Multimodel Inference and Geographic Profiling

Mike O'Leary
Department of Mathematics
Towson University

Abstract

The geographic profiling problem is to estimate the location of a serial criminal's anchor point from the locations of the crime sites in the series. The approach of O'Leary (2009) uses a model of offender behavior and Bayesian inference for this estimation, and the accuracy of this approach depends strongly on the appropriateness of the underlying model of offender behavior. We demonstrate how the Akaike information criterion can be used to evaluate candidates for models of offender behavior, and how multimodel inference can be used to combine estimates obtained from different models.

Keywords: Geographic profiling, Akaike information criterion, multimodel inference.

Introduction

The primary question of geographic profiling is, given the locations of a series of crimes committed by a single serial criminal to estimate the location of that offender's anchor point. A number of approaches have been developed to solve this problem; the most well known methods are those of Rossmo (2000), Levine (2009a), and Canter, Coffey, Huntley, and Missen (2000).

In some recent work O'Leary (2009, 2010) has developed a new approach to the geographic profiling problem that begins by showing how a model of offender behavior can be combined with the locations of the crime locations and Bayesian inference to produce estimates of the offender's anchor point. That work then continues by developing some reasonable models for offender behavior, and applies the technique to actual crime series. However, this approach to geographic profiling can be no more accurate than their underlying models of offender behavior, so it is essential that these models be carefully studied. In our continuing research, we are pursuing the use of model selection and multimodel inference based on the Akaike information criterion as a basis for comparing and evaluating models for offender behavior.

This paper will briefly review our geographic profiling method. We will introduce the relevant mathematical theory of model selection and multimodel inference then illustrate these ideas by applying them to a series of convenience store robberies in Baltimore County.

This work was supported by National Institute of Justice through grants 2009-SQ-B9-K014, 2007-DE-BX-K005 and 2005-IJ-CX-K036.

We express our gratitude to the Baltimore County Police Department and especially to Phil Canter.

Literature

The simplest method to estimate the anchor point of a serial offender from the locations of the crime sites is to use a centrographic method, for example by using the centroid or the center of minimum distance as the estimate (LeBeau, 1987). These types of methods are called spatial distribution strategies by Snook, Zito, Bennell, and Taylor (2005).

Another method used to estimate the anchor point of a serial offender is the circle method of Canter and Larkin (1993). They make a distinction between two types of offenders; marauders and commuters. A marauder is assumed to move from a home base, commit a crime, and then return, where the base acts as a focal point for the crime series. In this case, the offender's home range and criminal range overlap. A commuter, on the other hand, first moves from a home base to another area, then commits the crime. As a consequence, the focal point for the crime series is different from the home base, and the offender's home range and criminal range do not overlap

Their circle hypothesis is the following. Given a series of linked crimes committed by a marauder, draw a circle whose diameter are the two crimes locations that are the farthest apart. Then all of the crimes in the series would be in the resulting circle and the offender's home base would also lie in the circle so drawn.

There is evidence for the validity of the circle hypothesis. In their original paper, Canter and Larkin (1993) examined a collection of 45 male sexual assaults in Britain. In 41 of the 45 cases the circle correctly encompassed all of the crime sites and in 39 of the 45 cases the circle correctly contained a base for the offender. Kocsis and Irwin (1997) examined 24 rape series, 22 arson series, and 27 burglary series in Australia. The circle contained all of the crimes for 79% of the rape series, 82% of the arson series, and 70% of the burglary series, while the circle correctly contained the home base of the offender for 71% of the rape cases, 82% of the arson cases, and 48% of the burglary cases. This last result suggests that the marauder hypotheses may not necessarily be appropriate for burglary.

These results were amplified by Meaney (2004), who showed that burglars were more likely to act as commuters than non-burglars, while arsonists and sex offenders were more likely to act as marauders than non-arsonists / non-sex offenders. Similarly, Kocsis, Cooksey, Irwin, and Allen (2002) found in their study of 58 burglaries that occurred in rural Australian towns, that the circle theory was less effective. Laukkanen and Santtila (2006) examined 76 commercial robbery series, and found only 30 (=39%) that satisfied the circle hypothesis. Note however, that many of these series were very short; 62 of the 76 series analyzed contained either two or three crimes.

In contrast to the spatial distribution strategies are the probability distribution strategies (Snook et al., 2005); these methods are used in the the major computer programs for geographic profiling (CrimeStat, Dragnet, and Rigel). All have the common idea of constructing a hit score by summing the values of some decay function of the distances between a general point and the elements of the crime series. Rossmo's method, described in (Rossmo, 2000, Chp. 10) uses the Manhattan distance and an algebraic form for distance decay with a buffer zone. Canter et al. (2000) used Euclidean distance and an exponential distance decay with and without a buffer and/or plateau; the parameters were calibrated by comparing their results with the body disposal sites of 70 U.S. Serial killers. CrimeStat, from Ned Levine, allows the use of different distance metrics and different distance decay functions (Levine, 2009a).

