

RESEARCH

Open Access

# The spatio-temporal modeling for criminal incidents

Xiaofeng Wang\* and Donald E Brown

\* Correspondence:  
Predictive Technology Laboratory,  
University of Virginia,  
Charlottesville, Virginia, USA

## Abstract

Law enforcement agencies monitor criminal incidents. With additional geographic and demographic data, law enforcement analysts look for spatio-temporal patterns in these incidents in order to predict future criminal activity. When done correctly these predictions can inform actions that can improve security and reduce the impact of crime. Effective prediction requires the development of models that can find and incorporate the important associative and causative variables available in the data. This paper describes a new approach that uses spatio-temporal generalized additive models (ST-GAMs) to discover underlying factors related to crimes and predict future incidents. In addition, the paper shows extensions of the ST-GAM approach to produce local spatio-temporal generalized additive models (LST-GAMs). These local models can better predict criminal incidents conditioned on regions. Both models can fully utilize a variety of data types, such as spatial, temporal, geographic, and demographic data, to make predictions. We describe how to estimate the parameters for ST-GAM using iteratively re-weighted least squares and maximum likelihood and show that the resulting estimates provide for model interpretability. This paper also discusses methods to generate regions for LST-GAM. Lastly the paper discusses the evaluation of LST-GAM and ST-GAM with actual criminal incident data from Charlottesville, Virginia. The evaluation results show that both models from this new approach outperform previous spatial models in predicting future criminal incidents.

**Keywords:** spatio-temporal modeling, generalized additive model, criminal forecasting

## 1 Introduction

Law enforcement agencies have the need to model the spatio-temporal patterns of crimes. With a model of criminal incidents, they can study the causality of crimes and predict the locations and time of future criminal activity. If the model can predict future crimes accurately, law enforcement can deploy resources to improve security and reduce threats. Typical actions taken by law enforcement include walking and driving patrols, surveillance systems, and neighborhood watch programs.

Many types of data are available to assist building such models. Law enforcement agencies in the United States usually monitor criminal incidents as they occur. For example, they have locations and times of criminal incidents, as well as victim and perpetrator information. In addition to criminal incident data, most agencies can also

acquire spatial information from geographic information systems (GIS) and demographic and economic data from the census.

Several techniques and models have been developed to meet the need for predictive policing with available data. One of the most popular methods is the spatial hot spot model. In the hot spot model [1], current criminal incident data are collected and clustered over space. The locations of such clusters are so-called hot spots. The model assumes the current crime clusters to persist over the forecast horizon. Future criminal incidents are predicted to occur in these same areas. Methods to generate hot spots include spatial histograms, clustering, mixture models, scan statistics, and density estimation. The hot spot model only utilizes criminal incident data, such as types of crimes, locations and time of criminal incidents. It only shows the current patterns of crimes without the insight into the relationship between crimes and environment over time. As the local environment changes, the hot spot model cannot indicate the changes of crime patterns.

To address this problem, more sophisticated statistical models using both criminal data and environmental data have been built by researchers. Liu and Brown [2] applied a point pattern density model to criminal incidents. The spatial density of criminal incidents was assumed to be conditioned on features associated with locations. These features included geographic features, such as the distances to the nearest interstate highways, demographic features and consumer expenditure features. Xue and Brown [3] and Smith and Brown [4] developed a spatial choice model. They assumed criminals made choices to pick places that could be modeled by random utility maximization. This utility maximization is over all alternatives, where the utility is defined by the gain from crimes and the risk of being caught. Brown, Dalton, and Hoyle [5] then discussed a method that uses generalized linear models (GLM) to compute the risk over a territory. They first partitioned the space into grids. Each grid was associated with a response indicating whether incidents happened and features about the grid. Then, a spatial GLM was built with all grids. They applied the spatial GLM to predict terrorist events. Results showed the spatial GLM had better prediction performance than the density models. Rodrigues and Diggle [6] combined point process models and generalized additive models (GAM) to build a semiparametric point source model. In their model, features affected the risk nonlinearly. They applied the model to study the effect of installed security cameras on crimes.

None of the above models directly incorporate the temporal information of criminal incidents. For instance, Liu and Brown use Bayesian to model building that can include a variety of time series but no specific approach it recommended or tested. Other models usually estimate different parameter sets based upon coarse divisions of time. For example, these models use criminal incidents that happened within the most recent year to generate hot spots for this year. Another intuitive method discussed by Ivaha, Al-Madfai, Higgs, and Ware [7] first models the temporal behaviors of crimes with time series models and then models the spatial behaviors given the predicted number of incidents at a certain time. However, this approach does not model interactions between space and time. Recent research has developed spatio-temporal approaches that apply Generalized Additive Models (GAM) to combine spatial, temporal, and other (e.g., demographic) features for prediction. This approach with GAM has performed well in predicting continuous response variables such as matter

concentrations [8]. However, to date no similar approach has been attempted for criminal incident prediction.

In this paper, we describe a new approach to criminal incident prediction using spatio-temporal generalized additive model (ST-GAM) and local spatio-temporal generalized additive model (LST-GAM). The paper is organized as follows. Section 2 defines the problem of criminal incident modeling formally, describes ST-GAM and LST-GAM in detail, and discusses how to estimate parameters as well as evaluate spatial prediction performance. Section 3 applies both ST-GAM and LST-GAM to model the breaking and entering incidents in Charlottesville, VA and evaluates the prediction performance. Finally, section 4 gives conclusion and suggestions on future work.

## 2 The spatio-temporal modeling of criminal incidents

In this section, we first define the problem of criminal incident prediction formally. Then we describe our two models to model and predict criminal incidents. After that, we discuss how to estimate the model parameters. In the end of this section, we introduce a method to evaluate the performance of spatio-temporal predictions.

### 2.1 Definition of the problem of criminal incident prediction

As discussed in section 1, we have various features about an area where criminal incidents have occurred. We want to model the patterns of criminal incidents with these features and apply the model to predict the locations and times of future criminal incidents. Equivalently, we need to model the probability of a criminal incident happening at a certain location and time given all the features associated with this location and time.

Mathematically, we have an area of interest  $S \subset \mathbb{R}^2$ , a time period  $T \subset \mathbb{R}^+$ , and features  $\{X_{s,t} | s \in S, t \in T\}$  associated with  $S$  and  $T$ . To represent the area  $S$  and the time period  $T$ , we can partition  $S$  into grids  $\{s_i\}$  and  $T$  into time intervals  $\{t_j\}$ , where  $\cup s_i = S$ ,  $\cup t_j = T$ , and  $i, j \in N^+$  are indices. The features associated with a grid  $s_i$  and a time interval  $t_j$  can be represented by a vector  $X_{s_i, t_j}$ .

Our objective is to find a probability function:

$$p(\text{inci}_{s_i, t_j} = 1 | X_{s_i, t_j}) \quad (1)$$

and a decision function:

$$\delta_{s_i, t_j} = \delta[p(\text{inci}_{s_i, t_j} = 1 | X_{s_i, t_j})] : [0, 1] \rightarrow \{0, 1\} \quad (2)$$

such that:

$$L = \frac{1}{N_S N_{T^*}} \cdot \sum_{s_i \in S, t_j \in T^*} [w_0 \cdot I(\delta_{s_i, t_j} = 0 | \text{inci}_{s_i, t_j} = 1) + w_1 \cdot I(\delta_{s_i, t_j} = 1 | \text{inci}_{s_i, t_j} = 0)] < \epsilon \quad (3)$$

where  $\text{inci}_{s_i, t_j} = 1$  means at least one incident happens at the grid  $s_i$  and time  $t_j$ ;  $\text{inci}_{s_i, t_j} = 0$  means no incident happens at the grid  $s_i$  and time  $t_j$ ;  $p(\text{inci}_{s_i, t_j} = 1)$  is the probability that at least one incident happens at the location  $s_i$  and time  $t_j$ ;  $T^*$  is a set of time intervals in the future;  $L$  is a loss function;  $N_S$  is the total number of grids;  $N_{T^*}$  is the total number of time intervals;  $I(\cdot)$  is an indicator function;  $w_0, w_1$  are weights of different types of errors; and  $\epsilon$  is the tolerance threshold.

