Centroids, Clusters, and Crime: Anchoring the Geographic Profiles of Serial Criminals

Anil S. Damle
Colin G. West
Eric J. Benzel

University of Colorado–Boulder
Boulder, CO

Advisor: Anne Dougherty

Abstract

A particularly challenging problem in crime prediction is modeling the behavior of a serial killer. Since finding associations between the victims is difficult, we predict where the criminal will strike next, instead of whom. Such predicting of a criminal's spatial patterns is called *geographic profiling*.

Research shows that most violent serial criminals tend to commit crimes in a radial band around a central point: home, workplace, or other area of significance to the criminal's activities (for example, a part of town where prostitutes abound). These "anchor points" provide the basis for our model.

We assume that the entire domain of analysis is a potential crime spot, movement of the criminal is uninhibited, and the area in question is large enough to contain all possible strike points. We consider the domain a metric space on which predictive algorithms create spatial likelihoods. Additionally, we assume that the offender is a "violent" serial criminal, since research suggests that serial burglars and arsonists are less likely to follow spatial patterns.

There are substantial differences between one anchor point and several. We treat the single-anchor-point case first, taking the spatial coordinates of the criminal's last strikes and the sequence of the crimes as inputs. Estimating the point to be the centroid of the previous crimes, we generate a "likelihood crater," where height corresponds to the likelihood of a future crime at that location. For the multiple-anchor-point case, we use a cluster-finding and sorting method: We identify groupings in the data and build a likelihood crater around the centroid of each. Each cluster is given weight according to recency and number of points. We test single point vs. multiple points by

The UMAP Journal 31 (2) (2010) 129–148.

using the previous crimes to predict the most recent one and comparing with its actual location.

We extract seven datasets from published research. We use four of the datasets in developing our model and examining its response to changes in sequence, geographic concentration, and total number of points. Then we evaluate our models by running blind on the remaining three datasets.

The results show a clear superiority for multiple anchor points.

Introduction

The literature on geographic patterns in serial crimes shows a strong patterning around an anchor point—a location of daily familiarity for the criminal. We build prediction schemes based on this underlying theory and produce a surface of likelihood values and a robust metric.

The first scheme finds a single anchor point using a center-of-mass method; the second scheme assumes two to four anchor points and uses a cluster-finding algorithm to sort and group points. Both schemes use a statistical technique that we call *cratering* to predict future crime locations.

Background

The arrest in 1981 (and subsequent conviction) of Peter Sutcliffe as the "Yorkshire Ripper" marked a victory for Stuart Kind, a forensic biologist whose application of mathematical principles had successfully predicted where the Yorkshire Ripper lived.

Today, information-intensive models can be constructed using heat-map techniques to identify the hot spots for a specific type of crime, or to derive associations between the rate of criminal activity and attributes of a location (such as lighting, urbanization, etc.) [Boba 2005].

"Geographically profiling" the crimes of a single criminal has focused on locating the criminal's anchor points—locations (such as a home, workplace, or a relative's house) at which he spends substantial amounts of time and to which he returns regularly between crimes.

Canter and Larkin [1993] proposed that a serial criminal's home (or other anchor point) tends to be contained within a circle whose diameter is the line segment between the two farthest-apart crime locations; and this is true in the vast majority of cases [Kocsis and Irwin 1997]. Canter et al. [2000] found that for serial murders, generalizations of such techniques on average reduce the area to be searched by nearly a factor of 10.

By contrast, forecasting *where* a criminal will strike next has not been explored deeply [Rossmo 1999]. Paulsen and Robinson [2009] observe that for many U.S. police departments there are substantial practical, ethical, and legal issues involved in collecting the data for a detailed mapping of criminal tendencies, with the result that only 16% of them employ a computerized mapping technique.

Our treatment of the problem will employ anchor-point-finding algorithm. We generate likelihood surfaces that act as a prioritization scheme for regions to monitor, patrol, or search.

Assumptions

Domain is Approximately Urban

We use the word "urban" to denote features of an urbanized area that simplify our treatment: The entire domain is a potential crime spot, the movement of the criminal is completely unconstrained, and the area is large enough to contain all possible strike points. It is important to note, however, that even for serial crime committed in suburbs, villages, or spread between towns, the urbanization condition holds on the subset of the map in which crimes are regularly committed. To see this, consider the three urbanization conditions separately:

- **Entire domain is a potential crime spot.** Every neighborhood contains a possible crime location. Such an assumption is made by nearly all geographic profiling techniques [Canter et al. 2000; Rossmo 1999]

  It is obvious that every domain will violate these conditions to some extent: All but the most inventive serial killers, for example, will not commit a crime in the middle of a lake, or in the uninhabited farmland between small towns. Nevertheless, this observation simply requires that the output of the model be interpreted intelligently. In other words, while we assume for simplicity that the entire map is a potential target, police officers interpreting the results can easily ignore any predictions we make which fall into an obvious "dead zone."

- **Criminal's movement is unconstrained.** Because of the difficulty of finding real-world distance data, we invoke the "Manhattan assumption": There are enough streets and sidewalks in a sufficiently grid-like pattern that movements along real-world movement routes is the same as "straight-line" movement in a space discretized into city blocks [Rossmo 1999]. Kent [2006] demonstrated that across several types of serial crime, the Euclidean and Manhattan distances are essentially interchangeable in predicting anchor points.

- **Domain contains all possible strike points.** This condition says that the two conditions above hold on a sufficiently large area.

Taken together, these three conditions describe the region of interest as a metric space in which

- The subset of potential targets is dense,

- the metric is the L^2 norm, and
- the space is "complete": Sequences of crimes do not lead to predictions of crimes outside the space.

Violent Serial Crimes by a Single Offender

- **Focus on violent crimes.** Geographic profiling is most successful for murders and rapes, with the average anchor-point prediction algorithm being 30% less effective for criminals who are serial burglars or arsonists [Canter et al. 2000; Rossmo 1999].
- **Serial crimes.** We take serial killing (or violent crime) as involving "three or more people over a period of 30 or more days, with a significant cooling-off period between" [Holmes and Holmes 1998].
- **Single offender.**

Spatial Focus

Use of temporal data is problematic. Time data can be inaccurate. Also, while research has found cyclical patterns within the time between crimes, these patterns don't associate directly to predicting the next geographic location. What is useful is general trends in spatial movement over an ordering of the locations. We hence ignore specific time data in crime sets except for ordering of the crime sequence.

Developing a Serial Crime Test Set

Existing Crime Sets

Researchers have compiled databases of serial crimes for their own use: Rossmo's FBI and SFU databases [Rossmo 1999], LeBeau's San Diego Rape Case dataset [LeBeau 1992], and Canter's Baltimore crime set [Canter et al. 2000]. Each of these databases was developed with specific methods of integrity and specific source locations. These proprietary databases are not available to us, so we are faced with two options: simulate serial criminal data or find an indirect way of using the private data.

The Problem with Simulation

Simulation might seem like an attractive solution to the lack of data. However, utterly random crime-site generation would contradict the underlying assumption of a spatial pattern to serial crimes, while generating sites according to an underlying distribution would prejudge the pattern! Actual data must be used if there is to be any confidence in the model.

