A Statistical Threat Assessment

Samuel H. Huddleston and Donald E. Brown, *Fellow, IEEE*

***Abstract*—Criminal gangs, insurgent groups, and terror networks demonstrate observable preferences in selecting the sites where they commit their crimes. Accordingly, police departments, military organizations, and intelligence agencies seek to learn these preferences and identify locations with a high probability of experiencing the particular event of interest in the near future. Often, such agencies are keen not just to predict the spatial pattern of future events but even more importantly to conduct threat assessments of particular criminal gangs or insurgent groups. These threat assessments include identifying where each of the various groups presents the greatest threat to the community, what the most likely targets are for each criminal group, what makes one location more likely to experience an attack than another, and how to most efficiently allocate resources to address the specific threats to the community. Previous research has demonstrated that applying multivariate prediction models to relate features in an area to the occurrence of crimes offers an improvement in predictive performance over traditional methods of hot-spot analysis. This paper introduces the application of multilevel modeling to these multivariate spatial choice models, demonstrating that it is possible to significantly improve the predictive performance of the spatial choice models for individual groups and leverage that information to provide improved threat assessments of the criminal elements in a given geographic area.**

***Index Terms*—Crime analysis, multilevel modeling, predictive analysis, spatial data.**

I. Introduction

SINCE the end of the Cold War, an important phenomenon confronting security professionals has been the emergence of nonstate actors such as urban gangs, international terrorist organizations, and political insurgencies as the primary destabilizing forces in the world, particularly within urban centers. However, the technology that police and security agencies have often provides a limited understanding of the criminal groups they are targeting. A 2002 RAND study [1] commenting on the U.S. military's intelligence analysis process noted that "the sheer density and diversity of all features of an urban area—buildings, infrastructure, and people—flood extant technologies in ways that often make information superiority unreachable." In these complex environments, security agencies are often charged with identifying much more than where the next crime hot spot will be. They seek to understand and target the specific criminal groups responsible for the criminal acts. Important questions are raised such as the following: 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? With limited resources, how can I allocate my resources most efficiently? What makes one location more likely to experience a criminal or terror event over another?

Traditional methods of criminal hot-spot analysis such as kriging and kernel density estimation assume that the best predictor of future criminal incidents is a previous occurrence of a similar crime in close proximity. These approaches, which apply various density mapping techniques for crime clusters, are the most prevalent in crime analysis [2]. However, approaches based upon the clustering of crimes suffer when analyzing specific criminal groups for two reasons. First, the number of criminal incidents attributed to a specific criminal group tends to be limited. Second, criminal groups often attempt to randomize their patterns of behavior in order to avoid capture. The combination of these two effects means that, while criminal events such as burglaries may cluster in aggregate, the limited observations of events attributed to a single group may not. Thus, techniques that rely on the clustering of incidents in space can be of limited use for analyzing the actions of specific criminal groups.

In recent years, researchers have begun to apply data mining algorithms to anomaly detection and hot-spot identification in intelligence and crime analysis [3]–[8]. One approach that has been fielded is the use of clustering algorithms to identify regions of increased crime [4]. These approaches have been applied primarily to aggregate crime data rather than to predict the actions of a specific criminal group, and they suffer from the same performance shortcomings as the more traditional approaches as they rely on spatially clustered crimes and a large number of observations. Kianmehr and Alhajj [5] compare the performance of several other data mining techniques in predicting crime hot spots for homicide and other criminal events. In a performance comparison of support vector machines, neural networks, and spatial auto-regressive models, the authors argue that the support vector machine learning approach offers some performance benefit over the other two approaches and a possible new approach for crime mapping. However, these machine learning approaches also offer some significant limitations when applied to the crime patterns of specific criminal groups. The authors note that the preferred approach requires a large training data set, which, as previously discussed, is often not available for a specific criminal group. In addition, the purpose of intelligence analysis and threat assessment of criminal groups is to both predict the activities of the criminal groups and understand the conduct of the groups as much as possible. Machine learning techniques such as neural nets and support vector machines provide very limited insight into how the predictions are made, limiting their application in understanding the actions of the criminal groups under study.

Manuscript received November 17, 2008. First published September 11, 2009; current version published October 16, 2009. This paper was recommended by Associate Editor Q. Ji.

S. H. Huddleston is with the Department of Systems Engineering, United States Military Academy, West Point, NY 10996 USA (e-mail: samuel.huddleston@usma.edu).

D. E. Brown is with the Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA 22903-2600 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2009.2027611

There are proven methods that security agencies can use to exploit the patterns in behavior the criminal elements do present, even when their actions are not clustered in hot spots. Liu and Brown [6] and Brown *et al.* [7] have demonstrated that conducting spatial choice analysis can uncover the preferences criminals and terrorists use to select the locations for their crimes. They have also demonstrated that these spatial choice models offer significantly improved predictive performance over the traditional methods of hot-spot analysis in applications such as predicting burglaries [6] and suicide bombings [8]. However, these spatial choice models have been focused on addressing the general incidence of specific crime types rather than the problem of identifying the patterns of specific criminal groups. The U.S. Army's doctrine on targeting bomb cells [9] notes that "Intelligence must be focused on defeating the network or organization." Similarly, gang task forces in U.S. cities address their gang problems by individually targeting specific criminal groups.

The focus of this paper is on providing security agencies with improved threat assessments and predictive capability for targeting specific criminal groups rather than the general incidence of crimes by a set of criminal groups. This paper demonstrates that the application of multilevel modeling to spatial choice models of criminal gangs in Santa Ana, California, significantly improves the predictive performance for each of the criminal gangs by leveraging the additional information provided by the actions of other similar criminal gangs. This application makes a strong argument that this approach would offer improved threat assessments for distinct criminal gangs, insurgent elements, and terror networks when applied to more sensitive data sets by military and civilian security agencies.

