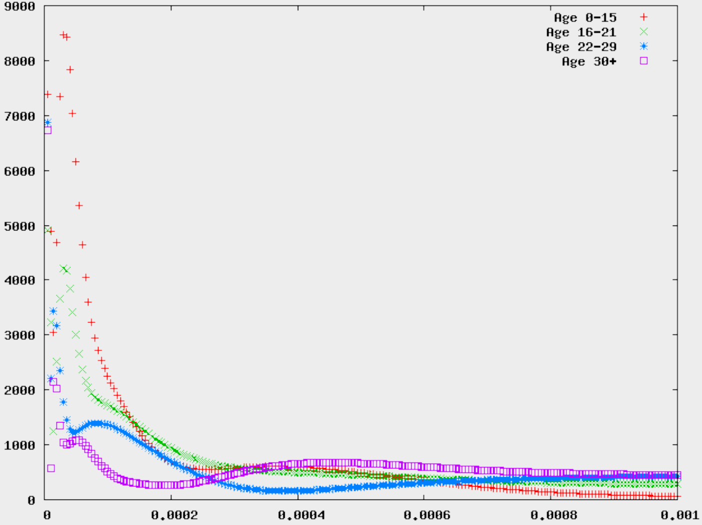
Distance Decay: Non-Residential Burglaries in Baltimore County- Model Fit



Distance Decay: Residential Burglaries in Baltimore County- Prior Distribution



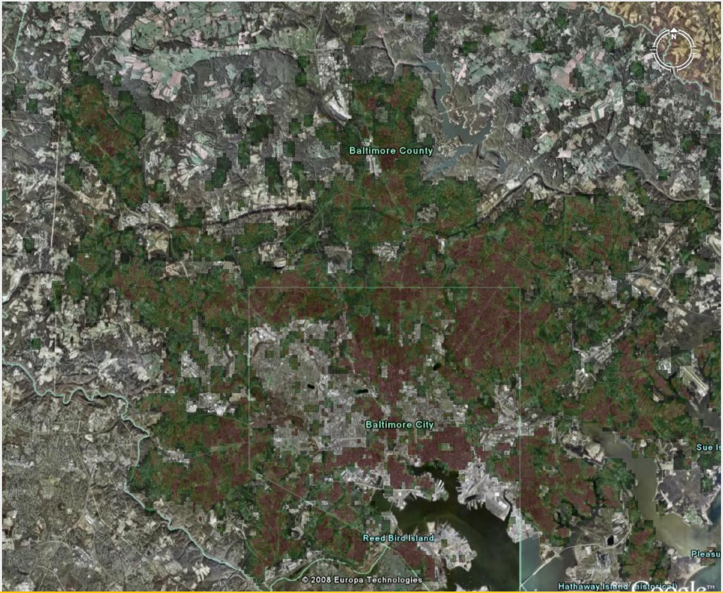
Distance Decay: Residential Burglaries in Baltimore County- Prior Distribution by Age



Anchor Points

- We have assumed
  - Each offender has a unique, well-defined anchor point that is stable throughout the crime series
  - The function $H(\mathbf{z})$ represents our prior knowledge of the distribution of anchor points before we incorporate information about the crime series.
- What are reasonable choices for the anchor point?
  - Residences
  - Places of work
- Suppose that anchor points are residences- can we estimate $H(\mathbf{z})$?
  - Population density information is available from the U.S. Census at the block level, sorted by age, sex, and race/ethnic group.
    - We can use available demographic information about the offender
    - Set $H(\mathbf{z}) = \sum_{i=1}^{N_{\text{blocks}}} = p_i K(\mathbf{z} - \mathbf{q}_i \mid \sqrt{A_i})$
    - Here block i has population p_i, center $\mathbf{q}_i$, and area A_i.
  - Distribution of residences of past offenders can be used.
    - Calculate $H(\mathbf{z})$ using the same techniques used to calculate $G(\mathbf{x})$

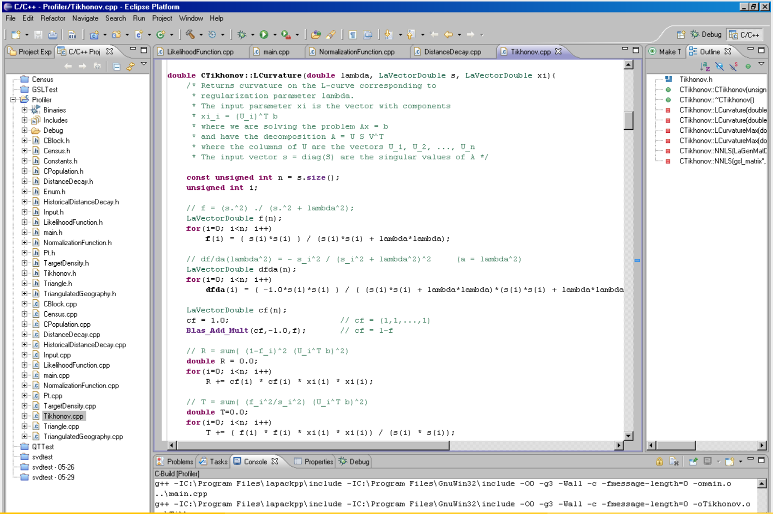
Baltimore 18-29 year old white men



Baltimore 18-29 year old black men

Software

- Code that implements this method is nearing completion, and will be released to police agencies.



Questions?

- Any questions?