The major difficulty of this problem is to find an accurate probability function  $p(\text{inci}_{s_i,t_j} = 1|X_{s_i,t_j})$ , such that it has high values for locations where criminal incidents will happen and low values for locations where criminal incidents will not happen for any future time intervals. Given a good probability function, the law enforcement agencies can easily choose their own decision functions based on resources and risk preferences. For example, they can choose a cutoff value  $p^* = 0.8$  to classify areas with predicted probabilities higher than  $p^*$  as high risk areas.

In this paper, we focus on the development of  $p(\text{inci}_{s_i,t_j} = 1|X_{s_i,t_j})$ .

## 2.2 The spatio-temporal generalized additive model (ST-GAM)

To model  $p(\text{inci}_{s_i,t_j} = 1|X_{s_i,t_j})$ , we developed a spatio-temporal generalized additive model (ST-GAM) [9]. Additive models usually perform well for problems having many predictors and provide interpretable results. As shown in [5], the spatial generalized linear model (GLM) had better predictability than other probability models. In ST-GAM, we used the generalized additive model (GAM) [10] instead of GLM, because GAM is more exible in the treatment of nonlinearity than GLM. GAM assumes additivity between predictors, but allows for local nonlinearity in each predictor. This exibility is helpful for modeling criminal incidents. For example, criminals may prefer to burgle richer houses. However, they might not choose expensive houses because these houses often have security systems. To include temporal information of previous criminal incidents, the ST-GAM borrowed the idea of the binary time-series-cross-sectional data (BTSCS) model from Beck, Katz, and Tucker [11]. In the BTSCS model, a dummy variable is used to indicate when the last incident happened.

Our ST-GAM has the following form:

$$\text{logit}[p(\text{inci}_{s_i,t_j} = 1)] = \sum_{n=1}^N f_n(x_{n,s_i,t_j}) + \kappa_{s_i,t_j} \quad (4)$$

where  $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$  is a logit function;  $N$  is the total number of features;  $x_{n,s_i,t_j}$  is the  $n^{\text{th}}$  feature associated with location  $s_i$  and time  $t_j$  ( $X_{s_i,t_j} = (x_{1,s_i,t_j}, \dots, x_{N,s_i,t_j})$ );  $f_n$  is the smooth function of the  $n^{\text{th}}$  feature to be estimated from data; and  $\kappa_{s_i,t_j}$  is the dummy variable indicating the length of the continuous zeros (no accident happens) that precede the current observation at location  $s_i$  and time  $t_j$ . An example of the values of  $\kappa_{s_i,t_j}$  is shown in table 1. Notice that  $\kappa_{s_i,t_j}$  is a dummy variable, and its values are factors instead of integers.

With features and past criminal incidents, we can estimate the smooth functions  $f_n$ , ( $n = 1, \dots, N$ ) and parameters of  $\kappa_{s_i,t_j}$  ( $\kappa_{s_i,t_j} = 1, \dots, K$ ) for the above model. Here,  $K$  is the maximum length of the continuous zeros considered. For example, if the last incident happened before  $K$  time intervals at location  $s_i$  and time  $t_j$ , then  $\kappa_{s_i,t_j} = K$ .

**Table 1 An Example of the Values of  $\kappa$**

$t_j$	1	2	3	4	5	6	7
$\text{inci}_{s_i,t_j}$	0	0	1	1	0	1	0
$\kappa_{s_i,t_j}$	1	2	3	1	1	2	1

### 2.3 The local spatio-temporal generalized additive model (LST-GAM)

ST-GAM assumes that all grids in the area  $S$  have the same underlying pattern. The probability of criminal incidents happened in  $S$  is computed by a single equation. In reality, it is possible to have multiple regions within  $S$ . For example, we have all incident data of a state, including big cities, small towns and rural counties. Different types of regions might have different criminal patterns. In addition, the same feature might impact high risk areas (such as crime hot spots) differently from low risk areas.

To account for this situation, we extend the ST-GAM to the local spatio-temporal generalized additive model (LST-GAM) as follows:

$$p(\text{inci}_{s_i, t_j} = 1) = \sum_{r=1}^R I(s_i \in S_r) \cdot p_r(\text{inci}_{s_i, t_j} = 1) \quad (5)$$

$$\text{logit}[p_r(\text{inci}_{s_i, t_j} = 1)] = \sum_{n=1}^N f_{r,n}(x_{n,s_i, t_j}) + \kappa_{r,s_i, t_j}, \text{ for all } r \in \{1, \dots, R\} \quad (6)$$

In the above LST-GAM, equation 5 models the probability of criminal incidents over the whole area  $S$  with  $R$  regions. Here,  $R$  is the total number of regions in  $S$ .  $S_r$  is the  $r^{\text{th}}$  region, where  $\{S_r | S_r \subset S, r \in \{1, \dots, R\}\}$  satisfies  $\cup_r S_r = S$  and  $S_{r_i} \cap S_{r_j} = \emptyset$  ( $r_i \neq r_j$ ).  $I(\cdot)$  is an indicator function with values of 0 and 1.  $p_r(\text{inci}_{s_i, t_j} = 1)$  models the probability of criminal incidents happened within the region  $r$ . Equation 6 defines  $p_r(\text{inci}_{s_i, t_j} = 1)$  for each region  $r$ . As we can see, it has the same form as equation 4. Notice that there are actually  $R$  equations in the form of equation 6 and each one may have different smooth functions  $f_{r,n}(\cdot)$  and parameters of  $\kappa$ .

LST-GAM can be considered as a two stage model. The first stage is to decide which region a grid  $s_i$  belongs to. The second stage is to build different ST-GAM to model the probability of criminal incidents for each region. Clearly, ST-GAM is the special case of LST-GAM with  $R = 1$ .

### 2.4 Estimation of ST-GAM and LST-GAM

The ST-GAM has the form of regular GAM. GAM has been studied extensively in many different disciplines. Therefore, it can be estimated efficiently by well developed methods and algorithms. Standard statistical softwares, such as R, S-plus and SAS, also have implements to estimate GAM.

We briefly review steps to estimate GAM here. Interested readers can refer to [12,13] for details.

To estimate the GAM model in equation 4, the smooth function  $f(x)$  is first represented by a sum of basis functions:

$$f(x) = \sum_{i=1}^B \beta_i \cdot b_i(x) \quad (7)$$

where  $b_i(x)$  is the  $i^{\text{th}}$  basis function; and  $\beta$  are the unknown parameters to be estimated.

A popular choice of basis functions is the cubic regression spline. The basis functions for this spline include:

$$b_1(x) = 1 \quad (8)$$

$$b_2(x) = x \tag{9}$$

$$b_{i+2}(x) = R(x, x_i^*) \tag{10}$$

where  $\{x_i^* | i \in \{1, \dots, B - 2\}\}$  are knots of the spline; and  $R(x, z)$  is as follows [13]:

$$R(x, z) = [(z - 1/2)^2 - 1/12][(x - 1/2)^2 - 1/12]/4 - [(|x - z| - 1/2)^4 - 1/2(|x - z| - 1/2)^2 + 7/240]/24 \tag{11}$$

By representing all smooth functions  $f(x)$  with basis functions, equation 4 becomes a GLM:

$$\text{logit}[p(\text{inci}_{s_i,t_j} = 1)] = \beta_0 + \sum_{n=1}^N [\beta_{n,1} \cdot x_{n,s_i,t_j} + \sum_{i=1}^{B_n-2} \beta_{n,i+1} \cdot R(x_{n,s_i,t_j}, x_{i,n,s_i,t_j}^*)] + \kappa_{s_i,t_j} \tag{12}$$

The above GLM can be estimated efficiently by maximizing the penalized likelihood using the penalized iteratively re-weighted least squares method (P-IRLS) [14].