Recent versions of CrimeStat (3.1+) introduce a new Bayesian approach to modeling the journey to crime. The region containing the crime series is subdivided, and historical crime data is used to count the number of crime trips from one subregion to another; this is then used as a basis for Bayesian inference; *c.f.* Levine (2009b); Levine and Lee (2009).

Methods

Before we begin our discussion of mathematical models of offender behavior, we adopt some common notation. A geographic point $\mathbf{x}$ will have two components $\mathbf{x} = \langle x^{(1)}, x^{(2)} \rangle$; these can be the longitude and latitude of the point referenced by some convenient datum or more simply the distances from a pair of perpendicular reference axes. We assume that the offender has a single well defined anchor point during the crime series denoted by $\mathbf{z}$, and we assume that the series under study has n linked crimes at the locations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$.

The fundamental assumption of the method of O'Leary (2009) is that there is a function $P(\mathbf{x}|\mathbf{z})$ that gives the probability density that an offender with anchor point $\mathbf{z}$ will commit a crime at the location $\mathbf{x}$. Though we are modeling the behavior of the offender with a probability density, this is not meant to imply that the offender necessarily has a random component in the selection of crime site locations. Rather, this randomness reflects the lack of knowledge that we have about the behavior of the offender.

In this approach, the problem of estimating the offender's anchor point $\mathbf{z}$ can be restated in the following mathematical form: Given a sample $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ chosen independently from a probability density $P(\mathbf{x}|\mathbf{z})$, estimate the value of the parameter $\mathbf{z}$. One simple approach to this problem is to use maximum likelihood estimation. In particular, this provides a point estimate $\hat{\mathbf{z}}$ of the anchor point $\mathbf{z}$ that is the value of $\mathbf{y}$ that maximizes the likelihood function

$$L(\mathbf{y}) = \prod_{i=1}^{n} P(\mathbf{x}_i|\mathbf{y}) = P(\mathbf{x}_1|\mathbf{y}) \cdot P(\mathbf{x}_2|\mathbf{y}) \cdots P(\mathbf{x}_n|\mathbf{y}).$$

Though mathematically reasonable, the primary limitation of the maximum likelihood estimator is that is only provides a single point estimate. A method that provides a prioritized search area would be much preferred by law enforcement agencies. This can be accomplished by using Bayesian methods.

To use Bayesian methods, we first need a prior estimate of the distribution of anchor points $H(\mathbf{x})$. This contains our presumed knowledge of the distribution of offender anchor points before we take the data from the crime series into account. If we assume that anchor points are residences, then making $H(\mathbf{x})$ proportional to the population density at $\mathbf{x}$ would be a reasonable choice; one could also hypothesize that the prior distribution of anchor points is proportional to the distribution of known past offender residences.

Bayesian analysis then tells us that the probability density distribution $P(\mathbf{z}|\mathbf{x})$ that the offender's anchor point is $\mathbf{z}$ given that they have offended at $\mathbf{x}$ satisfies

$$P(\mathbf{z}|\mathbf{x}) \propto P(\mathbf{x}|\mathbf{z}) H(\mathbf{z}).$$

To generalize this to a series of crimes, we again assume that the crimes $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are chosen independently from the probability density $P(\mathbf{x}|\mathbf{z})$; then the probability density distribution $P(\mathbf{z}|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ that the offender's anchor point is $\mathbf{z}$ given that they have offended at $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_n$ satisfies

$$P(\mathbf{z}|\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_n) \propto \left[\prod_{i=1}^{n} P(\mathbf{x}_i|\mathbf{z}) \right] H(\mathbf{z}).$$

Thus, given a particular model for offender behavior and an estimate of the prior distribution of anchor points, Bayesian analysis provides a mathematically rigorous method to construct an estimate for a search area for the anchor point of the offender. This approach can easily be extended to allow for the probability density that $\mathbf{x}$ is selected as a target to depend on other variables than just the offender's anchor point. One natural variable to include is the average distance α that the offender is willing to travel, as this may vary between offenders. For example, a young, inexperienced teenage offender may be less willing to travel a long distance than an older, more experienced offender.

The work in (O'Leary, 2009, 2010) concludes by by presenting reasonable potential models of offender behavior and describing how to implement the Bayesian analysis.

Though these methods allow us to construct mathematically rigorous estimates for the anchor point of the offender, they are suboptimal in a a number of ways. The most fundamental issue is that they posit the existence of a single, best model $P(\mathbf{x}|\mathbf{z})$ for offender behavior. However, it is unlikely that one can construct a single model that captures the full richness of human behavior. Indeed, offenders may commit their crimes with differing levels of skill and motivation, and a model that is appropriate for one offender may be totally inappropriate for another. Thus, to proceed we need to find a way to compare and contrast different models of offender behavior $P(\mathbf{x}|\mathbf{z})$.

