Bayesian Journey to Crime Modeling:
Improvements to Geographic Profiling Methodology

Ned Levine, PhD
Ned Levine & Associates
Houston, TX

Richard Block, PhD
Loyola University
Chicago, IL

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

Purpose:

- Describe the Bayesian Approach to Journey to Crime (JTC) Modeling
- Compare its Accuracy to Existing Methods
- Illustrate its use in CrimeStat III (ver 3.1)

Current Jtc Methodology

- Utilizes travel demand (distance) function
- Applies function to incidents committed by serial offender to produce density estimate
- Sums densities across all incidents
- Interpolates results to grid framework

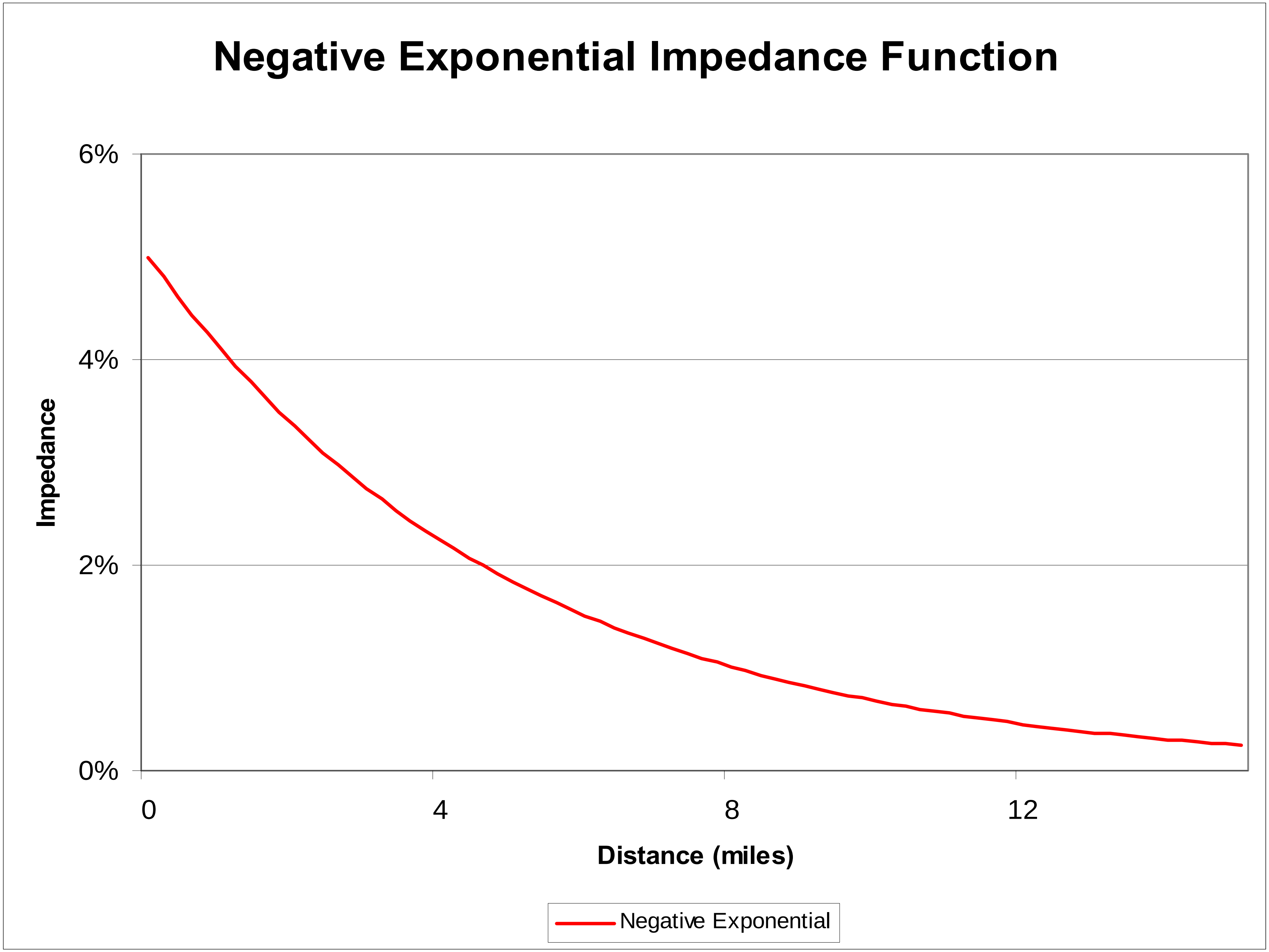
Existing Software Use Different Travel Demand Functions

- Type of mathematical function

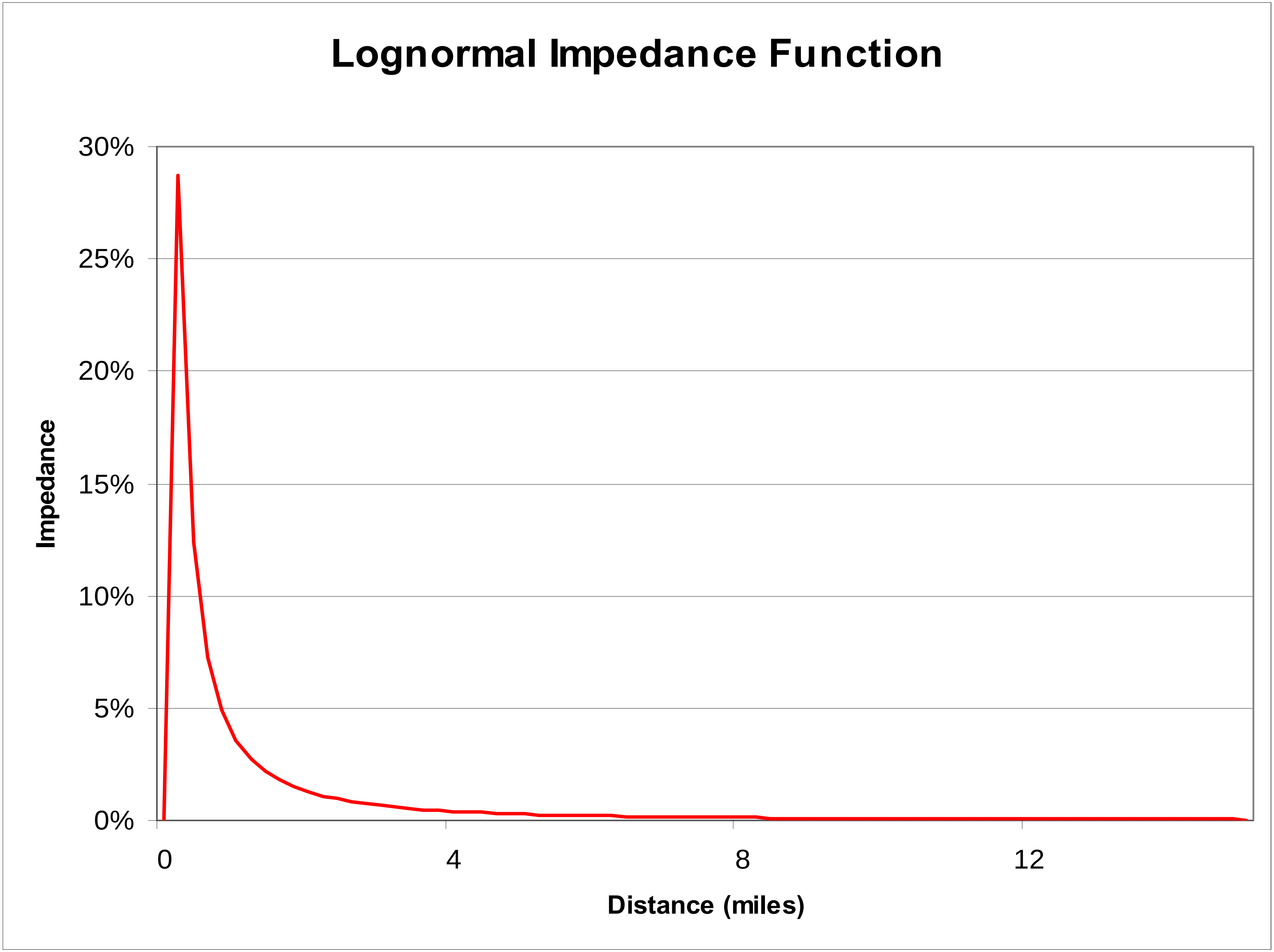
  Inverse distance
  Negative exponential
  Lognormal
  Single v. mixed mathematical distributions

- Fixed v. variable parameters in mathematical functions

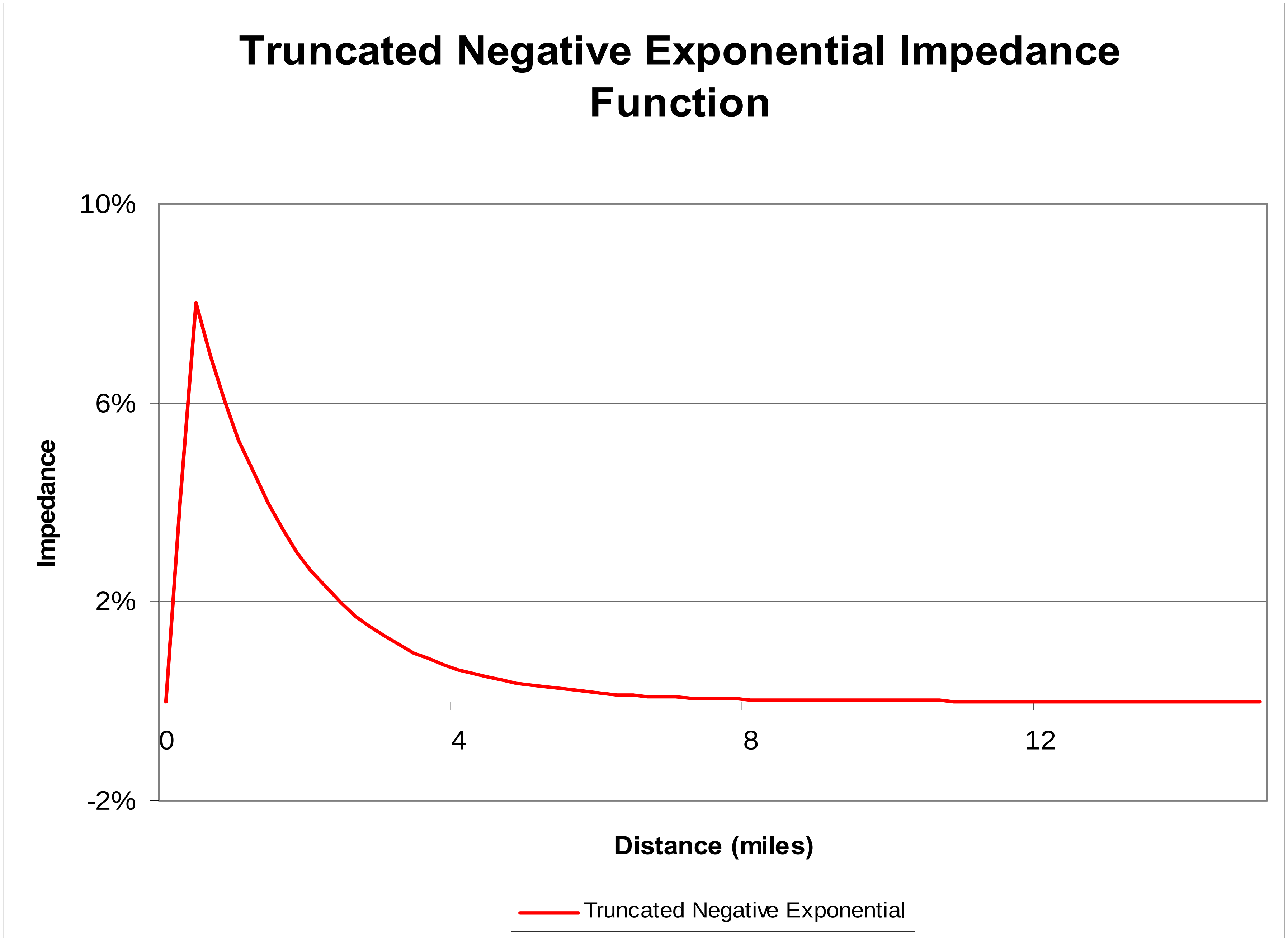
- Mathematical v. empirically-derived function

