

Spatial analysis with preference specification of latent decision makers for criminal event prediction

Yifei Xue, Donald E. Brown*

*Department of Systems and Information Engineering, University of Virginia, 151 Engineer's Way, P. O. Box 400747,
Charlottesville, VA 22904, USA*

Available online 1 September 2004

Abstract

Spatial analysis looks for statistically significant patterns in observed events that occur at specified locations. Most examples of spatial analysis consider aggregate characteristics over a number of coarsely defined regions rather than point processes. However, criminal events are point processes and should be modeled as such. In this paper, we combine recent advances in discrete choice theory and data mining to develop point process models for spatial analysis. We use this new methodology to analyze and predict the spatial behavior of criminals, and more generally, latent decision makers. The paper compares the performance of this methodology to more traditional hot spot methods of crime analysis.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Spatial choice; Feature selection; Preference specification; Model-based clustering

1. Introduction

Spatial data analysis is of considerable importance in a number of application areas from marketing to law enforcement. As with time series data, spatial data contain attributes and constraints that require specialized techniques. For example, in spatial analysis the locations and distances to features are often of great importance in developing statistical models and interpreting the results. The development of geographical information systems (GIS) has enabled both greater collection and analysis of spatial data than

ever before. GIS now exist in almost all major police departments and can even be found in ever increasing numbers in the smaller jurisdictions.

While GIS provide a foundation for spatial analysis, the actual analysis itself must still be performed with rather limited toolsets. Typically, GIS and their add-in components support visualization and exploration through the creation of map layers that allow the analyst to view spatial relationships. One of the most common spatial layers for visualization is called hot spots; these are concentrations of crime within indicated geographic areas. Hot spot techniques have applicability to law enforcement because they show high crime areas but the techniques are used in many other areas, such as epidemiology, public health, and

* Corresponding author. Tel.: +1 434 982 2074.

E-mail address: brown@virginia.edu (D.E. Brown).

environmental engineering. Examples of formal statistical models that support hot spot analyses include point pattern analysis [20,21], distance statistics [2] and area analysis [9,15,26].

Spatial analysis of crime requires more than visualization and exploration. Ideally we would also like to have predictive models. However, predictive capabilities in GIS and related toolsets consist of methods designed primarily for the earth sciences or economics. Spatial regression analysis provides an example of a method borrowed from economics and applied to crime analysis. At an aggregate level regression models do provide a way to discover correlations, if not causality, in criminal activities [17,22] and to predict future activities over large areas and time intervals. However, this approach is not well suited to predicting more localized activity.

To meet local needs for predictive models, law enforcement has turned to the above mentioned visualization and descriptive techniques of hot spot analyses. These approaches use the concentrations of crime incidents found from past data to predict that future incidents will occur in these same areas. There are many different approaches to generating hot spots such as spatial histograms, clustering, mixture models, scan statistics, and density estimation. The later technique provides a useful way to compare different methods for prediction and hot spot analysis [28].

To the extent that criminals continue to behave as they have in the past, the use of hot spots for local prediction is useful. However, as with regression, the technique works best as aggregate analysis tool. Hot spots show areas that commonly have high crime rates but provide no insight into the behavior of actual criminals. Further, as the local environment changes, hot spot techniques alone cannot indicate how crime patterns might change. What is needed is a predictive approach that models the micro behavior and spatial choices of criminals.

According to the rational choice perspective in criminology, criminal incidents, like many other human initiated events, involve a decision making and choice process [8]. According to this theory, criminals make decisions that maximize their gain and minimize their risk exposure to law enforcement.

In this paper, we exploit the rational choice perspective to model the site selection behavior of individual criminals. This is similar to methods used

in economics for consumer product selection. In both cases, the decision follows an evaluation by the decision maker of attributes of each alternative. Our model differs from those used in consumer theory in that the decision makers for us are active criminals and hence, not accessible by survey instruments. Further, from the analyst's perspective, the number of spatial alternatives available to the criminal is quite large and typically much larger than any alternative set used in analyzing consumer choices.

To account for these characteristics of criminal spatial choice we begin by developing models that represent criminal preferences. Since surveys are not possible, these models are built out of incident data, because these data show the actual behavior of the criminals. Essentially the patterns in the data show spatial preferences of the criminals. With preferences derived from spatial patterns we can then construct discrete choice models of criminal spatial choice using the theory of discrete choice pioneered by McFadden [18,19]. Discrete choice models have been applied to a number of areas, such as consumer destination selection [12,24], travel mode analysis [1,3,19], and recreational demand models [25].

As noted above, all of this previous work has been focused on predicting the decisions of a large number of known individuals with small numbers of alternatives. The approach we describe here for criminal event prediction using spatial choice analysis represents a new development in discrete choice theory. The alternative set is quite large and the decision makers are latent or unknown and cannot be surveyed for their preferences.

In the rest of this paper, we formally define criminal spatial choice in Section 2 and derive the resulting spatial choice models in Section 3. Section 4, shows the application of the spatial choice models to actual crime data to predict future criminal incidents and compares the results with those obtained from hot spot analysis using density estimation. Finally, Section 5 contains the conclusions.

2. Criminal spatial choice

Data items for criminal spatial choice have two components: a location component and an attribute component. They can be represented by a vector $\{Q,$

$S, k\}$. Q is the universe of the location component of the alternatives. Q is discrete and indexes all spatial alternatives by an ordered pair of coordinates $\{x, y\}$. S is the set of attributes associated with the spatial alternatives, where $S = \{s_1, s_2, \dots, s_S\}$. $k: Q \rightarrow S$ is a mapping which specifies the attribute values for the alternative.

With these data items, the spatial decision process can be represented by a vector $\{Q, S, k, A, D, u, P\}$. The set A is a subset of the spatial locations, Q , indicating finite choices available to all individuals decision makers D . So, $A = \{a_1, a_2, \dots, a_N\}$ represents the N available alternatives for the decision makers to choose from. For spatial analysis of crimes, N is a very large number.

D is the universe of individuals, criminals, who make choices over the available alternative set A . Each individual makes choices based on a decision process. u is the utility function mapping the preferences from individuals in D from the alternative set A to a real number known as the utility value, U . For an individual d , if any two spatial alternatives $a_i \in A_d, a_i' \in A_d'$ have same attribute values, then they will have same utility values $u(a_i) = u(a_i')$ for all $a_i \in A_d, a_i' \in A_d'$.