To begin to address this problem in our continuing and future work we plan to proceed using the mathematical techniques of model selection and multimodel inference. These techniques were first developed in a sequence of papers by Akaike (1973, 1974, 1977), and an excellent description of the relevant theory and applications can be found in Burnham and Anderson (2002). These methods have seen significant use in mathematical ecology, for example in Link and Barker (2006); Katsanevakis (2006); Stephens, Buskirk, and Martínez del Rio (2006). They have also been applied to many other areas; see for example Liddle (2007) who applies these ideas to astrophysics, or either Poeter and Anderson (2005) or Ye, Meyer, and Neuman (2008) who consider these ideas in the context of ground water modeling.

To apply these ideas to the geographic profiling problem, we first identify a collection of candidate models for the offender's behavior. We then use the crime series data and perform a standard maximum likelihood analysis on each model, obtaining estimates of the offender's anchor point and corresponding search area from each model. The Akaike Information Criterion is then calculated for each model; this is a quantitative technique that allows one to compare the amount of information a particular model provides about a data set. This calculation will then let us compare how well our candidate models give useful information about the crime series under study. These calculations are performed based solely on the characteristics of the models of offender behavior and the crime series under consideration; different crime series will likely return different best models. This enables us to allow for the possibility that not all offender's behavior can be modeled by the same relationship. Inferences about the best search area for the offender can then be made by the best model, or even better, by a weighted average of the better models, where the weights give the relative amount of information each model provides about the crime series; this process is called multimodel inference.

More precisely, suppose that we have R different models for offender's target selection density, say $P_1(\mathbf{x}|\mathbf{z},\sigma_1), P_2(\mathbf{x}|\mathbf{z},\sigma_2), \dots P_R(\mathbf{x}|\mathbf{z},\sigma_R)$ where $\mathbf{x}$ is still the crime site, $\mathbf{z}$ is still the offender's anchor point, and σ_r are (possibly empty) collections of additional parameters. The total number of parameters for each model is denoted by K_r; we have $K_r \geq 2$ for each of these geographic profiling models because we always have as parameters the two coordinates $\mathbf{z} = \langle z^{(1)}, z^{(2)} \rangle$

of the offender anchor point. The Akaike Information Criterion (AIC) for model r is calculated by

$$\text{AIC}_r = 2K_r - 2\sum_{i=1}^{n} \log\left(P_r(\mathbf{x}_i|\hat{\mathbf{z}}_r, \hat{\sigma}_r)\right). \tag{1}$$

Here $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are the crime sites in the series, while $\hat{\mathbf{z}}_r$ and $\hat{\sigma}_r$ are the maximum likelihood estimators for the anchor point $\mathbf{z}$ and the parameters σ generated for model r. Models with smaller values for AIC are considered to be more informative than models with large values for AIC. In particular, the important factor for inference are the differences in AIC values

$$\Delta_r = \text{AIC}_r - \text{AIC}_{\text{min}}. \tag{2}$$

Models in the collection under study with Δ_r satisfying $0 \le \Delta_r \le 2$ can be considered the most well supported by the data while a model with $\Delta_r \ge 10$ is essentially unsupported by the data. Note however, that the degree of support here is relative to the collection of models under study. In particular, every collection of models will have at least one model with the smallest value of AIC, and hence there will be at least one model with $\Delta_r = 0$ which is considered most supported by the data. This will occur regardless of the appropriateness of the original collection of models.

Another measure of the information content of a model is the Akaike weights; these are defined for model r by

$$w_r = \frac{\exp\left(-\frac{1}{2}\Delta_r\right)}{\sum_{j=1}^{R} \exp\left(-\frac{1}{2}\Delta_j\right)}. \tag{3}$$

These can be considered to be the relative likelihood of model r in the collection of models under study.

We can then compare these weights, and select and use as a basis for inference the model considered to be most informative. As a consequence, we no longer have to determine which model is the "best" model in advance; instead we can use the data from the series under consideration and the Akaike Information Criterion to choose which of the available alternatives is most informative. We can also draw inferences from more than model, by simply taking a weighted average of the estimates produced by each model, where the weighting factors are exactly the Akaike weights (Burnham & Anderson, 2002, Chp. 4). For example, the multimodel estimate of the for the offender's anchor point is simply

$$\hat{\mathbf{z}}_{\text{multimodel}} = w_1\hat{\mathbf{z}}_1 + w_2\hat{\mathbf{z}}_2 + \cdots + w_R\hat{\mathbf{z}}_R.$$

We see how this approach lets us use the information available from all of the potential models under consideration.

The first thing to note about the AIC technique is its parsimony with respect to the number of parameters in a model. We know that increasing the number of parameters in a model will allow it to better fit a given data set, but that this does not necessarily mean that this increases the explanatory power of the model. In particular, increasing the number of parameters in a model always allows to increase the value of the likelihood function for that model. By incorporating a penalty term for the number of model parameters, the AIC takes this into account when comparing the effectiveness of different models.