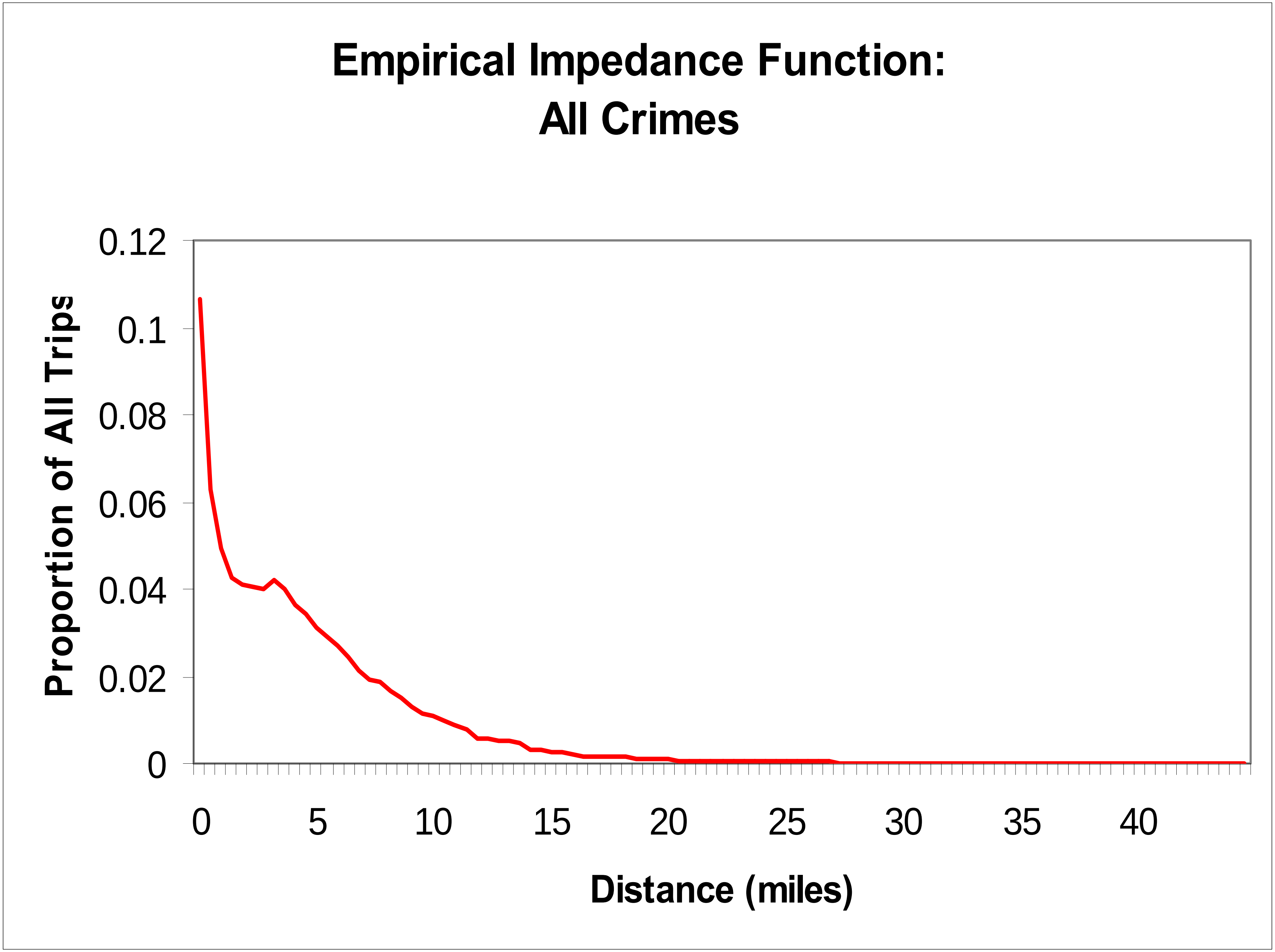


Negative Exponential Impedance Function
6%
4%
2%
0%
Impedance
0
4
8
12
Distance (miles)
Negative Exponential

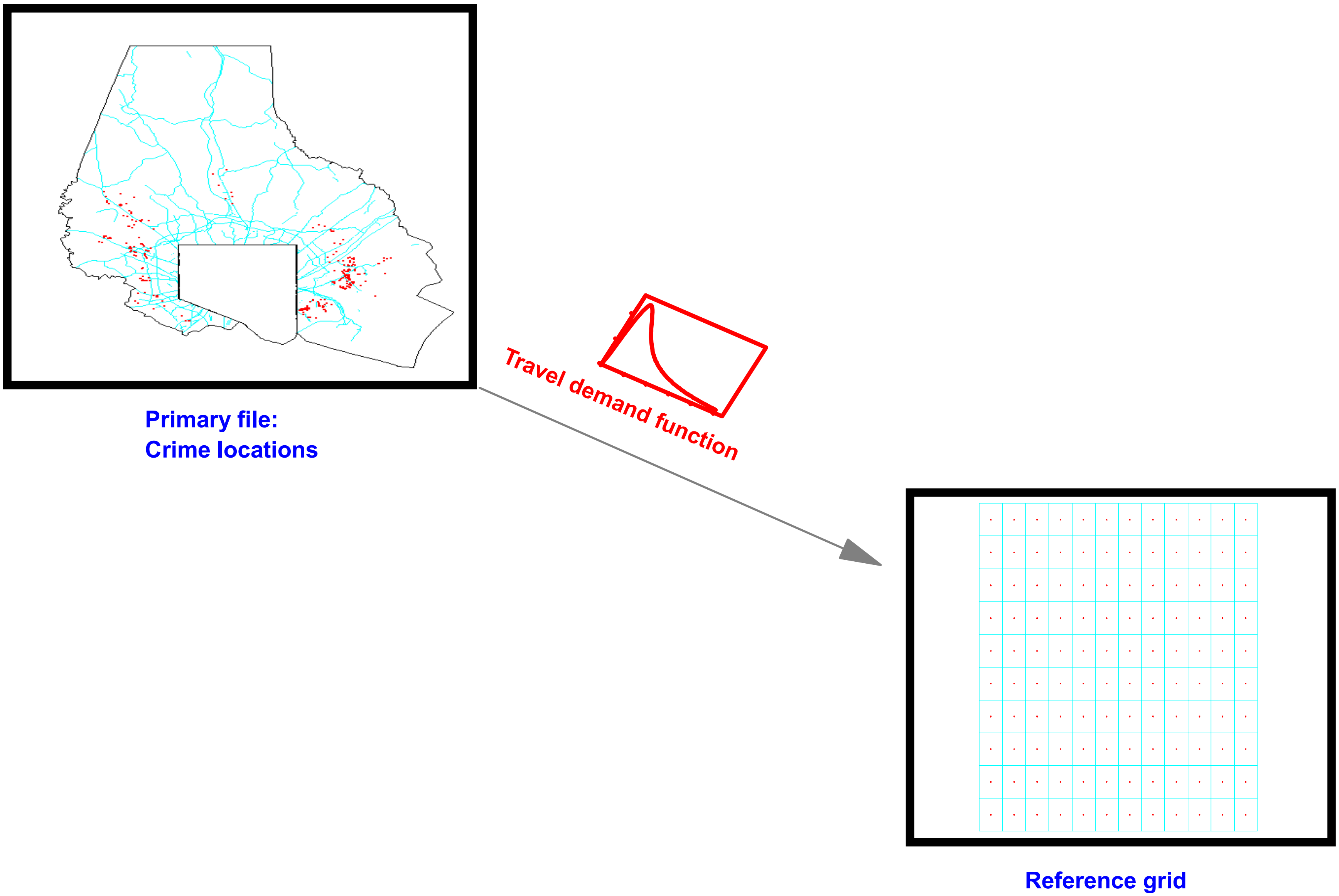


Lognormal Impedance Function
30%
25%
20%
15%
10%
5%
0%
Impedance
0
4
8
12
Distance (miles)


Truncated Negative Exponential Impedance Function
10%
6%
2%
-2%
Impedance
0
4
8
12
Distance (miles)
Truncated Negative Exponential


Empirical Impedance Function:
All Crimes
Proportion of All Trips
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
30
35
40
Distance (miles)

Journey to Crime Interpolation Routine

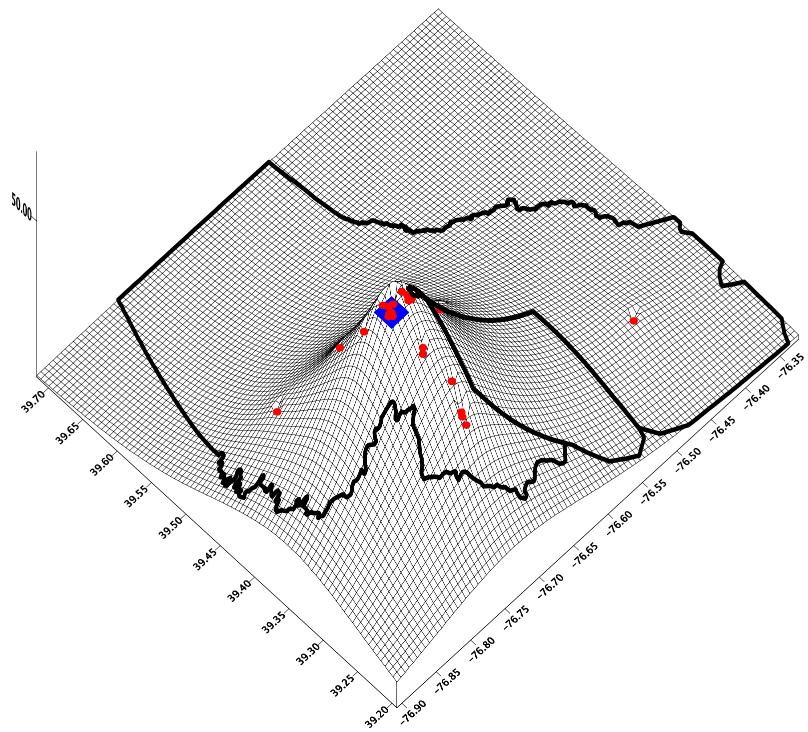


Predicted and Actual Location of Serial Thief

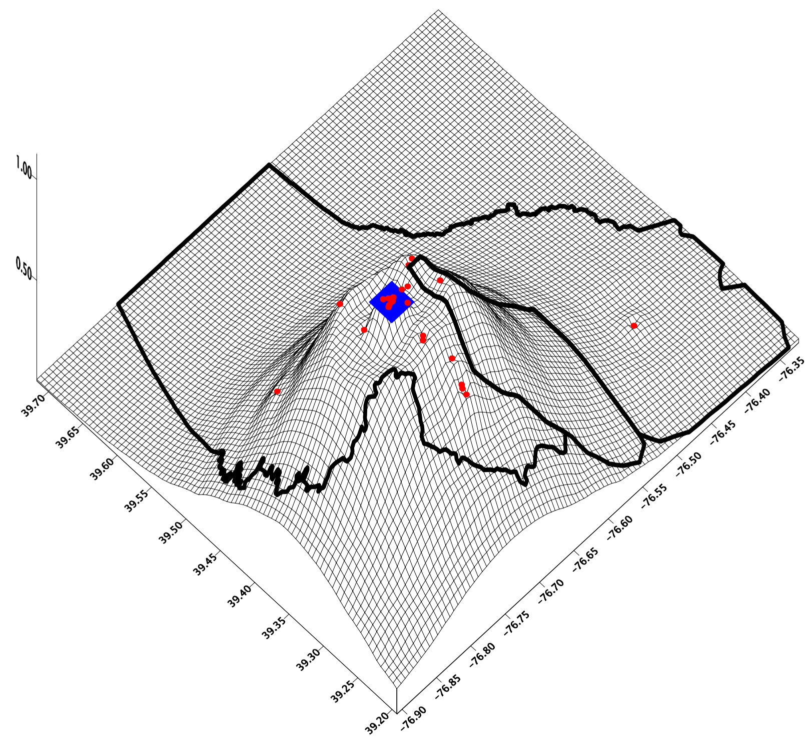
Man Charged with 24 Offenses in Baltimore County

