Mapping Gang Spheres of Influence

Samuel H. Huddleston, Jon Fox, Donald E. Brown
Predictive Technology Lab
University of Virginia

Abstract
Many urban environments have criminal gangs competing for control of available resources and gang territories. Research in criminology has linked the effect of these gang territories on criminal activity by criminal gangs. However, a great deal of a given criminal group’s activity occurs beyond the boundaries of that gang’s known territory but within a region we term a gang’s criminal sphere of influence. This paper documents the contributions of combining multilevel spatial choice modeling of gang behavior with a Geographic Information System (GIS) to develop a Sphere of Influence (SOI) analysis for criminal gangs in a given geographic region.
Keywords: Gang Analysis, Crime Analysis, Multilevel Modeling, Spatial Choice Modeling
Many urban environments have criminal gangs competing for control of available resources and territory. Intelligence analysis of these groups requires not only the prediction of future attack locations but also answers to questions such as: Who is the most likely perpetrator of a criminal incident at a given location? What is the most likely course of action for a given criminal group? What makes one location more likely to experience a gang incident over another? With limited resources, how can I best employ those resources most efficiently to engage specific criminal elements? Multilevel modeling of criminal site selection preference allows us to better answer these questions by linking the incidence of gang crime to the spatial, demographic, and socio-economic features of specific locations.

There is a rich body of literature that links gang activity spaces, socioeconomic conditions, and other factors to the incidence of crime committed by criminal gangs (Taniguchi, Ratcliffe, & Taylor, In Press; G. Tita & Ridgeway, 2007; G. E. Tita & Cohen, n.d.; R. Block, 2000). Many previous studies of gangs note that criminal gangs seize, control, and defend home territories from rival gangs (Thrasher, 1927; Whyte, 1937; Ley & Cybriwsky, 1974; Bernasco & Block, 2009). Other studies explore the effects that gang formation have on crime patterns and rates in the local communities where they form (G. Tita & Ridgeway, 2007). For instance, (G. E. Tita & Cohen, n.d.) demonstrate that areas where gangs congregate are highly correlated with high levels of gun violence and (J. H. Ratcliffe & Taniguchi, 2008) show that drug-gang street corners are highly correlated to the general incidence of crime. Several other studies also find that there exists a strong relationship between the number of gangs that are active in an area and the general level of criminal activity (R. Block, 2000) and that gang set spaces serve as crime attractors and crime generators (G. Tita & Ridgeway, 2007). However, analysis of at least one well-known gang data set (Meeker, Vila, & Parsons, 2002) demonstrates that many gangs commit as many as half of their crimes outside of their own controlled gang territories. An analysis of the six most active gangs in Santa Ana, California for the two year period from 1999-2000 reveals that these gangs committed 30% of their crimes outside any known gang territory and an additional 19% of their crimes in areas claimed by more than one gang, which we term gang territory conflicts. This indicates that an analysis of gang activity beyond their home territories is warranted. Multilevel modeling of criminal site selection provides a means for structuring this analysis and Geographic Information Systems (GIS) provide a means for communicating the results of this analysis in a format employable by police agencies.

Modeling Background

This paper merges research on criminal activity spaces with a multilevel modeling extension to previous criminal site selection methods and describes why the approach better predicts the actions of specific criminal groups and facilitates identification of gang criminal spheres of influence - the geographic regions in which each criminal group presents the greatest threat. This methodological approach rests on several foundations of research: data mining, criminal hot-spot prediction, criminal site selection modeling, and multilevel (hierarchical) modeling.

Criminal Hot-Spot Prediction

As noted above, many researchers have linked gang activity spaces, drug corners, set spaces, or turfs to concentrations of criminal activity, or hot-spots. The National Institute
of Justice defines a criminal hot-spot as “an area that has a greater than average number of criminal or disorder events, or an area where people have a higher than average risk of victimization (Eck, Chainey, Cameron, Leitner, & Wilson, 2005).” There are many techniques for identifying criminal hot-spots but they for the most part fall into one of two classes: those techniques that leverage only historical location data in the analysis (treating the problem as a spatial point pattern) versus those that treat the problem as a marked spatial point pattern. Point patterns are the type of spatial data that arise when the critical variable being analyzed is the location of events (Cressie, 1993). Most criminal incidents fall into this category of geographic analysis. A marked spatial point pattern is one in which the events in a point pattern are associated with features, measurements, or categorical marks. In crime analysis, examples include identification of the type of crime, the responsible party (if known), and the geographic features associated with the location of the criminal event.

Several techniques have been developed to identify criminal hot-spots using only the spatial point patterns generated by past observations. The most common approach for identifying criminal hot-spots is kernel density estimation because this approach is easily implemented in the Geographic Information Systems (GIS) most police agencies now employ (Eck et al., 2005; Boba, 2005). These techniques do not leverage the additional “marked” information associated with a criminal event but leverage only location (Latitude-Longitude or X-Y) data to estimate the relative risk associated with each X-Y coordinate on the map. Recently, (Mohler, Short, Brantingham, Schoenberg, & Tita, 2011) demonstrated that an approach using self-exciting point process models improves upon kernel density methods for predicting burglaries in Los Angeles in both space and time and (Lewis, Mohler, Brantingham, & Bertozzi, In Press) apply the same approach to modeling civilian deaths in Iraq.

In recent years, there has been a growing body of literature in which researchers identify criminal hot-spots by using the marks associated with crimes in police databases to identify criminal hot-spots. Social researchers tend to use various regression techniques in the application of environmental criminology to link social, economic, or spatial features to the incidence of crime (Brantingham & Brantingham, 1981). Examples of sociological analysis include identification of factors important in the occurrence of residential burglaries (Bernasco & Nieuwbeerta, 2005), robberies in Chicago (Bernasco & Block, 2009), the link between drug street corners and crime (J. H. Ratcliffe & Taniguchi, 2008), and several of the previously mentioned studies in gang activity (G. E. Tita, Cohen, & Engberg, 2005; G. Tita & Ridgeway, 2007; R. Block, 2000). Other researchers have begun to apply newly developed data mining techniques to the problem of identifying the areas most likely to see a criminal incident. Data mining approaches to hot-spot identification include machine learning techniques such as neural networks (Olligslaeger, 1997), fuzzy clustering (Grubesic, 2006), and support vector machines (Chang, Zeng, & Chen, 2005; Kianmehr & Alhajj, 2008).