According to the assumptions of rational decision theory, individuals make choices that maximize their utility U over all alternatives. However, inconsistency exists in the observed individuals' choice behaviors. This inconsistency can derive from an incomplete set of attributes available to the decision maker or the analyst. In general, it arises because of a lack of perfect knowledge about the characteristics of the decision problem. Thus, instead of selecting the alternative with the highest utility, the decision makers make selections stochastically, but biased toward the higher utility values. This means choice selection follows a probability distribution function over all alternatives and the distribution function includes the utilities of the alternatives as parameters. In this work, the probability that an individual d from D will choose alternative a_i from available choice set A can be specified as $P(a_i|A_d, d)$. The probability $P(a_i|A_d, d)$ is a mapping based on the preferences of individual d and the attributes of the alternatives in set A_d . The mapping can be stated as $P: A \times S \times D \rightarrow (0, 1)$, or indicated by the utility-based formulation: $P(a_i|A_d, d) = P\{u(a_i) \geq u(a_j) | d\}$ for all $a_j \in A_d$.

As with the utility values, some information needed for estimating the probabilities is not always available. In most cases for crime analysis, the characteristics of individuals are not directly observed by or available to analysts. The choice set A_d that individual d has considered is also unknown to the analysts. The unavailable information makes the probabilities difficult to estimate. Thus, we require some method to extract the preferences of the latent or unknown decision makers in order to improve the predictions of future spatial choices by these decision makers based on limited available information. An approach to this problem is given in the next section.

Continuing the development of the criminal spatial choice problem, the utility of spatial alternative a_i to individual d can be divided into two parts $U_{id} = V(d, S_i) + \varepsilon(d, S_i)$. $V(d, S_i)$ is the systematic component of the utility function, which is assumed to be deterministic. $\varepsilon(d, S_i)$ is the error term of utility function indicating the previously mentioned unobservable components of the decision makers' utility. The selection of a functional form of $V(d, S_i)$ is important. The function must reflect the theory of how the attributes influence choice and, ideally, provide computational convenience that makes the parameter estimation easier. In most cases, the functions chosen for the deterministic component are linear in parameters [18]: $V(d, s_i) = \sum_l \beta_l^i x_l^i \cdot x_l^i \in X = (S, D)$ represents the l th value of the combination of alternative a_i 's attribute values s_i for individual d . In this formulation, the parameters are dependent on the individual but we suppress this notational detail in the interest of clarity.

3. Criminal spatial choice model development

The problem formulation in Section 2 provides the basis for developing spatial choice models for criminal event prediction. In Section 3.1 we show how we use this formulation to construct a model for criminal site selection. A key component of this model is the method for handling alternative sets and we have developed two approaches to this problem. Section 3.2 shows first approach, uniform spatial choice model, which assumes a hierarchical information processing model and a density estimation

approach to finding the probabilities of alternative sets. Sections 3.3 and 3.4 describe the second approach which relaxes the strong assumption made in the uniform spatial choice model. The resulting approach is called the distinct spatial choice model.

3.1. A model for criminal site selection

The results in Section 2 show how discrete choice theory can describe spatial choices by criminals. In order to actually use this theory for criminal event prediction we need to modify it to allow for latent decision makers and large alternative sets.

From both the analyst's and the criminal's perspectives the alternative set is large. Consider a criminal intent on breaking and entering a property. Even in a modest sized town there may be hundreds, if not thousands, of residences and commercial properties that represent potential targets for this crime.

Looked at from the criminal's perspective, this large number of alternatives means that the criminal cannot evaluate all spatial alternatives before making a selection. The criminal can only compare part of the choice set and pick the spatial alternative with highest utility value. Formally, this can be stated as a sub-optimal or locally optimal problem. According to Fotheringham et al.'s [13] framework of individuals' hierarchical information processing, individuals make spatial choices from the alternatives they have evaluated. For individual d , the choice set will be $A_d \subset A$, which indicates all spatial alternatives that individual d really considered. With perfect information the choice that individual d makes would have the highest utility among all alternatives in choice set A_d . What is different from previous work in discrete choice theory is that in crime analysis the real choice set A_d is not known to the analyst. We need a method to identify or estimate the probability $P(a_i \in A_d)$ that an alternative a_i is considered by individual d .

Hence, unlike previous work in discrete choice, we need to estimate two components: (i) the utility of alternative a_i to individual d and (ii) the probability that alternative a_i is available or considered by individual d . Since the number of spatial alternatives is very large, it is possible that some alternatives can give higher utility values but they are never considered. In order to derive a model for the criminal's preferences, we make the following assumption.

Assumption 1. The two factors (i and ii) mentioned above are equally important to the individuals' choices. So, the combination of $P(a_i \in A_d)$ and the utility of alternative a_i to individual d , U_{id} , can better estimate the probability of the criminal choice behavior than using only the utility of the alternatives.

With Assumption 1, the probability that individual d chooses alternative a_i from A_d can be stated as $P(U_{id} > U_{jd} + \ln P(a_j \in A_d), \text{ all } a_j \in A_d) P(a_i \in A_d)$ [13]. This formulation models restricted choice behaviors by the criminal. From the original formulation, if the alternative i is not in the choice set A_d , $P(a_i \in A_d)$ will be zero and hence the probability that individual d chooses alternative a_i will be zero. The new formulation adds that if j is not in the restricted choice set and $P(a_j \in A_d) = 0$, then $U_{jd} + \ln P(a_j \in A_d) = -\infty$.

While this formulation gives us the method to model alternative sets considered by the criminal, we need a way to model the actual choice behavior. Recall that this choice behavior is stochastic, so we add the next assumption.

Assumption 2. The error term of an individual decision maker's utility function $\varepsilon(d, S_i)$ is independently and identically distributed with Type I extreme value distribution [18].

The Type I extreme value distribution is described as $P(\varepsilon_i < \varepsilon) = \exp(-\exp(-\varepsilon))$ [18]. ε_i represents $\varepsilon(d, S_i)$ here and ε is a constant. With this assumption we derive the spatial choice model following the approach in Refs. [13,18].

$$P(a_i | A_d, d) = \exp(V(d, s_i)) P(a_i \in A_d) / \left(\sum_{j \in A} \exp(V(d, s_j)) P(a_j \in A_d) \right) \quad (1)$$

This spatial choice model is a multinomial logit model where each spatial alternative's observable utility is weighted by the probability that the alternative is evaluated. It serves as the basis for predicting spatial choices by criminals.