The second thing to note is that this technique only relies on the data in the individual crime series under consideration when it measures the information content of a model. Indeed, the history

Date	Time	Location		Target
		Latitude	Longitude	
Mar. 8	12:30 pm	-76.7135	39.2985	Speedy Mart
Mar. 19	4:30 pm	-76.7499	39.3134	Exxon
Mar. 21	4:00 pm	-76.7620	39.3410	Exxon
Mar. 27	2:30 pm	-76.7135	39.2985	Speedy Mart
Apr. 15	4:00 pm	-76.7372	39.3174	Citgo
Apr. 28	5:00 pm	-76.7135	39.2985	Speedy Mart

Table 1:: A series of convenience store robberies in Baltimore County

of geographic profiling to date has been marked by a quest for the "one, true model" for offender behavior, and that there is a tendency towards a "one-size-fits-all" approach. In contrast, this approach will let us examine multiple models for offender behavior. By looking only at the data from the crime series and applying the parsimony of parameters described above, we can then select which models possesses more information, and then use these as our bases to predict the offender's anchor point.

Finally, we note that in general a modification of AIC is preferred when the number of elements of the sample is small relative to the number of parameters in the model. Indeed, if $n/K < 40$, Burnham and Anderson (2002, §2.4) suggest the use of a small sample correction

$$\text{AICc}_r = \text{AIC}_r + \frac{2K_r(K_r+1)}{n-K_r-1}$$

in place of AIC_r. Given the number of elements in a crime series is almost always small, the small sample correction will almost always be preferred.

Illustration

To see these ideas in action, let us apply them to a particular crime series. Table 1 contains basic information about a series of convenience store robberies that occurred in Baltimore County in Spring 2008; these are mapped in Figure 1 Note that crimes 1, 4 and 6 all occurred at the same location marked as Crime Site # 4 / Crime Site #6 on the map.

To apply these techniques of multimodel inference, we need to select two or more models of the offender's behavior. For clarity in this illustration, we shall compare four very simple models-two based on a normal distribution and two based on a negative exponential distribution. We emphasize that these simple models are chosen for clarity in exposition; our ongoing research is aimed at developing and analyzing better models.

Our first models use a normal distribution, and they assume that an offender with anchor point $\mathbf{z}$ chooses a target at the location $\mathbf{x}$ according to the probability density function

$$P(\mathbf{x}) = \frac{1}{4\alpha^2}\exp\left(-\frac{\pi}{4\alpha^2}|\mathbf{x}-\mathbf{z}|^2\right)$$

where Euclidean distance is used. This distribution is illustrated is Figure 2a. The parameter α that appears in this expression is the average distance that the offender is willing to travel to offend, and is proportional to the standard deviation of our normal distribution.