**Criminal Site Selection Models**

There is one modeling approach that provides both insight into the environmental processes that generate crime (the focus of sociological inquiries) and improved predictive performance: spatial choice modeling. Spatial choice models are based upon the work of Daniel McFadden’s development of discrete choice theory (McFadden, 1974). In McFadden’s
formulation, actors, indexed by $j$, evaluate the utility, $U$, that they would derive from choosing an alternative based upon the features or attributes of that alternative:

$$U_{ij} = \beta^T \tilde{X}_{ij} + \epsilon_{ij}$$  

(1)

In the above formulation, $\tilde{X}$ denotes the vector of features or attributes for alternative $i$. The $\epsilon$ term captures the error associated with each pair of actors and alternatives while $\beta$ records the regression coefficients of the model. McFadden established the theoretical foundation for the use of conditional logistic regression to model choice from a discrete set of alternatives. When actors are choosing from a discrete set of alternatives, then their probability of selecting alternative $i$, $\pi(y = i)$, can be modeled using the well-known logistic regression equation:

$$\pi_j(y = i) = \frac{e^{\beta^T \tilde{X}_{ij}}}{\sum_{i=1}^N e^{\beta^T \tilde{X}_{ij}}}$$  

(2)

Several groups of researchers have applied this approach in a spatial context for modeling criminal site selection preference. Several examples of the direct application of McFadden’s discrete choice theory to crime include an analysis of the target selection by burglars in The Hague, Netherlands (Bernasco & Nieuwbeerta, 2005) and several studies of robberies in Chicago (Bernasco & Block, 2009; Bernasco, Block, & Ruiter, 2012). (Xue & Brown, 2006) develop criminal site selection models that adapt the spatial choice modeling approach for conditions in which the individual discrete choices (crimes) cannot be attributed to individual criminals, which is the case for most of the crime data available to police for use in predictive policing. Their work provides an extensive discussion of the assumptions involved in this model adaption. In brief, their approach relies upon assuming that both the choice set and the decision-making preferences of all of the modeled actors (criminals) in the study domain are similar, and the model therefore describes what is generally true about the criminal preferences in a geographic region.

(Xue & Brown, 2006) also incorporate the idea of using feature-space rather than geographic coordinates to represent the locations of crimes. Feature-space is defined as the Euclidean distance to each of the features of interest such as various crime attractors and crime generators (Liu & Brown, 2004). Their research group has shown that various forms of these criminal site selection models significantly improve predictive performance over the traditional kernel density methods for predicting burglaries (Liu & Brown, 2004; Xue & Brown, 2006) and terrorist events (Brown, Dalton, & Hoyle, 2004) such as suicide bombings (Smith & Brown, 2007). One noted reason for this performance improvement is that these criminal site selection models can highlight high risk areas (those very likely to observe a future criminal incident based upon the features of that location) that kernel density approaches do not highlight because they are far from previously observed crimes (Liu & Brown, 2004). (Huddleston & Brown, 2009) extend these criminal site selection models using multilevel modeling to further improve performance for predicting the locations of crimes by specific criminal street gangs.
Multilevel Modeling

Multilevel models, sometimes called heirarchical models, extend traditional regression models by allowing regression coefficients to vary from group to group (Gelman & Hill, 2007). The two common alternatives to multilevel models are pooled models, in which all groups are pooled together and treated as one, and no-pooling models, which build a separate model for each group. No-pooling models often suffer by not considering generalities of behavior that are captured only when all incidents are included in the model development. On the other hand, modeling spatial behavior with a pooled model can apply generalities to groups to whom they may not apply (Fotheringham, Brunsdon, & Charlton, 2000). Multilevel modeling addresses both of these shortcomings simultaneously by partially pooling the results of both analyses (Gelman & Hill, 2007). The model performance of a multilevel model is bounded for each group by the pooled model (developed from all observations) and the no-pooling model for that group (developed only from observations from that group). The multilevel model will never do worse than the best of those two models. It can, however, significantly improve model performance over both the no-pooling and pooled models simultaneously by leveraging the strengths in each of these models. For example, (Huddleston & Brown, 2009) demonstrate that the multilevel modeling of criminal site selection preference significantly improves upon both pooled and no-pooling models in predicting criminal activity by specific gangs.

Data

Data for this analysis came from three sources, which are illustrated in Figure 1. First, the authors used the Gang Incident Tracking System (GITS) database to evaluate the performance of this new methodological approach. The GITS project was introduced in 1993 to help law enforcement officials in Orange County, California “make more informed decisions to counter gang activity, which had been on the rise in recent years (Meeker et al., 2002).” The authors used a subset of the data particular to the city of Santa Ana, California for the period from 1994 through 2000. Incidents from the period 1994-1998 were placed into a training set and data for the period 1999-2000 was held out to serve as a model performance test data set. This approach was taken to mimic the approach that would be taken by law enforcement agencies in using statistical software to predict gang activity in their jurisdictions. In order for an incident to be classified as gang activity and entered into the database, it went through a rigorous verification process described by (Meeker et al., 2002). Information about each gang incident included information on the responsible gang (if known), the specific crime (one of 21 different crimes such as felonious assault, homicide, burglary, etc.), criminal event type (violent, weapons, property, drug, or vandalism), and geographic information about the location of the crime. Table 1 provides a summary of the various crime counts by gang and crime type for the training and test data sets.

(Meeker et al., 2002) extensively document the three step verification process used to add gang-related incidents into the database. An incident was added to the database if it met one of four criteria for establishing a gang-related incident:

1. A suspect or suspects are identified as gang members or admit membership in a gang
2. A person becomes a victim due to his or her gang association
<table>
<thead>
<tr>
<th>Gang ID</th>
<th>Crime Type</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Violent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Property</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drugs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weapon</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vandalism</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Train</td>
<td>Test</td>
<td>Train</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>29</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>37</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Other Gangs</td>
<td>193</td>
<td>71</td>
</tr>
<tr>
<td>Unknown</td>
<td>2615</td>
<td>1204</td>
</tr>
<tr>
<td>Total</td>
<td>3096</td>
<td>1365</td>
</tr>
</tbody>
</table>
Figure 1. Data sources used for the analysis of gang spheres of influence. The left panel illustrates the locations of gang crimes by the six most active gangs in the city during the period 1994 - 2000. The center panel illustrates the 180 census block groups from the 2000 US Census containing census demographic and socio-economic data. The panel on the right is a digitized form of a gang intelligence map provided by the City of Santa Ana Police Department. The gang territories in the right panel overlap the boundaries of the city, but are illustrated as even gang territories outside the boundaries of the city serve as predictive features for activity within the boundaries of Santa Ana.