A possible difficulty with the estimation of ST-GAM is that the size of training dataset is usually huge and the response variable is a sparse vector. For example, we built a predictive model of criminal incidents for Charlottesville, Virginia (total area is 26.6 km<sup>2</sup>) using the grid size of 32m × 32m and the time interval of one month [9]. Then, there were 1,062,094 records for a year while about 1,700 records had the response of 1. It is time consuming to evaluate parameters using all the records. Therefore, we used subsampling. To generate a sample from the records, we included all the records with the response of 1 and a random sample from the records with the response of 0. Based on the analysis in [15], the effect from this biased sampling can be approximately corrected by adding an offset term  $\log(\text{sample size}/\text{total number of records})$  in the estimation process. Thus, the subsampling technique can reduce the size of training set and save estimation time. However, this method introduces stochastic effects to parameter estimates. If possible, we suggest to use all of training data to estimate parameters.

For LST-GAM, we can use the above method to estimate equation 6. In addition, we need to define or estimate regions  $\{S_r\}$ .  $\{S_r\}$  can be defined by domain knowledges. For example, if law enforcement agencies believe that criminal patterns are different in different cities, each  $S_r$  can be a different city. When no such knowledge is available, we can estimate  $\{S_r\}$  with features  $\{X_{s,\cdot}\}$  and past incidents  $\{\text{inci}_{s,t}\}$ . If we assume the region with high risk has different underlying patterns from the low risk region, we can use the following method to generate  $\{S_r\}$  based on the incident density:

1. Estimate the incident density over the whole area  $S : \{d_s | d_s \in [0, 1], s \in S\}$ ;
2. Pick points:  $\{d_1^*, \dots, d_{R-1}^* | d_0^* < d_1^* < d_2^* < \dots < d_{R-1}^* < d_R^*\}$ , where  $d_0^* = 0$  and  $d_R^* = 1$ ;
3.  $\{S_r\}$  based on the incident density are:  $S_r = \{s_i | d_{r-1}^* \leq d_{s_i} < d_r^*, s_i \in S\}$

### 2.5 A method to evaluate models: HRP vs. TIP

As discussed in section 2.1, we want to minimize the loss function  $L$  defined in equation 3. The first part of the function  $\sum_{s_i \in S, t_j \in T^*} w_0 \cdot I(\delta_{s_i,t_j} = 0 | \text{inci}_{s_i,t_j} = 1)$  is the

weighted sum of the times of incorrect predictions for grids where criminal incidents actually happen. The second part of the function  $\sum_{s_i \in S, t_j \in T^*} w_1 \cdot I(\delta_{s_i, t_j} = 1 | inci_{s_i, t_j} = 0)$  is the weighted sum of times of incorrect predictions for grids where no criminal incident happens. To minimize the first part, the probability model should predict high probabilities for the locations where incidents actually happen. To minimize the second part, the total area of the locations with high probabilities should be small at a given time because of the sparseness of criminal incidents over the whole area  $S$ .

Both of the two criteria are important. The first criterion means the model should not miss a high risk area so that law enforcement agencies can know all of future locations of crimes. The second criterion is important, because the police resources are limited and only a part of area can be patrolled at a given time. With a good model, they can better allocate limited resources to help prevent crimes. Based on these criteria, we proposed the HRP vs. TIP method to evaluate the performance of spatial predictions at a given time [9].

To measure the performance of a model at time  $t_j$ , the method first computes:

$$HRP_{\theta} = \frac{\| \{s_i | p(inci_{s_i, t_j} = 1) > \theta\} \|}{\| \{s_i\} \|} \quad (13)$$

$$TIP_{\theta} = \frac{\| \{inci_{s_i, t_j} = 1 | s_i \in \{s_i | p(inci_{s_i, t_j} = 1) > \theta\}\} \|}{\| \{inci_{s_i, t_j} = 1\} \|} \quad (14)$$

where  $\{p(inci_{s_i, t_j} = 1) | s_i \in S, t_j \in T^*\}$  are predictions from the model;  $\|\cdot\|$  is the size of a set; and  $\theta$  is a threshold ( $\theta \in [0, 1]$ ). Here,  $HRP_{\theta}$  represents the percentage of high risk area predicted by models; and  $TIP_{\theta}$  represents the percentage of incidents (from test set) happened within the high risk area given  $\theta$ .

Two vectors of HRP and TIP can be computed with different thresholds  $\{\theta_i | \theta_i \in [0, 1]\}$ . Then, TIP is plotted against HRP. The result plot looks like the receiver operating characteristic (ROC) curve [16]. Ideally, we hope as many as incidents happen within the high risk area with a given size. Therefore, the curve from a good model should be close to the upper left corner. Examples of this plot are shown in Figures 3, 4, 5, 6. Similar to ROC analysis, we can use area under the curve (AUC) to compare the performance of different models by a single score. Because a good model has the curve close to the upper left corner, AUC of a good model should be close to 1. It is easy to see that a random guess model has AUC= 0.5. Therefore, AUC of a bad model should be close to or less than 0.5.

### 3 Model evaluation: predicting criminal incidents in Charlottesville, Virginia

This section shows the application of our ST-GAM and LST-GAM to breaking and entering incidents in Charlottesville, Virginia. These two models, along with the spatial GLM [5] and the hot spot model, are evaluated and compared with the real incident data based on their performance of prediction.

#### 3.1 Data

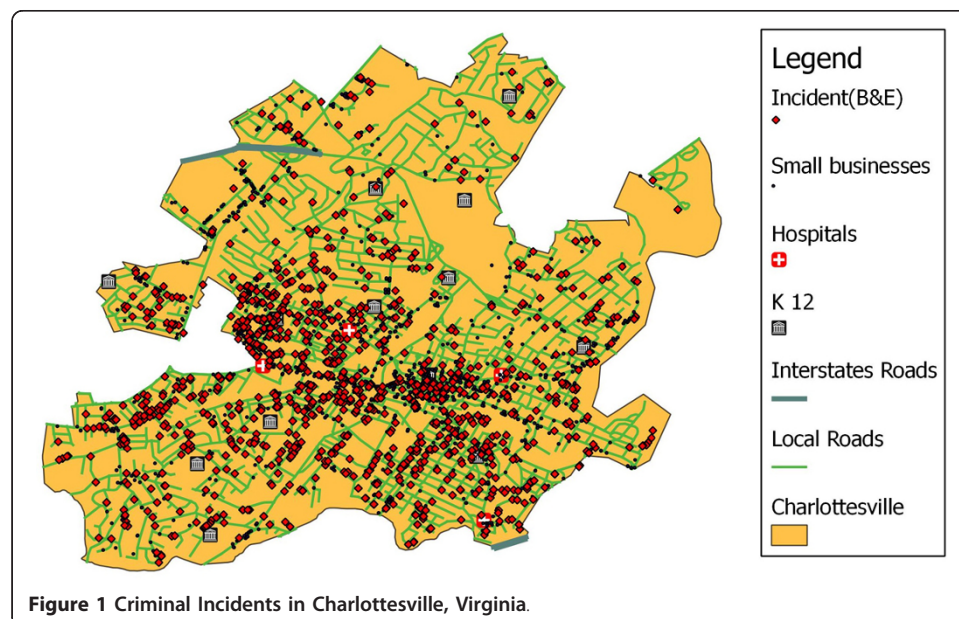
We used three data sets for this study. The first data set includes breaking and entering incidents in Charlottesville, Virginia from April 2001 to February 2005. In total, there were 1,795 incidents (58 incidents without the exact coordinates were excluded).

Each incident in this study had the coordinates of the incident and the time of when it happened. The second data set was the geographic information of the city in the form of GIS layers, such as locations of roads, interstate highways, small businesses and schools. The third data set had the demographic data of Charlottesville measured in census block groups, including population, median values of all houses, races, marriages and so on. Figure 1 shows a small number of geographic features of Charlottesville with all the breaking and entering incidents.