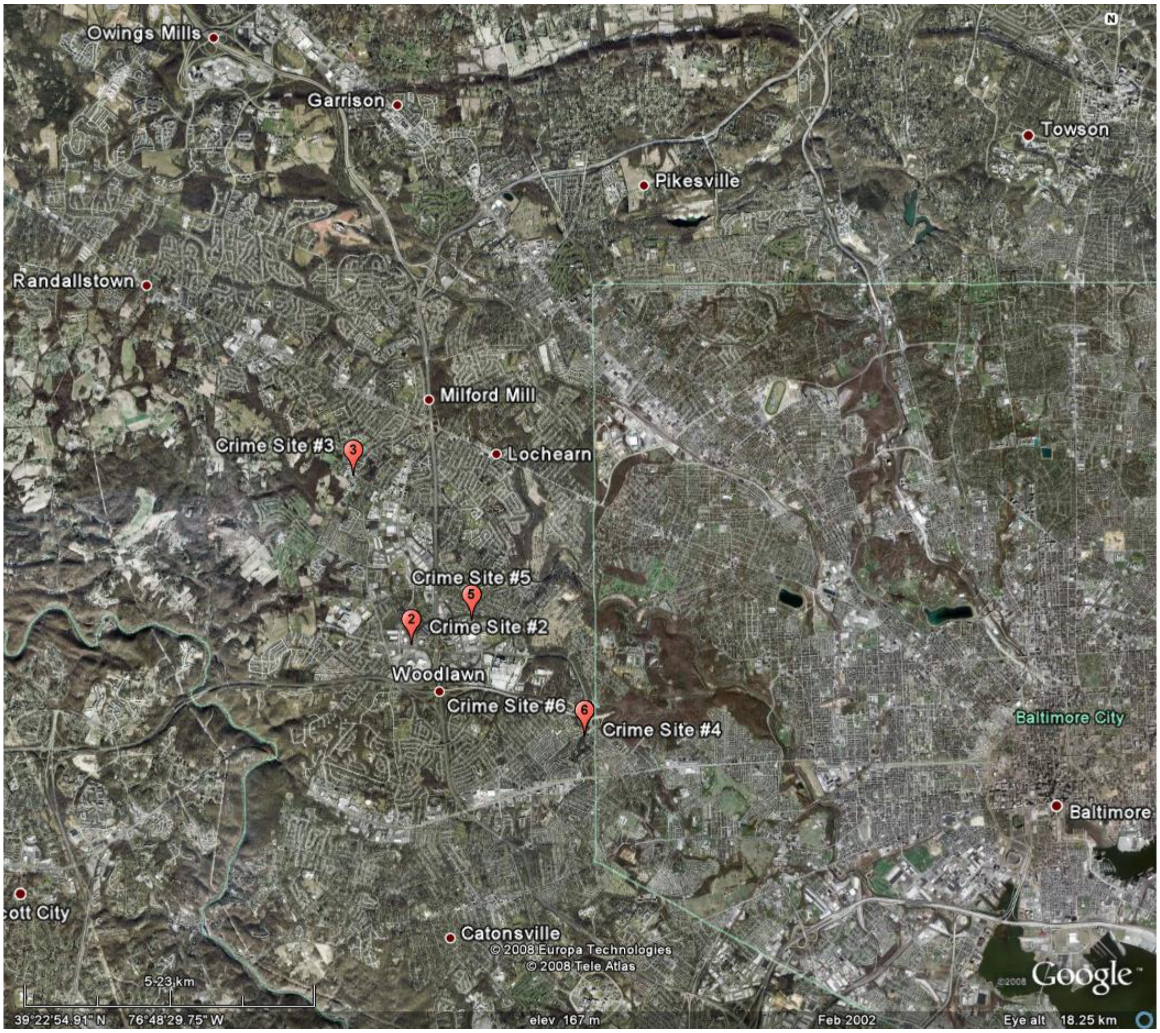


Figure 1. : Convenience store robbery series in Baltimore County

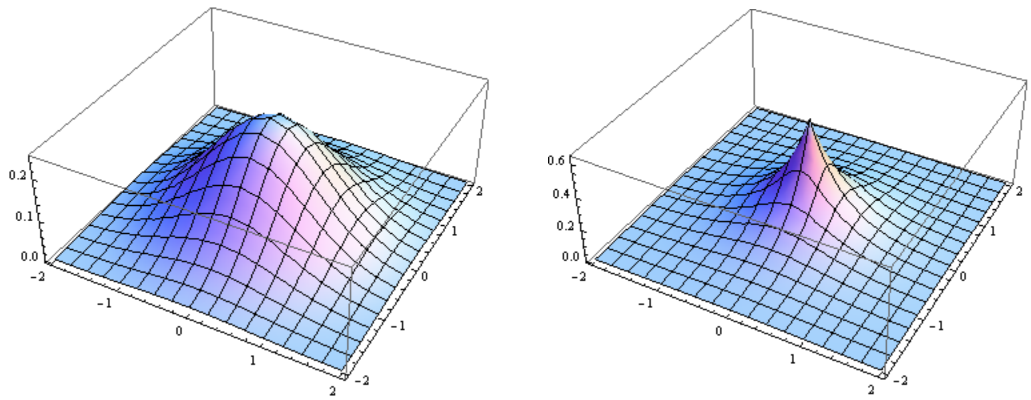
The value of the average offense distance α can be specified *a priori* as part of the model, or it can be considered an additional parameter to be calculated along with the anchor point $\mathbf{z}$. Examining a collection of 751 solved crimes in Baltimore County, we found that the average distance from the crime site to the offenders home was 1.716 miles, so we can consider two models:

$$P_1(\mathbf{x}|\mathbf{z}) = \frac{1}{4\alpha_0^2} \exp\left(-\frac{\pi}{4\alpha_0^2}|\mathbf{x}-\mathbf{z}|^2\right)$$

where $\alpha_0 = 1.716$ miles is fixed, or

$$P_2(\mathbf{x}|\mathbf{z},\alpha) = \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4\alpha^2}|\mathbf{x}-\mathbf{z}|^2\right)$$

where the average offense distance for the offender α is to be determined along with the anchor point $\mathbf{z}$ of the offender from the data in the series. Notice that model 1 has two unknown parameters $\mathbf{z} = \langle z^{(1)}, z^{(2)} \rangle$, while model 2 has three unknown parameters $\mathbf{z} = \langle z^{(1)}, z^{(2)} \rangle$ and α.



(a) Normal distribution (b) Negative exponential distribution

Figure 2. : Models for offender behavior. The offender has anchor point at the origin, and the height of the graph at a point represents the probability density that the offender selects that target location. The average offense distance is 1 for each graph.

We can also craft models from the negative exponential distribution. Consider

$$P_3(\mathbf{x}|\mathbf{z}) = \frac{2}{\pi\alpha_0^2} \exp\left(-\frac{2}{\alpha_0}|\mathbf{x}-\mathbf{z}|\right)$$

where $\alpha_0 = 1.716$ miles is fixed, or

$$P_4(\mathbf{x}|\mathbf{z},\alpha) = \frac{2}{\pi\alpha^2} \exp\left(-\frac{2}{\alpha}|\mathbf{x}-\mathbf{z}|\right)$$

where we want to use the crime series to calculate both the anchor point $\mathbf{z}$ and the average offense distance α. A representative graph of these distributions in Figure 2b. Note that model 3 has two unknown parameters $\mathbf{z} = \langle z^{(1)}, z^{(2)} \rangle$, while model 4 has three unknown parameters $\mathbf{z} = \langle z^{(1)}, z^{(2)} \rangle$ and α.