Predicted with Mathematical and Kernel Density Models for Larceny

Residence Location = Square
Crime Locations = Circles



Mathematical Model:
Truncated Negative Exponential



Kernel Density Model

Limitations of JTC Methodology

- Will always locate highest probability within convex hull of incidents.

- Cell with highest probability is approximated by Center of Minimum Distance (CMD)

- Travel demand function used is invariant. Does not distinguish between:

  Types of crime
  Types of offender
  Sub-regions of the study area
  Directionality in travel
  Time periods (e.g., 5 pm v. 3 am)

Distance is Not an Independent Variable

- Travel distance (or travel time) is the result of predispositions by offenders, opportunities, and the travel network
- People only set loose limits on travel distance/time
- Increasing mobility of American society has made automobiles almost universally available and travel very cheap

Bayesian Approach to JTC Modeling

- Incremental improvement to Journey to crime modeling
- Adds new information to update travel distance estimates
- Implicitly weights the travel distance by predispositions, opportunities, and the travel network

Bayes Theorem

- Relates marginal and conditional probabilities of two events together

- Marginal probability is the probability of an event independent of any other event

  P(A) and P(B) are two marginal probabilities

- Conditional probability is the probability of an event given that some other event has occurred

  P(A|B) is the probability of A given that B has occurred while P(B|A) is the probability of B given that A has occurred

Bayes Theorem (continued)

Relates two 'AND' conditions together:

$$P(A \text{ and } B) = P(A) * P(B|A) = P(B) * P(A|B)$$

Thus:

$$P(B|A) = \frac{P(B) * P(A|B)}{P(A)}$$

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$

Bayesian Inference

- In statistical interpretation of Bayes Theorem, probabilities are estimates of random variables

- Let θ be a parameter and let X be some data

$$P(\theta|X) = \frac{P(X|\theta) * P(\theta)}{P(X)}$$

where:

P(θ) is probability that θ has certain distribution
P(X|θ) is probability that data would have been obtained if θ is true
P(X) is the probability of obtaining the data

Bayesian Inference (continued)

Logically:

$$\text{Probability that } \theta \text{ is true given the data, } X = \frac{\text{Likelihood of obtaining the data, X, given } \theta \text{ is true} \;\; * \;\; \text{Prior probability of } \theta}{\text{Prior probability of the data, X}}$$

Since it's difficult to estimate the probability of obtaining the data under all circumstances:

$$\text{Probability that } \theta \text{ is true given the data, } X \approx \text{Likelihood of obtaining the data, X, given } \theta \text{ is true} \;\; * \;\; \text{Prior probability of } \theta$$

Application to Journey to Crime Estimation

- Update JTC estimate with information on where other offenders lived who committed crimes in same locations

- Sees behavior of offenders as being a mixture of:

    1. Unique tendencies

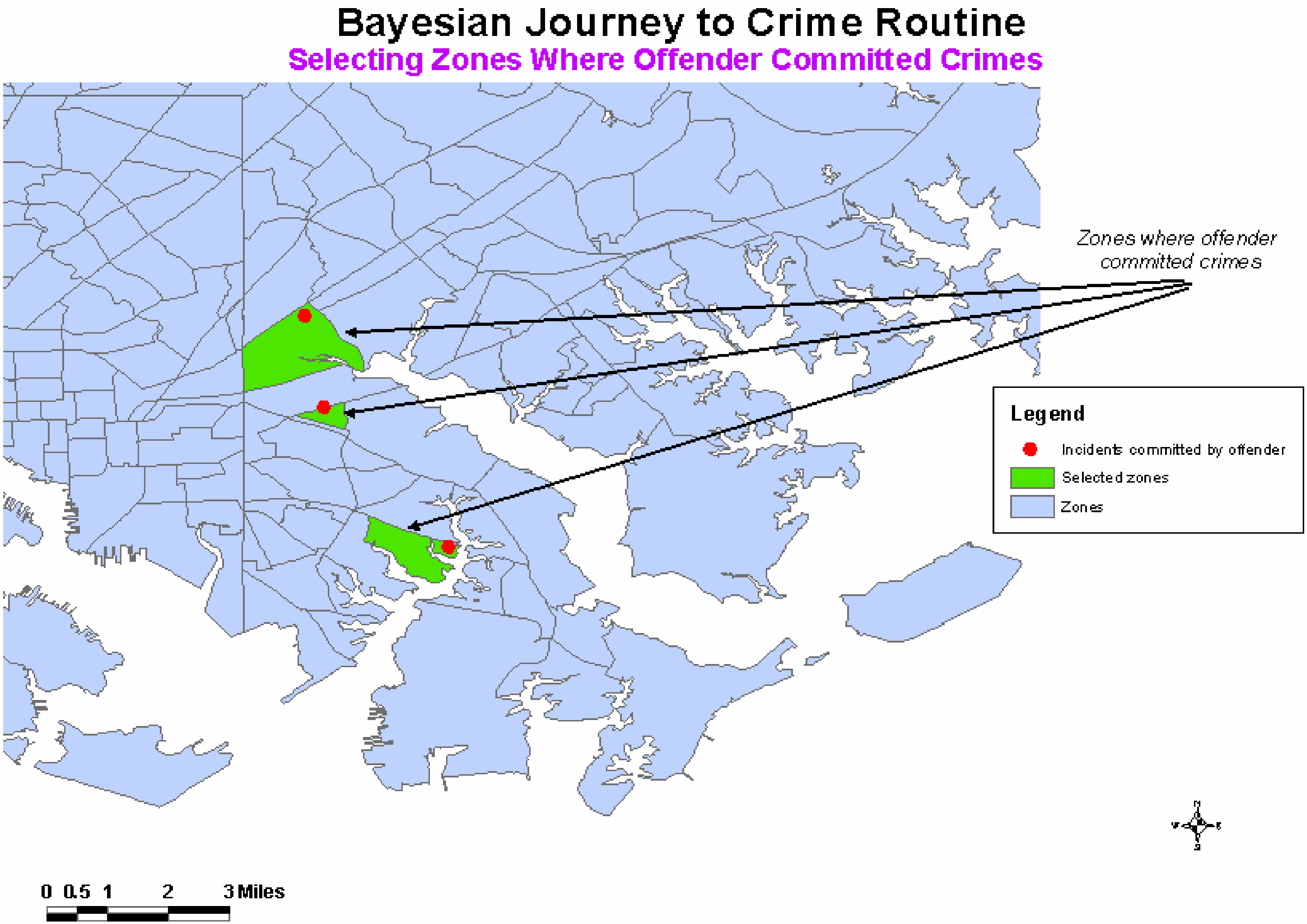
    2. The crime attractions associated with other offenders

Crime Origin-Destination Matrix

Crime destination zone

Crime origin zone

	1	2	3	4	5		N	Σ
1	**37**	15	21	4	3	. . . . . .	12	346
2	7	**53**	14	0	4	. . . . . .	15	1050
3	12	9	**81**	7	6	. . . . . .	33	711
4	4	10	6	**12**	1	. . . . . .	0	84
5	8	7	28	2	**24**	. . . . . .	14	178
	⋮	⋮	⋮	⋮	⋮		⋮	
M	12	5	43	3	10	. . . . . .	**92**	1466
Σ	153	276	1245	99	110		812	**43,240**


Bayesian Journey to Crime Routine
Selecting Zones Where Offender Committed Crimes
Zones where offender committed crimes
Legend
Incidents committed by offender
Selected zones
Zones
0 0.5 1 2 3 Miles

Conditional Origin-Destination Matrix

Destination zones where serial offender committed crimes

Crime destination zone

Marginal totals for selected zones only

Crime origin zone

	1	2	3	4	5		N	Σ
1		15		4		. . . . . .	12	121
2		**53**		0		. . . . . .	15	205
3		9		7		. . . . . .	33	65
4		10		**12**		. . . . . .	0	35
5		7		2		. . . . . .	14	40
		⋮		⋮			⋮	
M		5		3		. . . . . .	**92**	141
Σ		276		99			812	**1,597**

Application to Journey to Crime Estimation (cont)

Change definitions:

- P(JTC) is a probability estimate from Jtc method
- P(O|JTC) is a probability estimate based on the distribution of all offenders who committed crimes in same locations as JTC
- P(O) is a probability estimate based on the distribution of all offenders

Application to Journey to Crime Estimation (cont)

From:

$$P(\theta|X) = \frac{P(\theta) * P(X|\theta)}{P(X)}$$

We have an approximation:

$$P(JTC|O) \approx \frac{P(JTC) * P(O|JTC)}{P(O)}$$

Application to Journey to Crime Estimation (cont)