II. Background

A. Spatial Choice Modeling

The locations of criminal incidents, terrorist attacks, and insurgent weapon caches are all examples of spatial point patterns. Spatial data are categorized by a geographic or topological index such as latitude and longitude and can be mapped. Point patterns are the type of spatial data that arise when the critical variable being analyzed is the location of events. Because spatial point patterns are often clustered, kernel density estimation is the most frequent method used to predict the future locations of criminal events.

Liu and Brown [6] have provided an alternate method for forecasting criminal events to the traditional approach of kernel density estimation. Building upon Daniel McFadden's Nobel Prize-winning work in discrete choice theory [10] and theories from environmental criminology, they assert that criminals have a set of preferences that are taken into account when deciding the location and time to commit a criminal act. They introduce the concept of modeling the criminal's intelligent site selection using feature-space analysis. Feature space is described as the distance from a point to a selected set of features which make up the criminal's preferences such as roads, bridges, etc. Rather than an incident being described in geographic space as a latitude and longitude, the description of a location in feature space would be the geographic Euclidean distance to each of the many features in the preference set. The analyst discovers this set of preferences through analysis of historical incident data.

The payoff for identifying and modeling feature-space preferences over traditional methods of spatial prediction is that potential high-risk areas that have not yet been targeted by the criminal can be identified. As a result, the predictive performance of feature-space models is often better because kernel density estimation only provides a high estimate of probability for an event if it is near previous events. Because criminals, terrorists, and insurgents are thinking actors, they tend to adjust their spatial patterns in order to avoid capture. However, it is reasonable to assume that the criminals maintain their site selection preferences based upon the features of a location even as they change the physical location they choose for their crimes. Previous research has demonstrated how this spatial mapping of preference significantly improves upon kernel density hot-spot analysis for predicting the incidence of burglary in Richmond, VA, [6] and for predicting suicide bombing attacks in Israel [7].

B. Multilevel Modeling

Multilevel models, sometimes called hierarchical models, are extensions to regression models in which data are organized into groups and the regression coefficients are allowed to vary from group to group. There are two common alternatives to multilevel modeling: ignoring group level variation (pooled models) and developing different regression models for each group (no-pooling models) [11]. No-pooling models can suffer by not considering the context in which the group behavior occurs. In other words, there may be generalities of behavior that are not captured by analyzing only one group at a time. On the other hand, modeling spatial behavior with a pooled model that does not consider a group indicator can apply generalities to groups to whom they may not apply [12]. A multilevel model serves as a tradeoff between these two extremes by partially pooling the results of both analyses [11].

A multilevel model generalizes when there are few observations of a specific group and builds a more detailed model of the group as the sample values for that group increase in the data set. For example, a group that has a great deal of samples in the database will have multilevel coefficients closer to the coefficients of a no-pooling model than a group that has fewer samples. Basic multilevel modeling extends classical regression inference to develop different coefficients for each group, using maximum likelihood estimation to estimate the regression coefficients by minimizing the sum of squared error. The coefficients which vary by group are themselves modeled, usually using a Gaussian distribution to statistically model the random effects of the variation between groups and variation within groups [11].

C. Data Sources

To illustrate the application and benefit of applying the multilevel models to predicting criminal events, we demonstrate the use of this model to predict criminal incidents by gangs in Santa Ana, California, for the years 1999–2000. The gang incident data used in this paper are from the Gang Incident Tracking System (GITS) project funded by the National

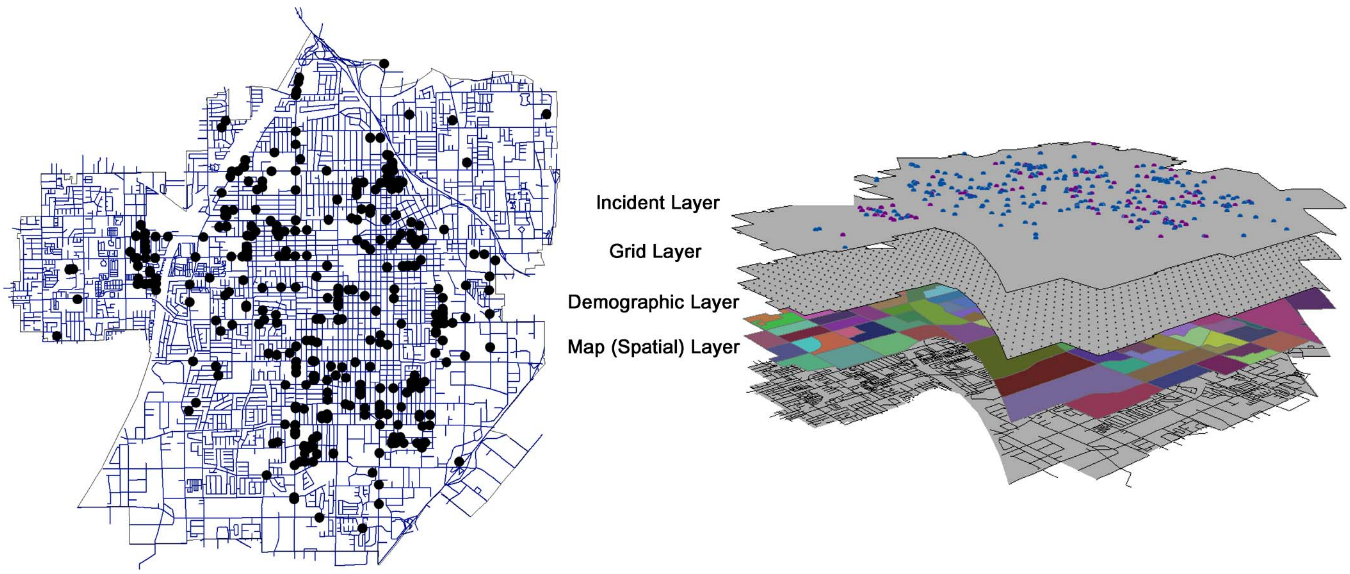

Fig. 1. Incident data for the six most active gangs in Santa Ana, California, during the period from 1994 to 2000. To build the predictive model, the incident data for the training period (1994–1998) are overlaid on top of demographic census blocks from 2000 Census and spatial map data to build a feature-space data set which includes categorical, demographic, and spatial data into one data set for analysis. A null grid is also superimposed over the layers in order to develop the null occurrence data set for the regression.