An Alternative: Pixel Point Analysis

Instead, we "mine" the available data, in Rossmo [1995] and in the spatial analysis of journey-to-crime patterns in serial rape cases in LeBeau [1992]. LeBeau depicts the data as scatterplots, which we rasterize and re-render with scaling. **Figure 1** shows an example of this process, which we applied to seven criminals' data (four killers, three rapists). The rape sequences have an explicit ordering, while the murder sequences are unordered.

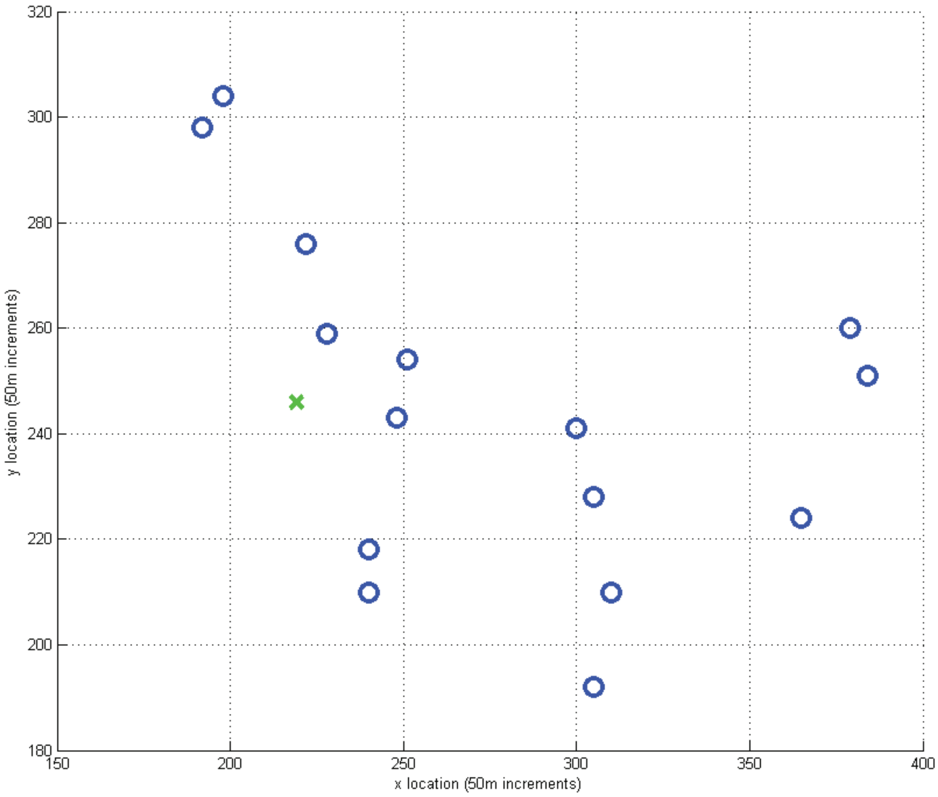


Figure 1. Re-rendering of scatterplot in LeBeau [1992] for Offender B.

Metrics of Success

A successful model must outperform random predictions.

The Effectiveness Multiplier

We assume that police effectiveness is proportional to resources allocated, and that the resources allocated at a location are proportional to the likelihood given by the model. We say that one model outperforms another if it recommends allocating more police resources to where the next crime is actually committed.

We assess how much one model outperforms another by the *effectiveness multiplier* κ, the ratio of resources allocated at the crime point under each model. Alternatively, κ is the ratio of the percentage of total department resources allocated to the point. Since the total resources are the same in both models, we can evaluate κ simply as

$$\kappa = \frac{Z_1(CrimePoint)}{Z_2(CrimePoint)}$$

where Z_i is the likelihood function of model i and $CrimePoint$ is the actual location of the next crime.

A randomly guessing algorithm will have a uniform distribution over locations and hence a flat likelihood plot. We compare our model to such a random guess by computing a *standard effectiveness multiplier* κ_s:

$$\kappa_s = \frac{Z_{\text{our model}}(CrimePoint)}{Z_{\text{flat}}(CrimePoint)}$$

A value of 1 would indicate that the model was no better than a random guess, and a value less than 1 would indicate that the model *misled* the police.

Robustness of the Metric

We also want to compare the success of our algorithm across multiple datasets. It is legitimate to compare κ_s values between two datasets only if they have the same ratio of the killer's active region to the total area. The size of the killer's active region cannot be precisely known; however, we employ a standard technique to make this condition approximately true. According to Canter and Larkin [1993] and Paulsen [2005], in more than 90% of cases all future crimes fall within a square whose side length is the maximum distance between previous crime points and whose center is the centroid of the data. For each of our datasets, we construct such a square and then multiply its side length by 3, thereby creating an overall search area nearly 9 times as large as the criminal's active area. This ratio is nearly constant for all datasets, allowing us to compare effectiveness multipliers.

Two Schemes for Spatial Prediction

Journey-to-crime research for violent serial crimes strongly suggests that serial crime is patterned around a criminal's home, workplace, or other place of daily activity [Godwin and Rosen 2005; Holmes and Holmes 1998; Kocsis and Irwin 1997; Rossmo 1999; Snook et al. 2005]; so researchers have developed and evaluated methods of finding such a crime centroid and investigated it as an anchor point in the criminal's activity. In most research, this anchor point is the serial criminal's home. This method has been tested and found to reduce the necessary search area by a factor of 10.

We develop two schemes, one for a single anchor point and the other for multiple anchor points.

Single Anchor Point: Centroid Method

Figure 2 shows our algorithm to predict likely crime locations using a single anchor point.

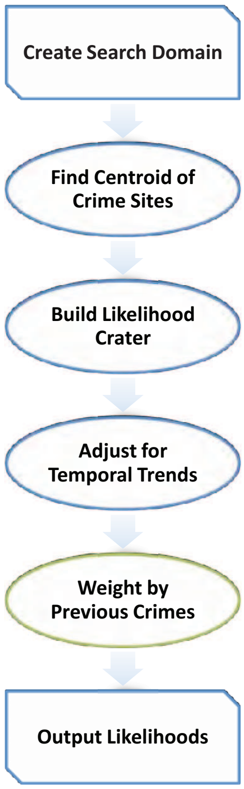


Figure 2. Flow chart for centroid method.

Algorithm

Create Search Domain

We construct the smallest square that contains every previous crime, then scale up each dimension by a factor of 3. This ensures that all of our fundamental assumptions about the underlying domain are satisfied, and the consistent scale factor of 3 allows us to compare the algorithm between datasets.

Find Centroid of Crime Sites

The anchor point is the average of the n crime coordinates (x_i, y_i).

Building a Likelihood Crater

We predict future crime locations using the "journey-to-crime" model, which says that the criminal's spatial pattern of crime around an anchor point does not change. A rough first prediction might be to draw a large shape (circle, square, polygon, etc.) around this anchor point based on the largest distance from a crime point to the anchor point; such a method is incredibly ineffective compared to the largest-circle guess that we described earlier [Paulsen 2005].