### 3.2 Model construction and estimation

To model the criminal incidents in Charlottesville, we first partitioned the city into spatial grids with the size of 32m × 32m. The total number of grids covering the area was 23,089. We used the time interval with the length of one month and there were 46 months in the data set. Therefore, we had 1,062,094 (= 23,089 × 46) records. Each record had a response variable indicating whether at least one incident happened within the grid and the time period. There were also two types of features associated with each record as explanatory variables. The first type was the distance feature. We calculated the shortest distance between the centroid of a grid and a certain geographic landmark, such as the distance to the nearest road. This calculation was done by a toolkit programmed in Visual C# and PostGIS [17]. The second type was the demographic feature, such as the population, marriage status, house values. Because we used the demographic data measured in census block groups, the demographic features of a grid actually measured the properties of the neighborhood where the grid was located. There were 14 distance features and 20 demographic features. For this study, we only kept the most important 11 features out of 34 features as did in our previous study [9]. Those 11 features were selected by the stepwise selection of GLM. Table 2 shows the description of the features for modeling.

To test our models, we kept the incident data that happened in the last 12 months as the test data. Thus, the training data set included incidents between April 2001 and February 2004. The test data set included incidents between March 2004 and February 2005.





**Table 2 Features Used for Modeling**

Feature	Type	Description	Significance*
$\kappa$	Temporal	the dummy variable indicating the time of the previous incident	2,3,4
college_univ_dist	distance	the distance to the nearest college or university	1,2,3
k_12_dist	distance	the distance to the nearest K-12 school	1,2,4
roads_all_dist	distance	the distance to the nearest road	1,2,3,4
roads_interstates_dist	distance	the distance to the nearest interstate highway	2,3
small_businesses_dist	distance	the distance to the nearest small business	1,2,3,4
median_val	demographic	median value of all housing units	1,2,3
males	demographic	number of males	1,2,3
widowed	demographic	number of people whose spouse died	
divorced	demographic	number of people who are divorced	1,2,3,4
owner_occ	demographic	count of owner-occupied households	1,2,4
medianrent	demographic	median rent charged for all housing units that are rented	1

\* The significance column describes whether a feature is significant (at the level of 95%) in the following models: 1-Spatial GLM, 2-ST-GAM, 3-LST-GAM (in region  $S_1$ ), and 4-LST-GAM (in region  $S_2$ ).

We first built ST-GAM and LST-GAM as described in section 2. To build ST-GAM and LST-GAM, we chose the parameter  $K = 13$ , which means incidents happened before one year would not be considered. To build LST-GAM, we defined two regions,  $S_1$  and  $S_2$ , using the method discussed in section 2.4. The high risk region  $S_2$  included 10% of the area with the highest incident density. The low risk region  $S_1$  included the other 90% of the area. We used the package “mgcv” in R [18] to estimate the smooth functions and parameters in ST-GAM and LST-GAM. This package implemented the estimation method of GAM described in section 2.4. To avoid stochastic effects from subsampling, we used all the training data to estimate models.

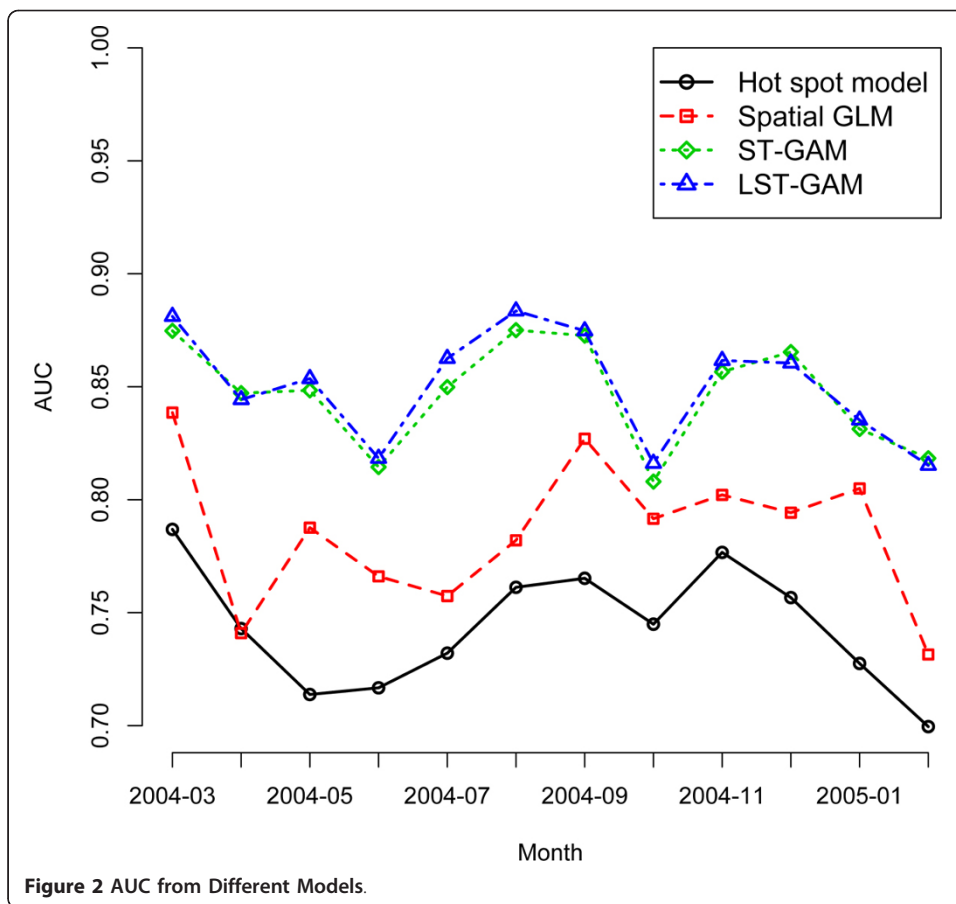
To compare our models with the previous work, we also built a spatial GLM and a hot spot model. The spatial GLM used the same features in table 2 and parameters were estimated with all the training data. The hot spot model estimated the density with all the incidents in the training data set using Gaussian kernels. Both models were estimated by the software R.

### 3.3 Results

#### 3.3.1 Prediction performance

We applied ST-GAM, LST-GAM, the spatial GLM and the hot spot model to predict the probability of criminal incidents in Charlottesville from March 2004 to February 2005 using the test data set. Then, we compared those four models with the metric described in section 2.5.

Figure 2 shows the AUC of 12 month predictions using the four models. The larger the AUC value is, the better the model predicted. Apparently our two models, ST-GAM and LST-GAM, performed better than the previous work, the spatial GLM and the hot spot model. The performance of LST-GAM was a little better than the performance of ST-GAM. To test whether the difference between any two curves in Figure 2 was significant, we performed paired Wilcoxon significance tests on groups of AUC values. Table 3 shows the test results. Small p-values mean the differences are significant ( $p < 0.05$ ). We can see that the difference between any two curves was significant. Therefore, LST-GAM was significantly better than ST-GAM. Both LST-GAM and ST-GAM were significantly better than the spatial GLM and the hot spot model.