Institute of Justice [13]. The scope of our analysis was reduced to conducting a study of the six most active gangs in the city. Incidents attributed to these six gangs for the period from 1994 to 1998 were placed in a training data set. Incidents for the period from 1999 to 2000 were placed into a test data set. Fig. 1 shows a visualization of the criminal incidents in the data set and the approach used to develop the data set for the predictive models. Demographic information from the 2000 census was also incorporated to model the possible socioeconomic preferences for gang activities. Demographic information was represented as an irregular surface with discrete demographic values recorded at the census block level. All incident locations that fell within a given census block were assigned the discrete value of the demographic information for that census block. In addition, the Santa Ana Police Department provided group-specific spatial information in the form of a known gang territory map current as of 1998. As this gang territory map was the only one available, we assumed that the gang boundaries in Santa Ana remained stable throughout the training and testing periods. These spatial data were incorporated into the data set by extracting feature-space calculations for all incidents as described in Section II-A. We incorporated null observations into the data set by laying a point grid with points separated by 200 ft over the study area as shown in Fig. 1. The same data extraction steps were performed for the grid layer to assign demographic and spatial features. The inclusion of the null grid serves three purposes in the data set. First, it serves to convert the irregular demographic surfaces in the census data to a regular lattice of point sites with discrete variables [14]. Second, the null grid allows us to fit a regression model by providing null occurrence observations in geographic space in rough proportion to the surface area in the city that did not observe a criminal incident. Finally, once a regression model has been developed, predicted likelihoods for event occurrences for each of the observations in the grid can be extrapolated to build a continuous threat surface, as shown in Fig. 2.

III. Spatial Preference Models

As documented by Brown *et al.* [7], 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 for a location, including the feature-space distance to important features, demographic information available from census data, and categorical variables such as indicators to determine whether a particular location is within a known gang territory. A logistic regression also provides a closed-form solution for modeling the criminal preference by returning a value between zero and one, indicating the relative likelihood of an event occurring at a given location.

A. Pooled Model

To formulate our first logistic regression model, we state that $\pi(y_i = 1)$ is the probability of a criminal event by any group at location y_i given a vector of features for that location X_i. This likelihood will hereafter be referred to as the general likelihood of an incident. We model this general criminal preference as follows:

$$\pi(y_i = 1|X) = \text{logit}^{-1}(\alpha + \beta X_i). \quad (1)$$

This model specifies the pooled model, in which all incidents, regardless of group, are used to fit the regression based upon the features of that location. This model attempts to capture the general preferences of all the groups together, and models the overall pattern of criminal incidents. This model represents the method currently used to build criminal preference models for the general incidence of crime in military and police crime analysis software that employ the spatial choice method.

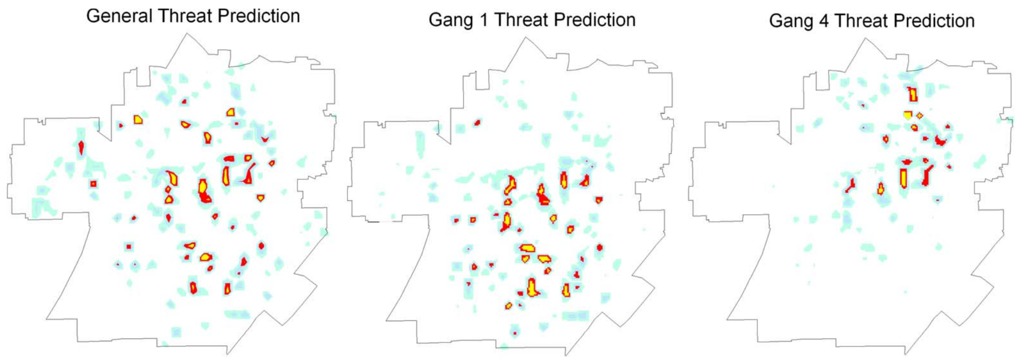

Fig. 2. Threat surfaces for Santa Ana from a multilevel model. The general prediction maps the threat surface for an incident by any of the groups. The other two maps represent the threat surfaces for two specific gangs. The ability of the multilevel model to provide distinct threat surfaces for each of the gangs is evident in these threat maps.

B. No-Pooling Model

The next model we fit is a group-specific model, known as the no-pooling model. The no-pooling model actually consists of J different models, with the index j indexing each group. Only the incidents from the individual group are used to fit the regression for that group. Thus, we get J models of the form

$$\pi(y_i = 1|X) = \text{logit}^{-1}\left(\alpha_{j[i]} + \beta_{j[i]} X_i\right). \tag{2}$$

The notation $j[i]$ indicates that the observations for each individual group j are indexed during model fit by the known group for the actual event observation i. In other words, when the model is fit using actual incidents, model π_j is developed using only incidents identified in the training data set as having been committed by group j. In this manner, the no-pooling model fits the group-specific criminal preference. This model represents the method currently used to build criminal preference models for specific groups in military and police crime analysis software that employ the spatial choice method.

C. Multilevel Model

The multilevel model is of the same logistic regression form as the aforementioned models. However, the multilevel model has two parts. The first part consists of a set of estimated coefficients for each group. In addition, each of the coefficients is also modeled. Coefficients that vary by group and are also themselves modeled are sometimes referred to as *random effects*, referring to the randomness in the probability model for the group level coefficients [11]. This two-part structure can be seen in the following multilevel model:

$$\pi_j\left(y_{[i]} = 1|X\right) = N\left(\text{logit}^{-1}\left[\alpha_{j[i]} + \beta_{j[i]} X_i\right], \sigma_y^2\right) \tag{3a}$$

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

In this model, the intercepts and vector of regression coefficients for the groups are modeled as Gaussian-distributed random variables with mean, variance, and covariance parameters. The model requires that the coefficients for all groups, for example, the intercepts α_j, all come from a common distribution with mean μ_α. The probability of an event by group j, $\pi_j(y_i = 1)$, is Gaussian distributed with a mean determined via the logistic regression and variance σ_y^2.