3.2. Uniform spatial choice model

Section 3.1 has provided us with a spatial choice model for criminals that uses restricted alternative sets. However, since the analyst does not know the

restricted alternative sets considered by the criminal, we require some mechanism to estimate them. We have two approaches to this problem and this section describes the first of these, which we call the uniform spatial choice model.

In this approach we derive estimates for the probabilities that spatial alternatives are considered by the criminals. We start by assuming that criminals use hierarchical information processing to make their spatial choices. Hierarchical information processing for decision making is a common assumption for consumer choice analysis [13]. Individuals first evaluate sets of alternatives and only alternatives within the sets can be selected. So, if A_d is the set of pre-evaluated spatial alternatives, the probability that an individual will evaluate a specific alternative is $P(a_i \in A_d)$. To capture this probability we make the following strong assumption in this approach.

Assumption 3. For criminal decision making the preferences of all individuals $d \in D$ are same. The pre-evaluated spatial alternative set A_d for different individuals is also same. We use M to represent the set of pre-evaluated spatial alternatives for all individuals.

Under Assumption 3, the spatial choice model in Eq. (1) for criminal site selection changes to

$$P(a_i | A_d, d) = \exp(V(d, s_i)) P(a_i \in M) / \left(\sum_{j \in A} \exp(V(d, s_j)) \cdot P(a_j \in M) \right) \quad (2)$$

The definition of $P(a_i \in M)$ is important in Eq. (3) since it gives the probability that spatial alternative a_i is evaluated by the criminals. We need a method to estimate this probability, and for the approach in this section we use kernel density estimation.

From the study of Brown et al. [7], we know that location components of spatial alternatives alone do not provide enough information about the criminals' preferences. There are many other feature values associated with the spatial alternatives that influence criminal choices. For example, the criminal will estimate the contents of residence based on the appearance of the house, items in the yard, etc.

Again we do not know which non-spatial attributes of the alternatives the criminal uses in making a decision. However, we can discover the criminal's

preferences from all attribute values of past criminal incidents. To do this we use a feature selection process to find the smallest attribute subset from the available space of all attributes. We call the attributes selected by this process key features. We then look for patterns of criminal activity in the space of key features, since these patterns indicate criminal preferences. Using the selected key features, we derive the prior evaluation probabilities $P(a_i \in M)$ as shown below [5].

$$P(a_i \in M) = \frac{1}{L} \sum_{l=1}^L T \times \left(\frac{s_i^1 - s_l^1}{h_1}, \frac{s_i^2 - s_l^2}{h_2}, \frac{s_i^3 - s_l^3}{h_3}, \dots \right) \quad (3)$$

where, $s_i^1, s_i^2, s_i^3, \dots$ are the key features of spatial alternative a_i . L is the total number of observations; T is a function to specify the kernel estimator. We use a Gaussian function here because the Gaussian function provides very smooth estimates of the density values. h 's are bandwidth parameters used in the kernel estimation. The change of bandwidths will influence the effect of density estimation. The choice of bandwidths is important and the literature in this area offers a variety of choices. We use a recommended bandwidth selection method from Bowman and Azzalini [5], $h_i = \left(\frac{4}{(p+2)L} \right)^{1/(p+2)} \times \sigma_i$ for i th dimension. p is the number of dimensions for density estimation. We call the spatial choice model in Eq. (2) with the estimated prior probability $P(a_i \in M)$ as shown in Eq. (3) the uniform spatial choice model. It is called this because we assume uniform decision making by all criminals using the key feature space [28].

3.3. Distinct spatial choice model

The uniform spatial choice model makes a strong assumption about the criminal behavior: their pre-evaluated set of spatial alternatives is the same and their preference structure is the same. The model described in this section relaxes this assumption of uniform criminal decision making. The resulting model is called the distinct spatial choice model.

The problem with trying to make a model that is distinct to each criminal is that there is a very large number of both alternatives and alternative attributes

that may be considered by the criminals. Since the criminals are unknown to the analyst we have no direct way to limit these numbers. It is quite possible that any model we develop will not include the actual attributes used by the criminals in the choices because they are not even measured by anyone but the criminal.

In our development of the uniform spatial choice model, Assumption 3 indicates that the preferences of all individuals $d \in D$ are same. The pre-evaluated choice set A_d is also the same for all individuals. This makes it possible to estimate the pre-evaluated choice set A_d . However, it also makes the estimated individuals' preferences biased due to the lack of specific information about the decision makers' preferences. For crime analysis, it is not possible to include all individual criminal preference information in the spatial choice model. However, we can recover the preferences of decision makers from records of their past criminal incidents.

Obviously we do not know which criminals committed which crimes, until they are caught. At that point they cease to be of interest to us in predicting future crimes. However, we can modify the procedure we used earlier to discover preferences to find individual criminal preferences. Again we apply attribute selection to determine the key features used by criminals in making their choices. Instead of treating the preferences in key features as the same for all criminals, we now look for individual differences. This means we look for clusters or high density areas in feature space that suggest distinct preferences. These clusters show then represent preferences for single criminals or groups of criminals who behave the same. We call the resulting model the Distinct Spatial Choice Model. The next section provides details on our approach to clustering in key feature space to support this model.

3.4. Clustering methods for the distinct spatial choice model

As noted above, we use clustering to identify distinct criminals or groups of criminals who share a common preference structure. This section shows how this clustering is performed.

Clustering is one of the most useful tasks in data mining for discovering groups and identifying inter-

esting distributions and patterns of an underlying data set. Clustering involves partitioning a given data set such that the data points in a cluster are more similar to each other than points in different clusters. Researchers have extensively studied clustering since it occurs in many applications in engineering and science.

Clustering may result in a different partitioning of a data set, depending on the specific criterion used for clustering. The basic steps to develop clustering can be summarized as feature selection, clustering, validation of the results, and interpretation of the results. Feature selection chooses the features on which clustering is to be performed so as to encode as much information as possible. For our purposes this step was used to find the key features for criminal site selection. By removing all features that are irrelevant to classification, the small feature space subset provides enough information for criminal preference discovery thereby reducing the cost and improving the quality of the results [23]. The clustering algorithm is the most important part of the clustering process, and it includes similarity measures, partitioning methods, and stopping criteria. Each of these is described in Refs. [11,16].