3. A reliable informant identifies an incident as gang activity.

4. An informant of previously untested reliability identifies an incident as gang activity, and this identification is corroborated by other independent information.

More than 75% of the incidents in the data set have no identified gang affiliation for the perpetrator. While there are more than 132 unique gangs identified in the database, the vast majority of them are attributed very few crimes. For instance, 97 gangs in the dataset commit an average of less than one crime a year, and 59 gangs are attributed only one crime over the entire eight year time span.

The US 2000 Census provided the second data source, used for demographic information. Demographic information from the 2000 census was represented as an irregular surface with discrete demographic values recorded at the census block group level. Each criminal incident that fell into a given census block was given the socio-demographic information of the census block group it fell within. Socio-economic and demographic features found to be relevant included: median income, property, and rental values; racial demographics; the percentage of males in the population; the percentage of the population on public assistance; and the percentage of residents who own their homes.

The final data source used for this research was a gang intelligence map developed by the Santa Ana Police department in 1998. The gang intelligence map detailed gang territories of the city’s largest gangs and point locations (addresses) for many of the smaller gangs in the city. They also showed regions of the city claimed as gang territory by more than one gang - areas we termed “territory conflicts.”
The data available presents some constraints and limitations on the analysis. First, as seen in Table 1, the crime counts for the property, drug, weapon, and vandalism offenses are very limited, with null observations for many crime type and gang combinations. In a study of gang street crime in Chicago, (C. R. Block & Block, 1993) found that the spatial distribution of drug crimes and gang turf-motivated violent crimes differed. It may be true that the spatial pattern of the various crime types differ in Santa Ana, but the limited number of crimes by type, especially in the test data set, prevent this consideration. Due to the limited number of observations, crime types were pooled together by gang.

The second significant limitation was that the only gang intelligence product still available for this time period was a gang intelligence map for the city of Santa Ana from the year 1998. Thus, we made the assumption that the gang territories remained static throughout the study period (1994-2000). Some authors have noted that gang boundaries in other cities tend to shift relatively frequently (R. Block, 2000), so this assumption may not be valid but was necessary for the purposes of this study. More accurate intelligence products that incorporate shifting gang boundaries (if necessary) should improve the accuracy of this approach in police applications because the multilevel regression models used map the relationship between the probability of crime by a particular gang and distance to their gang territory or known address. Therefore, the models demonstrated here can adapt the model prediction (and the predicted sphere of influence for a criminal gang) if gang territory boundaries change over time.

Results are presented in this paper for the analysis of three different data sets. First, the Sphere of Influence (SOI) analysis is demonstrated on an analysis of the six largest and most active gangs in the city. These six gangs account for 33% of the crimes in the data set for which the perpetrator’s gang was identified. All six of these gangs have large gang territories mapped on the police intelligence map displayed in the right panel of Figure 1. The second data set extends the first SOI analysis to mapping the spheres of influence for the 15 most active gangs in the city. Incidents by these 15 gangs comprise 58% of the crimes in the data base for which the perpetrator’s gang is known and all of these gangs have either a gang territory or gang point location (address) identified for them on the police gang intelligence map discussed below. The last SOI analysis discussed uses all of the incidents in the dataset.

Methodology

Developing the Sphere of Influence (SOI) analysis for criminal gangs consists of three steps: data-set preparation, statistical modeling of gang criminal preference, and communication of results in a GIS system. Figure 2 provides a work-flow diagram for completing this analysis.

Developing the dataset for the statistical model is not possible without the use of a GIS system to conduct data preparation. During this step, inputs from many different data sources are fused together into a single data table that is exported into a statistical software package. The final product of the data preparation stage is a GIS point layer containing descriptive information about both the locations where the crime occurred and a “null grid.”

In order to provide representation for all of the locations not chosen by the criminal gangs, we incorporated a null grid by laying a point grid spaced at 200 feet over the study
area. This use of a null grid converts the irregular census block data surface into a regular lattice of point sites with discrete variables (Besag, 1974). This approach allows us to fit the regression model by providing null occurrence observations in geographic space in very rough approximation to the proportion of the surface area of the city that did not observe criminal incidents. It also allows us to develop the predicted continuous threat surfaces illustrated in Figure 4 by mapping the predictions over the null grid surface.

Additional GIS data preparation processes include: geocoding the data set of criminal incidences, developing a point grid layer to represent locations where a crime did not occur as previously described, merging these two point layers together, conducting a spatial join to give the point layer socio-demographic features from census blocks, and calculating the Euclidean distance from each individual point to important spatial features (sometimes referred to as feature space modeling) in the environment such as gang territories or addresses identified in police intelligence products. The final product of this step is a data table produced from a GIS point layer that contains the following information for each point: binary incident marker (0=null, 1=incident); Census 2000 socio-economic and demographic data; and the Euclidean distance to the nearest: gang territory, gang territory boundary, known gang address, and gang territory conflict. In the case that a crime occurred in one of the
geographic regions such as a gang territory, then the feature-space distance to that predictor was recorded as 0.

Once the analysis data table is built in a GIS, the results are exported into the statistical software R for statistical modeling. Using a multilevel generalized linear model, we examine not only the relationships between the features of interest and the incidence of crime, but we also identify how those relationships vary across the modeled groups. Thus, the multilevel modeling approach allows us to identify the generalities of criminal behavior across all of the gangs while still modeling the uniqueness of how each gang interacts with the environment. Capturing this trade-off between what is generally true of all gangs and what is unique to each gang significantly improves our ability to predict the criminal behavior of each individual gang in the future. This stage of the analysis includes developing models of the criminal site selection behavior of the various criminal gangs including: feature selection, performance assessment of the various models (on both training and test data sets), and using the multilevel modeling to develop predictions for future behavior by the criminal gangs. These statistical models answer one of the research questions posed in the introduction: What makes one location more likely to experience a gang incident than another location? The final output of the statistical modeling stage is a “threat data table” which records the risk of criminal behavior by each criminal gang at every point in the original data table. This threat data table is then exported back into a GIS system and appended to the original data point layer.