The requirement for each coefficient for the groups to come from a common distribution generates a tradeoff between the two extremes of the pooled model and no-pooling model. It incorporates what is known about the overall criminal preferences but is weighted toward the individual-group no-pooling model in rough proportion to each groups' contribution to the general model. The coefficients for each of the groups in the multilevel model will lie somewhere between the coefficients of the pooled model seen in (1) and the group-specific no-pooling model seen in (2). As we expect some autocorrelation in the predictor matrix, we also estimate the covariance between the predictors as well. An additional feature of a multilevel model is that it also produces a general model in the form of (1) in which the coefficients α and β of the general model are the means μ_α and μ_β.

IV. Fitting the Spatial Preference Models

A. Fitting the Pooled and No-Pooling Models

The coefficients for the logistic regression models mentioned earlier are fit using maximum likelihood estimation. The vector Y records the responses in the training data set as 1 if an incident occurred at the given location and 0 if an incident did not occur. If we set the first column of the predictor set X equal to 1 to account for the intercept α, we can rewrite (1) as

$$\pi(y|\beta, X) = \prod_{i=1}^{N} \begin{cases} \text{logit}^{-1}(\beta^{\mathrm{T}} X_i), & \text{if } y_i = 1 \\ 1 - \text{logit}^{-1}(\beta^{\mathrm{T}} X_i), & \text{if } y_i = 0. \end{cases} \tag{4}$$

The log-likelihood of this regression equation can be simplified to

$$l(\beta) = \sum_{i=1}^{N} \left\{ y_i \beta^{\mathrm{T}} X_i - \log\left(1 + \exp[\beta^{\mathrm{T}} X_i]\right)\right\}. \tag{5}$$

We solve for our maximum likelihood estimate by taking the derivative of (5), setting it equal to zero, and solving for β. As there is no closed-form solution to the derivative of (5) with respect to β, we estimate a solution for β using the converging *iteratively reweighted least squares algorithm* in which the coefficient matrix is iteratively approximated as follows [15]:

$$\beta^{\text{new}} = (X^{\text{T}} W X)^{-1} X^{\text{T}} W z \tag{6}$$

where

$$z = X\beta^{\text{old}} + W^{-1}\left[y_i - \pi(y_i = 1; \beta^{\text{old}})\right] \tag{7}$$

and W is an $N \times N$ diagonal matrix of weights with the ith diagonal element, i.e., w_{ii}, defined as

$$w_{ii} = \pi(y_i = 1; \beta^{\text{old}})\, \pi(y_i = 0; \beta^{\text{old}}). \tag{8}$$

B. Fitting the Multilevel Model

The coefficient vector for group j, i.e., β_j, is made up of both a fixed effect and a random effect such that

$$\beta_j = \mu + \gamma_j \tag{9}$$

when all coefficients vary by group. In this notation, μ denotes the vector of fixed effects (which is equivalent to the vector of coefficients in the pooled model) and γ denotes the vector of random, or group-specific, effects. If not all coefficients vary by group, then we must define an additional predictor matrix A which defines a matrix of predictors that vary by group. Then, the expected value of our response, given a vector μ of fixed effects and a vector γ of group effects, is

$$E(y_i|\mu, \gamma) = \widehat{y_i} = \text{logit}^{-1}\left(X_i^{\text{T}}\mu + A_{j[i]}^{\text{T}}\gamma\right). \tag{10}$$

The variance of $\widehat{y_i}$ is denoted as $v(\widehat{y_i})$, and γ is normally distributed with mean zero and covariance matrix Σ. We also define a function for notational convenience

$$g(\widehat{y_i}) = X_i^{\text{T}}\mu + A_i^{\text{T}}\gamma_{j[i]}. \tag{11}$$

Estimating the coefficient vectors for the multilevel model requires the use of several approximation algorithms developed and implemented in software by Bates and Pinheiro [16] and recorded by Jiang [17]. These approximations use the Laplacian approximation to approximate the integrals in the likelihood function and minimize Green's penalized quasi-likelihood function [18]. In-depth discussion of the steps involved in approximating the maximum likelihood estimates for the coefficients of a multilevel model is available in [16] and [17].

Assuming that the covariance matrix Σ is known, the approximated maximum likelihood solutions for the coefficients of the model are

$$\mu^{\text{new}} = (X^{\text{T}} V^{-1} X)^{-1} X^{\text{T}} V^{-1} z \tag{12}$$
$$\gamma^{\text{new}} = \Sigma A^{\text{T}} V^{-1}(z - X\mu) \tag{13}$$

where

$$z_i = \gamma_{j[i]}^{\text{old}} + g'(\widehat{y_i})(y_i - \widehat{y_i}) \tag{14}$$
$$V = W^{-1} + A^{\text{T}}\Sigma^{-1}A. \tag{15}$$

Note that the form of the solution for the mean coefficient μ in (12) is similar to that of the pooled and no-pooling model coefficients in (6). The multilevel model mean coefficient is of the same form with a different weighting matrix.

Another approach for fitting multilevel models involves using Markov Chain Monte Carlo (MCMC) techniques [11]. However, this approach requires much more computational power and time and an analyst to monitor the model fit and actively assess convergence of the algorithm. As the ultimate goal of this paper is to develop predictive software for police and intelligence agencies, the requirement to set up and monitor the MCMC algorithm for convergence is problematic. The maximum likelihood approach identified earlier delivers automated convergence of the model fitting process, facilitating the use of the approach by nonstatisticians.