We instead use a cratering technique first described by Rossmo [1999]. The two-dimensional crime points x_i are mapped to their radius from the anchor point a_i, that is, we have $f : x_i \to r_i$, where $f(x_i) = \|x_i - a_i\|_2$ (a shifted modulus). The set r_i is then used to generate a crater around the anchor point.

There are two dominating theories for the pattern serial crimes follow around an anchor point:

- **There is a buffer zone around the anchor point.** The criminal commits crimes in an annulus centered at the anchor point [Kocsis and Irwin 1997]. This theory is often modeled using the positive portion of a Gaussian curve with parameters the mean and the variance of the $\{r_i\}$.
- **Crimes follow a decaying exponential pattern from the anchor point.**

Both theories have been substantiated by journey-to-crime research. We seek a distribution that would model either theory, depending on the pattern in the crime sequence. For this we turn to the flexibility of the gamma distribution, which offers a "shifted-Gaussian"-like behavior when points lie farther away but a curve similar to a negative exponential when the parameters are small.

Define the random variable X_i to be the distance between the ith crime point and the anchor point r. We let each X_i have a gamma distribution with parameters k and θ: $X_i \sim \Gamma(k, \theta)$, with probability density function (pdf)

$$f(x; k, \theta) = \frac{\theta^k}{\Gamma(k)}\, x^{k-1} e^{-\theta x}.$$

We assume independence of the X_i and use the maximum likelihood estimates of the parameters k and θ as calculated by the *gamfit* function in MatLab.

We then build the crater of likely crime locations using the resulting distribution. For every point in the search region, we evaluate the pdf. We then normalize so that the volume under the likelihood surface is exactly 1.

Applying this method to the set of crime locations of Peter Sutcliffe, the "Yorkshire Ripper," we get the heat map of **Figure 3**.

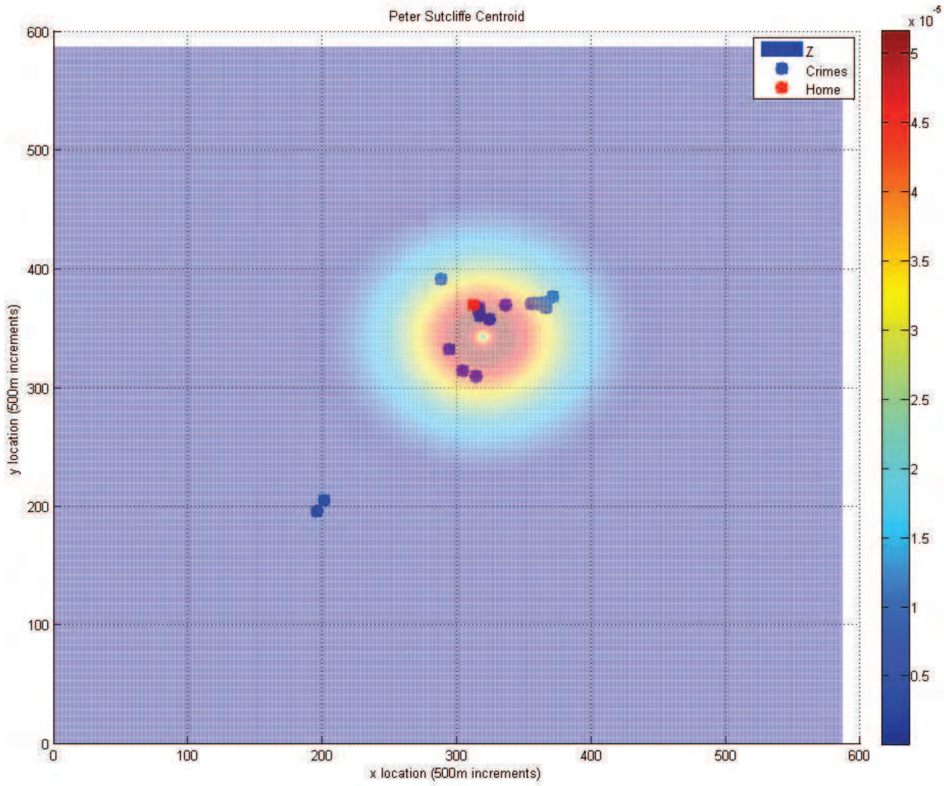


Figure 3. Heat map showing cratering technique applied to the crime sequence of Peter Sutcliffe.

Adjust for Temporal Trends

We would like our prediction to account for any radial trend in time (the criminal becoming more bold and committing a crime closer or further from home): An outward or inward trend in r_i may suggest that the next crime will follow this trend [Kocsis and Irwin 1997]. We let $\tilde{X} = X + \overline{\triangle r}$, where $\triangle r = r_n - r_{n-1}$. The new random variable $\tilde{X}$ gives our intended temporal adjustment in expected value:

$$E[\tilde{X}] = E[X + \overline{\triangle r}] = E[X] + \overline{\triangle r}.$$

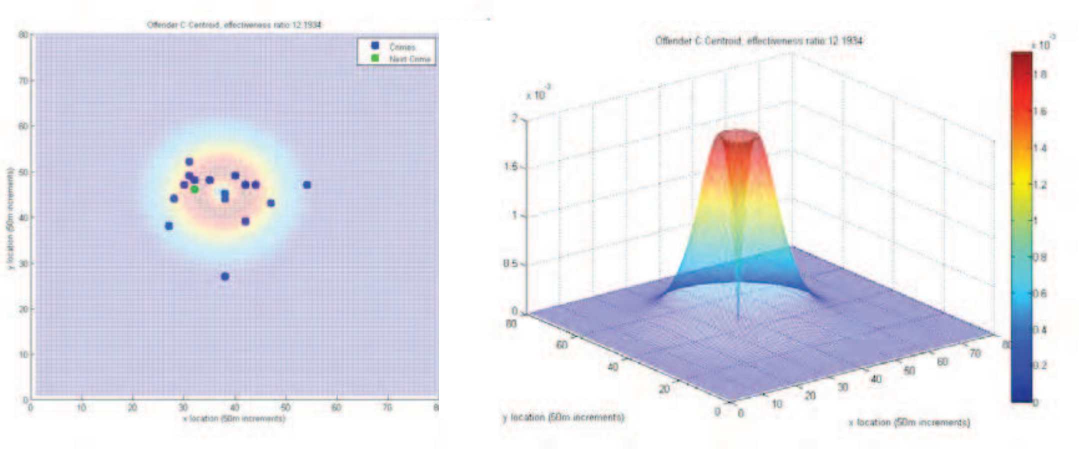
Results and Analysis

To evaluate our method, we feed it data from three serial-rape sprees.

In each test case, we remove the data point for the final crime and produce a likelihood surface $Z(x, y)$. We then estimate the location of the final crime and compute the standard effectiveness multiplier κ_s.

Offender C

Our first test dataset, for Offender C, is a comparative success (**Figure 4**). With $\kappa_s \approx 12$, it is a full order of magnitude better to distribute police resources using this model instead of distributing them uniformly.



a. Heat map. **b.** Surface plot.

Figure 4. Offender C predictions of location of final crime, from centroid model of previous crimes.

The next-crime estimate falls satisfyingly near the isoline of maximum height; but there 120 grid squares are rated greater or equal in likelihood, meaning 0.3 km^2 must be patrolled at the same or greater intensity. This area is small in an absolute sense; but it is comparatively large given that the vast majority of the crimes in this case were committed within an area of 1 km^2.