The last stage consists of using a GIS system to develop products for use by law-enforcement agencies. First, the threat data table can be used to plot “threat surfaces” for each of the criminal gang in the GIS system (see Figure 4). These threat surfaces show areas where each criminal gang presents the greatest threat and are analogous to the “hot-spot maps” commonly developed using kernel density estimation. The point layer can also be used to map the regions where each of the criminal gangs presents the greatest threat (see Figure 5), a product we term the “sphere of influence” map. This product illustrates where each criminal gang is the most likely to commit a crime. Finally, the threat surfaces can be used to develop products useful for allocating police resources to specific high threat areas (see Figure 6). This product shows the highest probability areas in the city to observe a crime (for example the highest 5 percent risk areas) and the criminal gang most likely to commit a crime there.

\textit{Multilevel Modeling of Gang Criminal Site Selection}

The following notation captures the results of the data preparation stage:

\begin{align*}
N &= \text{the number of observations in the data set} \\
J &= \text{the number of groups represented in the data set} \\
K &= \text{the number of predictors} \\
X &= \text{the predictor matrix} \\
\vec{Y} &= \text{the response vector}
\end{align*}
\[ y_i = \begin{cases} 1 & \text{if an incident occurred at the location} \\ 0 & \text{if an incident did not occur at the location} \end{cases} \]

for \( i = 1, \ldots, N \)

\[ X = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NK} \end{bmatrix} \]

This takes into account that criminal actors can select more than one location for their crimes. Thus, each actor can make several selections out of the set of available locations, each of which will be coded with the dummy variable 1, while all locations not selected by that actor would be coded with a 0. Note that in this application, we are modeling each criminal gang as an “actor.” Thus, all individuals within the gang are assumed to have the same criminal preference set and the same choice set. These criminal preferences and choice sets are assumed to vary between gangs.

As documented by (Brown et al., 2004), generalized linear models using a logit link function can account for the feature space distances to key features as well as categorical variables. Applying a logistic regression to model the criminal preferences of the studied group allows us to incorporate all different types of available data discussed above: feature-space data, categorical data (such as presence in a gang territory), and socio-demographic information from the census. This methodological approach is also beneficial in that it allows us to incorporate the idea of a criminal’s journey to crime, a theory that assumes that the likelihood of an offender’s target selection decreases with the distance to the target from his home (Bernasco & Nieuwbeerta, 2005; Rengert, 2004). In this case, the perpetrator’s “home” is represented by the gang’s home territory or known point site (address).

A logistic regression also provides a closed form solution to modeling the criminal preference by returning a value between 0 and 1 indicating the conditional likelihood of an event occurring at a given location. The conditional likelihood expresses the following idea: given that a criminal event has occurred, the probability that it occurred at this location is [a number between 0 and 1]. Although the geographic spatial choice methodology developed by (Xue & Brown, 2003) can incorporate temporal considerations, we have ignored temporal variations and constraints as is often done in criminological research (Bernasco & Block, 2009; J. Ratcliffe, 2006). (Fox & Brown, 2012) and (Fox, Huddleston, Gerber, & Brown, 2012) provide approaches for incorporating temporal considerations using a multilevel modeling approach.

We model the criminal site selection process for the criminal gangs in Santa Ana using a multilevel logistic regression model. As seen in Equation 3a, the multilevel model uses the standard form of a logistic regression model from Equation 2. However, the key difference in the multilevel model is that an additional constraint is placed on the coefficients for each of the different groups. We require that the coefficients for each feature of interest across the modeled criminal groups come from a common distribution that is estimated at the time of model fit. Thus, we model both the relationship between the features of interest
and the incidence of crime and also how those relationships vary across the modeled groups. This two part structure can be seen in the multilevel model below:

\[
\pi_j(y_i = 1|X) = N \left( \logit^{-1} \left[ \alpha_j + \beta_j X_i \right], \sigma_y^2 \right) \tag{3a}
\]

\[
\left( \frac{\alpha_j}{\beta_j} \right) \sim N \left( \left( \mu_\alpha, \mu_\beta \right), \begin{pmatrix} \sigma_\alpha^2 & \rho \sigma_\alpha \sigma_\beta \\ \rho \sigma_\alpha \sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right) \tag{3b}
\]

The probability of an event by group \( j \) at location \( i \), \( \pi_j(y_i = 1) \), is Gaussian distributed with a mean determined via the logistic regression and variance \( \sigma_y^2 \). Additionally, we have a model for how the different gangs interact with each feature of interest, with the assumption that there are differences in how the various gangs interact with each feature of interest and that those differences can be modeled with a Gaussian distribution. As an example, every gang will have an intercept term in the logistic regression equation which we model as the term \( \alpha \). We model that all \( \alpha_j \) come from a Gaussian distribution with estimated mean \( \mu_\alpha \), variance \( \sigma_\alpha^2 \), and covariance parameters to the other features of interest. These coefficients that vary by group are sometimes referred to as random effects, referring to the randomness in the probability model for the group level coefficients (Gelman & Hill, 2007).

The requirement for each coefficient for the groups to come from a common distribution generates a tradeoff between the two extremes of a general (pooled) model and group-specific (no-pooling) models. It incorporates what is known about the general criminal preferences but is weighted towards the group specific model in rough proportion to each groups’ contribution to the general model. This approach allows us to identify the generalities of behavior across all of the gangs while still modeling the uniqueness of each gang.

Model Fitting and Feature Selection

To fit the models for this analysis, we used the popular and freely available statistical package R (Crawley, 2007), which provides a computing package for fitting multilevel models based on the algorithm developed by (Bates & Pinheiro, 1998). Detailed information on algorithms for fitting the multilevel models using a maximum likelihood approach can be found in (Huddleston & Brown, 2009), (Jiang, 2007), and (Bates & Pinheiro, 1998). We also unsuccessfully sought an automated process for feature selection. In fitting traditional regression models, stepwise regression is a popular choice for automated selection of features from a set of possible predictors (Hastie, Tibshirani, & Friedman, 2001). During each “step”, the algorithm adds or drops one predictor and calculates both the statistical significance of all the predictors and a performance statistic such as model deviance or Aikake Information Criterion (AIC) (Akaike, 1974). The subset of features that provides the best performance for a given performance statistic is the subset selected for use. Feature selection for a multilevel model is more difficult because we are simultaneously modeling several (or many) groups at a time. A predictive feature which is important for the performance of one group may be unimportant for others, and insignificant in the pooled model from which the multilevel models are extended. The solution we used to address this problem was to first conduct a stepwise regression of all available features for each of the gangs (i.e. develop
Figure 3. The figure in the left panel illustrates the modeled effect of gang territories on the incidence of gang crime by any of the groups. The plotted points indicate the locations where a gang crime did (the top of the graphic) or did not (the bottom of the graphic) occur. The solid line graphs the modeled change in probability of observing a crime at a given location as distance to a gang territory increases. As seen in the figure, there are thousands of locations where a crime did not occur. The exponential drop in probability (which models the effect of gang territories after all other factors have been considered) mathematically models the fact that the vast majority of crimes occur within 1/2 mile of a gang territory. The figure in the right panel illustrates how the log-odds change for the incidence of crime by specific criminal gangs with respect to distance from the gang territory of Gang 1 after all other predictive features (including gang territories in general) have been considered.