All four of these models can be used as a reasonable basis for inference about the anchor point of the offender. The classical approach would be to determine which of these four models is the "best" and use that model alone. Moreover, the decision about which model is most appropriate would usually be made via some heuristics that do not account for the particular data in the series under study. Instead, we will measure which of our four candidate models gives us the most information about the observed behavior of the offender while using the fewest parameters.

To proceed, we start by placing a coordinate system on our crime sites; for simplicity we will take crime site 1 = crime site 4 = crime site 6 as our origin, and calculate the distances along the north-south and the east-west direction for each crime; we then obtain the locations in table 2 where distances are measured in miles

To apply our tools of model selection and multimodel inference, we first need to determine the maximum likelihood estimators for the parameters in each model. For model 1, the likelihood function is

$$L_1(\mathbf{y}) = \prod_{i=1}^{n} P_1(\mathbf{x}_i|\mathbf{y}) = \prod_{i=1}^{n} \frac{1}{4\alpha_0^2} \exp\left(-\frac{\pi}{4\alpha_0^2}|\mathbf{x}_i-\mathbf{y}|^2\right)$$

Crime	x	y
1	0	0
2	-1.94390	1.03087
3	-2.59456	2.93647
4	0	0
5	-1.26650	1.30725
6	0	0

Table 2:: Scaled coordinates of the crime series, measured in miles east-west (x) and north-south (y) from crime site 1 = crime site 4 = crime site 6. Distances north or east are positive and distances south or west are negative.

where the series crimes are at the locations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ (with $n = 6$). The maximum likelihood estimator $\hat{\mathbf{z}}_1$ is the value of $\mathbf{y}$ that makes $L_1(\mathbf{y})$ as large as possible. It is simple to verify that this will occur at the mean center of the crime series, which for this series occurs at $\hat{\mathbf{z}}_1 = \langle -0.967494, 0.879098 \rangle$.

For model 2, the likelihood function is now

$$L_2(\mathbf{y}, a) = \prod_{i=1}^{n} P_2(\mathbf{x}_i | \mathbf{y}, a) = \prod_{i=1}^{n} \frac{1}{4a^2} \exp\left(-\frac{\pi}{4a^2} |\mathbf{x}_i - \mathbf{y}|^2 \right)$$

The maximum likelihood estimator is now the pair $(\hat{\mathbf{z}}_2, \hat{\alpha}_2)$ that makes L_2 as large as possible. As before, $\hat{\mathbf{z}}_2$ is still the mean center of the crime sites $\hat{\mathbf{z}}_2 = \langle -0.967494, 0.879098 \rangle$, while $\hat{\alpha}_2 = 1.31726$.

For model 3, we have the likelihood function

$$L_3(\mathbf{y}) = \prod_{i=1}^{n} P_3(\mathbf{x}_i | \mathbf{y}) = \prod_{i=1}^{n} \frac{2}{\pi \alpha_0^2} \exp\left(-\frac{2}{\alpha_0} |\mathbf{x}_i - \mathbf{y}| \right).$$

Now the maximum likelihood estimator of the anchor point $\hat{\mathbf{z}}_3$ turns out to be the center of minimum distance of the crime sites which occurs at $\hat{\mathbf{z}}_3 = \langle 0, 0 \rangle$.

Model 4 is handled similarly, with the likelihood function

$$L_4(\mathbf{y}, a) = \prod_{i=1}^{n} P_4(\mathbf{x}_i | \mathbf{y}, a) = \prod_{i=1}^{n} \frac{2}{\pi a^2} \exp\left(-\frac{2}{a} |\mathbf{x}_i - \mathbf{y}| \right).$$

The maximum likelihood estimate $\hat{\mathbf{z}}_4$ of the offender anchor point is again the center of minimum distance of the crime series $\hat{\mathbf{z}}_4 = \langle 0, 0 \rangle$ while the maximum likelihood estimate of the average offense distance is $\hat{\alpha}_4 = 1.32316$.

With these maximum likelihood estimators, we can then simply calculate the values of AIC, Δ, and w for each of the four models; the results are summarized in Table 3

This table contains a great deal of information. First, the models that fit the data best are models 2 and then 4, as they have the largest values of the likelihood function. This is not surprising however, as these models also have more parameters (3) to use to fit the data. When we use AIC to also evaluate the effectiveness of each model while applying parsimony of parameters, we discover instead that models 1 and 3 are the best choices, followed by models 2 and 4.