V. Model Feature Selection and Evaluation

Stepwise regression is one of several known algorithms for finding feature subsets in a satisfactory manner. The statistical significance of each of the predictors is calculated as the model is fit, and greater emphasis is placed upon features which are statistically significant as indicated by a p-value. By using the statistical significance of the predictors and a measure of predictive performance such as a receiver operating characteristic (ROC) curve or model deviance, an analyst can quickly converge on a set of features that offer good performance.

Feature selection for a multilevel model is more complex. The complexity is introduced because we are simultaneously modeling several groups at one time. Often in multilevel modeling, predictive features that are significant for one group are insignificant for another and are reflected as insignificant in the pooled model from which the no-pooling and multilevel models are extended. However, removing that predictor will severely hinder the predictive performance for the specific group for which that predictor is significant.

A correctly fit model should offer good performance both in general prediction and for each of the modeled groups. Fig. 2 shows a visualization of a correctly fit multilevel geospatial predictive model. In Fig. 2, one can see that the threat surfaces for the different groups are distinct from the general prediction surface and from each other.

Fitting a multilevel model which delivers good performance in both general and group-specific applications requires a two-step process. First, we use stepwise regression on the pooled model to identify a subset of features which are necessary for a good fit for general prediction (a threat surface predicting criminal incidents by any of the modeled groups). The stepwise regression analysis provides a quick method to reduce the number of predictors to a subset of variables that provides satisfactory predictive performance. We used the Akaike Information Criteria (AIC) to conduct the stepwise regression feature selection, seeking to find a tradeoff between model size (number of features) and model accuracy. We check this performance using a surveillance plot to measure the accuracy and precision of the pooled models as they are fit. The surveillance plot is discussed in depth hereinafter, but it serves to apply the goal of an ROC curve in a spatial context by measuring the accuracy and precision of the threat surfaces produced by the models. We then conduct a stepwise regression analysis of J no-pooling models, one for each group. Each group will have a separate list of features required to provide a good fit.

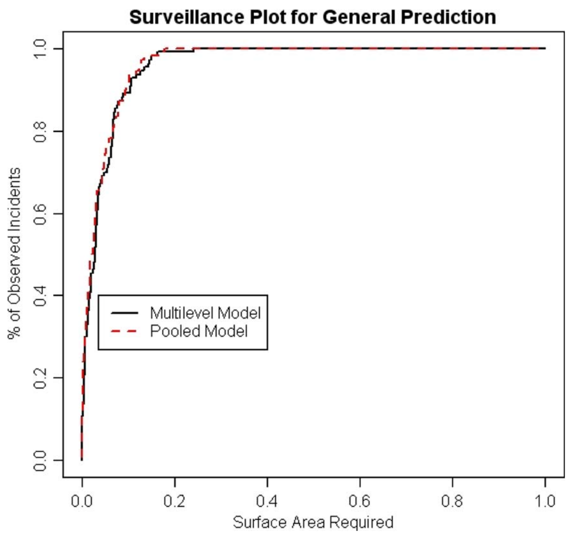

Fig. 3. Surveillance plot comparisons of (solid line) the multilevel model and (dashed line) the pooled model. The general threat surfaces produced by the two models for general prediction are comparable.

The second step of the process is to conduct a stepwise regression analysis for the multilevel model. This analysis is done across the set of features identified as important in either the pooled model or any one of the J no-pooling models. We conduct stepwise regression on the various combinations of the set of features important in either no-pooling or pooled models, once again seeking to find the model which has the best AIC score. An AIC score in a multilevel model reflects the performance across all of the modeled groups, again seeking to find a good tradeoff in model size and accuracy.

A. Evaluating Performance With a Surveillance Plot

A surveillance plot records the amount of surface area predicted by the model a security agency would need to monitor in order to observe a corresponding percentage of incidents occurred and is an adaptation of a model comparison metric introduced by Smith and Brown [8]. A model's tradeoff in accuracy and precision as shown in a surveillance plot is a very important characteristic for evaluating geospatial model performance because we desire a model that focuses security efforts in as small a geographic area as possible.

It is easiest to explain the surveillance plots in Figs. 3 and 4 by referencing the threat surfaces in Fig. 2. Fig. 3 is created by rank ordering all of the points on a model's general predicted threat surface in descending order from the most likely to the least likely location. The horizontal axis of the surveillance plot records the percentage of surface area, referenced in descending order of threat surface, needed to observe the corresponding percentage of actual incidents in the data set. Fig. 3 shows the performance of a model in general prediction. Fig. 4 shows the group-specific performance for the various models. The surveillance plot evaluates the performance of the model in the same way that an ROC curve does, except that it couches the evaluation in terms of evaluating the efficiency of a model in allocating resources expended by a security agency in attempting to observe or interdict the crimes as they occur.

The best model is the one that achieves the highest percentage of observed incidents for a given percentage of high-probability surface area required. The appropriate point of evaluation on the surveillance plot graph is contingent upon the resources available to the evaluating agency. For instance, if it is possible to monitor only 5% of the surface area in the research area with your available resources, then the model that provides the highest number of observed incidents in the test set in 5% or less of the high-probability surface area observed is the best one for your organization.

B. Evaluating Performance With Sphere of Influence Analysis

The predicted sphere of influence for a gang is the area where a given model predicts that each gang is the most likely to commit a crime. This sphere of influence can be mapped by calculating $\pi_j(y_i = 1)$ for all J's for all grid points and then mapping the areas where each of the J groups is dominant as in Fig. 5. We can quickly evaluate the accuracy of the sphere of influence predicted by a model by predicting the individual gang likelihood, $\pi_j(y_i = 1)$, for each analyzed group for every incident in the evaluation set and evaluating the percentage of time that the gang with the highest predicted likelihood matches the actual offender in the test data set. A well-fit model should correctly predict which group committed the crimes in the evaluation data set most of the time. The model that most reliably predicts which group committed the crimes in the test data set has the best combination of reliable likelihood surface models for all of the groups. Sphere of influence performance allows a quick evaluation of the performance of the fit of a particular multilevel model or set of no-pooling models over the set of all groups considered and reduces performance on all groups to a single metric of overall performance.