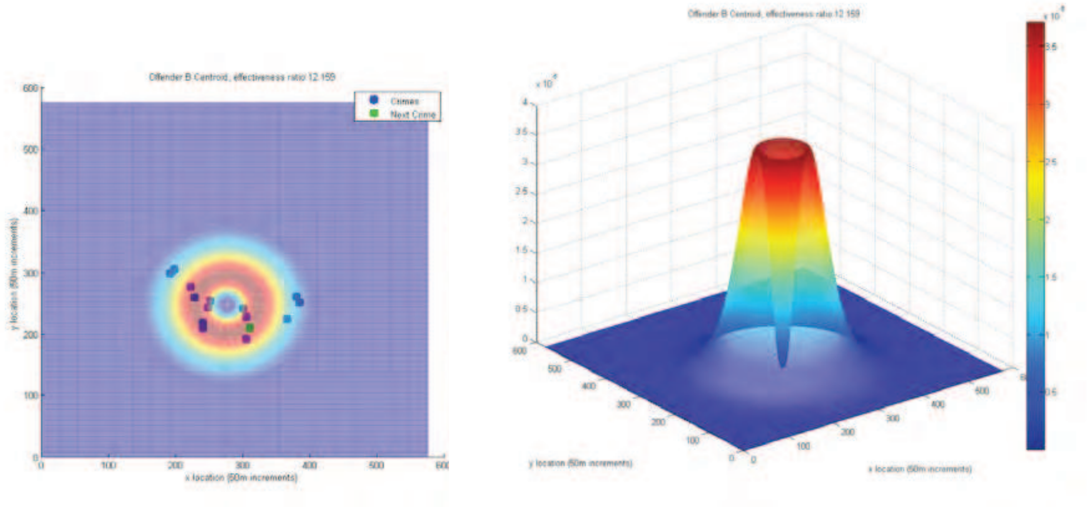
With $\overline{\Delta r} = -0.276$, the temporal corrections in this distribution are negligible; our projection is simply a radially symmetric fit to the geographic dispersion of previous crimes. The surface plot also shows the steepness of the "inside" of the crater, as the geographic distribution apparently suggests a small buffer zone around the centroid.

Offender B

Our second test dataset, for Offender B (see **Figure 5**), is similarly successful, with $\kappa_s \approx 12$.

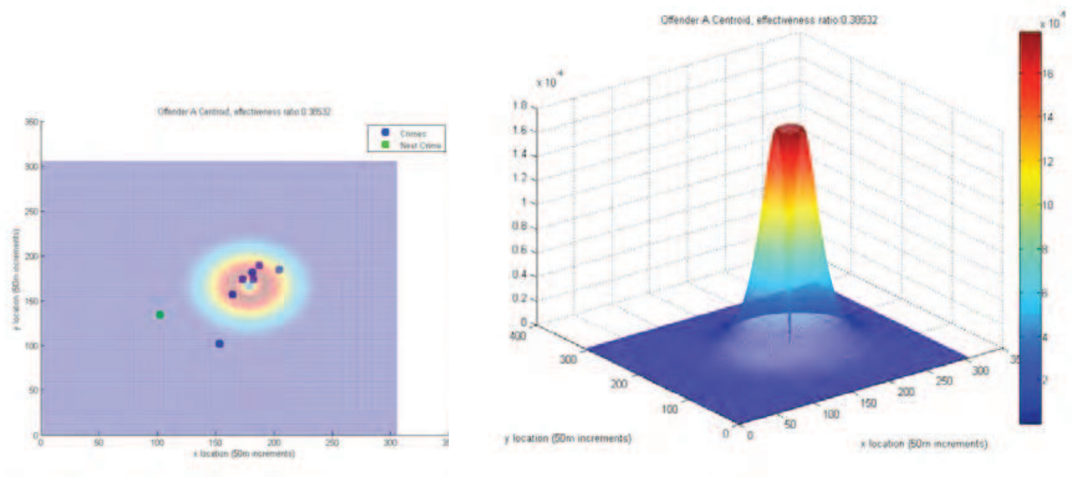
Offender A

For Offender A, we find a clear example of how our model can fail. The last crime (see **Figure 6**) is one of two substantial outliers, and in fact, with a standard effectiveness multiplier of $\kappa_s \approx 0.4$, our model is less helpful than random guessing. However, the majority of previous crimes are still well described by the model, so the assumption of an anchor point somewhere within the crater region does not appear to have been a poor one. Some scheme, whim, or outside influence caused the criminal to deviate from his previous pattern.



a. Heat map. **b.** Surface plot.

Figure 5. Offender B predictions of location of final crime, from centroid model of previous crimes.



a. Heat map. **b.** Surface plot.

Figure 6. Offender A predictions of location of final crime, from centroid model of previous crimes.

Multiple Anchor Points: Cluster Method

Our second method explicitly assumes at least two anchor points (for example, a home and a workplace) and treats each as the centroid of its own local cluster of crimes. This method requires determining an appropriate number of clusters, which we derive from the locations of the previous crimes.

Algorithm

The basis of this algorithm is a hierarchical clustering scheme [Jain, Murty, and Flynn 1999]. Once clusters are found, the previous algorithm is applied at each cluster centroid.

Finding Clusters in Crime Sequences

We force a minimum of 2 clusters and a maximum of 4. The clustering algorithm is accomplished in a 3-step process.

1. Compute the distances between all crime locations, using the Euclidean distance.

2. Organize the distances into a hierarchical cluster tree, represented by a *dendrogram*. The cluster tree of data points $P_1, \ldots, P_N$ is built up by first assuming that each data point is its own cluster. The dendrogram for Offender B is shown in **Figure 7**.

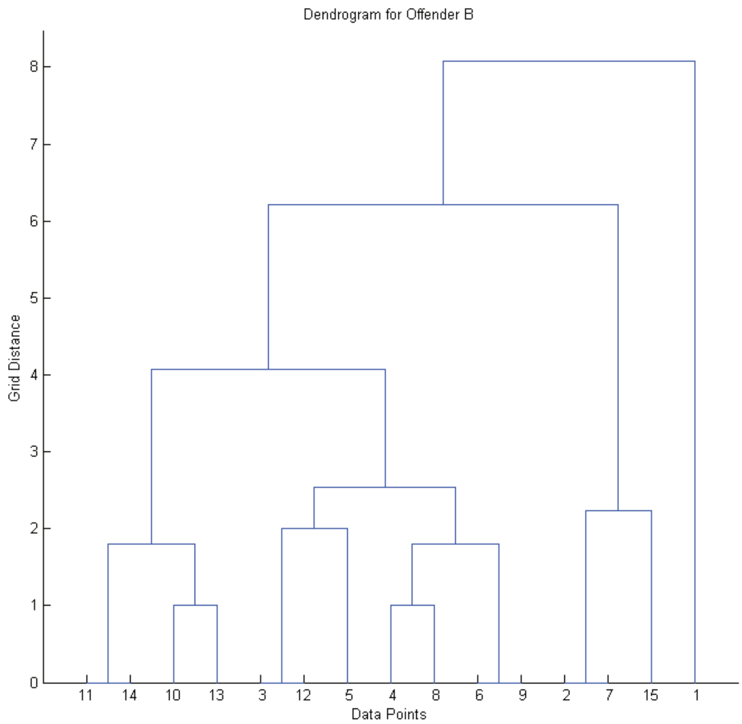


Figure 7. Dendrogram for Offender B.