No matter which clustering algorithms are used, it is important to find a way to define a stopping criterion or define how many clusters are in the data set. Various strategies for simultaneous determination of the number of clusters and cluster membership have been proposed, such as those in Engelman and Hartigan [10], Bock [4], Bozdogan [6], and Fraley and Raftery [14].

The approach we employ is similar to that found in Ref. [14]. In Ref. [14], the data are viewed as coming from a mixture of probability distributions, each representing a different cluster. The resulting mixture density is shown below.

$$f(a_i) = \sum_{k=1}^K \pi_k f_k(a_i|\theta_k) \quad (4)$$

$f_k(a_i|\theta_k)$ is the density of an incident a_i from the k th cluster. θ_k are the corresponding parameters. K is the number of clusters in the mixture. π_k is the probability that an observation belongs to the k th cluster. ($\pi_k \geq 0$ and $\sum_{k=1}^K \pi_k = 1$). By maximizing the likelihoods of the mixture model, each incident

is labeled with different component numbers. The resulting mixture likelihood function is indicated below.

$$L_m = (a_i, \Theta, \Pi) = \prod_{i=1}^n \sum_{k=1}^K \pi_k f_k(a_i | \theta_k) \quad (5)$$

The density function $f_k(a_i | \theta_k)$ is frequently assumed to be a multivariate Gaussian distribution and that is the assumption we will make. This density function has the form as

$$f_k(a_i | \mu_k, \Sigma_k) = \frac{\exp\{\frac{1}{2}(a_i - u_k)^T \Sigma_k^{-1} (a_i - u_k)\}}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} \quad (6)$$

where μ_k is the mean vector, Σ_k is covariance matrix of incidents. These are the parameters of the density distribution. The clusters are assumed to be ellipsoidal and centered at the means μ_k . The parameterization of covariance matrix Σ_k decides the characteristics (orientation, volume and shape) of the distributions of clusters. These characteristics can be allowed to vary between clusters or constrained to be the same for all clusters.

With this formulation expectation maximization (EM) is used to find the clustering solution or the maximum mixture likelihood. When EM is used for this purpose, a reliable approximation, twice the log BIC, is provided to compare different models. The BIC is called Bayesian Information Criterion, which has the definition below.

$$\text{BIC} = 2l(a_i, \hat{\theta}, \Pi) - m \log(n) \quad (7)$$

Here, m is the number of parameters. Bayesian Information Criterion (BIC) is used as a criterion to compare different models and decide the number of clusters in the data set.

4. Evaluation of the spatial choice models

The models of Section 3 provide a method for predicting criminal incidents. This section examines the performance of these models on real data as compared to hot spot methods using density estimation.

4.1. Crime data set

The data for model evaluation came from the Regional Crime Analysis Program (ReCAP). ReCap was created as a cooperative project between local police departments in Northern and Central Virginia and the researchers in the University of Virginia. ReCAP system uses databases, a geographic information system (GIS), and statistical tools to analyze, predict, and display future crime patterns. It is a data integration system, an interactive information sharing and decision support system, and a dynamic reporting and visualization system.

The specific data set used for evaluating this work came from the criminal incident reporting mainframe system (RAMS). It is the data set used by the police department in Richmond, VA. The main table in RAMS is the “Report” table. The information about the time, location, and narrative description of criminal incidents is stored in table “Report”. The features of each spatial alternative come from the combination of census data and calculated distance values. The feature values are appended to the crime records by locations of criminal incidents. As noted above, all features are possibly related to the decision process of criminals.

We extracted the Residential “Breaking and Entering” (B & E) criminal incidents between July 1, 1997 and October 31, 1997 in the city of Richmond, VA for model estimation and validation. There are more than 1200 crime observations. The data from the first three months are used for model estimation. The observations of October 1997 are used for model validation. Using the crime incidents in the training data set, we got locations of all incidents on a geographic map. The sub regions shown in Fig. 1 are block groups, which are the smallest areas, for which census counts are recorded.

The analysis of B & E is related to locations of households in a city. However, it is difficult to represent all locations of individual houses in even a modest sized city, such as Richmond. Therefore, we aggregated alternatives using 2517 regular grids, which were assumed to be fine enough to represent all spatial alternatives within this area.

4.2. Feature selection by similarities

Since the attributes of spatial alternatives came from census data and calculated distance values, it is

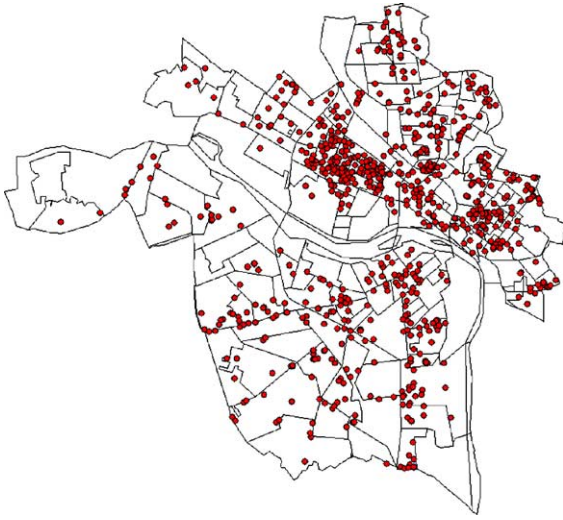


Fig. 1. Breaking and Entering criminal incidents between July 1, 1997 and September 30, 1997 in Richmond, VA.

possible that some values of these attributes are correlated. Using the calculated correlation values as similarities, we computed hierarchical clusters for all features of observed spatial incidents. The resulting clustering structure of features of observed spatial alternatives is shown as Fig. 2.

From the clustering tree, we divided the features into five clusters. Each cluster included correlated

features. After checking the distribution of the feature values, we found that COND1.DST (number of condition 1 houses per unit area) is almost uniformly distributed. So, it is not a good feature for our analysis. There are two choices for feature selection of the rest of the features: randomly picking from each cluster or combining the features into the same clusters. We picked the features D.HIGHWAY (distance to highway), FAM.DENSITY (Family density per unit area), P.CARE.PH (personal care expenditure per household) and D.HOSPITAL (distance to hospital). The first three were used by Brown et al. [7]. These are the key features for use in the spatial choice models and each represents all other features in its cluster.