C. Performance Comparison of Models

We assess and compare the performance of the three types of models using surveillance plots and a sphere of influence analysis. The multilevel model can both build a general predictive surface to model the likelihood of incidents by any gang and also provide threat surfaces for each gang individually. The other two types of models are designed to provide only one type of surface. The pooled model provides a general threat surface. Each no-pooling model provides a threat surface only for the group that it models.

First, we compare the performance of the pooled model and the multilevel model. Fig. 3 compares the performance of the pooled and multilevel models in general prediction, predicting the general incidence of gang crimes. One can observe that, in the general case, the multilevel model performs comparably to the pooled model. This result is expected because the fitted general model for the multilevel model is very similar to the pooled model. The only difference in the two models in the general case is that the stepwise regression analysis for the two models resulted in slightly different features selected, which results in slight differences of performance in general prediction.

We next compare the performance of all three models on group-specific prediction. Although the pooled model provides a threat surface that represents the likelihood of a crime event by any group, some security agencies use the pooled model to predict the likelihood of incidents by the individual groups as well, in effect assuming that all groups have the same

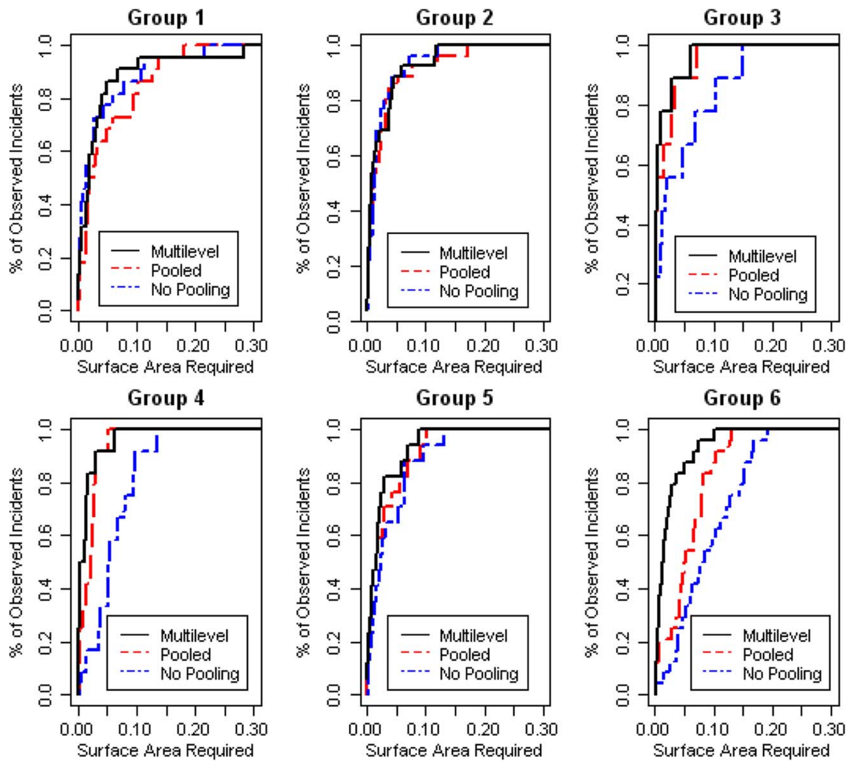

Fig. 4. Surveillance plot comparison of the three spatial models' performance for each of the six analyzed gangs. Note that the multilevel model offers much better performance than the no-pooling model for all groups but 1 and 2 where it is comparable. The multilevel model offers better performance than the pooled model for all groups, with significant improvement for groups 1 and 6.

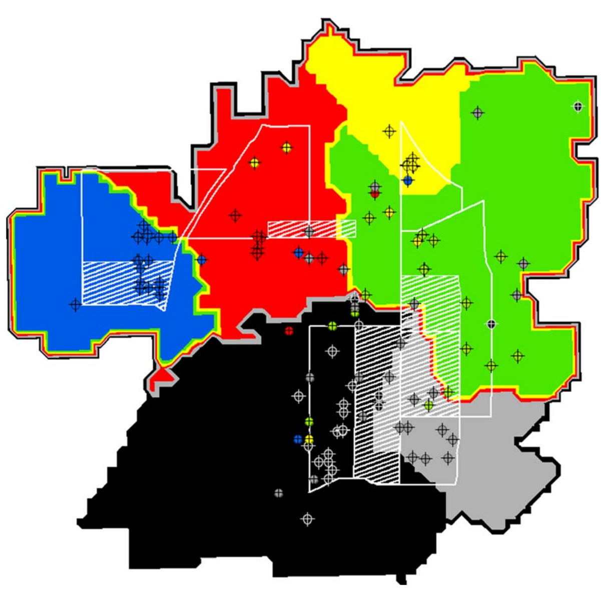

Fig. 5. Multilevel model predicted sphere of influence versus test data set incidents by gang. Incidents and sphere of influence are color coded by gang. Known gang territories for the six groups are plotted (white outline), as well as areas claimed by more than one gang (cross-hatched). This multilevel model of sphere of influence correctly predicts which of the six analyzed gangs committed the crime 69% of the time.

preferences. As can be seen in Fig. 4, the multilevel model provides the best overall fit and provides a significant lift in performance over both the pooled and no-pooling models for some of the groups.

We also compared the performance of the no-pooling models and the multilevel model on the sphere of influence analysis. The pooled model cannot provide a sphere of influence as it makes no differentiation between the modeled groups. The no-pooling model's sphere of influence correctly identifies the gang who committed the crimes in the test data set 51% of the time. The multilevel model sphere of criminal influence improves significantly on this by correctly identifying the gang who committed the crimes in the test set 69% of the time. The multilevel model better predicts a sphere of criminal influence because it models all groups simultaneously and thus is able to better identify the areas where the probability of a crime by one group dominates another. The threat surfaces of the six no-pooling models are not coordinated with each other in the same way.