3. Merge the two clusters that are the closest (in distance between their centroids), and continue such merging until the desired number of clusters is reached. These cluster merges are plotted as the horizontal lines in the dendrogram, and their height is based on the distance between merged clusters at the time of merging.

To determine the optimal number of clusters, we use the notion of *silhouettes* [Rousseeuw 1987]. We denote by $a(P_i)$ the average distance from P_i to all other points in its cluster and by $b(P_i, k)$ the average distance from P_i to points in a different cluster C_k. Then the silhouette of P_i is

$$s(P_i) = \frac{\left[\min\limits_{k|P_i \notin C_k} b(P_i, k)\right] - a(P_i)}{\max\left(a(P_i), \min\limits_{k|P_i \notin C_k} b(P_i, k).\right)} \tag{1}$$

The silhouette s can take values in $[-1, 1]$: The closer $s(P_i)$ is to 1, the better P_i fits into its current cluster; and the closer $s(P_i)$ is to -1, the worse it fits within its current cluster.

To optimize the number of clusters, we compute the clusterings for 2, 3, and 4 clusters. Then for each number of clusters, we compute the average silhouette value across every point that is not the only point in a cluster. (We ignore silhouette values at single-point clusters because otherwise such clusters influence the average in an undesirable way.) We then find the maximum of the three average silhouette values. For Offender B, we found average silhouette values of 0.52, 0.50, and 0.69 for 2, 3, and 4 clusters. So in this case, we go for four clusters. The cluster groupings computed by the algorithm are shown in **Figure 8**. Because the average silhouette value tends to increase with the number of clusters, we cap the possible number of clusters at 4.

Cluster Loop Algorithm

We compute the likelihood surface for the centroid of each cluster.

If a cluster contains a single point, we do not assume that this cluster represents an anchor point; instead, we treat this point as an outlier. We use a Gaussian distribution centered at the point as the likelihood surface, with mean the expected value of the gamma distribution placed over every anchor point of a cluster that has more than one point.

Combining Cluster Predictions: Temporal and Size Weighting

Using the separate likelihood surfaces computed for each cluster, we create our final surface as a normalized linear combination of the individual surfaces, using weights for the number of points in the cluster (to weight

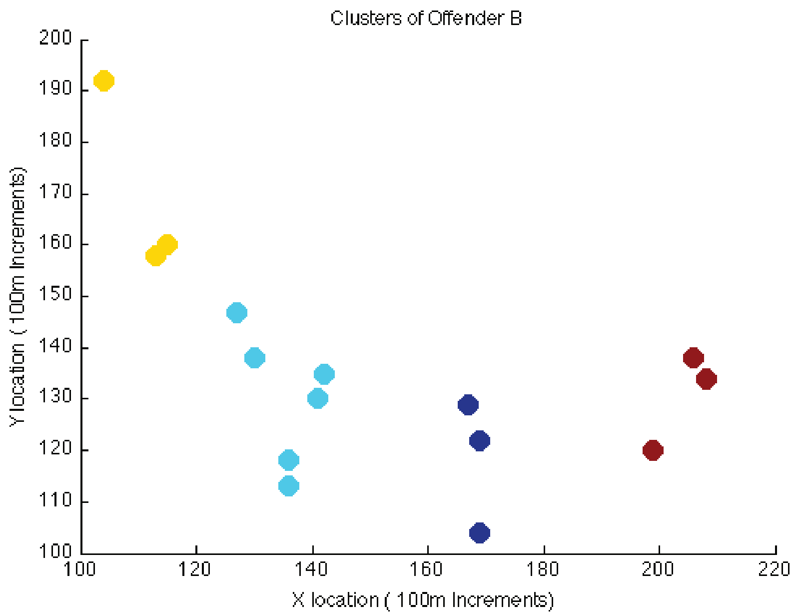


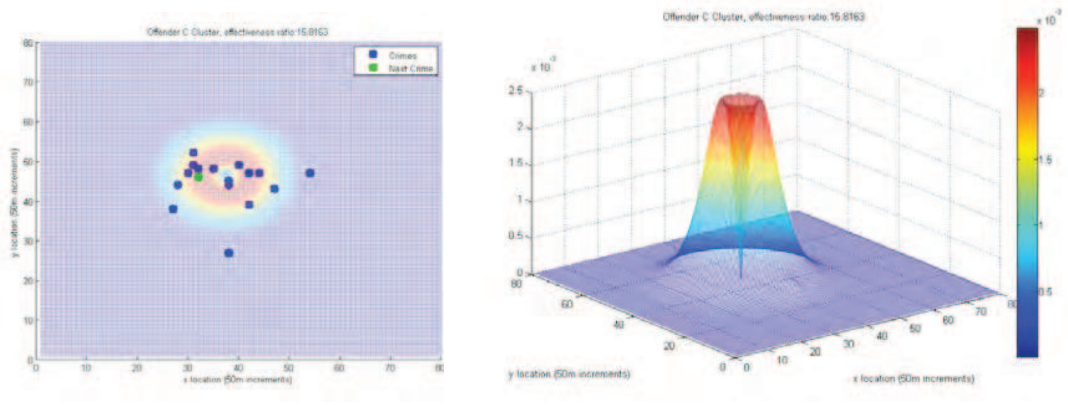
Figure 8. Offender B crime points sorted into 4 clusters; the clusters are colored differently and separated by virtual vertical lines into 3, 6, 3, and 3 locations.

more-common locations) and for the average temporal index of the events in the cluster (to weight more-recent clusters).

Results and Analysis

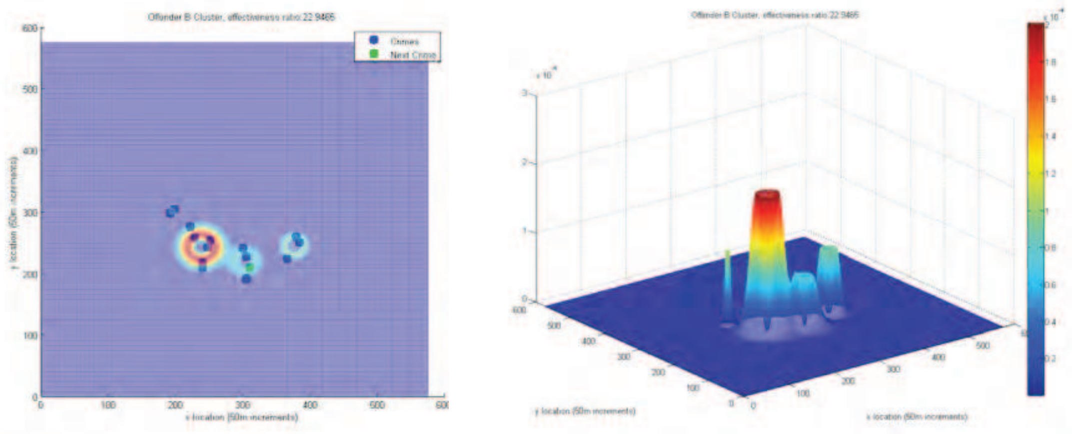
The three test datasets conveniently display the cluster method's superior adaptability.