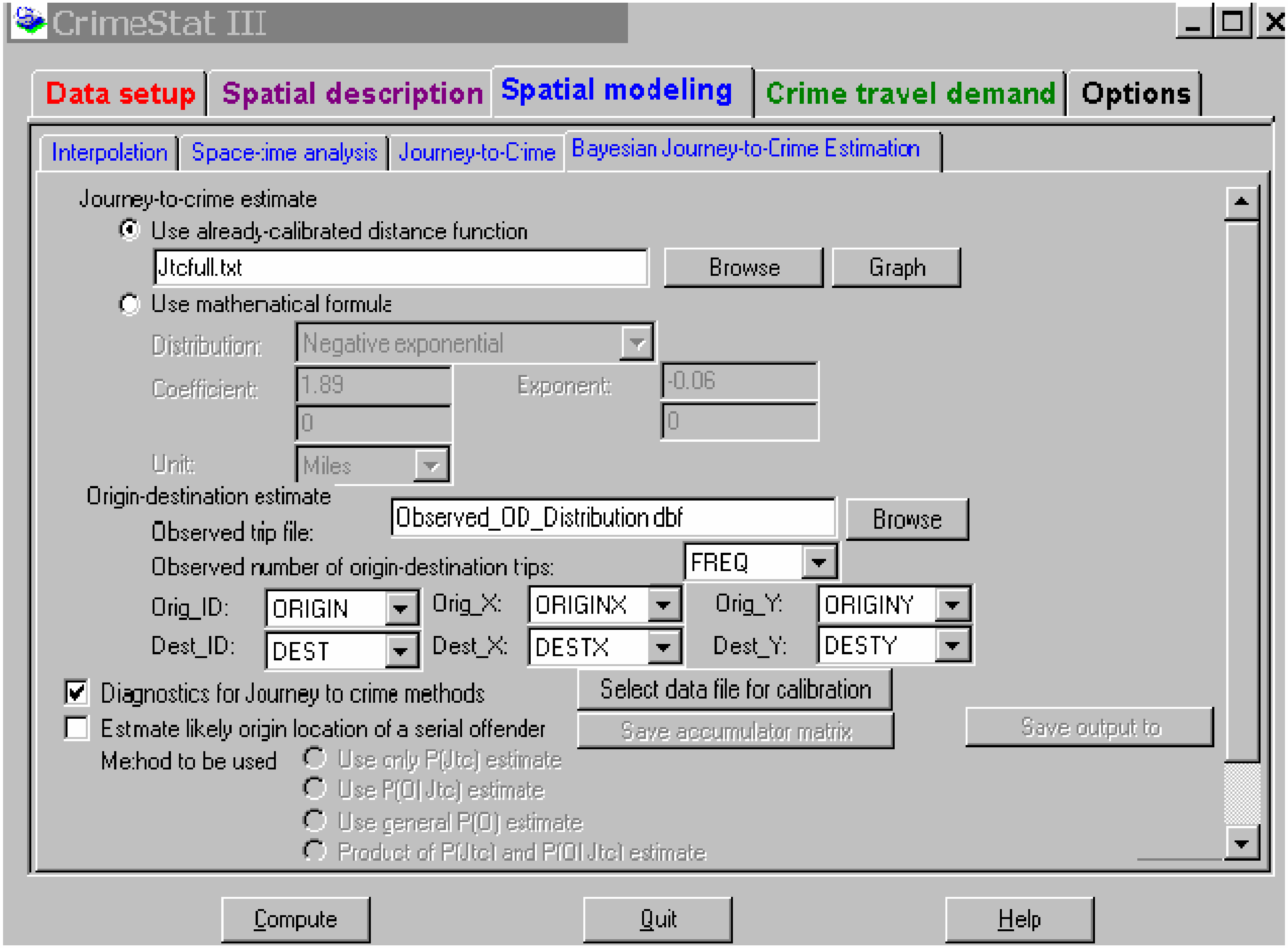
$$P(JTC|O) \approx \frac{\text{Prior probability of JTC * Likelihood of origins given crime locations}}{\text{Prior probability of all crime locations}}$$

$$\approx \frac{\text{“Product probability”}}{\text{“General probability”}}$$

“*Bayesian Risk*”

Tests:

- Whether Bayesian methodology is more accurate than JTC alone
- Whether Bayesian methodology is more accurate than Center of Minimum Distance (CMD)
- Whether a combination of methods is more accurate than one method only


CrimeStat III
Data setup
Spatial description
Spatial modeling
Crime travel demand
Options
Interpolation
Space-time analysis
Journey-to-Crime
Bayesian Journey-to-Crime Estimation
Journey-to-crime estimate
Use already-calibrated distance function
Jtcfull.txt
Browse
Graph
Use mathematical formula
Distribution:
Negative exponential
Coefficient:
1.89
0
Exponent:
-0.06
0
Unit:
Miles
Origin-destination estimate
Observed trip file:
Observed_OD_Distribution.dbf
Browse
Observed number of origin-destination trips:
FREQ
Orig_ID:
ORIGIN
Orig_X:
ORIGINX
Orig_Y:
ORIGINY
Dest_ID:
DEST
Dest_X:
DESTX
Dest_Y:
DESTY
Diagnostics for Journey to crime methods
Select data file for calibration
Estimate likely origin location of a serial offender
Save accumulator matrix
Save output to
Method to be used
Use only P(Jtc) estimate
Use P(O|Jtc) estimate
Use general P(O) estimate
Product of P(Jtc) and P(O|Jtc) estimate
Compute
Quit
Help

Data Sets:

- 88 serial offenders from Baltimore County (Md) who committed various crimes from 1993-1997
- 103 serial offenders from Chicago (IL) who committed robberies from 1996-1998

For Each Offender:

Each of seven measures are calculated:

• P(JTC)	"JTC probability"
• P(O)	"General probability"
• P(O|JTC)	"Conditional probability"
• P(JTC) * P(O|JTC)	"Product probability"
• [P(JTC) * P(O|JTC)]/P(O)	"Bayesian risk probability"
• Average P(JTC) & P(O|JTC)	"Average probability"
• CMD	"Center of minimum distance"

Accuracy Assessed with Four Measures:

- Probability in cell where offender actually lived
- Percentage of study area that has to be searched to find offender (% of cells with probabilities higher than cell where offender lived)
- Distance between cell with highest probability and cell where offender actually lived
- Percent of offenders who live within one (1) mile of cell with the highest probability

Results for Baltimore County Data Set:

(Average of 88 Serial Offenders)

Method	Probability in Offender Cell	% of Study Area with Higher Probabilities	Distance from Highest Prob. Cell to Offender Cell (mi)
JTC	0.00084	4.6%	2.78
General	0.00025	16.7%	8.21
Conditional	0.00052	4.6%	3.12
Product	**0.00176**	4.2%	2.65
Bayesian risk	0.00134	4.6%	3.23
Average	0.00068	**4.1%**	2.70
CMD	n.a.	n.a.	**2.62**

Results for Baltimore County Data Set:

(Average of 88 Serial Offenders)

Method	% of Offenders Who Live Within 1 Mile
JTC	56.8%
General	2.3%
Conditional	47.7%
Product	59.1%
Bayesian risk	51.1%
Average	60.2%
CMD	54.5%

Results for Chicago Data Set:

(Average of 103 Serial Robbers)

Method	Probability in Offender Cell	% of Study Area with Higher Probabilities	Distance from Highest Prob. Cell to Offender Cell (mi)
JTC	0.0073	4.1%	1.99
General	0.0011	13.6%	3.98
Conditional	0.0056	**1.4%**	1.95
Product	**0.0274**	2.6%	**1.86**
Bayesian risk	0.0202	3.2%	1.93
Average	0.0065	2.2%	1.93
CMD	n.a.	n.a.	1.89

Results for Chicago Data Set:

(Average of 103 Serial Robbers)

Method	% of Offenders Who Live Within 1 Mile
JTC	42.7%
General	11.7%
Conditional	52.4%
Product	46.6%
Bayesian risk	45.6%
Average	45.6%
CMD	45.6%

Separate Methods for Different Crimes?

By Specific Crime Types: Baltimore County

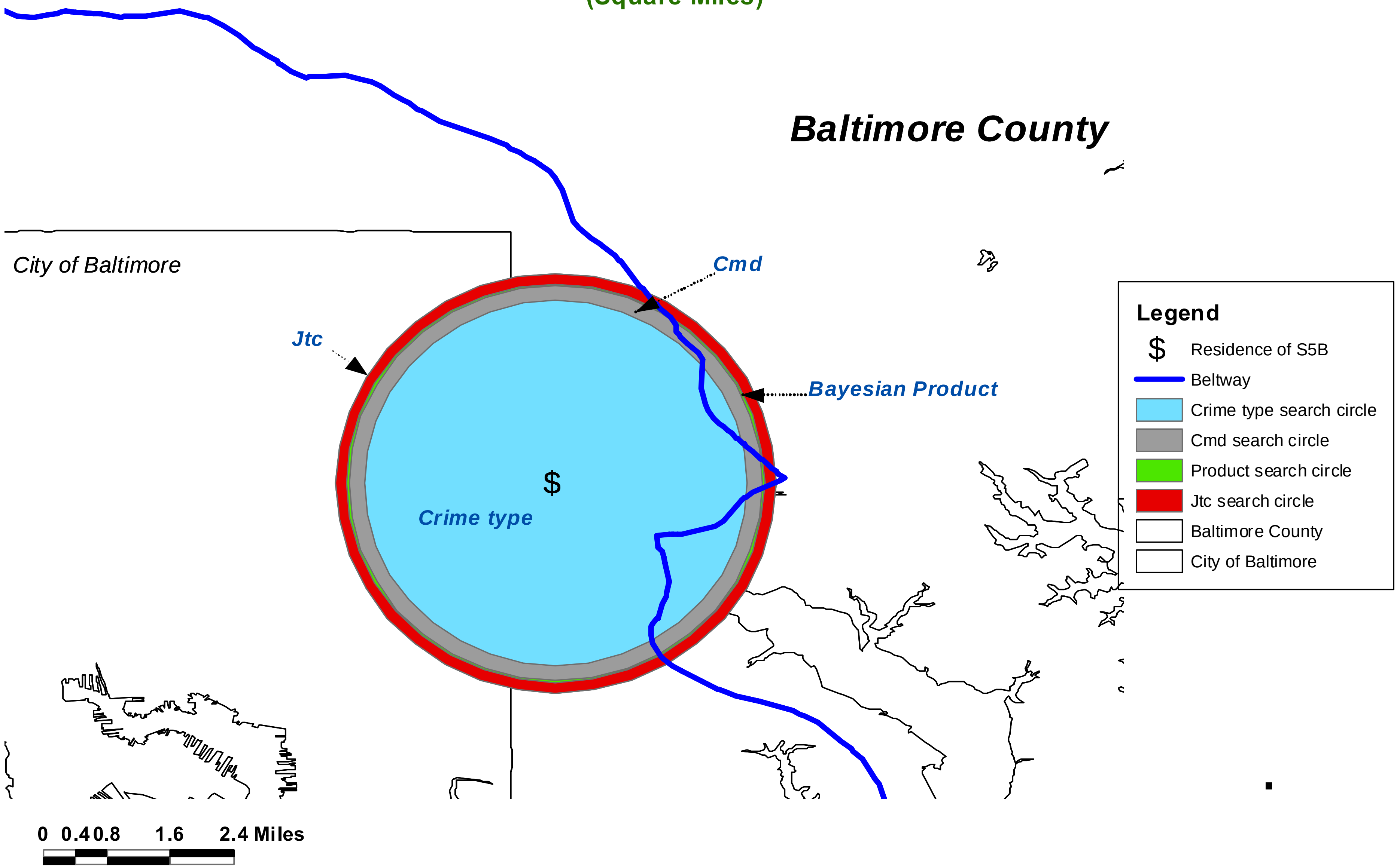
(Distance Measure Only)

	Distance from Highest Prob. Cell to Offender Cell (mi)	Best Method
Assault	0.99	"CMD"
Burglary	2.33	"Conditional"
Larceny	3.22	"CMD"
Robbery	0.85	"Average"
Vehicle theft	2.43	"Conditional"
Other	0.35	"Product"/"Average"
Weighted mean =	**2.40**	

Bayesian Journey to Crime Routine

Average Search Circle of Jtc, Bayesian Product and Crime Type Minimum Estimates

(Square Miles)



Conclusion:

- Bayesian "product" probability had the highest probability in the cell where the offender lived
- For search area, "product" probability was very efficient for about half the cases
- Bayesian "product" probability was about as accurate as the Center of Minimum Distance
- Using crime-specific methods *may* increase accuracy
- "General" probability is very inaccurate