Based on the selected features, we applied the model based clustering methods [14] to the crime data for analysis. As indicated in Section 3, the number of clusters was decided by the calculated BIC values. The larger BIC values indicate better clustering results. The trends of BIC values are indicated by Fig. 3. From Fig. 3, it is clear that the BIC values become stable or decreasing after six clusters.

The results shown in Fig. 3 suggest there are six clusters in this crime data set. Each cluster corresponds to a group of criminals that have similar preferences in their choices of spatial alternatives. The

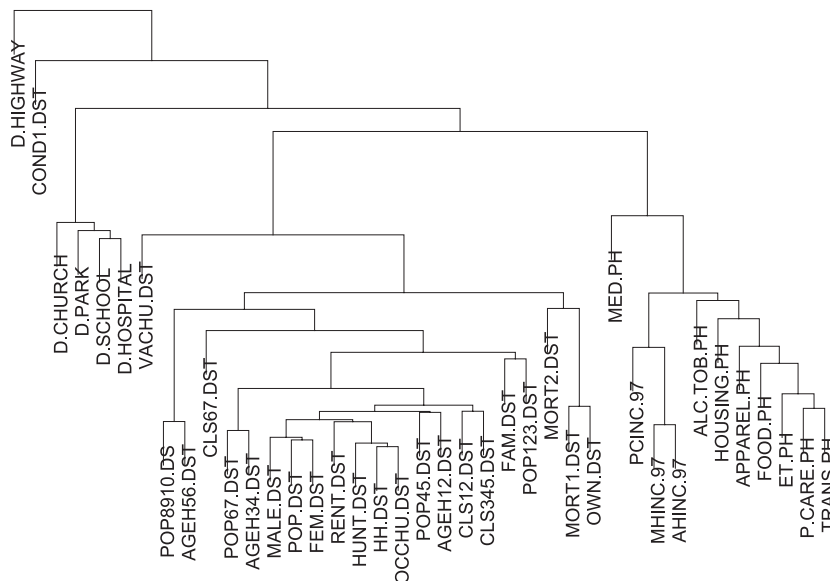


Fig. 2. Clusters of features of observed spatial alternatives.

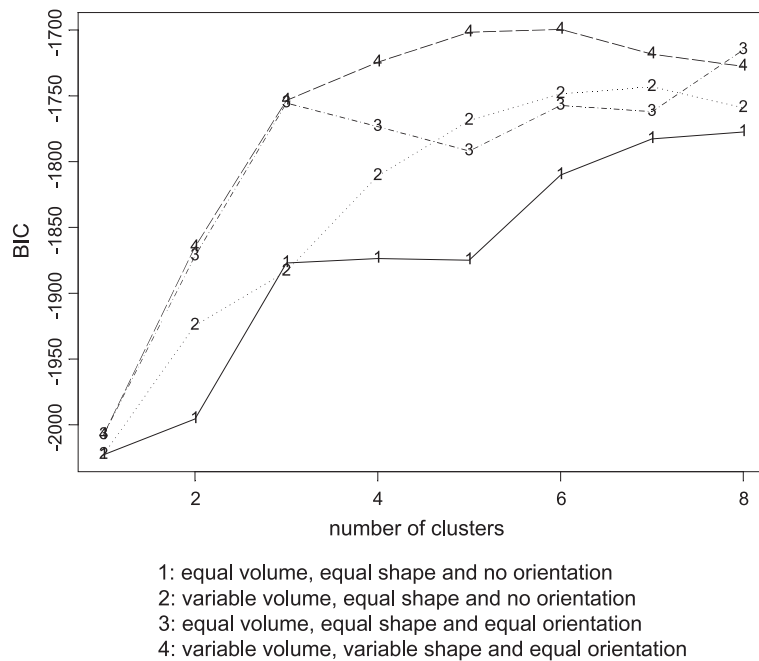


Fig. 3. The trends of BIC values of different parameterized model-based clustering algorithms.

numbers of criminal incidents within the clusters are listed in Table 1.

4.3. Model estimation and prediction

As we have noted several times, the number of spatial alternatives for crime analysis is very large. From an implementation standpoint, this makes the data preparation and computation time prohibitively expensive. To handle this problem, we adopted an importance sampling technique suggested by Ben-Akiva [1]. Sampling alternatives is a commonly applied technique for reducing the computational burden involved in the model estimation process.

Next we consider the model estimation and prediction step. The prior probabilities $P(a_i \in M)$ of the uniform spatial choice models were calculated as in Section 3.2 for each cluster. The key features are

the features coming out from the feature selection process above.

Using the training data set of B & E incidents of each cluster, we obtained the estimation of the uniform spatial choice model for each cluster $P(a_i|A_d, d \in M_k)$. M_k indicates the presence of criminals with preferences in the k th cluster. The final prediction of a future crime's spatial distribution is the combination of the predicted probabilities of all clusters. The combination method is also very important. Given the conditional probability that spatial alternative a_i will be picked by criminals within cluster M_k , $P(a_i|A_d, d \in M_k)$ and the chance that criminals $d \in M_k$ will commit the next crime within the study region, the probability that spatial alternative a_i is picked by any criminal will be $P(a_i|A_d, d \in M) = \sum_{k=1}^K P(a_i|A_d, d \in M_k)P(M_k)$. K is the total number of clusters within the crime data set. The probability $P(M_k)$ can be determined by many methods. Here we used a ratio as

$$P(M_k) = \frac{P(a_i \in M_k)}{\sum_{j=1}^K P(a_i \in M_j)} \quad (8)$$

$P(a_i \in M_k)$ is the probability that an individual $d \in M_k$ pre-evaluate spatial alternative a_i .