For each of the features selected, we then have a mathematically defined relationship for how that feature is related to the incidence of crime. For example, Fig. 3(a) illustrates the effect that gang territories have on the incidence of gang crime. This relationship is modeled as an exponential drop in probability as you move away from gang territories and confirms previous research on the effect of gang territories on the incidence of crime (J. H. Ratcliffe & Taniguchi, 2008; G. Tita & Ridgeway, 2007; G. E. Tita & Cohen, n.d.). The advantage of the multilevel model is that we can also observe the differences in how various gang respond to individual features. Fig. 3(b) illustrates the differing effects of
Gang 1’s territory on the incidence of crime by various criminal groups. As one would expect, Gang 1’s territory is correlated with incidences of crime by that group, after all other factors have been considered. The overall effect of Gang 1’s territory on the incidence of crime is minimal (this is the “fixed effect”). The multilevel model does identify that increasing distance from Gang 1’s territory increases the log-odds for a crime by Gang 5.

The relationships that the multilevel model identifies for the incidence of crime by Gang 5 with respect to Gang 1’s territory merits additional discussion. These two gang territories overlap and therefore there is some co-linearity between these two predictive features. Likewise, each individual gang’s territory has some co-linearity with the predictive feature “Distance from Gang Territory.” The most important predictive feature for each gang in the multilevel model is the distance to its’ own gang territory, so each gang’s territory must appear in the model to preserve model predictive performance for that gang. As a result, we must be very careful not to draw general conclusions about the effects of the individual gang territories as predictors, especially those with known co-linearity to other territories. Because we are concerned primarily with predictive performance, and predictive performance suffers greatly with the removal of these features, we have continued to simultaneously use the feature-space distances to each of the considered gang territories and the feature-space distance to the nearest gang territory as a general predictive variable in our analysis. This use facilitates predictive performance but limits our ability to draw strong conclusions about the significance and importance of the various features. Because of the intended application for building predictive software, this is acceptable. It would not be acceptable for a sociological examination of features important to the incidence of crime.

Kernel Density Mapping of Gang Criminal Site Selection

As previously noted, kernel density estimation is the most frequently used approach for developing predictive threat surfaces in police applications. The kernel density approach is briefly presented here to provide a point of reference for performance assessment of the multilevel criminal site selection methods presented in the previous section. The kernel density method estimates the crime intensity at each location for gang $j$. Let $s_{jn}$ denote the location of the $n$th crime by group $j$ and $N_j$ the total number of crimes in the training data set attributed to gang $j$. The crime intensity for gang $j$ at location $i$, $\lambda_{ij}$, is calculated using a kernel smoothing function.

\[
\lambda_{ij} = \tilde{f}_{jh}(y_i) = \frac{1}{N_j} \sum_{n=1}^{N_j} K \left( \frac{\|y_i - s_{jn}\|}{h} \right) \tag{4}
\]

In Equation 4, the notation $\|y_i - s_{jn}\|$ denotes the Euclidean norm (distance) between locations $y_i$ and $s_{jk}$. Model fitting requires the selection of the kernel function $K$ and the bandwidth parameter $h$. In this application, we used the quartic kernel function and selected the optimal bandwidth parameter for each gang model using maximum likelihood estimation on the training data set as implemented in R software by the splancs package (Rowlingson et al., 2012; Bivand, Pebesma, & Gomez-Rubio, 2008).
Results

After fitting the models, we build predictive threat surfaces for each of the modeled groups by calculating the conditional probability for every point on the null grid (or mapping the crime intensity in the case of the kernel density method). The null grid, coded with the conditional probabilities at each grid point, is then exported to a GIS system as a point shapefile. This point layer is converted into a raster threat surface by interpolating between points. The final product of the predictive models is a threat surface created in a GIS system such as those in Figure 4. The ability of the multilevel model to provide distinct threat surfaces for each of the gangs is evident in Figure 4. Note that many of the gangs share high-probability areas. Note also that many of the high-threat areas for these gangs lie outside of their claimed gang territories. These threat surfaces address the most likely courses of action for each of the studied criminal groups, facilitating observation, targeting, or interdiction of suspected gang members.

A threat surface produced by a predictive algorithm can be thought of as a binary prediction of the probability of criminal incident at each individual location. Thus, threat surfaces can be evaluated using methods developed to assess the performance of binary classifiers. Recent years have seen the Receiver Operating Characteristic (ROC) curve become particularly popular for evaluating and comparing predictive algorithms in the machine learning community (Fawcett, 2006, 2004; Spackman, 1989). The ROC curve plots the cost-benefit trade-off for a classifier at all possible classification thresholds (Fawcett, 2004). The cost, plotted on the horizontal axis, is the model’s false positive rate. The benefit, plotted on the vertical axis, is the model’s true positive rate (also called model
Table 2: Predictive performance summarized using the Area Under the Curve (AUC) statistic

<table>
<thead>
<tr>
<th>Gang</th>
<th>Multi-Level Model</th>
<th>Kernel Density Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.97</td>
<td>0.82</td>
</tr>
<tr>
<td>2</td>
<td>0.98</td>
<td>0.76</td>
</tr>
<tr>
<td>3</td>
<td>0.99</td>
<td>0.88</td>
</tr>
<tr>
<td>4</td>
<td>0.99</td>
<td>0.78</td>
</tr>
<tr>
<td>5</td>
<td>0.98</td>
<td>0.72</td>
</tr>
<tr>
<td>6</td>
<td>0.98</td>
<td>0.85</td>
</tr>
</tbody>
</table>

ROC curves are a two-dimensional representation of classifier performance and often researchers would like to reduce performance to a single (scalar) statistic. The most common approach for summarizing ROC performance is to calculate the area under the ROC curve (denoted AUC) as a scalar value representing model classification performance (Fawcett, 2006; Hanley & McNeil, 1982). This statistic represents the probability that a randomly chosen positive incidence (in this application a randomly chosen location where a crime has occurred) will score higher than a randomly selected negative instance (i.e. a randomly selected location from the “null grid” created to represent non-incidents) (Fawcett, 2006). The AUC is equivalent to the Wilcoxon test of ranks commonly used in categorical data analysis (Fawcett, 2006; Hanley & McNeil, 1982) and is also directly related to the Gini coefficient (Breiman, Fiedman, Olshen, & Stone, 1984). The AUC is often considered to be the standard method to assess the accuracy of binomial classifiers.