Model	K	$\ln L$	$\hat{\mathbf{z}}$	AIC	Δ	w
1	2	-18.3333	$\langle -0.967494, 0.879098 \rangle$	40.6666	0	0.320212
2	3	-17.6245	$\langle -0.967494, 0.879098 \rangle$	41.2489	0.582278	0.23933
3	2	-18.4423	$\langle 0, 0 \rangle$	40.8847	0.218024	0.28714
4	3	-18.0698	$\langle 0, 0 \rangle$	42.1396	1.47294	0.153318

Table 3:: Summary of results for Models 1-4 using AIC for all six crimes in the series

Model	K	$\ln L$	$\hat{\mathbf{z}}$	AICc	Δ	w
1	2	-18.3333	$\langle -0.967494, 0.879098 \rangle$	44.6666	0	0.521056
2	3	-17.6245	$\langle -0.967494, 0.879098 \rangle$	53.2489	8.582278	0.00713291
3	2	-18.4423	$\langle 0, 0 \rangle$	43.8847	0.218024	0.467241
4	3	-18.0698	$\langle 0, 0 \rangle$	54.1396	9.47294	0.00456943

Table 4:: Summary of results for Models 1-4 using AICc for all six crimes in the series

We also notice that the values of Δ for all four models satisfy $0 \le \Delta \le 2$, which indicates that all four models are reasonably supported by this data. In particular, this means that AIC does not conclude that one of the four models is dramatically better than the other three. Thus, it is reasonable to base our inference not on any single model, but on the combination. For example, a point estimate for the offender's anchor point would then be

$$w_1\hat{\mathbf{z}}_1 + w_2\hat{\mathbf{z}}_2 + w_3\hat{\mathbf{z}}_3 + w_4\hat{\mathbf{z}}_4 = \langle -0.541354, 0.491892 \rangle.$$

Because the number of elements in the series (6) is small relative to the number of parameters in our model (2 or 3), the small sample correction is a better choice than the full AIC. Table 4 provides the corresponding results when AIC is replaced by AICc.

In this table, the effect of the small sample correction is apparent. As before, models 1 and 3 are still considered to be the best models, with AICc values near 43, but now the values of Δ for models 2 and 4 are nearer to 10, which suggests that they are significantly less informative than models 1 and 3. In particular, the small sample correction is saying that, in effect, there is insufficient information in the data to estimate both the anchor point $\mathbf{z}$ and the average offense distance α for the models in this collection.

The values of w give nearly 99% of their total weight to models 1 and 3, and the corresponding multimodel point estimate for the offender's anchor point is now

$$w_1\hat{\mathbf{z}}_1 + w_2\hat{\mathbf{z}}_2 + w_3\hat{\mathbf{z}}_3 + w_4\hat{\mathbf{z}}_4 = \langle -0.51102, 0.46433 \rangle.$$

Our calculations have shown us that, for this crime series and these models, the normal distribution provides a slightly more informative model for offender behavior than does a negative exponential model. It is instructive however, to repeat these calculation using only the first five elements of the crime series, as would be done for example by an analyst analyzing the series in the nearly two weeks after the fifth crime in the series, but before the sixth crime. The results of this analysis are summarized in Table 5

The most interesting thing to note about this case is that now model 3 is considered to be better than model 1, the opposite of the case when all six crimes in the series were considered. In

Model	K	$\ln L$	$\hat{\mathbf{z}}$	AICc	Δ	w
1	2	-15.3201	$\langle -1.16099, 1.05492 \rangle$	40.6402	0.235503	0.47057
2	3	-14.7584	$\langle -1.16099, 1.05492 \rangle$	60.7484	20.3437	0.00002
3	2	-15.2023	$\langle -1.27051, 1.28143 \rangle$	40.4047	0	0.529375
4	3	-14.8402	$\langle -1.27051, 1.28143 \rangle$	59.6803	19.2756	0.000035

Table 5:: Summary of results for Models 1-4 using AICc for only the first five crimes in the series

particular, the choice of the best of these four models clearly depends on the particulars of the series under consideration. We also notice that both models were considered nearly equally good. When all six crimes were considered, model 1 was preferred to model 3 by a score of 52% to 47%, but when the first five crimes were considered model 3 was preferred to model 1 by a score of 53% to 47%. This suggests that, rather than trying to identify the "best" model, a better approach would be to identify the good models and weight them accordingly.

Also apparent here is the value of multimodel inference. Our analysis here has shown that both model 1 (normal distribution) and model 3 (negative exponential distribution) when compared are of nearly equal explanatory power. The point estimate of the offender's anchor point for the normal distribution is the mean center of the crime series, while the point estimate from the negative exponential distribution is the center of minimum distance of the crime series. When all six crimes are considered, these two points had the values $\hat{\mathbf{z}}_{\text{mean center}} = \langle -0.967494, 0.879098 \rangle$ and $\hat{\mathbf{z}}_{\text{cmd}} = \langle 0, 0 \rangle$ which are more than a mile and a quarter apart. If we limited our analytic techniques to only selecting the "best" model then necessarily we can select no more than one of these choices with their very different predicted anchor points. On the other hand, we have also seen that the "best" model changed depending on how far along the crime series has progressed. Thus using just one model would have put ourselves in a position where we must make very different predictions for the anchor point depending on the choice of the model while knowing that these models have essentially the same explanatory power.