Conclusion: (continued)

- The method can predict both ‘marauders’ and ‘commuters’ since the anchor point can be outside the convex hull

- The method still has a lot of error

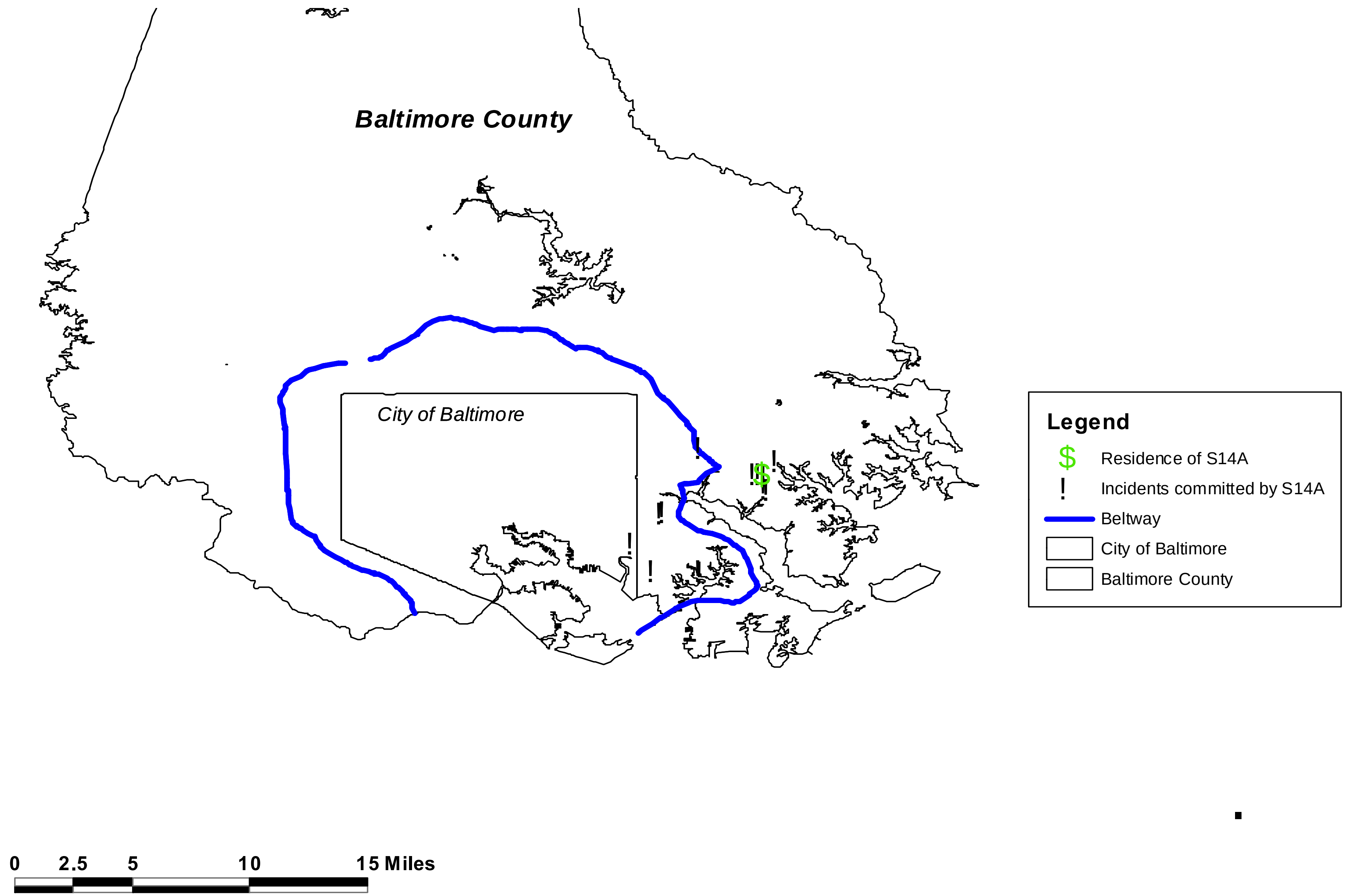
Illustrations from Baltimore County

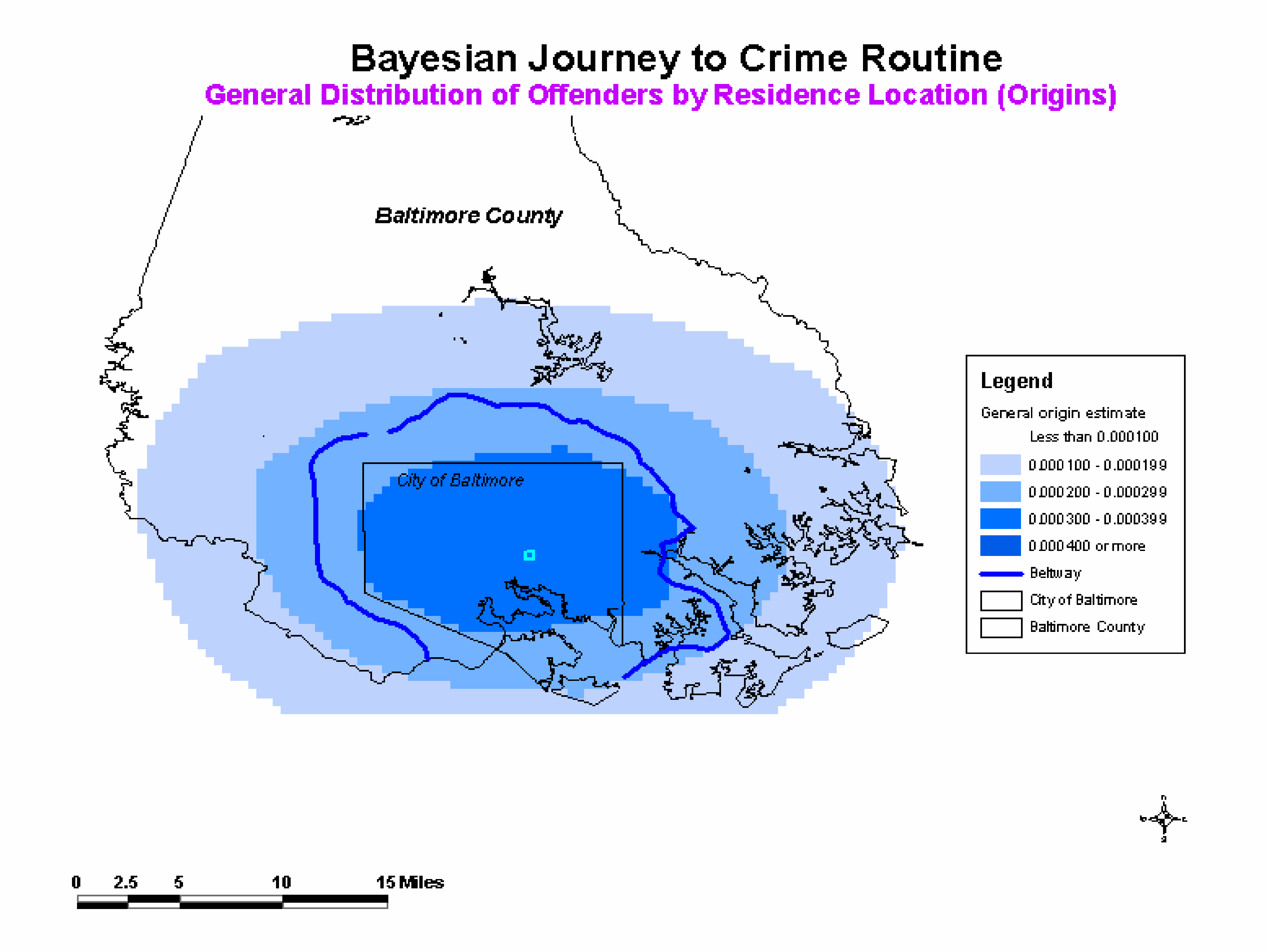
Offender S14A:

7 larceny thefts
4 aggravated assaults
2 robberies
1 burglary

Bayesian Journey to Crime Routine

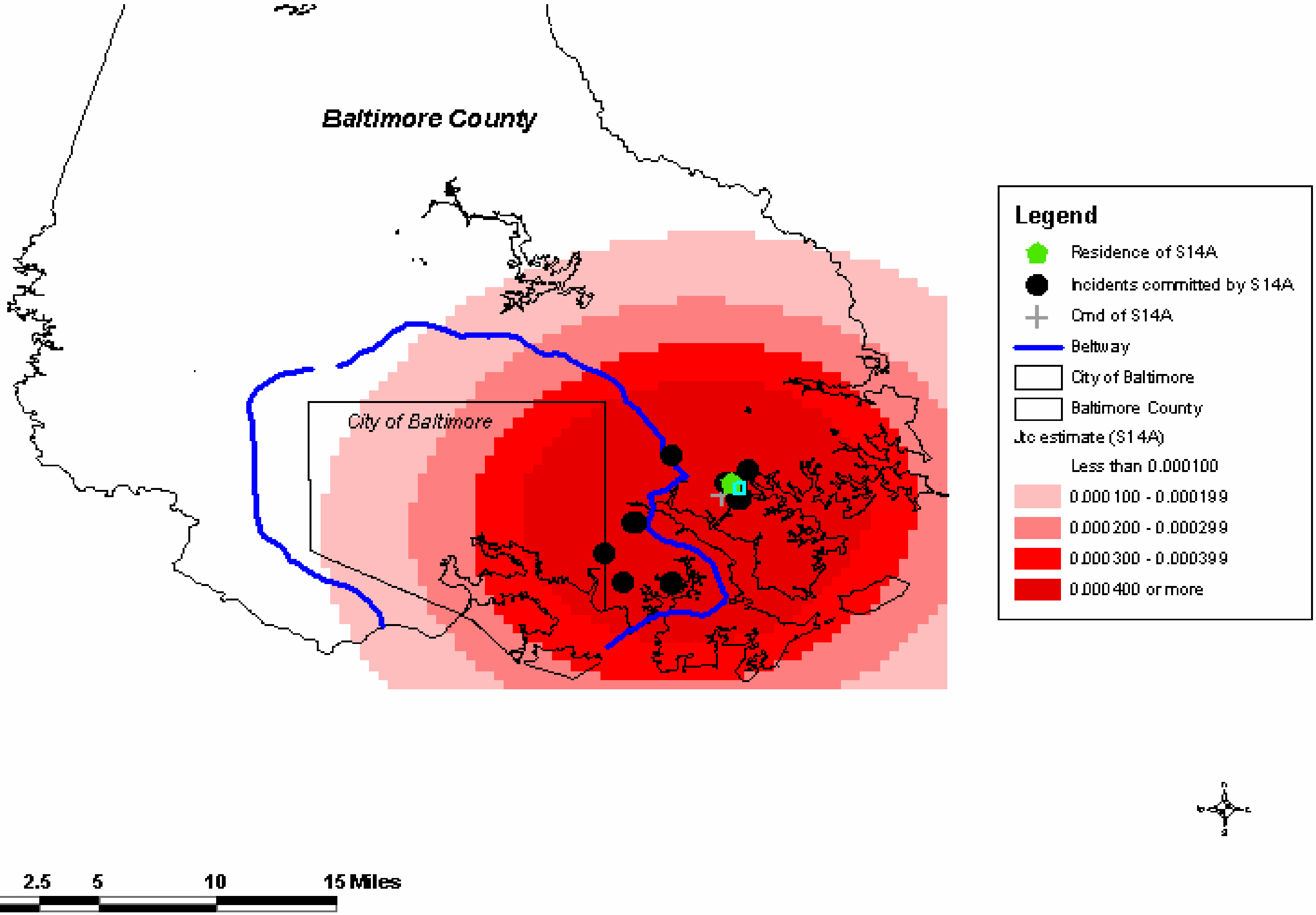
Location of Incidents and Residence of Offender S14A