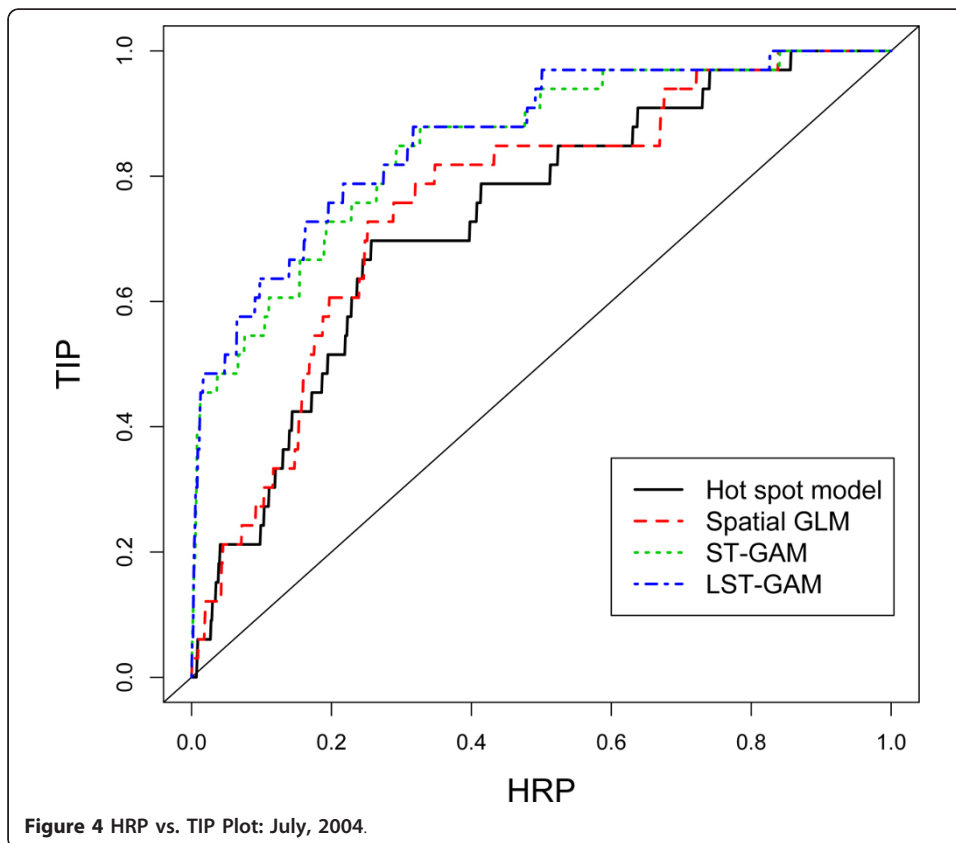
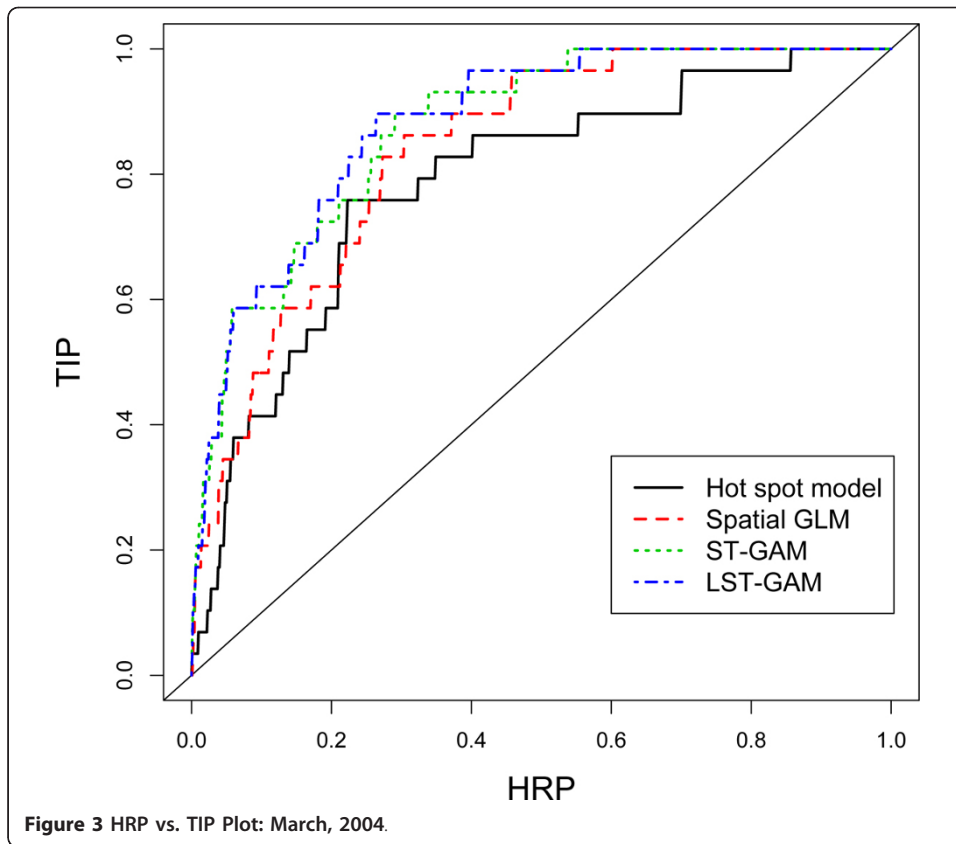


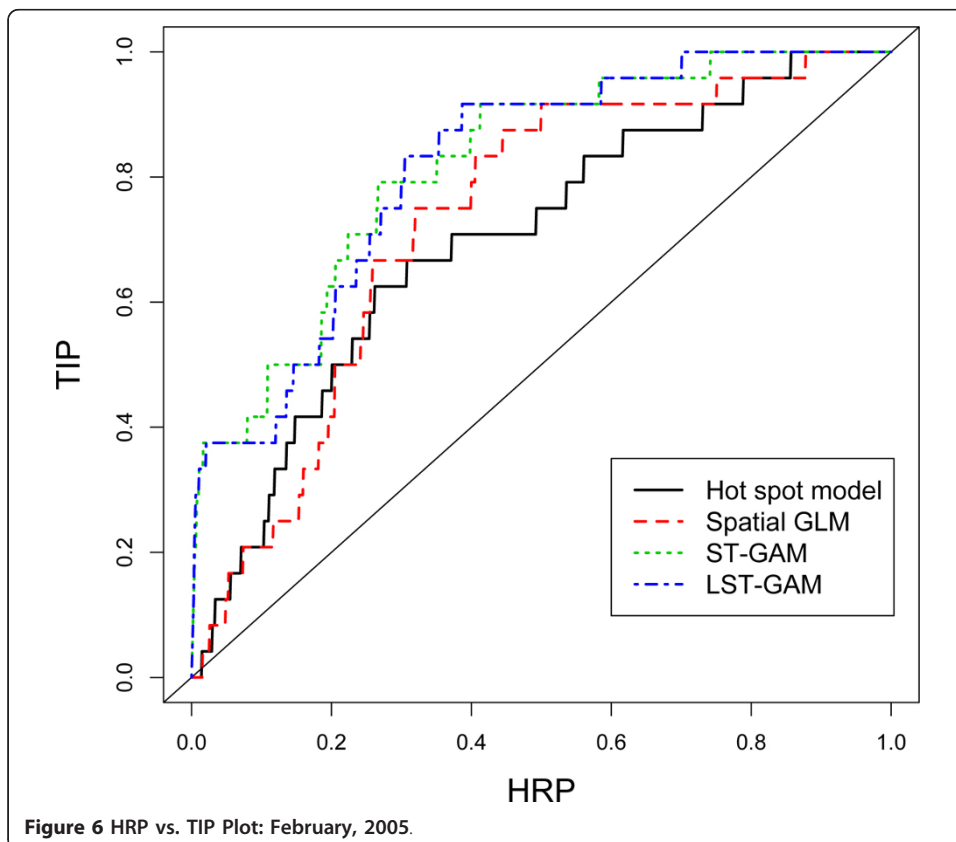
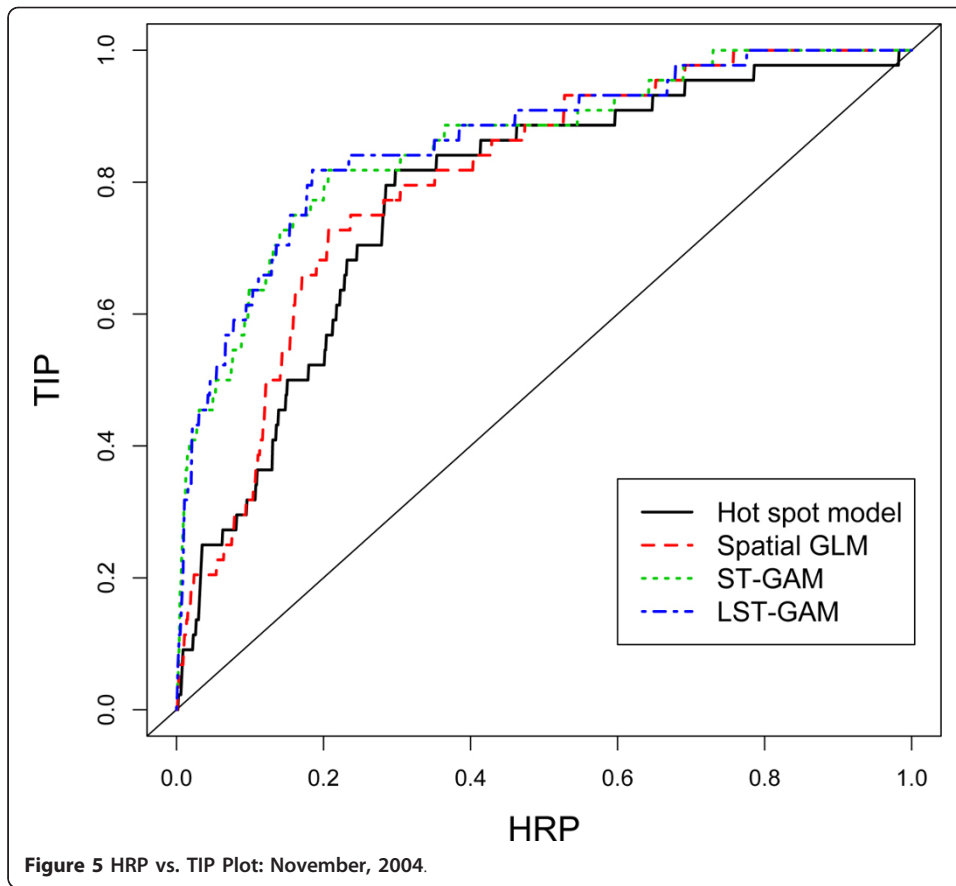
Figures 3, 4, 5, 6 show HRP vs. TIP plots for the predictions in March 2004, July 2004, November 2004 and February 2005. In the plots, HRP and TIP are the percentage of high risk area and the percentage of incidents happened within the high risk area respectively, as defined in section 2.5. From these plots, we can confirm that ST-GAM and LST-GAM had better prediction performance in those four months. Especially, ST-GAM and LST-GAM can capture about half of the real incidents happened with a very small high risk area in each case. For example, about 50% of real incidents happened within the top 2% area with the highest risk predicted from LST-GAM in July 2004. The police department can use this prediction to patrol more efficiently.