Another way to visualize the performance of the models in the sphere of influence prediction is to measure the percentage of incidents in each predicted sphere of influence actually committed by the predicted group. Table I records the percentage of crimes committed by the predicted group in each sphere of influence. As can be seen in Table I, the multilevel model's spheres of influence better capture the regions of dominance for each gang as a much higher percentage of incidents in each predicted sphere of influence are committed by the predicted group.

D. Further Analysis

An additional contribution of building spatial preference models over density estimation is that the correlation of the modeled effects of the various features can be analyzed to suggest crime reduction strategies. Fig. 6 shows a visual illustration of the modeled effect of the general spatial

TABLE I
PERCENTAGE OF TEST INCIDENTS COMMITTED BY EACH GROUP IN PREDICTED SPHERE OF INFLUENCE

Sphere of Influence	Multilevel Model	No-Pooling Model
Gang 1	61%	48%
Gang 2	80%	78%
Gang 3	50%	25%
Gang 4	78%	58%
Gang 5	55%	45%
Gang 6	100%	85%

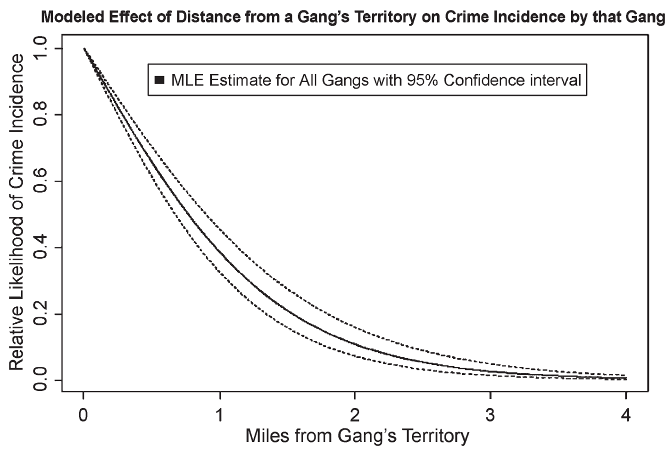

Fig. 6. (Solid line) Modeled effect and (dashed lines) 95% confidence interval for the predictor "Distance from Home Territory." This figure illustrates the modeled effect of the distance from gang's turf on likelihood of observing a gang incident by that gang. At one mile from the gang's turf, the likelihood of observing a gang crime by that group is about 40% of what it is within their gang territory, after all other factors have been taken into consideration.

predictive feature Distance to Gang Turf. This figure illustrates the modeled effect of this predictor and provides a visual mapping of the effect that proximity to a known gang turf has on the likelihood of a criminal incident by one of the modeled gangs. With the notation that ϕ_{β_i} denotes the standard error for the coefficient β_i, we can produce Fig. 6 by plotting the following three equations:

$$\pi(y_i = 1) = \text{logit}^{-1}(\beta_1 x_1) \tag{16a}$$

$$\pi_{\text{upper}}(y_i = 1) = \text{logit}^{-1}\left[(\beta_1 + 2\phi_{\beta_1})x_1\right] \tag{16b}$$

$$\pi_{\text{lower}}(y_i = 1) = \text{logit}^{-1}\left[(\beta_1 - 2\phi_{\beta_1})x_1\right] \quad \text{for } x_1 = 0, 1, \ldots 4 \text{ miles.} \tag{16c}$$

The modeled relationship in Fig. 6 quantifies the nonlinear correlation between the predictor and the incidence of crime. This correlation does not establish causality, but it can serve to identify likely approaches to take in reducing crime. Environmental criminologists often use such correlations in situational crime prevention, in which security agencies attempt to prevent a crime by changing the features of a location highly correlated with criminal incidents [19]. Implementing operations to change the features at certain locations would amount to implementing a treatment on the environment. Additional data collection and analysis over time would be needed to establish a causal relationship between the predictor variables and the incidence of crime. However, the modeled relationships between the predictive features and crime rates serve as an excellent

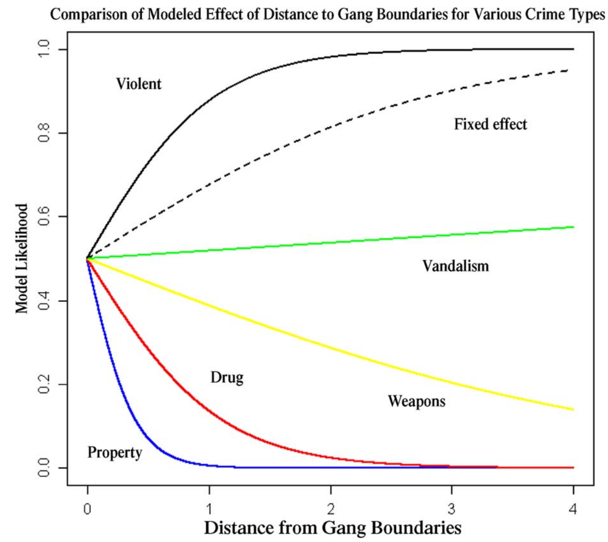

Fig. 7. Comparison of modeled effect of "Distance to Gang Boundaries" for various crime types. This plot illustrates the change in log odds due to increasing distance from the boundaries of the gang territories. As can be seen, property and drug crimes are highly correlated with proximity to gang boundaries, while violent crimes are negatively correlated with the gang boundaries.

starting point for identifying treatments that offer the potential to change an environment so that it is not that conducive to crimes.

As an additional example, the researchers developed a non-nested multilevel model which varied the model coefficients by both group and crime type because crime and intelligence analysts often desire to predict the likelihood of a particular type of criminal act by a particular criminal organization. The GITS project [13] categorized the crimes entered into the database into one of the five types: violent crimes, property crimes, drug crimes, weapons crimes, and vandalism. We leveraged that information to build threat models of each crime type for each of the six most active gangs.