- **Offender C:** The highly-localized nature of the data points in **Figure 9** means there is little difference between the centroid-method results and the cluster-method results. The only difference is that the cluster method identifies the point directly below the centroid as a cluster of a single point (an outlier) and therefore excludes it from the computation of the larger cluster's centroid (a slight Gaussian contribution from this point's "own" cluster can be seen in the surface plot). This has the effect of slightly reducing the variance and therefore narrowing the fit function; consequently, the standard effectiveness multiplier rises slightly, from about 12 to almost 16.
- **Offender B:** By contrast, **Figure 10** shows the cluster method operating at the other edge of its range, as the silhouette-optimization routine produces four clusters. It might appear that the centroid method outperforms the cluster algorithm for this dataset; after all, the actual crime point no longer appears in the band of maximum likelihood. This is true and intentional, since the model weights the largest cluster most strongly



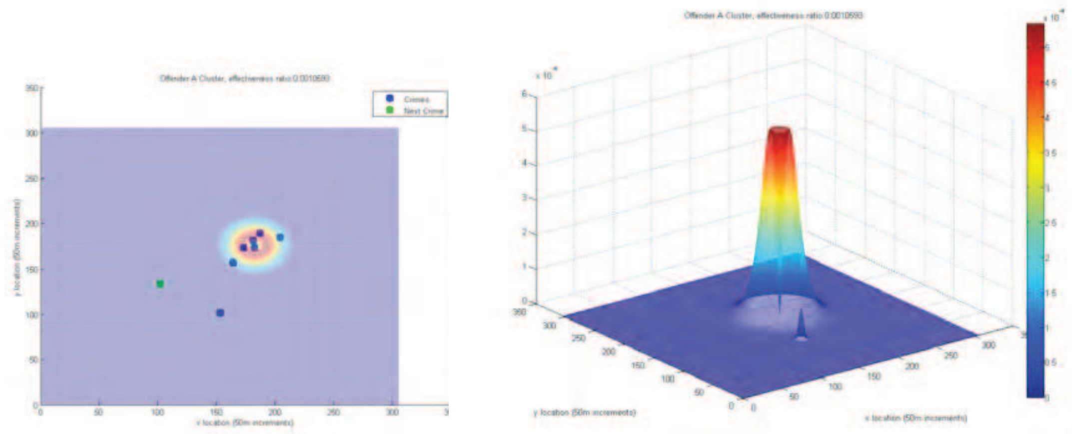
a. Heat map. **b.** Surface plot.

Figure 9. Offender C predictions of location of final crime, from cluster model of previous crimes.



a. Heat map. **b.** Surface plot.

Figure 10. Offender B predictions of location of final crime, from cluster model of previous crimes.



a. Heat map. **b.** Surface plot.

Figure 11. Offender A predictions of location of final crime, from cluster model of previous crimes.

and the "freshest" cluster next-most. Nevertheless, while not accurately predicting that the offender returns to an earlier activity zones, the cluster method *still* outperforms the centroid method with a $\kappa_s \approx 23$. This is because the craters generated by the cluster method are sharper and taller for this dataset, so fewer resources are "wasted" at high-likelihood areas where no crime is committed.

- **Offender A:** Unsurprisingly, the cluster model fares no better than the centroid method (**Figure 11**). Since the outlier points are excluded from the centroid calculation for the larger cluster, the model bets even more aggressively on this cluster, with a resulting $\kappa_s \approx 0$.

Combining the Schemes

Based on further work not shown in this briefer version of our contest paper, we developed a combined method to decide for a given situation whether the centroid method or the cluster method is better, based on calculating running means of the effectiveness multipliers of the two methods as we predict each of the crime locations from the previous ones.

Our combined method would use the clustering algorithm for all three Offenders A, B, and C. One would have to have a a highly-cohesive single cluster of data with no outliers in order for the centroid method to ever prevail. This is as it should be, since even when there appears to be only one true anchor point, the cluster method can reject up to three statistical outliers before computing the centroid, which capability should in general improve the fits and consequently improve the results.

The Model vs. Random Guess, Intuitive Cops, etc.

Although our scant datasets do not allow anything conclusive, our model is a strong candidate over other alternatives, for several reasons:

- The predictions are based on the assumption of trends in serial crime behavior which *have* been tested on large sets of real-world data [Canter et al. 2000; Kocsis and Irwin 1997; Paulsen and Robinson 2009].
- Similar mathematical techniques are used in the anchor-point estimation schemes currently employed, which consistently outperform random guesses when tested across data samples.
- The model is successful in the two of the three real-life datasets on which we tested it.
- Several crimes in each dataset can be predicted well, even in the dataset where our model fails to predict the final, outlier crime point—i.e., in a dataset of 16 crimes, we are also often able to predict the 15th crime

using the preceding 14, the 14th using the preceding 13, etc., all with more success than a random guess.

We do not claim that the model will do better than a police department in assigning resources based on knowledge of the area, a sense of patterning in the offender's crimes, and any "gut feelings" developed about the offender's psychology, based on previous experience in law enforcement.

Additionally, research suggests that with a little informal training, any layperson can perform nearly as well as anchor-point-prediction algorithms in guessing a criminal's home location [Snook, Taylor, and Bennell 2004]. As a result, one might well expect an "intuitive cop" to outperform the model overall, by approximating its mathematical strengths through intelligent estimation plus bringing a breadth of knowledge and experience to bear.

Executive Summary

Overview: Strengths and Weaknesses of the Model

We present a model of where a violent serial criminal will strike next, based on the locations and times of previous crimes. Our algorithm creates a color-coded map of the area surrounding the criminal's previous strikes, with the color at each point indicating the likelihood of a strike there. The model has several key strengths:

- **The model contains no arbitrary parameters**. In other words, most aspects of the model are determined simply by trends observed in datasets about many serial criminals.
- **The model can estimate the level of confidence in its predictions.** Our model first checks how well it would have predicted the criminal's previous crimes, in order to provide an estimate of how well it can predict future crimes.
- **The model understands that police have limited resources**. In particular, the confidence level described above becomes large if we are making good predictions but will shrink again if our predicted areas become so large as to become unhelpful.

At the same time, our model contains some fundamental limitations:

- **The model is applicable only to violent serial criminals**. We claim applicability only for serial killers and rapists, since our research shows that serial burglars and arsonists are more unpredictable and influenced by non-geographic factors.
- **The model cannot predict when a criminal will strike**. While we consider the order of previous crimes in order to predict locations, we do *not* predict a strike *time*.

- **The model cannot make use of underlying map data**. To maintain generality, we do not make any assumptions about the underlying physical geography. A human being must interpret the output (for example, choosing to ignore a prediction in the middle of a lake).
- **The model has not been validated on a large set of empirical data**. Sizable sets of data on serial criminals are not widely available.