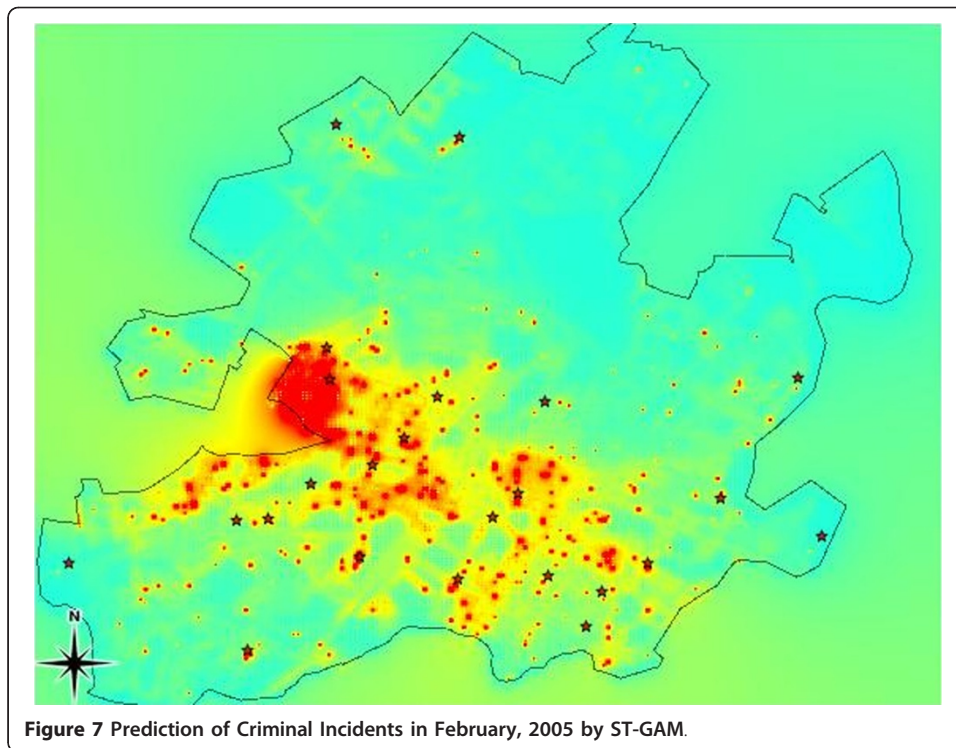
Predictions from our models were probabilities on spatial grids. This type of data can be visualized easily with available GIS softwares. Figure 7 shows the heat map of the prediction from ST-GAM in February 2005 generated by Quantum GIS [19]. We used a kernel density with 3 standard deviation to smooth the prediction. On this map, red color means high predicted probabilities while the light blue color means low predicted

**Table 3 Wilcoxon Significance Test Results**

P-value	Hot Spot Model	Spatial GLM	ST-GAM	LST-GAM
Hot Spot Model	-	0.0009766	0.0004883	0.0004883
Spatial GLM	0.0009766	-	0.0004883	0.0004883
ST-GAM	0.0004883	0.0004883	-	0.02686
LST-GAM	0.0004883	0.0004883	0.02686	-