Table 2 provides a performance comparison of the predictive performance between the multilevel criminal site selection model and the kernel density model on the test data set using the Area Under the Curve (AUC) statistic derived from the Receiver Operating Characteristic (ROC) curve. AUC scores are bounded between 0 and 1. AUC scores of 0.5 (or less) provide no discriminatory value. An AUC score above 0.75 is considered to provide enough discriminatory power to be clinically useful in the medical community, and AUC scores above 0.97 are considered to provide excellent discriminatory power (Fan, Upadhye, & Worster, 2006). The kernel density method provides moderately useful discriminatory power and this performance combined with its ease of use explains its wide-spread use in policing applications. The multilevel criminal site selection model significantly improves upon the predictive performance of the kernel density model, providing excellent discriminatory power for all six gangs. Thus, the threat surfaces illustrated in Figure 4 are much more accurate than those produced using kernel density estimation, the most frequently used approach in crime analysis. However, in addition to providing significant improvement in the predictive performance of the threat surfaces (hot-spot maps) for each of the gangs, the multilevel criminal site selection model also provides another important benefit: the ability to map a gang’s criminal sphere of influence.
Figure 5. Gang Sphere of Influence (SOI) Maps comparing the criminal spheres of influence for an analysis of the six largest gangs in Santa Ana using the multilevel model (left panel) and kernel density estimation (right panel).

Criminal Spheres of Influence

(Soukhanov, 1999) defines a sphere of influence as “a region of dominance; a geographic region or area of activity in which a state, organization, or person is dominant.” We define the predicted criminal sphere of influence for a group as the geographical area where a given model predicts that each group is the most likely to commit a crime (i.e. the geographic regions where one gang’s threat surface is higher than all other gang’s threat surfaces). This sphere of influence is created by mapping the calculated \( \pi_j(y_{ij} = 1) \) (in the case of the multilevel model) or the crime intensity \( \lambda_{ij} \) (for the kernel density method) for all \( j \) groups and all grid points, \( y_i \). By comparing each group level threat surface, we determine where each of the \( J \) groups is dominant.

Fig. 5 illustrates the results when sphere of influence maps are built with the two modeling approaches discussed in this paper. The SOI map produced with the kernel density method has regions of the city which fall within no predicted sphere of influence because these areas have no predicted gang crime intensity due to being far from previous crimes. During the test period, both Gang 1 and Gang 6 both commit a crime in these non-attributed regions. When the threat surfaces (hot-spot maps) produced by kernel
density estimates for the different gangs are compared, the resulting sphere of influence maps are non-contiguous representations with high-intensity areas for each gang highly concentrated around previous criminal activity by that gang (note the small concentrations spread throughout the city that show hotspots around Gang 6’s previous criminal activity).

In contrast, the multilevel modeling approach simultaneously models the relationships of all of the criminal gangs to the various gang territories and other important predictive features. The multilevel regression model asserts at least some probability of a crime by every criminal gang for every location in the city, and therefore maps every part of the city into one of the six groups’ spheres of influence. The resulting sphere of influence map plots a contiguous sphere of influence for each of the six gangs centered around their claimed gang territories. In some cases, a more dominant gang’s sphere of influence encroaches upon a neighboring gang’s territory. All gang “territory conflicts” have been divided into competing spheres of influence. Not only is the multilevel model’s mapped sphere of influence more visually appealing due to providing contiguous regions surrounding gang’s territory, it is a more accurate representation of future gang activity by the criminal gangs.

Table 3: Multilevel Model Sphere of Influence Predictive Performance

<table>
<thead>
<tr>
<th>Sphere of Influence</th>
<th>Gang 1</th>
<th>Gang 2</th>
<th>Gang 3</th>
<th>Gang 4</th>
<th>Gang 5</th>
<th>Gang 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gang 1 SOI</td>
<td>61%</td>
<td>19%</td>
<td>3%</td>
<td>3%</td>
<td>10%</td>
<td>3%</td>
</tr>
<tr>
<td>Gang 2 SOI</td>
<td>7%</td>
<td>80%</td>
<td>0%</td>
<td>0%</td>
<td>7%</td>
<td>7%</td>
</tr>
<tr>
<td>Gang 3 SOI</td>
<td>0%</td>
<td>21%</td>
<td>50%</td>
<td>14%</td>
<td>0%</td>
<td>14%</td>
</tr>
<tr>
<td>Gang 4 SOI</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>78%</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>Gang 5 SOI</td>
<td>9%</td>
<td>23%</td>
<td>5%</td>
<td>9%</td>
<td>55%</td>
<td>0%</td>
</tr>
<tr>
<td>Gang 6 SOI</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 3 displays the results obtained when we investigate the veracity of the gang SOI analysis by examining the predictive performance of the SOI analysis against a test set. Table 1 records the percentage of incidents during the test period (1999-2000) which were committed in each multilevel model predicted criminal sphere of influence by the various gangs. As seen in the table, the multilevel model accurately predicts the gang most likely to commit a crime in each sphere of influence.