Bayesian Journey to Crime Routine
General Distribution of Offenders by Residence Location (Origins)
Baltimore County
City of Baltimore
Legend
General origin estimate
Less than 0.000100
0.000100 - 0.000199
0.000200 - 0.000299
0.000300 - 0.000399
0.000400 or more
Beltway
City of Baltimore
Baltimore County
0
2.5
5
10
15 Miles

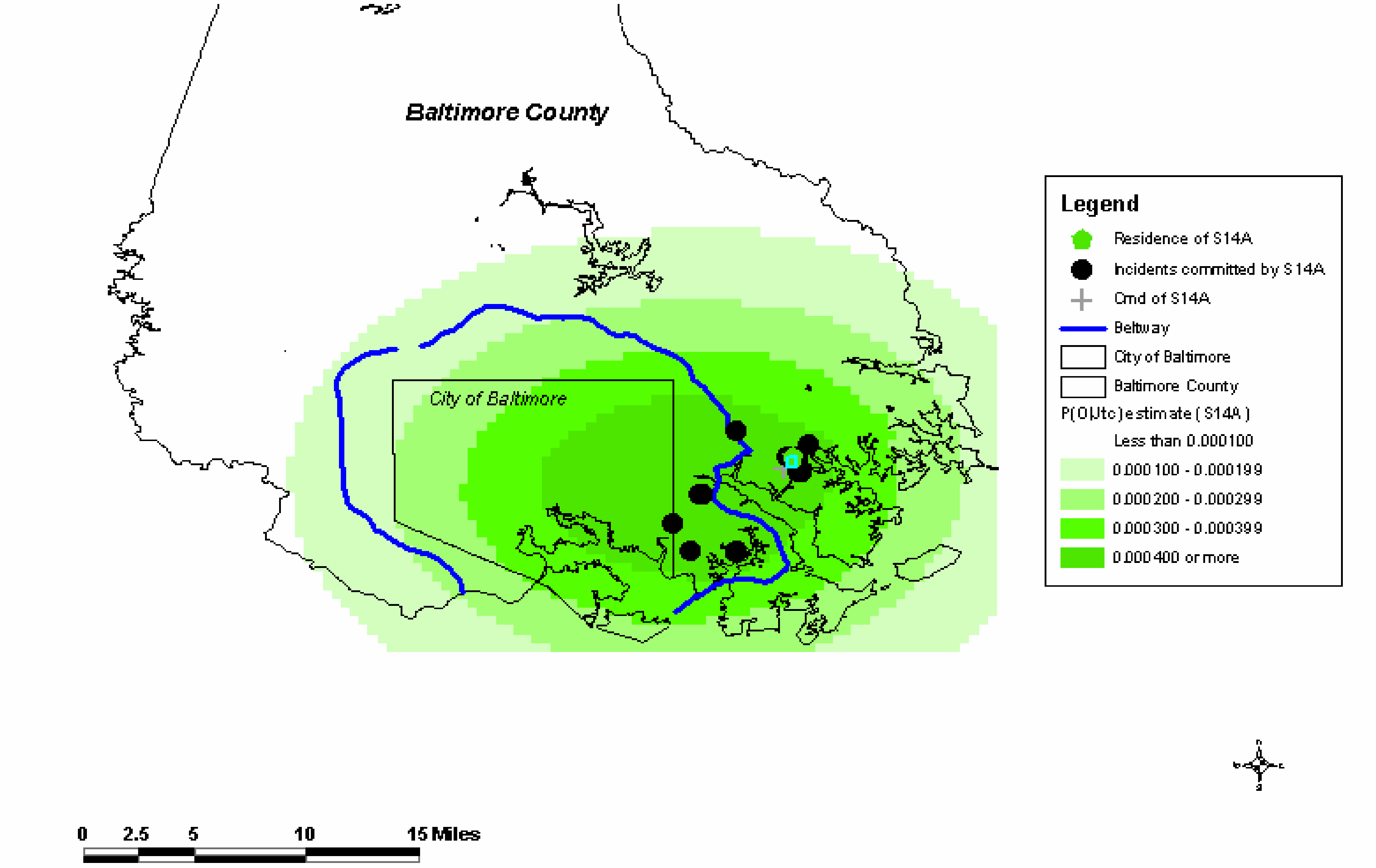
Bayesian Journey to Crime Routine

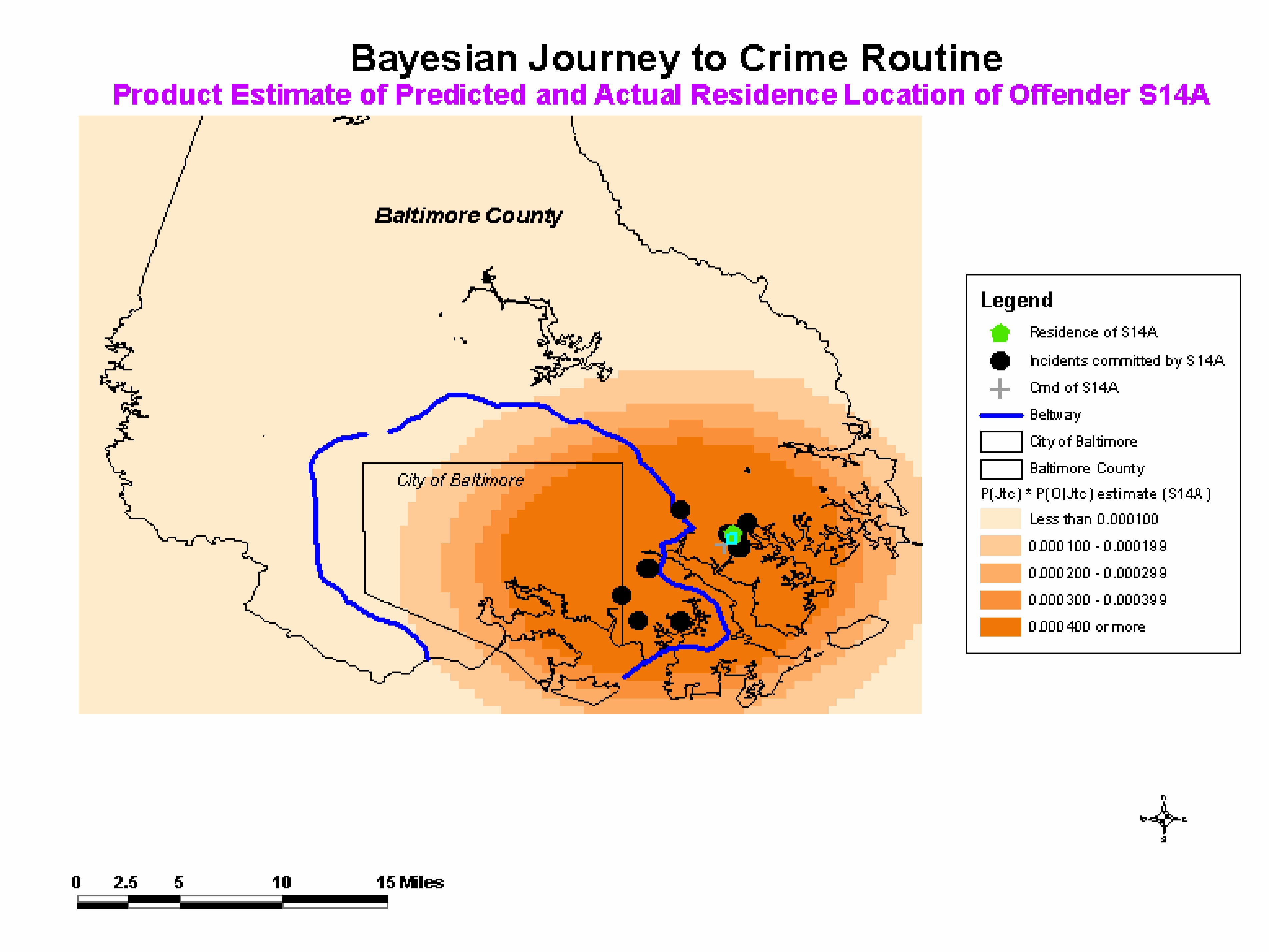
Jtc Estimate of Predicted and Actual Residence Location of Offender S14A



Bayesian Journey to Crime Routine

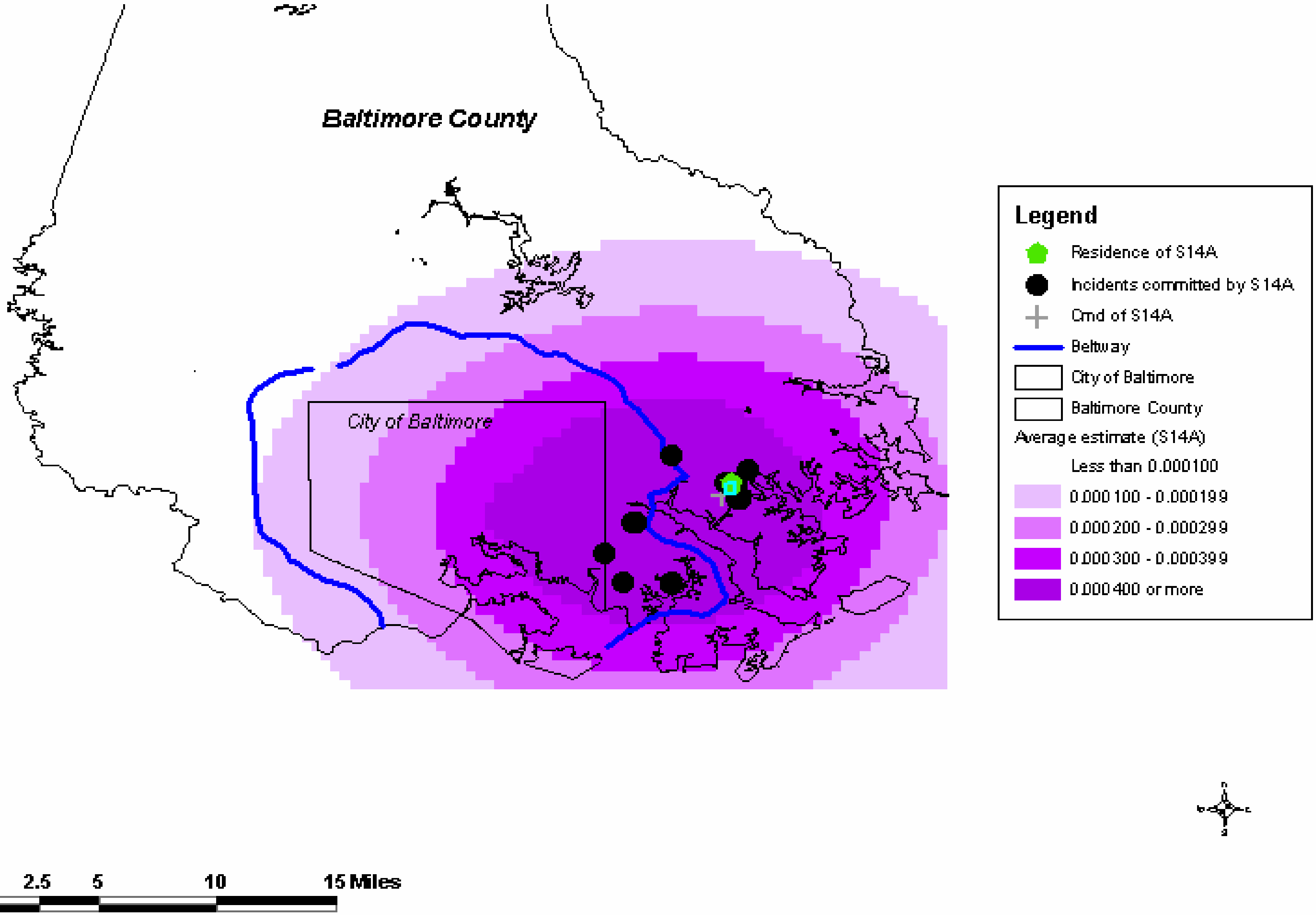
Conditional Estimate of Predicted and Actual Residence Location of Offender S14A