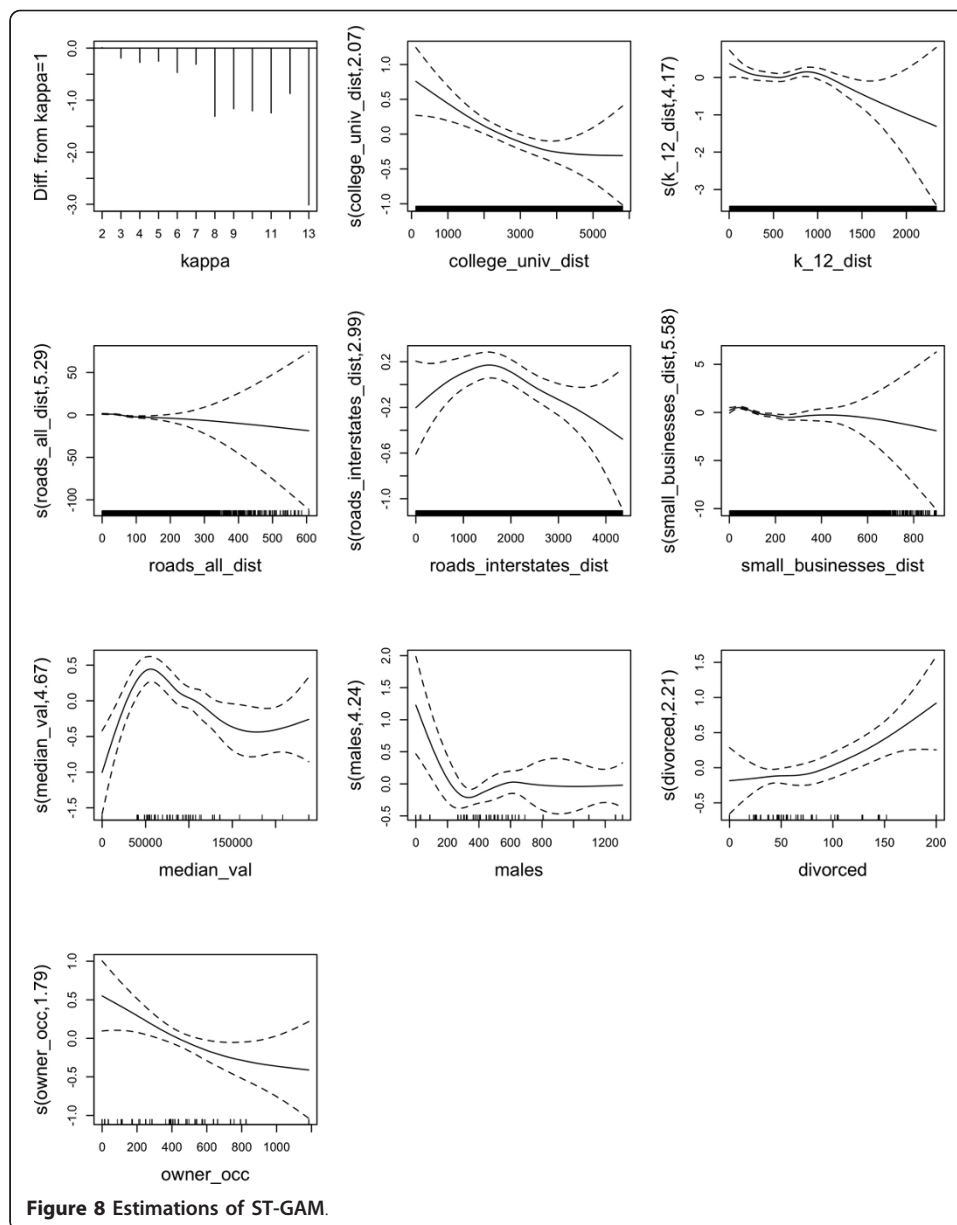


probabilities. The red stars are the real criminal incidents happened in February 2005. As we can see from this map, most of the real criminal incidents were located within the high probability area.

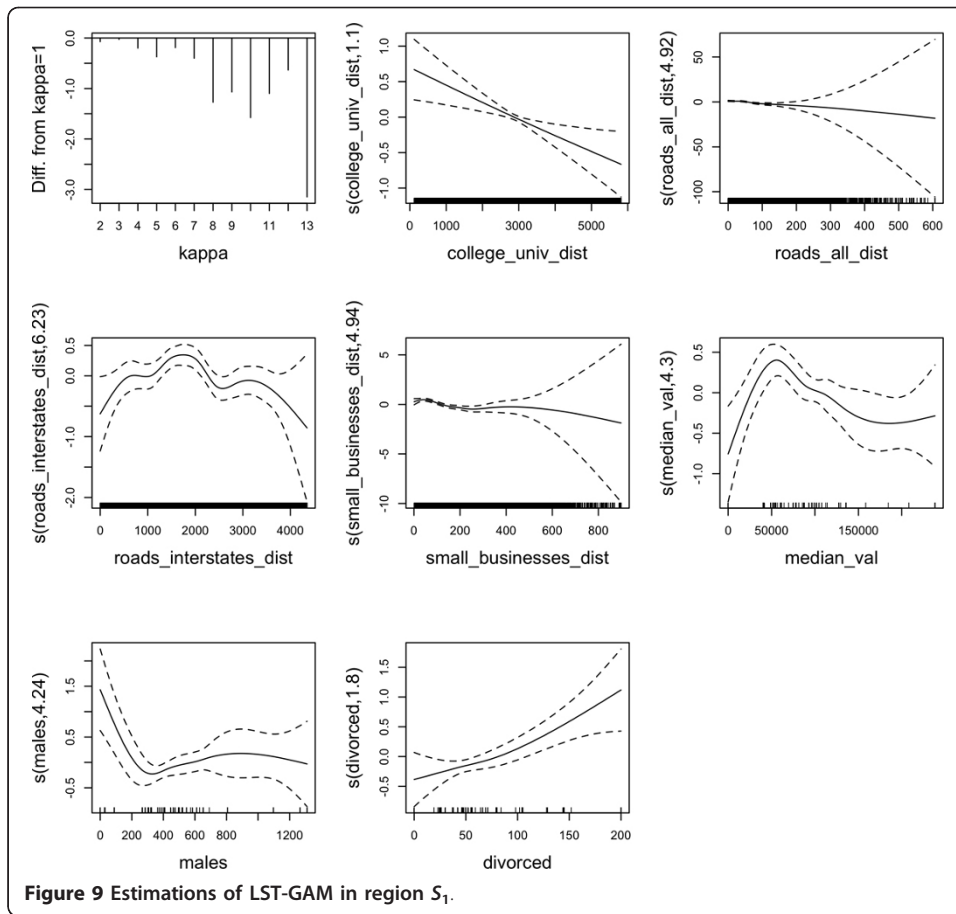
### 3.3.2 Model interpretation

Table 2 shows feature significance in different models. As we can see, the temporal dummy variable  $\kappa$  was significant in both ST-GAM and LST-GAM. It was helpful to explain the variance of criminal incident probability. All selected features were significant in at least one model, except the feature *widowed*. Features *roads all dist*, *small businesses*, and *divorced* were significant in all the models. Comparing features in LST-GAM in region  $S_1$  and  $S_2$ , we can see the different regions had the different sets of significant features. For example, *median\_val* was important to explain the variance in the low risk area  $S_1$ , but not in the high risk area  $S_2$ .

Figures 8, 9 and 10 show the estimated parameters and smooth functions of ST-GAM, LST-GAM in region  $S_1$  and LST-GAM in region  $S_2$  respectively. Only significant features were plotted. In the figures, solid lines represent the estimated smooth functions while the dotted lines are 95% confidence intervals. Clearly, we can see the nonlinear effects of features on the crime probability. Based on Figure 8, locations with no incident happened in the previous year were less likely to have a new incident. Out of the locations where incidents happened in the previous year, the locations with incidents just happened in the past half year were more likely to have a new incident. Incidents were more likely to happen at locations closer to schools, roads, and small businesses. As we expected, the neighborhoods with the least and the most expensive median house value were less likely to be broken and entered. The neighborhood with the median house value of about \$60,000 was the most likely to have such incidents. The number of males in the neighborhood also impacted crimes. It was more likely to

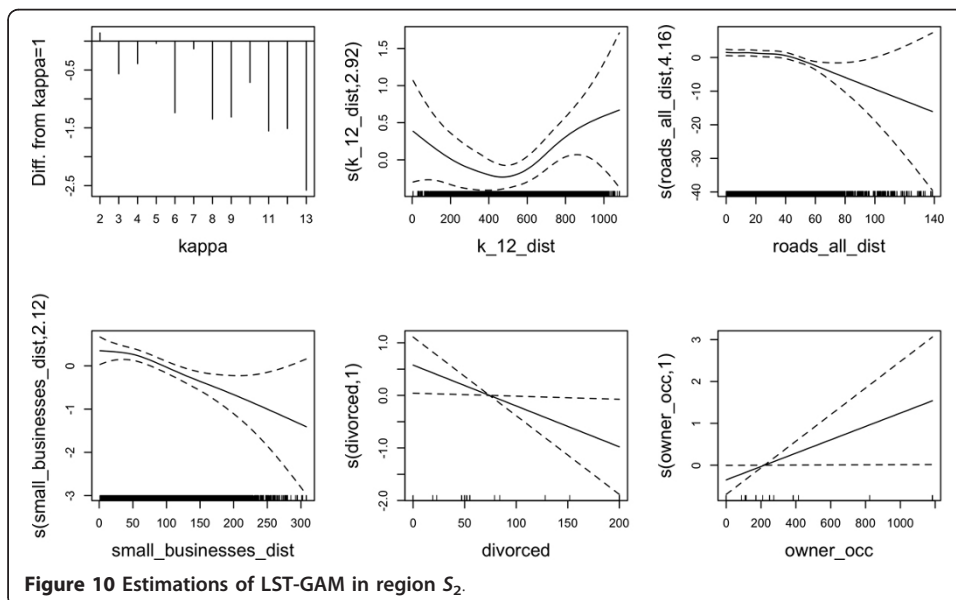


have incidents in the neighborhoods with less males, but this effect was not significant after the neighborhoods had more than 350 males. In addition, breaking and entering incidents were less likely to happen in the neighborhoods with less divorced and more owner occupied houses. Figures 9 and 10 show the different patterns in the different regions. In the low risk region  $S_1$ , features had similar effects on crimes as in the ST-GAM, but the number of significant features was less. In the high risk region  $S_2$ , locations with no incident happened in the previous year were still less likely to have a new incident. However, out of the locations where incidents happened in the previous two months, the locations with incidents just happened in the past month less likely to have a new incident in the following month. Different from the low risk region, incidents were more likely to happen in the neighborhoods with less divorced and more owner occupied houses.