Fig. 7 shows how each of the five crime types differs in its correlation with a predictor variable. This plot illustrates the change in log odds due to increasing distance from the boundaries of the gang territories (the lines designating the borders of the territories as opposed to the general areas claimed by the gangs). As can be seen, property and drug crimes are highly correlated with proximity to these boundaries, while violent crimes are negatively correlated with the gang boundaries (but are positively correlated with presence in a gang territory—indicating that the deeper you are into a gang territory, the more likely you are to experience a violent crime).

VI. CONCLUSION

Multilevel modeling of the spatial preferences of criminal groups offers the opportunity to better answer the questions posed in the introduction. The sphere of influence map indicates which of the modeled groups is most likely to commit a crime at a given location. The threat surfaces like those in Fig. 2 indicate the most likely course of action for each of the modeled criminal groups. Fig. 4 shows that the threat surfaces produced by multilevel models can offer significantly improved performance over those produced with the no-pooling models, the method currently employed by several predictive software packages. The general threat surfaces produced by multilevel models would allow security agencies to more efficiently allocate their resources to those areas most likely to see crime, and the performance of these models is comparable to that of currently used methods.

The multilevel model therefore offers improved performance in group-specific prediction with no cost in general prediction. Finally, in-depth analysis of the modeled effects of the predictors can model the correlation between features and the rates of crime, providing an opportunity to identify likely strategies for situational crime prevention and predict the effects on the environment resulting from the implementation of those strategies.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Meeker of the University of California, Dr. B. Vila of Washington State University, P. M. Walters, and the dedicated professionals of the Santa Ana Police Department for their significant support and insight provided in this work.

REFERENCES

[1] J. J. Medby, *Street Smart: Intelligence Preparation of the Battlefield for Urban Environments*. Santa Monica, CA: RAND Corporation, 2002.

[2] J. E. Eck, S. Chainey, J. G. Cameron, M. Leitner, and R. E. Wilson, "Mapping crime: Understanding hot spots," Nat. Inst. Justice, Washington, DC, Tech. Rep., 2005.

[3] S. Singh, H. Tu, W. Donat, K. Pattipati, and P. Willett, "Anomaly detection via feature-aided tracking and hidden Markov models," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 1, pp. 144–159, Jan. 2009.

[4] T. H. Grubesic, "On the application of fuzzy clustering for crime hot spot detection," *J. Quant. Criminol.*, vol. 22, no. 1, pp. 77–105, Mar. 2006.

[5] K. Kianmehr and R. Alhajj, "Effectiveness of support vector machine for crime hot-spots prediction," *Appl. Artif. Intell.*, vol. 22, no. 5, pp. 433–458, May 2008.

[6] H. Liu and D. E. Brown, "A new point process transition density model for space-time event prediction," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 3, pp. 310–324, Aug. 2004.

[7] D. E. Brown, J. Dalton, and H. Hoyle, "Spatial forecast methods for terrorist events in urban environments," in *Proc. 2nd Symp. Intell. Security Informatics*, Jun. 2004, pp. 426–435.

[8] M. A. Smith and D. E. Brown, "Application of discrete choice analysis to attack point patterns," *Inf. Syst. e-Bus. Manage.*, vol. 5, no. 3, pp. 255–274, Jun. 2007.

[9] "Intelligence analytical support to counter improvised explosive device operations," US Army Train. Doctrine Command, Fort Monroe, VA, Tech. Rep., 2007.

[10] D. McFadden, "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, P. Zarembka, Ed. New York: Academic, 1974.

[11] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge Univ. Press, 2007.

[12] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Quantitative Geography: Perspectives on Spatial Data Analysis*. London, U.K.: Sage, 2000.

[13] J. W. Meeker, B. Vila, and K. J. Parsons, "Developing a GIS-based regional gang incident tracking system," in *Responding to Gangs: Evaluation and Research*. Washington, DC: Nat. Inst. Justice, 2002.

[14] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. R. Stat. Soc.*, vol. 36, no. 2, pp. 192–236, 1974.

[15] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001.

[16] D. M. Bates and J. C. Pinheiro, "Computational methods for multilevel models," Bell Labs Tech. Memo., 1998. [Online]. Available: http://stat.bell-labs.com/NLME/CompMulti.pdf

[17] J. Jiang, *Linear and Generalized Linear Mixed Models and Their Applications*. New York: Springer-Verlag, 2007.

[18] P. J. Green, "Penalized likelihood for general semi-parametric regression models," *Int. Stat. Rev.*, vol. 55, no. 3, pp. 245–259, Dec. 1987.

[19] R. Boba, *Crime Analysis and Crime Mapping*. London, U.K.: Sage, 2005.

Samuel H. Huddleston received the B.S. degree in mechanical engineering from the United States Military Academy, West Point, NY, and the M.S. degree in systems engineering from the University of Virginia, Charlottesville.

He is currently an Instructor with the Systems Engineering Department, United States Military Academy, and also a Combat Veteran of Operation Iraqi Freedom. His research interests include statistical learning, agent-based simulation, and crime analysis.

Donald E. Brown (F'01) received the B.S. degree from the United States Military Academy, West Point, NY, the M.S. and the M.Eng. degrees in operations research from the University of California, Berkeley, and the Ph.D. degree in industrial and operations engineering from the University of Michigan, Ann Arbor.

He is currently a Professor with the Department of Systems and Information Engineering, University of Virginia, Charlottesville. He is also a Director of the Critical Incident Data Analysis Center. His research interests include statistical learning, data fusion, predictive modeling, pattern recognition, and spatial-temporal analysis.

Dr. Brown is a former Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A. He is the 2002 winner of the Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Society for his outstanding contributions to research and education in systems engineering. He is also a recipient of the Millennium Medal from the IEEE.