Table 1
Distribution of crime incidents in clusters

Crime incidents		Crime incidents	
Cluster 1	109	Cluster 4	202
Cluster 2	180	Cluster 5	133
Cluster 3	200	Cluster 6	55

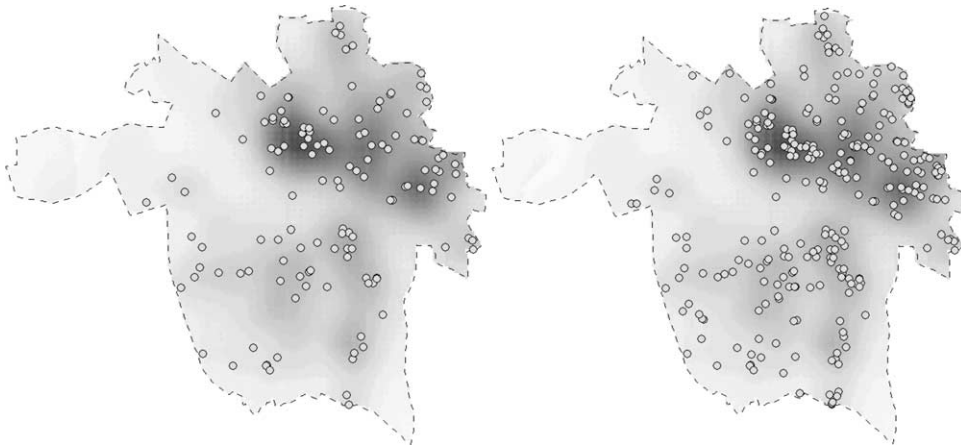


Fig. 4. Prediction of hot spot model with crime incidents from 10/01/97 to 10/15/97 and incidents from 10/01/97 to 10/31/97.

For the distinct spatial choice model described in Section 3, we developed predictions using the mixture modeling technique given in that section. We also used hot spot models as the comparison models to evaluate the uniform spatial choice model and distinct spatial choice model. The residential B & E incidents between October 1, 1997 and October 31, 1997 were used as testing data set. The predicted spatial distributions of future crimes and the testing incidents are shown in Figs. 4–6.

The dark areas in Figs. 4–6 indicate high crime densities. The circular spots represent the locations of crimes that took place between October 1 and October 15 or between October 1 and October 31. The distributions of the predicted crime densities are smoother for the hot spot model and uniform spatial choice model. However, the distinct spatial

choice model provides more accurate and distinct predictions.

4.4. Formal model comparisons

The prediction results of the spatial choice models are the probabilities of spatial alternatives for the next criminal incident. The sum of the probabilities for all spatial alternatives in the choice set A is 1. For incident a_i , let the predicted probability given by the spatial models be $p_{a_i}^p$ and that given by the comparison hot spot model be $p_{a_i}^c$. Then $\sum_{i=1}^N p_{a_i}^p = \sum_{i=1}^N p_{a_i}^c = 1$. The evaluation hypothesis is that for the population of all future criminal incidents, the proposed model will statistically outperform the comparison model. Assume that the testing data set contains m incidents that occurred at locations a_1, a_2, \dots, a_m , respectively.



Fig. 5. Prediction of uniform spatial choice model with crime incidents from 10/01/97 to 10/15/97 and incidents from 10/01/97 to 10/31/97.

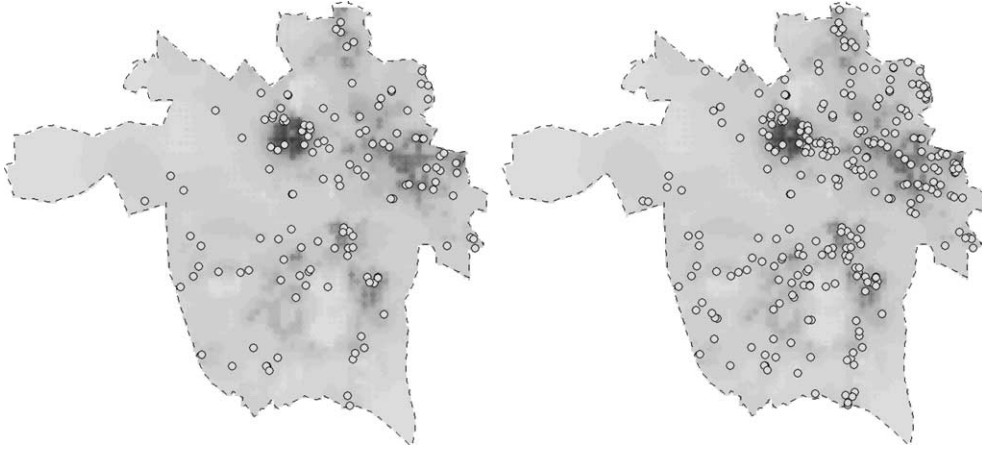


Fig. 6. Prediction of distinct spatial choice model with crime incidents from 10/01/97 to 10/15/97 and incidents from 10/01/97 to 10/31/97.

These data and the corresponding predictions are considered as sampling results here.

To compare the proposed models and the comparison model, we need to specify a hypothesis and a test statistic. The differences between the predicted probabilities of the models can provide us the test statistic. Since the distribution of the differences of predictions on sampled locations obviously do not follow Gaussian or other parametric distributions, we need to use nonparametric statistics for our evaluation.

The data for model comparison are predicted probabilities of different models in fixed locations. In other words, they are paired observations from the predictions. We use $(p_{a_i}^p, p_{a_i}^c)$ to represent the paired predictions at spatial alternative a_i . The pair can be reduced to a single sample by considering differences. The Wilcoxon signed rank test is designed to test whether a particular sample comes from a population with specified median [27] and we use it here for our evaluation.

The difference between the paired predictions is represented by D_{a_i} . The absolute differences is indicated as:

$$|D_{a_i}| = |p_{a_i}^p - p_{a_i}^c| \quad (9)$$

The absolute differences are computed for each $i=1, 2, \dots, m$. Since the predictions are probabilities of different models, it is unlikely the differences are zero. So we define the number of pairs as m . Ranks from 1 to m are assigned to these m pairs according to the relative size of the absolute differences. The rank 1 is

given to the pair $(p_{a_i}^p, p_{a_i}^c)$ with the smallest value on $|D_{a_i}|$; the rank 2 is given to the pair with the second smallest absolute difference; and so on with the rank m being assigned to the pair with the largest absolute difference. One assumption is required here before the definition of the hypothesis and the test statistic.

Assumption 4. Each D_{a_i} is a continuous random variable and the distribution of D_{a_i} is symmetric. The D_{a_i} 's are mutually independent and have the same median.