Table 4 expands this analysis by comparing the performance of the sphere of influence analysis conducted with kernel density to that developed using the multilevel model. While the gang spheres of influence from both approaches often correctly identify a geographic area in which the predicted group is the most likely to commit a crime, the multilevel model is much more accurate. The gang spheres of influence developed using the multilevel model contain a higher percentage of incidents committed by the predicted group in every case and this model provides significantly better overall performance.
Table 4: Percentage of Test Incidents Committed by Each Group in Predicted SOI

<table>
<thead>
<tr>
<th>Sphere of Influence</th>
<th>Multi-Level Model</th>
<th>Kernel Density Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gang 1</td>
<td>61%</td>
<td>56%</td>
</tr>
<tr>
<td>Gang 2</td>
<td>80%</td>
<td>55%</td>
</tr>
<tr>
<td>Gang 3</td>
<td>50%</td>
<td>35%</td>
</tr>
<tr>
<td>Gang 4</td>
<td>78%</td>
<td>50%</td>
</tr>
<tr>
<td>Gang 5</td>
<td>55%</td>
<td>53%</td>
</tr>
<tr>
<td>Gang 6</td>
<td>100%</td>
<td>94%</td>
</tr>
<tr>
<td>Overall Performance</td>
<td>69%</td>
<td>57%</td>
</tr>
</tbody>
</table>

The kernel density model performs as well as it does on this analysis only because most of the crimes by these gangs fall into areas with high predictions for the respective gangs. These areas represent *hot-spots* for each group and the peak risk areas for each gang are still accurately mapped. The multilevel model better sorts out who presents the greater threat in the regions that aren’t hot-spots for the different gangs.

The predicted criminal sphere of influence provides important insight for an analyst that is not available to analysts who currently only have access to incident hot spots and known gang areas. This is because criminal gangs often commit crimes outside of their known territories or in geographic areas contested between gangs. For example, in the two year test period (1999-2000) of this study, nineteen percent of the crimes attributed to the six most active gangs were committed in areas claimed by more than one gang, which we term “territory conflicts.” Thirty percent of the crimes committed by these six gangs occur outside any of their known territories. The sphere of influence prediction maps which of the groups is most likely to commit a crime in these geographic areas which lay away from the assumed home territories depicted in the gang territory map used by the Santa Ana police department. Figure 6 provides an illustration of how this information can be leveraged. It provides a “gang resource map” that illustrates the highest five percent risk area for gang activity in the city and identifies the gang most likely to commit a crime in those areas. This map can be developed either by directly identifying the highest probability risk areas and responsible party in the statistical package and exporting that information in the threat data table or by leveraging the three-dimensional modeling capabilities of GIS systems that use the “heights” of the threat surfaces and an intersecting plane to produce this reference product (the authors developed this illustration using the ArcGIS ArcScene software).

*Extending the Analysis to More Gangs*

To further explore the potential of multilevel modeling, we extended the SOI analysis for Santa Ana to consider more than the six largest and most active gangs. First, we extended the analysis of the gangs in Santa Ana to include all incidents from the 15 most active gangs in Santa Ana. These 15 gangs commit 58% of the crimes in the data base for which the gang affiliation of the perpetrator was known. The sphere of influence continues
Figure 6. Gang Resource Map showing the highest probability areas for gang activity in the city and the gang most likely to commit a crime at that location.

to be a very accurate predictor of where each gang is the most likely offender for a gang related crime incident, even as more gang are added to the analysis. Extending the analysis to include more gangs allows us to more accurately assess the situation in a given geographic region with a large number of criminal elements.

The 15 gang sphere of influence map in Figure 7 identifies where in the city each of the 15 most active gangs presents the dominant threat. Although there are more gangs active within the city of Santa Ana, the remaining gangs all commit less than five crimes per year, and thus have comparatively little influence. Figure 7 serves as a reference document for a law enforcement analyst, identifying the dominant threats in each area of the city. There are some edge effects visible when comparing the change from the six gang SOI to the fifteen gang SOI. Work continues on how to reduce the edge effects near the boundaries of the city that are most easily visible in the changes to gang boundaries across the southern border of the city. Gang 10 has no visible sphere of influence since it is dominated by larger and more active nearby gangs.

Table 5 records the predictive accuracy of the 15 gang sphere of influence analysis. The multilevel model sphere of influence accurately predicts the gang most likely to commit a crime in the given sphere of influence for 10 out of 12 spheres of influence in which a crime occurred during the test period. No crimes occurred in the spheres of influence for gangs 7, 10, and 15. In Gang 9’s predicted sphere of influence, both Gangs 8 and 9 had an equal number of crimes. In Gang 8’s sphere of influence, Gang 10 (which has no predicted sphere of influence) committed the majority (three out of four) of the crimes. Given the limited number of crimes in these two smaller spheres of influence, it is difficult to determine whether the SOI analysis did poorly for these two regions due to changing conditions during the test period (1999 - 2000) or because it simply did not model the SOI for these two gangs.
well. Overall, the SOI analysis appears to provide a fairly accurate map of the regions of dominance for these criminal gangs.

The last analysis conducted used all incidents in the data set, including the incidents in which the offending gang was not identified. These incidents were classified as belonging to the “unknown” gang. This analysis revealed several important insights into conducting an SOI analysis. First, the sphere of influence for the “unknown gang”, which is attributed more than 75% of the crimes in the data set, extends to cover the entire city. Thus, the sphere of influence analysis developed from this scenario is uninformative. Second, the predictive performance for the known gangs already modeled suffered significantly when the “unknown gang” incidents were included in the training data set. This almost certainly happens because the model structure assumes that all members of a gang operate from a similar preference set and that the model coefficients for all gangs come from a common distribution (see Equation 3). Because of the large amount of incidents attributed to the “unknown” gang, this gang’s incidents provide most of the information for the model fitting, resulting in a model skewed towards providing a good fit for a gang that does not exist. This insight generates the recommendation that when developing predictive models or a sphere of influence analysis using multilevel models, the appropriate dataset should contain only
incidents in which the discrete choices (for example crimes) can be attributed to individual
groups (or individual actors).

The second insight developed from analyzing the entire data set concerns the need
for location data and sufficient sample sizes for the modeled criminal groups. The gang
addresses or territories for most of the “other gangs” weren’t known and most of these
gangs commit very few crimes. These gangs did not have enough observations during the
testing period to conduct performance assessment of the generated threat surfaces. They
also have no mapped sphere of influence because their threat surfaces were dominated by
at least one of the 15 most active gangs in the city. Note that multilevel modeling can
be conducted even in the case in which there are as few as one or two observations for
individual groups so long as there is a sufficient number of groups but this exponentially
increases the computational burden for fitting the models (Gelman & Hill, 2007). In this
case, including the remaining gangs into the analysis generated no more insights for a police
agency. Thus, we recommend the SOI analysis be limited to active gangs for which specific
location data (addresses or mapped gang territories) is available.