#### 4 Conclusions

In this paper, we describe the spatio-temporal generalized additive model (ST-GAM) and show its extension into the local spatio-temporal generalized additive model (LST-GAM)



to predict criminal incidents. Both ST-GAM and LST-GAM can fully utilize many different types of data, such as spatial and temporal data, geographic data, demographic data, etc. ST-GAM can be easily estimated by available algorithms and has good interpretability. LST-GAM models the different criminal patterns based on different regions, and thus is more flexible than ST-GAM. We also showed that ST-GAM is a special case of LST-GAM. Methods to estimate LST-GAM are described in Section 2.4.

To evaluate this new approach we applied both models to predict breaking and entering incidents in Charlottesville, Virginia. Based on our assessments with the real criminal incident data, both models can predict future incidents accurately. Results showed that our two models outperformed the previous spatial GLM and the hot spot model. Compared with ST-GAM, LST-GAM had better performance in prediction. Law enforcement agencies can use ST-GAM and LST-GAM to model criminal incidents, predict future incidents and prevent crimes. In addition, those two models can be applied to other areas with the need to study the spatio-temporal patterns and predict future incidents. For example, we can use ST-GAM and LST-GAM to predict terrorist events and car accidents.

Several methods can be investigated to improve both models. To select features for modeling, more sophisticated methods like penalized regression methods [10] can be incorporated into ST-GAM to choose features automatically. For LST-GAM, better methods can be developed to generate optimal regions. For example, regions can be generated to minimize the loss function discussed in section 2.1. Our previous study [20] showed narratives provided important information about crimes. We can use text mining models to extract text features from narratives and add these features in ST-GAM and LST-GAM.

#### **Authors' contributions**

XW developed the models in this paper, tested them, and drafted the manuscript. DB developed the models and revised the manuscript. Both authors read and approved the final manuscript.

#### **Authors' information**

X. Wang is a PhD candidate in the Department of Systems and Information Engineering at the University of Virginia. His research interests include statistical learning with high dimensional data, spatio-temporal modeling and text mining. He has applied his research to solve real problems, such as criminal incident prediction and asymmetric threat tracker. He won the best paper honorable mention award at the 2011 IEEE International Conference on Intelligence and Security Informatics. He holds the MS degree in Systems and Information Engineering from the University of Virginia, and the BS degree in Management Information Systems from the Central University of Finance and Economics, China.

D. Brown is William Stansfield Calcott Professor of Engineering and Applied Science and Director of the Applied Predictive Technology Laboratory at the University of Virginia. Prior to joining the University of Virginia, Dr. Brown served as an officer in the U.S. Army and later worked at Vector Research, Inc. on projects in medical information processing and multi-sensor surveillance systems. He is President of Commonwealth Computer Research, Inc. This company provides products and services in prediction analytics for numerous private and governmental organizations. He served on the National Research Council (NRC) Committee on Transportation Security also on the National Academy of Sciences panel on High Performance Computing and Crisis Management, as well as, the NRC Committee on Surface Transportation Infrastructure Security. Dr. Brown is a Fellow of the National Institute for Aerospace. He is a past member of the Joint Directors of Laboratories Group on Data Fusion and a former Fellow at the National Institute of Justice Crime Mapping Research Center. Dr. Brown is the recipient of the IEEE Joseph Wohl Career Achievement Award for his work in systems engineering and data fusion. He is also the recipient of the IEEE Norbert Wiener Award for Outstanding Research in the areas of systems engineering, data fusion, and information analysis. He received the IEEE Intelligence and Security Informatics Award for outstanding research achievements in information for security, law enforcement, and intelligence. The Governor of Virginia presented him with the Governor's Technology Award for his achievements in providing the technology to enable rapid crime analysis by local law enforcement agencies. He has also received an IEEE Outstanding Contribution Award and the IEEE Millennium Medal. Dr. Brown is a Fellow of the IEEE, the past Editor-in-Chief of the *IEEE Transaction on Systems, Man, and Cybernetics, Part A: Systems and Humans*, and a past President of the IEEE Systems, Man, and Cybernetics Society.



### Competing interests

The authors declare that they have no competing interests.

Received: 2 September 2011 Accepted: 27 February 2012 Published: 27 February 2012

### References

1. Block C: **STAC hot-spot areas: A statistical tool for law enforcement decisions.** *Crime analysis through computer mapping.* Washington, DC: *Police Executive Research Forum, Citeseer* 1995, 15-32.
2. Liu H, Brown D: **Criminal incident prediction using a point-pattern-based density model.** *International journal of forecasting* 2003, **19**(4):603-622.
3. Xue Y, Brown D: **Spatial analysis with preference specification of latent decision makers for criminal event prediction.** *Decision support systems* 2006, **41**(3):560-573.
4. Smith MA, Brown DE: **Application of Discrete Choice Analysis to Attack Point Patterns.** *Information Systems and e-Business Management* 2007, **5**(3):255-274.
5. Brown D, Dalton J, Hoyle H: **Spatial Forecast Methods for Terrorist Events in Urban Environments.** *Intelligence and security informatics: Second Symposium on Intelligence and Security Informatics, ISI 2004, June 10-11, 2004; Tucson, AZ, USA* Springer-Verlag New York Inc; 2004, 426.
6. Rodrigues A, Diggle P, Assuncao R: **Semiparametric approach to point source modelling in epidemiology and criminology.** *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2010, **59**(3):533-542.
7. Ivaha C, Al-Madfai H, Higgs G, Ware A: **The Dynamic Spatial Disaggregation Approach: A Spatio-Temporal Modelling of Crime.** *Proceedings of the World Congress on Engineering, Volume 2, Citeseer* 2007.
8. Paciorek C, Yanosky J, Puett R, Laden F, Suh H: **Practical large-scale spatio-temporal modeling of particulate matter concentrations.** *Ann Appl Stat* 2009, **3**:369-396.
9. Wang X, Brown D: **The spatio-temporal generalized additive model for criminal incidents.** *Proceedings of the IEEE International Conference on Intelligence and Security Informatics: 9-12 July 2011; Beijing, China, IEEE* 2011.
10. Hastie T, Tibshirani R, Friedman J: *The elements of statistical learning: data mining, inference, and prediction* Springer Verlag; 2009.
11. Beck N, Katz J, Tucker R: **Taking time seriously: Time-series-cross-section analysis with a binary dependent variable.** *American Journal of Political Science* 1998, **42**(4):1260-1288.
12. Hastie T, Tibshirani R: *Generalized additive models* Chapman & Hall/CRC; 1990.
13. Wood S: *Generalized additive models: an introduction with R, Volume 66* CRC Press; 2006.
14. Wood S: **Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models.** *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011.
15. King G, Zeng L: **Logistic regression in rare events data.** *Political analysis* 2001, **9**(2):137.
16. Fawcett T: **An introduction to ROC analysis.** *Pattern recognition letters* 2006, **27**(8):861-874.
17. **PostGIS.** [http://postgis.refrains.net/].
18. R Development Core Team: *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria; 2011 [http://www.R-project.org/], [ISBN 3-900051-07-0].
19. Quantum GIS Development Team: *Quantum GIS Geographic Information System* Open Source Geospatial Foundation; 2009 [http://qgis.osgeo.org].
20. Wang X, Brown D, Conklin J: **Crime Incident Association with Consideration of Narrative Information.** *Processings of the IEEE Systems and Information Engineering Design Symposium: April 2007; Charlottesville, VA, IEEE* 2007, 1-4.

doi:10.1186/2190-8532-1-2

**Cite this article as:** Wang and Brown: The spatio-temporal modeling for criminal incidents. *Security Informatics* 2012 1:2.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)