In addition, the standard warnings that would apply to any geographic profiling scheme apply: The output should not be treated as a single prediction but rather as a tool to help prioritize areas of focus. It is designed to do well on average but may fail in outlier cases. And to implement it reliably, it must be paired with a human assessment. Please note also that models for predicting the offender's "home base" are much more well researched and in general will be more accurate than any algorithm claiming to predict the next strike point. A police department should choose on a case-by-case basis which model type to use.

Internal Workings of the Model

Inputs

Our model requires the coordinate locations of a serial criminal's previous offenses, as well as the order in which these crimes were committed.

Assumptions

1. The criminal will tend to strike at locations around one or more anchor points (often, the criminal's home).
2. Around this anchor point, there may be a "buffer zone" within which he will not strike.
3. If the criminal has multiple anchor points, the regions around those from which he has struck most often or most recently are more likely.

Method

The algorithm implements two different models, and then decides which is better.

- The first method assumes that the criminal has a single anchor point and builds likelihood regions around the anchor point based on the distribution of his past crimes.
- The second method assumes multiple anchor points, calculates the best number of anchor points to use, determines likelihood around each point individually, and weights the area around each by the criminal's apparent preferences.

Finally, the algorithm tests *both* models to see how well they would have predicted the previous crimes and uses the model with the better track record.

Summary and Recommendations

While our model needs more real-world testing, its strong theoretical basis, self-evaluation scheme, and awareness of practical considerations make it a good option for a police department looking to forecast a criminal's next strike. Combining its results with the intuition of a human being will maximize its utility.

References

Boba, Rachel. 2005. *Crime Analysis and Crime Mapping*. Thousand Oaks, CA: Sage Publications.

Canter, David, and Paul Larkin. 1993. The environmental range of serial rapists. *Journal of Environmental Psychology* 13: 63–63.

Canter, David, Toby Coffey, Malcolm Huntley, and Christopher Missen. 2000. Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology* 16(4): 457–478.

Godwin, Maurice, and Fred Rosen. 2005. *Tracker: Hunting Down Serial Killers*. Philadelphia, PA: Running Press.

Holmes, Ronald M., and Stephen P. Holmes. 1998. *Contemporary Perspectives on Serial Murder*. Thousand Oaks, CA: Sage Publications.

Jain, A.K., M.N. Murty, and P.J. Flynn. 1999. Data clustering: A review. *ACM Computing Surveys* 31(3): 264–323.

Kent, Joshua David. 2006. Using functional distance measures when calibrating journey-to-crime distance decay algorithms. Master's thesis, Louisiana State University. `http://etd.lsu.edu/docs/available/etd-1103103-095701/` .

Kocsis, R.N., and H.J. Irwin. 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 Law* 4(2): 195–206.

LeBeau, James L. 1992. Four case studies illustrating the spatial-temporal analysis of serial rapists. *Police Studies* 15: 124–145.

Minka, Thomas P. 2002. Estimating a Gamma distribution. `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.142.7635&rep=rep1&type=pdf` .

Paulsen, Derek J. 2005. Predicting next event locations in a time sequence using advanced spatial prediction methods. http://www.jdi.ucl.ac.uk/downloads/conferences/third_nat_map_conf/derek_paulsen.pdf .

________, and Matthew B. Robinson. 2009. *Crime Mapping and Spatial Aspects of Crime*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Rossmo, D. Kim. 1995. Geographic profiling: Target patterns of serial murderers.. Ph.D. thesis, Simon Fraser University.

________. 1998. Geographic profiling (NCIS Conference 1998). http://www.ecricanada.com/geopro/krossmo.pdf .

________. 1999. *Geographic Profiling*. Boca Raton, FL: CRC Press.

Rousseeuw, Peter J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20(1): 53–65. doi:10.1016/0377-0427(87)90125-7. ftp://ftp.win.ua.ac.be/pub/preprints/87/Silgra87.pdf .

Snook, Brent, Paul J. Taylor, and Craig Bennell. 2004. Geographic profiling: The fast, frugal, and accurate way. *Applied Cognitive Psychology* 18(1): 105–121. http://www.mun.ca/psychology/brl/publications/Snook12.PDF .

Snook, Brent, Richard M. Cullen, Andreas Mokros, and Stephan Harbort. 2005. Serial murderers' spatial decisions: Factors that influence crime location choice. *Journal of Investigative Psychology and Offender Profiling* 2(3): 147–164. http://www.mun.ca/psychology/brl/publications/Snook11.pdf .

Advisor Anne Dougherty with team members Anil Damle, Colin West, and Eric Benzel.

Judges' Commentary: The Outstanding Geographic Profiling Papers

Marie Vanisko
Dept. of Mathematics, Engineering, and Computer Science
Carroll College
Helena, MT 59625
mvanisko@carroll.edu

Introduction

The stated problem this year dealt with the issue of geographical profiling in the investigation of serial criminals. International interest in this topic has led to numerous publications, many of which present mathematical models for analyzing the problems involved. Although it was entirely appropriate and expected that teams working on this problem would review the literature on the subject and learn from their review, teams that simply presented published schemes as their mathematical models fell far short of what was expected. The judges looked for sparks of creativity and carefully explained mathematical model building with sensitivity analysis that went beyond what is found in the literature. This factor is what added value to a paper.

Documentation and Graphs

We observed a noticeable improvement in how references were identified and in the specific precision in documenting them within the papers. Considering the numerous online resources available, proper documentation was an especially important factor in this year's problem.

Despite the improvement, many papers contained charts and graphs from Web sources with no documentation. All graphs and tables need labels and/or legends, and they should provide information about what is referred to in the paper. The best papers used graphs to help clarify their results and documented trustworthy resources whenever used.

The UMAP Journal 31 (2) (2010) 149–151.

Assumptions

In many cases, teams made tacit assumptions about the criminals being considered but did not state or justify critical mathematical assumptions that were later used implicitly. Assumptions concerning probability distributions, anchor points, distances, units, mathematical procedures, and how to measure results were generally not discussed or justified.

Since this is a modeling contest, a lot of weight is put on whether or not the model could be used, with modification, in the real world. Also, clear writing and exposition is essential to motivate and explain assumptions and to derive and test models based on those assumptions.

Summary

The summary is of critical importance, especially in early judging. It should motivate the reader and be polished with a good synopsis of key results. For this problem, teams were asked to add to their one-page summary (which can have some technical details) also a two-page executive summary appropriate for the Chief of Police. Many teams seemed to assume that the Chief of Police would have impressive mathematical credentials.

The Problem and Its Analysis

Teams were asked to develop at least two different schemes for generating geographical profiles and then to develop a technique for combining the results of the different schemes in such a way as to generate a useful prediction for law enforcement officers. Although the papers designated as Meritorious generally developed interesting schemes, very few papers did an adequate job of testing their results and doing sensitivity analysis.

Most papers dealt with issues associated with the serial criminal's home base, usually referred to as the anchor point, and the buffer zone around that point within which the criminal is unlikely to commit crimes. Locations were identified using latitude and longitude and sometimes a time factor. Weights were frequently assigned to data points, sometimes taking more recent crimes into account more heavily and sometimes incorporating qualitative factors into the scheme. Teams used various metrics in describing "distances" between the anchor point and crime locations. Papers that rose to the top used well-defined metrics that were clearly explained. One cannot measure the reliability or validity of a model without clearly defined metrics.