Since D_{a_i} is assumed to be symmetric, this indicates that the expectation and median of D_{a_i} are same. Then the null hypothesis for our model comparison is that the mean of the predictions of proposed spatial choice model will be less than or equal to the mean of the predictions of the comparison model. The alternative hypothesis states that the null hypothesis is not true.

$$\begin{aligned} H_0 : E(p_{a_i}^p) &\leq E(p_{a_i}^c) \\ H_a : E(p_{a_i}^p) &> E(p_{a_i}^c). \end{aligned} \quad (10)$$

The test statistic is defined as the sum of the ranks assigned to the pairs $(p_{a_i}^p, p_{a_i}^c)$ that have positive values. Let R_i be the rank assigned to $(p_{a_i}^p, p_{a_i}^c)$. If $p_{a_i}^p, p_{a_i}^c > 0$, $R_i' = R_i$. Otherwise, $R_i' = 0$. The test statistic Z is defined as

$$Z = \sum_{i=1}^m R_i' \quad (11)$$

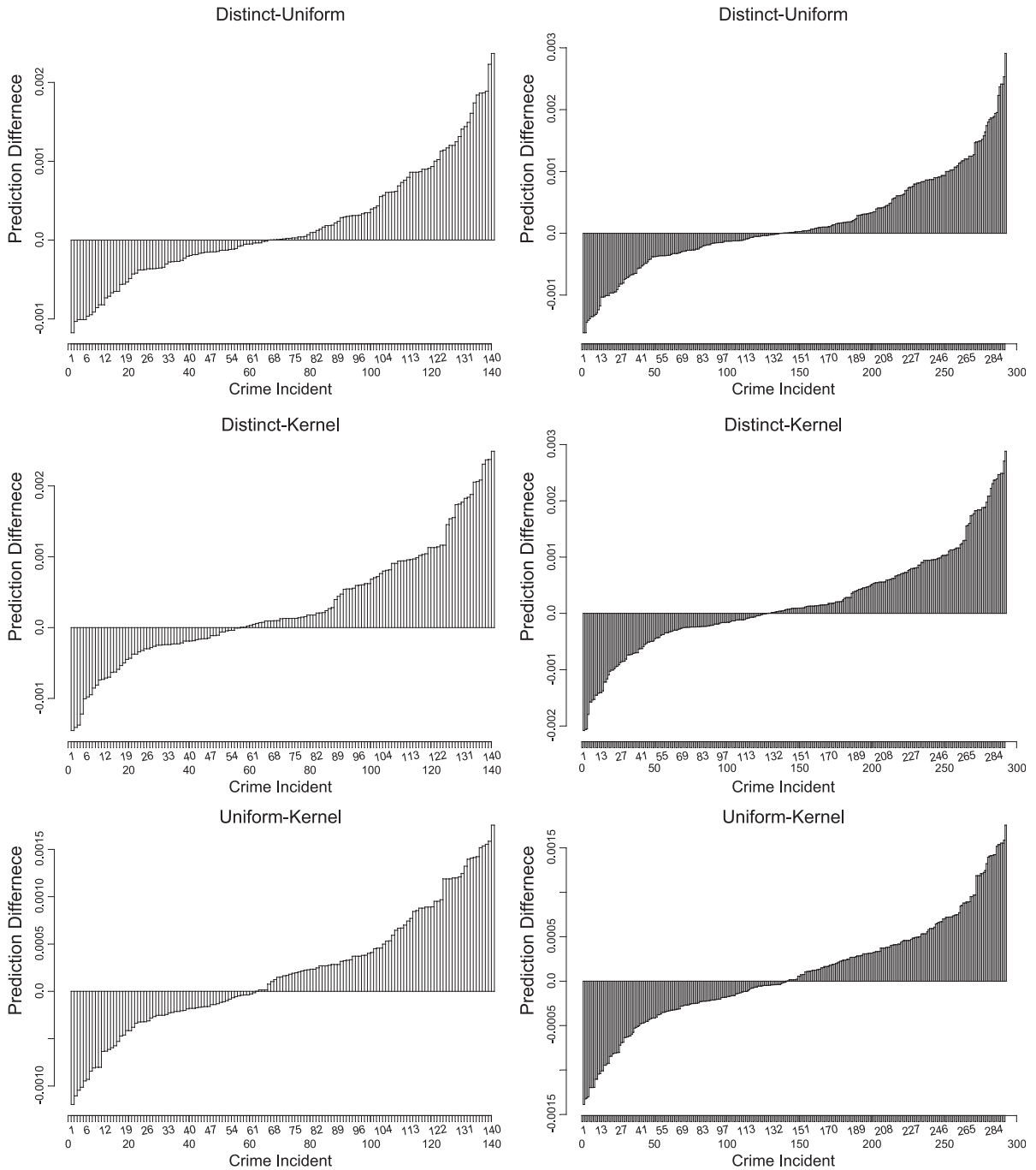


Fig. 7. Comparisons between distinct spatial choice model, uniform spatial choice model, and hot spot model for two testing data sets (from 10/01/97 to 10/15/97 and from 10/01/97 to 10/31/97).

Assume that the new models have better prediction results than the comparison model for all future crimes. The larger value of Z indicates that the hypothesis H_0 is false. With certain significance level, we can reject the H_0 hypothesis and accept the alternative hypothesis H_a . As indicated above, the testing and comparison results of the new models and the comparison model are accomplished by using the Wilcoxon's signed rank test.

The predictions of the compared models are paired probabilities over the spatial alternatives in choice set A . The testing data set provides locations or spatial alternatives at the sites that the crimes really happened. During the model comparison process, we are trying to discover if the new models can provide higher probabilities on the locations that really contain criminal incidents. The comparisons of the predictions are represented by bar plots in Fig. 7. Each plot provides a comparison between new models and the hot spot model on one of the two testing data sets. The vertical axis indicates the differences between the predictions of the two compared models. The higher bar for each incident represents the bigger difference between the predicted probabilities of the compared models. The horizontal axis stands for the incident numbers, which have been sorted by the differences of the predicted probabilities of future crimes. The larger areas above the horizontal axis indicate the new model outperform the hot spot model. The hot spot model used in the comparison was a kernel density estimation model. Hence, the label in the figure for the hot spot method is kernel to indicate that this particular method for hot spots was used.

Using the Wilcoxon signed rank test to compare the models, we get the testing results in Table 2.

Table 2
The model comparison results

	Z statistic	p value
<i>Testing data set 1 (10/01/97–10/15/97)</i>		
Distinct vs. uniform	1.8055	0.0355
Distinct vs. hot spot	3.2876	0.0005
Uniform vs. hot spot	3.0619	0.0011
<i>Testing data set 2 (10/01/97–10/31/97)</i>		
Distinct vs. uniform	2.407	0.008
Distinct vs. hot spot	3.1871	0.0007
Uniform vs. hot spot	2.1259	0.0168

The Z statistic is the Wilcoxon test statistic. The p value is the probability that is the significance level for the test.