Bayesian Journey to Crime Routine
Product Estimate of Predicted and Actual Residence Location of Offender S14A
Baltimore County
City of Baltimore
Legend
Residence of S14A
Incidents committed by S14A
Cmd of S14A
Beltway
City of Baltimore
Baltimore County
P(Jtc) * P(O|Jtc) estimate (S14A)
Less than 0.000100
0.000100 - 0.000199
0.000200 - 0.000299
0.000300 - 0.000399
0.000400 or more
0 2.5 5 10 15 Miles

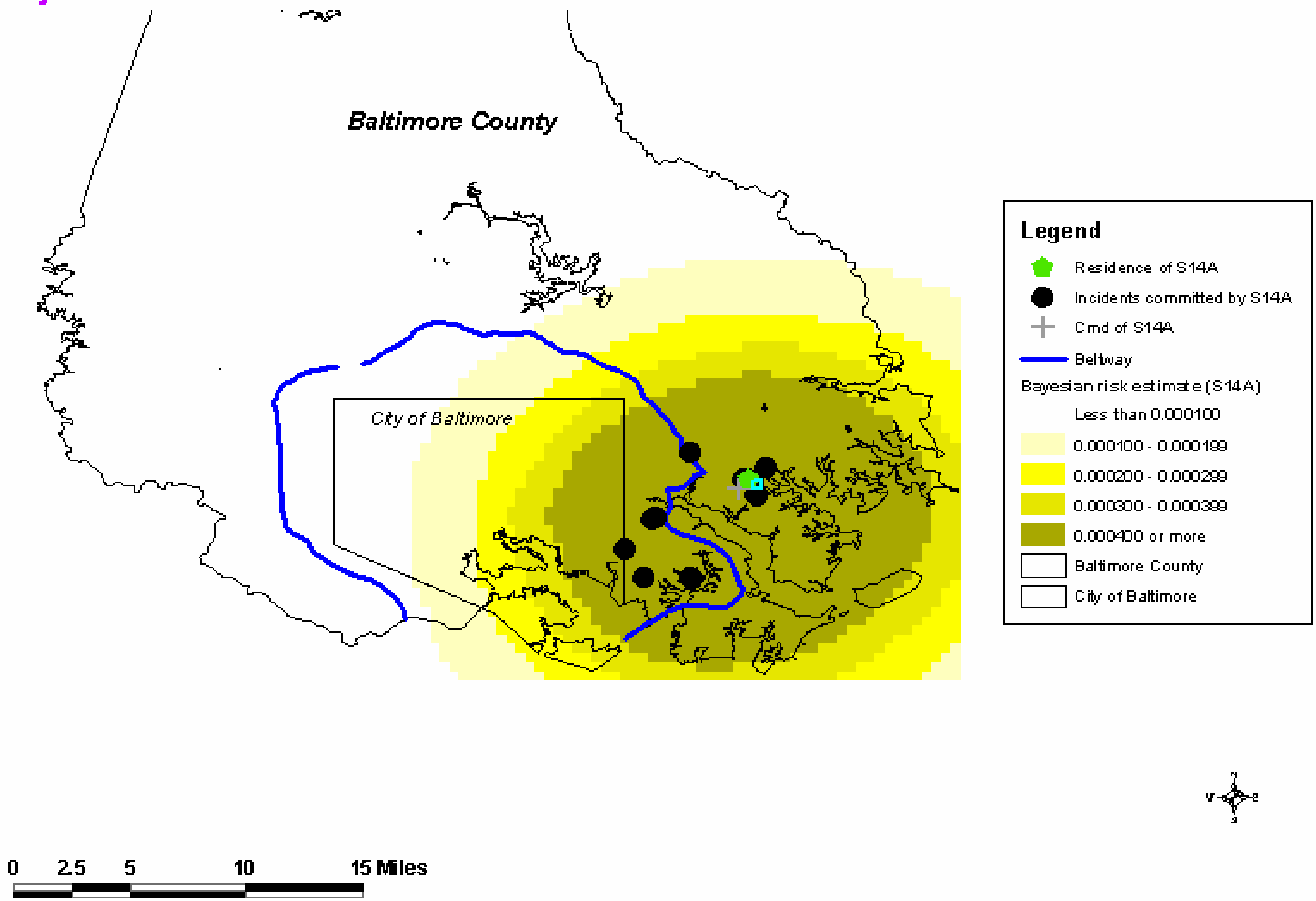
Bayesian Journey to Crime Routine

Average Estimate of Predicted and Actual Residence Location of Offender S14A



Bayesian Journey to Crime Routine

Bayesian Risk Estimate of Predicated and Actual Residence Location of Offender S14A



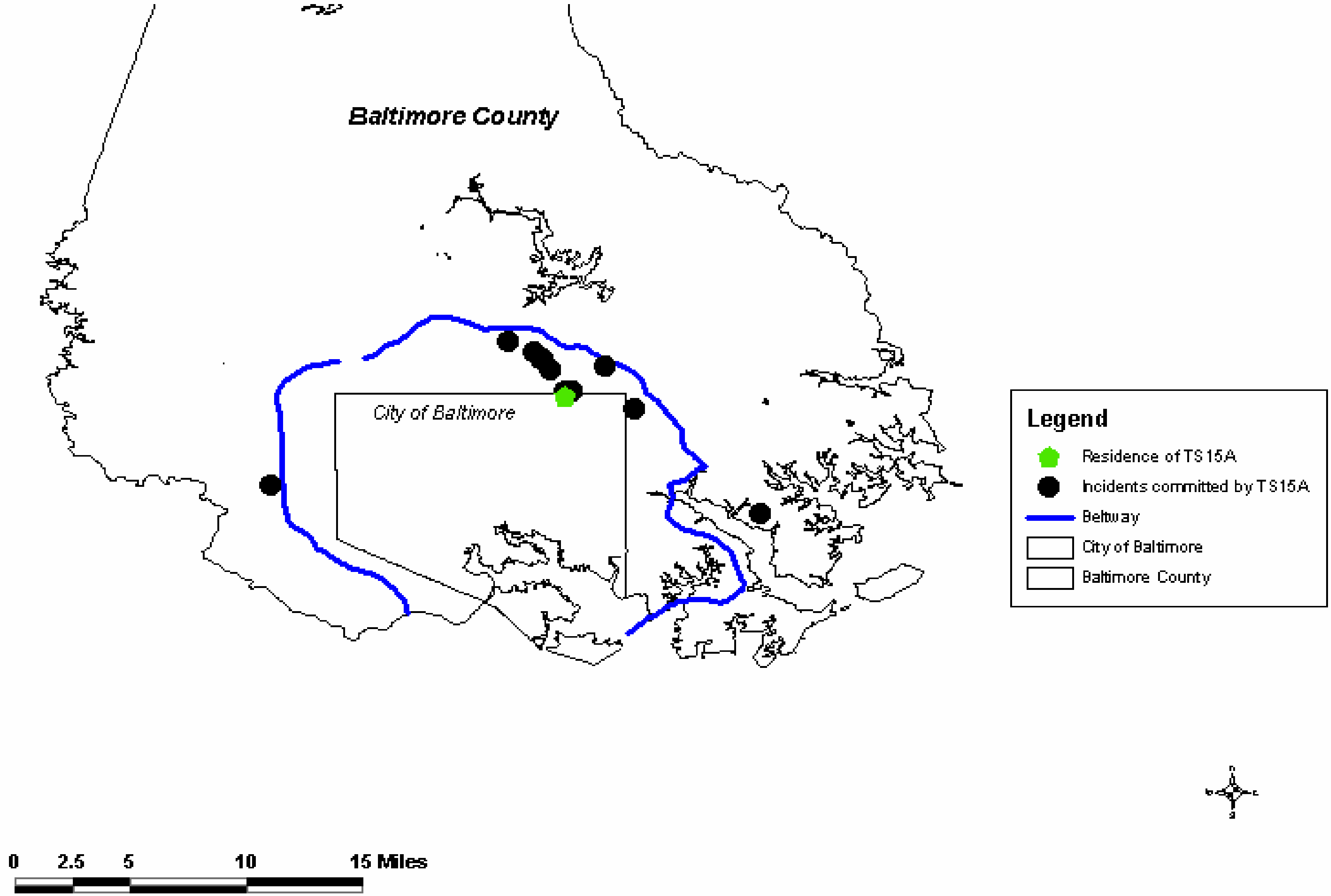
Illustrations from Baltimore County (continued)