The advantage of multimodel inference then, is that instead of forcing us to select just one of these very comparable models, we can meaningfully combine them.

Conclusions and Future Work

Our illustration has shown us that we can apply model selection and multimodel inference to a real problems in geographic profiling. We also saw how these methods can help us make better informed decisions about how to draw inference about the offender's anchor point from the locations of the crime series.

On the other hand, the illustration does not show that the models selected as most informative (models 1 and 3) are the "best" choices for geographic profiling in general. All that can be concluded from the analysis is that these were the most informative of the four choices studied for the crime series under consideration. In particular, it is possible and even likely that there are better models both for this particular data set and for other serial crimes.

Instead we have presented a technique that can be used to compare and evaluate models of offender behavior against one another; this method can also be used to draw inferences from multiple models in a reasonable fashion for data taken from a real crime series.

Although these methods allow us to compare and evaluate models, they do not specify a way to craft or create these models. In particular, it is an important open question to determine what sorts of models should be used in these comparisons, and then to see if the mathematically "best" models yield useful predictions for law enforcement.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *Second international symposium on information theory: Tsahkadsor, Armenia, U.S.S.R., September 2-8, 1971* (pp. 267–281). Bucharest: Akadémiai Kiadó.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Akaike, H. (1977). On entropy maximization principle. In P. R. Krishnaiah (Ed.), *Applications of statistics: Proceedings of the symposium held at Wright State University, Dayton, Ohio, 14-18 June 1976* (pp. 27–41). Amsterdam: North Holland Pub. Co.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY: Springer Verlag.

Canter, D., Coffey, T., Huntley, M., & Missen, C. (2000). Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, *16*(4), 457–478.

Canter, D., & Larkin, P. (1993). The environmental range of serial rapists. *Journal of Environmental Psychology*, *13*, 63-69.

Katsanevakis, S. (2006). Modelling fish growth: Model selection, multi-model inference and model selection uncertainty. *Fisheries Research*, *81*, 229-235.

Kocsis, R. N., Cooksey, R. W., Irwin, H. J., & Allen, G. (2002). A further assessment of circle theory for geographic psychological profiling. *The Australian and New Zealand Journal of Criminology*, *35*(1), 43–62.

Kocsis, R. N., & Irwin, H. J. (1997). An analysis of spatial patterns in serial rape, arson and burglary: The utility of the circle theory of environmental range for psychological profiling. *Psychiatry, Psychology, and the Law*, *4*(2), 195–206.

LeBeau, J. L. (1987). The methods and measures of centrography and the spatial dynamics of rape. *Journal of Quantitative Criminology*, *3*(2), 125–141.

Levine, N. (2009a). *CrimeStat: A spatial statistics program for the analysis of crime incident locations (v. 3.2).* Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. Retrieved July 2009, from `http://www.icpsr.umich.edu/crimestat`

Levine, N. (2009b). *Update notes to version 3.2. CrimeStat: A spatial statistics program for the analysis of crime incident locations.* Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. Retrieved July 2009, from `http://www.icpsr.umich.edu/crimestat`

Levine, N., & Lee, P. (2009). Bayesian journey-to-crime modelling of juvenile and adult offenders by gender in Manchester. *Journal of Investigative Psychology and Offender Profiling*, *6*, 237–251.

Liddle, A. R. (2007). Information criteria for astrophysical model selection. *Monthly Notices of the Royal Astronomical Society: Letters*, *377*, L74-L78.

Link, W. A., & Barker, R. J. (2006). Model weights and the foundations of multimodel inference. *Ecology*, *87*(10), 2626-2635.

Meaney, R. (2004). Commuters and marauders: An examination of the spatial behavior of serial criminals. *Journal of Investigative Psychology and Offender Profiling*, *1*, 121–137.

O'Leary, M. (2009). The mathematics of geographic profiling. *Journal of Investigative Psychology and Offender Profiling*, *6*, 253–265.

O'Leary, M. (2010). Implementing a Bayesian approach to criminal geographic profiling. In *COM.Geo '10: Proceedings of the 1st international conference and exhibition on computing for geospatial research & application.* New York, NY, USA: ACM.

Poeter, E., & Anderson, D. (2005). Multimodel ranking and inference in ground water modeling. *Ground Water*, *43*(4), 597-605.

Rossmo, K. (2000). *Geographic profiling*. Boca Raton, FL: CRC Press.

Snook, B., Zito, M., Bennell, C., & Taylor, P. J. (2005). On the complexity and accuracy of geographic profiling strategies. *Journal of Quantitative Criminology*, *21*(1), 1–26.

Stephens, P. A., Buskirk, S. W., & Martínez del Rio, C. (2006). Inference in ecology and evolution. *Trends in Ecology and Evolution*, *22*(4), 192–197.

Ye, M., Meyer, P. D., & Neuman, S. P. (2008). On model selection criteria in multimodel analysis. *Water Resources Research*, *44*(3), W03428.