Discussion and Conclusions

Although this new approach does provide a way of automatically developing a threat
assessment product for use by law enforcement or intelligence agencies, there are some
important caveats to the use of this methodological approach. The first issue concerns
identification of features of interest for use in the multilevel models. The approach relies on
manual stepwise regression because the statistical package used did not provide a convenient
way to automate this. This would be a key feature needed to provide maximum benefit in the
application to crime analysis software. It took a great deal of time to find appropriate models
for increasing numbers of criminal gangs. This process was very iterative and required some
understanding of statistical significance and the ability to fit and interpret mathematical
models. We were not able to find an approach to easily automate this process, limiting
immediate application in law enforcement software. We also did not have access to all of
the features that we would have liked to include in the analysis. There were many crime
generators and crime attractors (Brantingham & Brantingham, 1981) we did not include in
the analysis that could significantly improve predictive performance.

Another concern with the sphere of influence approach relates to a potential use by
law enforcement personnel. Given a sphere of influence map, many law enforcement officials
are likely to want to use the map to identify the most likely perpetrators of criminal events
after the fact. However, when including all of the incident data from the GIT database, the
“unknown” gang becomes the most dominant sphere of influence throughout the geographic
region. This result is not unexpected since about 75% of the incidents in the GIT database
are attributed to an unknown party. Since the “unknown” gang commits 75% of the crimes
and these crimes are spread throughout the spatial region of study, the “unknown” gang’s
sphere of influence is modeled as most dangerous throughout the city. This underscores the
fact that in spite of the precise boundaries established between the groups by the sphere of
influence map, there is still a good deal of uncertainty inherent in the models because the
police don’t know who committed the majority of gang-related crimes.

The sphere of influence map serves as a priori predictors of criminal activity but
should not be used as a posterior identifier of criminal responsibility. Instead, the sphere
of influence map should be a useful tool for resource allocation decisions, assigning areas of responsibility as part of counter-gang initiatives, or prioritizing investigative efforts after the occurrence of a crime. One scenario might be that given a suspected gang-related event at a specific location, we can identify which of the gangs are the most likely to have perpetrated the crime and allocate investigative resources to the most likely culprit(s). The sphere of influence map is simply identifying the most likely criminal gang to commit a crime in a given area. It is not predicting that the gang is responsible for all crime within their sphere of influence. The multi-level model can be used to calculate a confidence interval for the stated conditional probability. Especially in areas near the sphere of influence boundaries, these confidence intervals overlap for two or more groups. The correct approach would be to iteratively assign investigative resources in a manner consistent with the probabilities proposed by the model. When no other information about an incident is known, start with the most probable culprits and work your way through the list of predicted suspects in decreasing order of the probability of activity at that location.

This paper demonstrates that multilevel modeling of criminal site selection for criminal gangs offers some significant contributions. This modeling approach significantly improved the predictive performance of threat maps (or hot-spot maps) for individual criminal gangs over kernel density estimation, the most commonly used predictive policing method. In addition, the multilevel modeling structure also allows us to map criminal spheres of influence for the various gangs, highlighting the geographic regions where specific gangs present the dominant threat. This ability can be leveraged to produce analytic products for police use such as the Gang Resource Map illustrated in Figure 6 and the Gang Sphere of Influence Map illustrated in Figure 7. In addition, the sphere of influence analysis is generally applicable for any domain in which we want to compare some spatial choice behavior (criminal, consumer, etc.) across individuals or groups and map the probability that a group’s or individual’s target selection behavior will dominate all others at a given geographic point. Future research efforts in the applications of this multilevel modeling approach include identifying improved approaches for automated feature selection, the modeling of temporal choice behavior, and the application of multilevel criminal site selection models to the spatial choice behavior of insurgent/terror groups, retail customers, and corporate real estate.

References


<table>
<thead>
<tr>
<th>SOI</th>
<th>Percentage of Test Data Set Incidents Committed by Each Gang Which Occur in the Selected Sphere of Influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38.3% 27.7% 0.0% 2.1% 6.4% 4.3% 4.3% 2.1% 2.1% 4.3% 0.0% 2.1% 0.0% 2.1% 4.3%</td>
</tr>
<tr>
<td>2</td>
<td>0.0% 71.4% 0.0% 0.0% 14.3% 0.0% 14.3% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%</td>
</tr>
<tr>
<td>3</td>
<td>0.0% 8.7% 30.4% 8.7% 0.0% 8.7% 0.0% 8.7% 0.0% 13.0% 0.0% 0.0% 0.0% 13.0%</td>
</tr>
<tr>
<td>4</td>
<td>0.0% 0.0% 0.0% 63.6% 9.1% 9.1% 9.1% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 9.1%</td>
</tr>
<tr>
<td>5</td>
<td>8.5% 12.8% 2.1% 4.3% 25.5% 0.0% 21.3% 4.3% 2.1% 2.1% 0.0% 4.3% 4.3% 0.0%</td>
</tr>
<tr>
<td>6</td>
<td>0.0% 0.0% 0.0% 0.0% 0.0% 89.5% 0.0% 0.0% 5.3% 0.0% 0.0% 0.0% 0.0% 5.3%</td>
</tr>
<tr>
<td>7</td>
<td>- - - - - - - - - - - - - - -</td>
</tr>
<tr>
<td>8</td>
<td>0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 25.0% 0.0% 75.0% 0.0% 0.0% 0.0%</td>
</tr>
<tr>
<td>9</td>
<td>0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 25.0% 25.0% 16.7% 16.7% 8.3% 0.0%</td>
</tr>
<tr>
<td>10</td>
<td>- - - - - - - - - - - - - - -</td>
</tr>
<tr>
<td>11</td>
<td>0.0% 0.0% 0.0% 0.0% 28.6% 0.0% 0.0% 0.0% 0.0% 0.0% 57.1% 0.0% 0.0% 14.3%</td>
</tr>
<tr>
<td>12</td>
<td>0.0% 0.0% 33.3% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 66.7% 0.0% 0.0%</td>
</tr>
<tr>
<td>13</td>
<td>0.0% 25.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 75.0% 0.0%</td>
</tr>
<tr>
<td>14</td>
<td>0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0%</td>
</tr>
<tr>
<td>15</td>
<td>- - - - - - - - - - - - - - -</td>
</tr>
</tbody>
</table>