Many teams mentioned that there was not a lot of data with which they could validate their model, although they did find some specific location information that included from 13 to 20 crimes in a given series. Some teams used as their only example the Sutcliffe case cited in the problem. In almost all cases, teams used their model to predict the location of the final crime based on all of the previous locations for that criminal. They could easily have had many more data points with which to validate their models. For example, if 13 crime locations were available, they could have used the first n locations to predict the location of crime $n + 1$, for each $n = 7, \dots, 12$. The judges agreed that this problem did not lend itself to validation by simulation, as many other problems do.

In describing the reliability of predicted results for proposed models, it was sometimes difficult to determine precisely how teams had arrived at their results. Since the literature is full of models and even computer models, it would have been worthy if teams had solved a problem via one of these methods and used that as a baseline to compare the results of original models that they proposed. Not a single team did this to the judge's satisfaction. Judges do not generally look for computer code, but they definitely look for precise algorithms that produce results based on a given model.

Concluding Remarks

Mathematical modeling is an art. It is an art that requires considerable skill and practice in order to develop proficiency. The big problems that we face now and in the future will be solved in large part by those with the talent, the insight, and the will to model these real-world problems and continuously refine those models. Surely the issue of solving crimes involving serial killers is an important challenge that we face.

The judges are very proud of all participants in this Mathematical Contest in Modeling and we commend you for your hard work and dedication.

About the Author

Marie Vanisko is a Mathematics Professor Emerita from Carroll College in Helena, Montana, where she taught for more than 30 years. She was also a Visiting Professor at the U.S. Military Academy at West Point and taught for five years at California State University Stanislaus. In both California and Montana, she directed MAA Tensor Foundation grants on mathematical modeling for high school girls. She also directs a mathematical modeling project for Montana high school and college mathematics and science teachers through the Montana Learning Center at Canyon Ferry, where she chairs the Board of Directors. She has served as a judge for both the MCM and HiMCM.

Judges' Commentary: The Fusaro Award for the Geographic Profiling Problem

Marie Vanisko
Dept. of Mathematics, Engineering, and Computer Science
Carroll College
Helena, MT 59625
mvanisko@carroll.edu

Peter Anspach
National Security Agency
Ft. Meade, MD
anspach@aol.com

Introduction

MCM Founding Director Fusaro attributes the competition's popularity in part to the challenge of working on practical problems. "Students generally like a challenge and probably are attracted by the opportunity, for perhaps the first time in their mathematical lives, to work as a team on a realistic applied problem," he says. The most important aspect of the MCM is the impact that it has on its participants and, as Fusaro puts it, "the confidence that this experience engenders."

The Ben Fusaro Award for the 2010 Geographic Profiling Problem went to a team from Duke University in Durham, NC. This solution paper was among the top Meritorious papers that this year received the designation of Finalist. It exemplified some outstanding characteristics:

- It presented a high-quality application of the complete modeling process.
- It demonstrated noteworthy originality and creativity in the modeling effort to solve the problem as given.
- It was well-written, in a clear expository style, making it a pleasure to read.

The UMAP Journal 31 (2) (2010) 153–156.

Criminology and Geographic Profiling

Each team was asked to develop a method to aid in the investigations of serial criminals. The team was to develop an approach that makes use of at least two different schemes and then combine those schemes to generate a geographic profile that would be a useful prediction for law enforcement officers. The prediction was to provide some kind of estimate or guidance about possible locations of the next crime. based on the time and locations of the past crimes. In addition to the required one-page summary, teams had to write a two-page less-technical executive summary for the Chief of Police.

In doing Web searches on this topic, teams found many publications and many proposed models. While it was important to review the literature, to receive an Outstanding or Meritorious designation, teams had to address all the issues raised and come up with a solution that demonstrated their own creativity.

The Duke University Paper

Abstract (One-Page Summary)

The Duke team did an excellent job with their abstract. In one page, they motivated the reader and provided the reader with a good sense of what the team had accomplished. It gave an overview of everything from the assumptions, to the modeling and how it was done, to the testing of their models, and finally, to the analysis of the accuracy of their results and limitations of their models. It was well-written and a great example of what an abstract should be.

Executive Summary (for the Police Chief)

The executive summary too was well-written and gave an overview of the team's approach, acknowledging limitations of their models. However, it was a little too vague in providing a precise idea of exactly what information would need to be collected and how to go about implementing the proposed models. Because the executive summary is a critical part of the requirements, this was part of what kept the Duke paper from being designated as Outstanding.

Assumptions

The team began with a brief survey of previous research on geographic profiling and used the information that they had gathered to decide what assumptions seemed appropriate. As a result, their list of assumptions was well-founded. The team exemplified one of the most important aspects in mathematical modeling by demonstrating precisely how their assumptions were used in the development of their models and how the assumptions enabled them to determine parameters in their models.

The Models

The team's first model involved a geographic method that took into account not only the location of crimes but also population densities, crime rates, and selected psychological characteristics. They used a bivariate Gaussian probability function and numerous parameters associated with previous crime locations. They did a very good job of showing how their assumptions and previous crime scenes led to the computation of these parameters and then using these parameters to estimate the probability function to be used in their model.

The team's second model involved a risk-intensity method and made use of the geographic method but extended it to make different projections with different probabilities associated with each of those projections.

Testing the Models

The Duke team was among the top papers, not only because of their well-based models, but because they tested their models—not with just one serial crime case, but with many cases. Their parameters allowed them to consider crimes other than murder, and they were able to examine how good their models were in several real-life situations. By analyzing their results, they were able to comment on the sensitivity and robustness of their models. This was something that very few papers were able to do, and a very important step in the modeling process.

Recognizing Limitations of the Model

Recognizing the limitations of a model is an important last step in the completion of the modeling process. The teams recognized that their models would fail if their assumptions did not hold—for example, if the criminal did not have a predictable pattern of movement.

References and Bibliography

The list of references consulted and used was sufficient, but specific documentation of where those references were used appeared only for a few at the start of the paper. Precise documentation of references used is important throughout the paper.

Conclusion

The careful exposition in the development of the mathematical models made this paper one that the judges felt was worthy of the Finalist designation. The team members are to be congratulated on their analysis, their clarity, and for using the mathematics that they knew to create and justify their own model for the problem.

About the Authors

Marie Vanisko is a Mathematics Professor Emerita from Carroll College in Helena, Montana, where she taught for more than 30 years. She was also a Visiting Professor at the U.S. Military Academy at West Point and taught for five years at California State University Stanislaus. In both California and Montana, she directed MAA Tensor Foundation grants on mathematical modeling for high school girls. She also directs a mathematical modeling project for Montana high school and college mathematics and science teachers through the Montana Learning Center at Canyon Ferry, where she chairs the Board of Directors. She has served as a judge for both the MCM and HiMCM.

Peter Anspach was born and raised in the Chicago area. He graduated from Amherst College, then went on to get a Ph.D. in Mathematics from the University of Chicago. After a post-doc at the University of Oklahoma, he joined the National Security Agency to work as a mathematician.