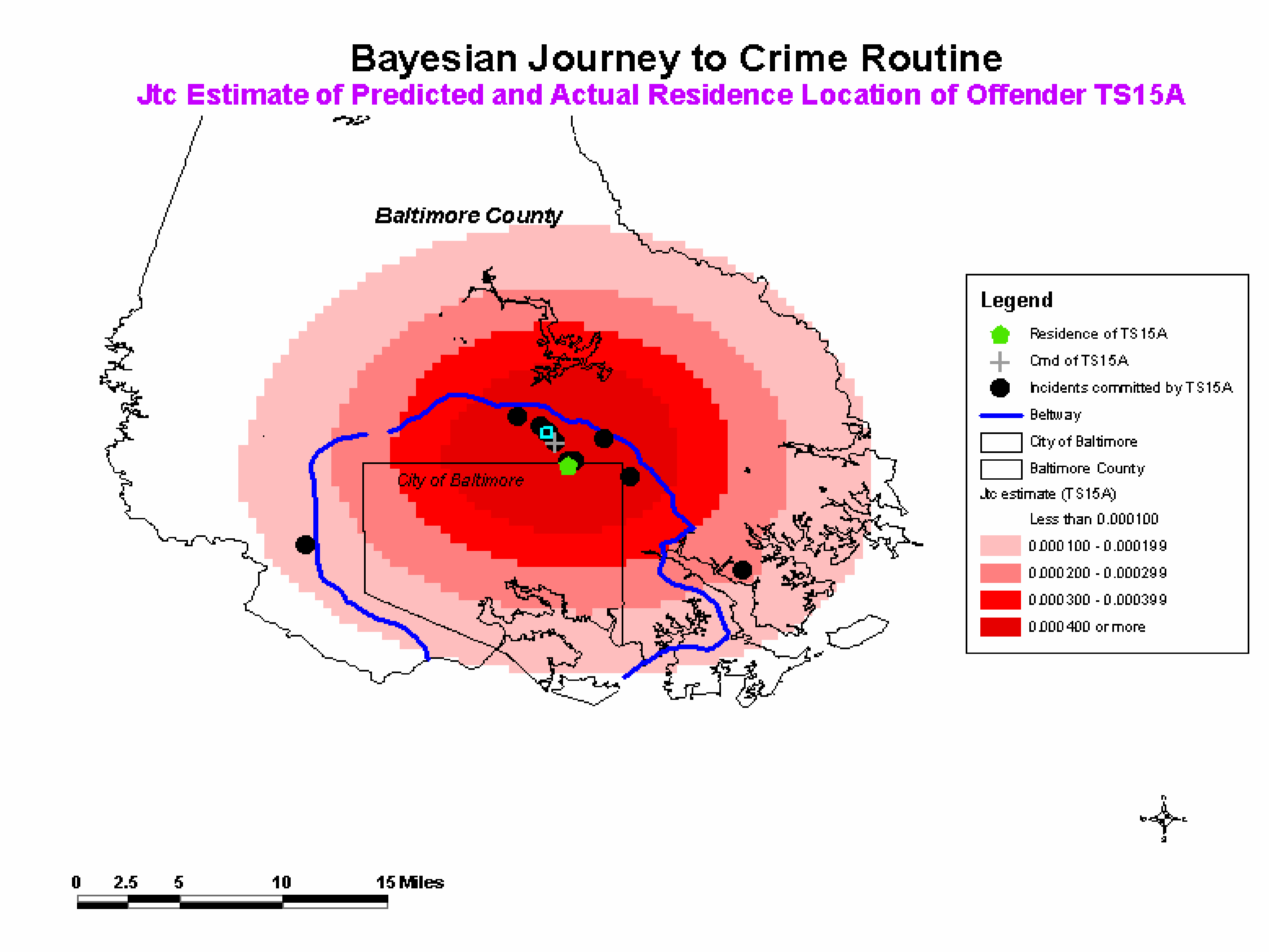
Offender TS15A:

6 larceny thefts
2 aggravated assaults
2 vehicle thefts
1 robbery
1 burglary
3 arson incidents

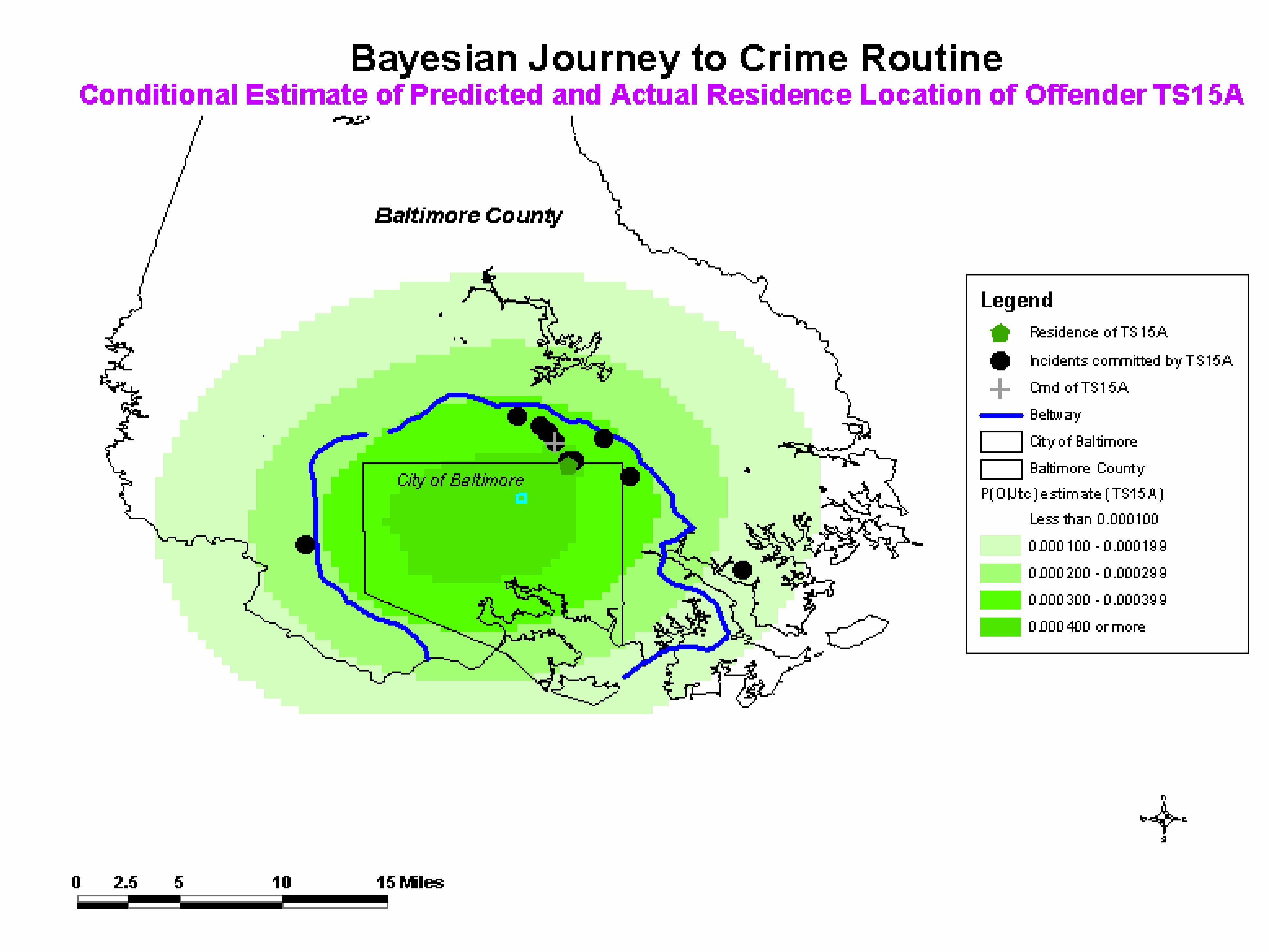
Bayesian Journey to Crime Routine

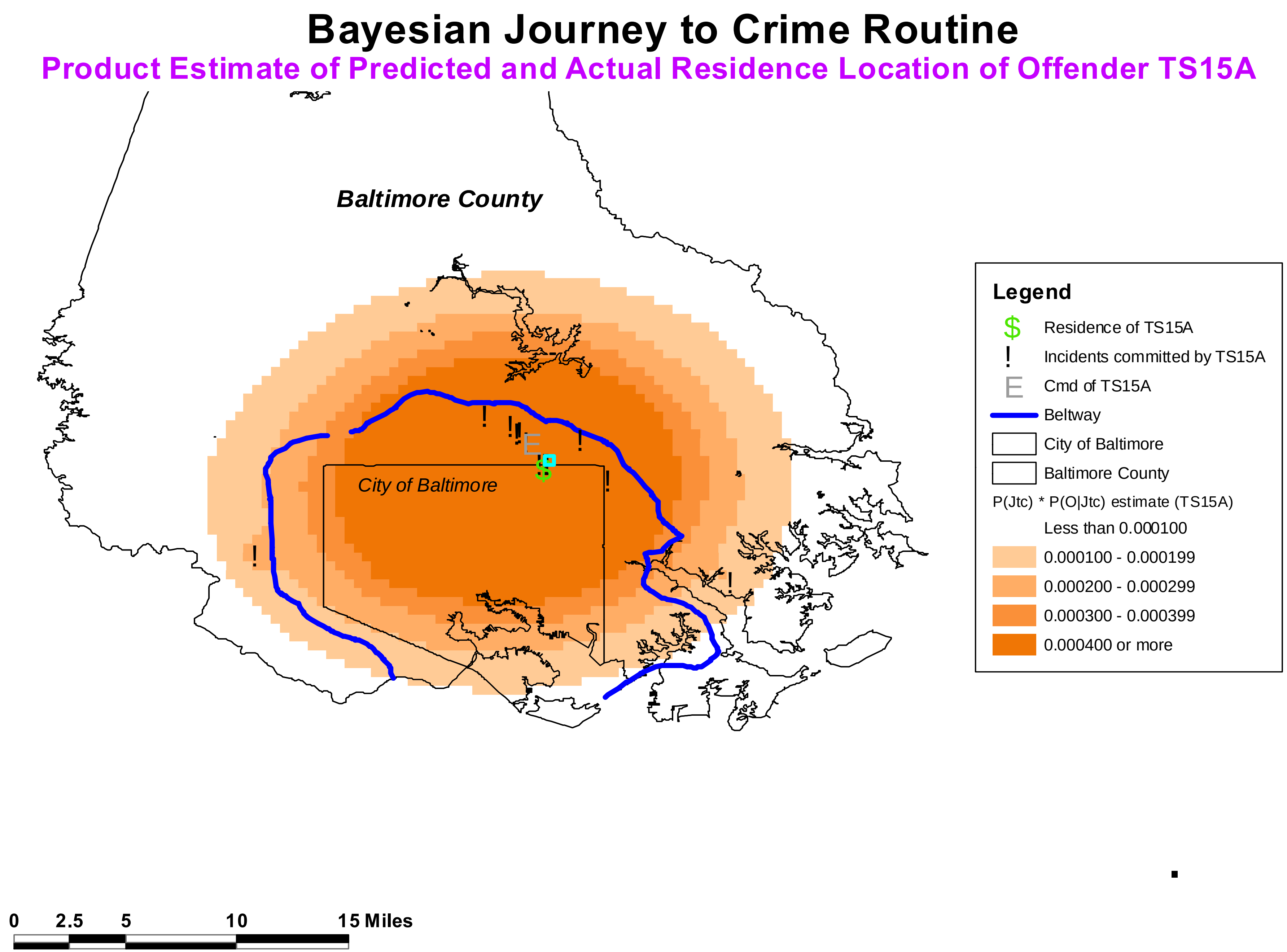
Residence and Location of Incidents Committed by TS15A





Bayesian Journey to Crime Routine
Jtc Estimate of Predicted and Actual Residence Location of Offender TS15A
Baltimore County
City of Baltimore
Legend
Residence of TS15A
Cmd of TS15A
Incidents committed by TS15A
Beltway
City of Baltimore
Baltimore County
Jtc estimate (TS15A)
Less than 0.000100
0.000100 - 0.000199
0.000200 - 0.000299
0.000300 - 0.000399
0.000400 or more
0
2.5
5
10
15 Miles


Bayesian Journey to Crime Routine
Conditional Estimate of Predicted and Actual Residence Location of Offender TS15A
Baltimore County
City of Baltimore
Legend
Residence of TS15A
Incidents committed by TS15A
Cmd of TS15A
Beltway
City of Baltimore
Baltimore County
P(O|Jtc) estimate (TS15A)
Less than 0.000100
0.000100 - 0.000199
0.000200 - 0.000299
0.000300 - 0.000399
0.000400 or more
0
2.5
5
10
15 Miles


Bayesian Journey to Crime Routine
Product Estimate of Predicted and Actual Residence Location of Offender TS15A
Baltimore County
City of Baltimore
Legend
Residence of TS15A
Incidents committed by TS15A
Cmd of TS15A
Beltway
City of Baltimore
Baltimore County
P(Jtc) * P(O|Jtc) estimate (TS15A)
Less than 0.000100
0.000100 - 0.000199
0.000200 - 0.000299
0.000300 - 0.000399
0.000400 or more
0
2.5
5
10
15 Miles