The comparison results in Table 2 indicate that the two new spatial choice models significantly outperform the hot spot model. The distinct spatial choice model also significantly outperforms the uniform spatial choice model. The results demonstrate that the analysis of feature values attached to all spatial alternatives and the analysis of specified preferences of decision makers lead to improvement in the prediction of future crime locations.

In addition to these formal tests, the analysis results have been presented to crime analysts and community residents. These evaluators agreed that the spatial choice models predict some crime areas that did have high Breaking and Entering crime rates in the past several years. These results suggest that the estimation of criminals' preferences and incorporation in spatial choice models provides a method for improving criminal event prediction.

5. Conclusions

Spatial analysis is of critical importance to law enforcement. It enables better planning and the use of scarce resources. It has become more important as a way to address the variety of threats facing modern communities. Past work in this area has concentrated on aggregated approaches to understanding criminal behavior and displayed results of this analysis as hot spots. In this paper, criminal incidents are described as product of spatial choice processes of criminals or latent decision makers. The preferences and characteristics of criminals are impossible to obtain directly as was done in other choice scenarios such as transportation and marketing. Further the number of alternatives available to criminals is much larger than that considered in these other applications.

Two models of spatial choice are described in this paper. The first model, uniform spatial choice model, makes a strong assumption on criminal preferences: criminals have the same choice sets and preferences. The second model, distinct spatial choice model, relaxes this assumption. Together the models show how the preferences of criminals can be modeled to better understand the spatial patterns of crime.

When evaluated with actual breaking and entering data, these method increased the accuracy of the prediction of future criminal locations significantly. The differences of the paired predictions between the hot spot approach and the new models are visualized by bar plots. The testing results and the visualization show that the new models perform much better than the hot spot methods using density estimation. In addition, the method also provides a way to interpret the relationship between criminal decision making and spatial attributes.

References

- [1] M. Ben-Akiva, S. Lerman, *Discrete Choice Analysis, Theory and Application to Travel Demand*, MIT Press, Cambridge, MA, 1985.
- [2] J. Besage, J. Newell, The detection of clusters in rare diseases, *Journal of the Royal Statistical Society A* 154 (1991) 143–155.
- [3] C. Bhat, Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling, *Transportation Science* 34 (2) (1998 May) 228–238.
- [4] H.H. Bock, Probability models and hypothesis testing in partitioning cluster analysis, in: P. Arabie, L. Hubert, G. DeSorte (Eds.), *Clustering and Classification*, World Science Publishers, River Edge, NJ, 1996.
- [5] A. Bowman, A. Azzalini, *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, Oxford Statistical Science Series, Oxford Science Publications, Oxford, NY, 1997.
- [6] H. Bozdogan, Choosing the number of component clusters in the mixture model using a new informational complexity criterion of the inverse Fisher information matrix, in: O. Opitz, B. Lausen, R. Klar (Eds.), *Information and Classification*, Springer-Verlag, New York, 1993, pp. 40–54.
- [7] D. Brown, H. Liu, Y. Xue, Mining preferences from spatial-temporal data, *Proceedings of First SIAM Conference*, 2001.
- [8] R. Clarke, D. Cornish, Modeling offenders' decisions: a framework for research and policy, in: M. Tonry, N. Morris (Eds.), *Crime Justice: An Annual Review of Research*, vol. 6, University of Chicago Press, Chicago, IL, 1985.
- [9] A.D. Cliff, J.K. Ord, *Spatial Processes, Models, and Applications*, Pion, London, 1981.
- [10] L. Engelman, J.A. Hartigan, Percentage points of a test for clusters, *Journal of the American Statistical Association* 64 (1969) 1674.
- [11] B. Everitt, *Cluster Analysis*, John Wiley & Sons, New York, 1993.
- [12] S. Fotheringham, Consumer store choice and choice set definition, *Marketing Science* (1988 Summer) 299–310.
- [13] S. Fotheringham, C. Brunsdon, M. Charlton, *Quantitative Geography*, SAGE Publications, Thousand Oaks, CA, 2000.
- [14] C. Fraley, A.E. Raftery, How many clusters? Which clustering method?—Answers via model-based cluster analysis, *The Computer Journal* 41 (8) (1998) 578–588.
- [15] U. Graham, B. Fingleton, *Spatial Data Analysis by Example*, John Wiley & Sons, New York, 1985.
- [16] A.K. Jain, M.N. Murty, P.J. Flynn, *ACM Computing Surveys* 31 (3) (1999 September) 264–323.
- [17] A. Kposowa, K.D. Breault, Reassessing the structural covariates for U.S. homicide rates: a county level study, *Sociological Forces* 26 (1993) 27–46.
- [18] D. McFadden, Conditional logit analysis of qualitative choice behavior, *Frontiers in Econometrics* (1973) 105–142 (New York).
- [19] D. McFadden, K. Train, The goods/leisure tradeoff and disaggregate work trip mode choice models, *Transportation Research* 12 (1978) 349–353.
- [20] S. Openshaw, M. Charlton, C. Wymer, A. Craft, A mark 1 geographical analysis machine for the automated analysis of point datasets, *International Journal of Geographical Information Systems* 1 (1987) 335–358.
- [21] S.A. Openshaw, A. Craft, M. Carlton, J. Birch, Investigation of leukemia clusters by use of a geographical analysis machine, *Lancet* 1 (1988) 272–273.
- [22] W. Osgood, Poisson-based regression analysis of aggregate crime rates, *Journal of Quantitative Criminology* 16 (1) (2000).
- [23] B.D. Ripley, *Spatial Statistics*, John Wiley and Sons, New York, 1981.
- [24] R. Rust, N. Donthu, Capturing geographically localized misspecification error in retail store choice models, *Journal of Marketing Research* XXXII (1995) 103–110.
- [25] K. Train, Recreational demand models with taste differences over people, *Land Economics* 74 (1998) 230–239.
- [26] G. Upton, B. Fingleton, *Spatial Data Analysis by Example*, John Wiley & Sons, New York, 1985.
- [27] F. Wilcoxon, Individual comparisons by ranking methods, *Biometrics* 1 (5.1, 5.3) (1945) 80–83.
- [28] Y. Xue, D. Brown, A decision model for spatial site selection by criminals: a foundation for law enforcement decision support, *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews* 33 (2003 February) 78–85.