Investigating a Bayesian Hierarchical Framework for Feature-Space Modeling of Criminal Site-Selection Problems

Jon Fox, Samuel H. Huddleston, Matthew Gerber, Donald E. Brown
Department of Systems and Information Engineering, University of Virginia
151 Engineer's Way
Charlottesville, Virginia 22904–4747
jmf3a@virginia.edu

Abstract

A significant amount of academic research in criminology focuses on spatial and temporal event analysis. Although several efforts have integrated spatial and temporal analyses, most previous work focuses on the space-time interaction and space-time clustering of criminal events. This research expands previous work in geostatistics and disease clustering by using a Bayesian hierarchical framework to model criminals' spatial-temporal preferences for site-selection across a continuous time horizon. The development of this Bayesian hierarchical feature-space model (BHFSM) offers law enforcement personnel a method for accurate crime event forecasting while improving insight into criminal site-selection at the strategic level. We compare the BHFSM to other feature-space modeling techniques using both a long range and short range criminal event dataset collected from police reporting. While the BHFSM remains sufficiently accurate for event prediction, current computational requirements limit the applicability for "just-in-time" crime modeling.

Introduction

Although much theoretical and practical work has been done on the use of Bayesian hierarchical modeling for geostatistics and disease clustering, applications within the criminal site-selection problem have been limited. This article merges the feature-space model of Liu and Brown (1998) with the Markov random field construct of Zhu, Huang, and Wu (2006) to model the criminal's preference for initiating a crime within a specific spatial-temporal zone. By adapting theoretical and computational work from disease mapping and environmental studies, we develop a Bayesian hierarchical feature-space model (BHFSM) for the criminal event prediction problem in order to examine both parameter estimation and predictive inference. The remainder of this section provides a quick review of applicable crime theory and feature-space modeling for criminal site-selection problems. The subsequent sections discuss the Bayesian hierarchical framework, introduce the dataset used for this article, and review the performance of the BHFSM against the dataset for both a long term and a short term temporal study horizon. In the final section, we review conclusions from this initial research and propose paths for future research.

Crime Theory

Much of the work in crime studies proceeds from a frame of reference built upon the location of the crime (Townsley, Homel, and Chaseling 2000; Groff and LaVigne 2002). This frame of reference is conditioned upon Tobler's first law of geography: "everything is related to everything else, but near things are more related than distant things" (Tobler 1970). For the crime analyst, this means that if a crime happened yesterday at the corner of Main Street and Broadway, then the most likely location for a crime tomorrow is the corner of Main and Broadway. Hotspotting and crime clustering are built upon the assumption that future crimes are likely to occur at the same location as past crimes (Ratcliffe 2004; Cohen, Gorr, and Olligschlaeger 2007).

Rational criminal theory assumes that individuals have specific reasons for committing a crime at a certain time and a certain location (Clark 1980). By examining the historical criminal activity data within a spatial region, we can discover patterns that might indicate criminals' preferences for executing crimes at certain locations (Brantingham and Brantingham 1984). Spatial choice models offer analysts a methodology for identifying a criminal's preference for one site over another within a spatial region.

Spatial choice models assume an actor will select a site (e.g., for migration, retail establishment, or criminal event) based on the perceived utility, or worth, of that site from a set of alternatives (Ewing 1976; McFadden 1986). The use of spatial choice models nests well within the rational criminal theory since it assumes that spatial point processes involving actors are a result of the actors' mental processes and perceptions (Burnett 1976). This article expands on the spatial choice problem to examine the impact of both geographic and temporal features on the criminal's site-selection process.

Feature-Space and Criminal Site-Selection

This work is inspired by the following question: What if, instead of focusing on where the crime happened on the ground, we focus on where the crime initiation took place in the mind of the criminal? The idea of spatial choice presents a framework of decision processes for a rational actor to choose a location based on the perceived value of that location. Consider a criminal who wants to steal a car. Will he choose a parking garage at the center of town with restrictive traffic flows or will he choose the mall parking lot near a major freeway on the outskirts of town? Previous work has shown that the car thief will take the car from the mall since features surrounding a location are as critical as the location itself (Rengert 1997). Brown, Liu, and Xue (2001) showed that data mining previous criminal events provides insight to what spatial features might be considered by a criminal in selecting a location to commit a crime. We define this set of spatial considerations to be the feature-space. Several investigations have shown that feature-space modeling performs as well, or better, than density based methods (Brown, Dalton, and Holye 2004; Smith and Brown 2004).

Criminal site-selection is the process by which a criminal selects the time and space to execute an event based on their feature-space preferences (Porter 2006). Rather than using a latitude and longitude to describe each location in a study region, we use spatial distances to environmental features — such as schools, streets, or stadiums — and spatial representations of social demographics — such as population, percent rental properties, and household income — to examine which locations are preferred by criminals for certain types of crimes (Bannatyne and Edwards 2003; Liu and Brown 2003; Huddleston and Brown 2009).

Bayesian Hierarchical Modeling

Hierarchical models allow us to deconstruct complex problems into a series of smaller tractable problems. Using the methodology developed by Wickle (2003), we formulate the criminal site-selection problem into three basic stages: a data model, a process model, and a parameter model. Our data model accounts for our knowledge of the spatial-temporal patterns of crime within the study region. The process model provides insight to the criminal site-selection process while accounting for spatial and temporal effects. Finally, our parameter model accounts for the uncertainty in both the data and process models (Wickle 2003).

Formulation

The goal of this article is to develop a Bayesian hierarchical model that uses the feature-space methodology to accurately predict crime events across an irregular lattice while providing insight into the criminal site-selection process. To estimate the criminal's spatial preferences, our ***data model*** represents the criminal's site-selection process as a binary random variable where $Y_{s,t} \in 0, 1$ is the observation of the presences, or absence, of crime at location s at time t given a set of features $\boldsymbol{X}$.

$$Y_{s,t} | \boldsymbol{X} \sim \ Bern(\mu_{s,t}) \tag{1}$$

For our least complex model, we assume that the probability $\mu_{s,t}$ is a function of the criminal's preferences for certain features and a random effects term. Mathematically, we represent the ***process model*** as:

$$\begin{gathered}\mu_{s,t} = \text{logit}^{-1}\left(\beta_0 + \beta_1 X_{s1} + \ldots + \beta_k X_{sk} + \theta_{s,t}\right), \\ \text{for } s = 1, ..., S \text{ and} \\ \text{for } t = 1, ..., T \ .\end{gathered} \tag{2}$$

Equation 2 uses a set of features $\boldsymbol{X}$ as a vector of length k for each location s combined with the estimated $\boldsymbol{\beta}$ values from the parameter model to estimate the probability $\mu_{s,t}$. For this article, we use a set of demographic variables to represent a portion of the feature-space considered by the criminal in their site-selection process. Analyzing previous criminal event data gives us a method to account for the criminal's site-selection process. By modeling the relationship between the features and the probability of crime, we estimate the preferences criminals have for locations with a specific set of features. However, just as the criminal's preferences for certain locations might change depending on proximity to freeways or vacant houses, the criminal site-selection process can also change depending on the time of day or other seasonal events (Rossmo, Laverty, and Moore 2005; Gorr 2009a). The variable $\theta_{s,t}$ provides a method for including other random effects.

The first random effect considered is the temporal component. We consider a temporal effect $g_t \sim N(g_{t-1}, \tau_g)$. Based on previous research (Gorr, Olligschlaeger, and Thompson 2003; Eck et al. 2005), we believe that criminal activity often preceeds criminal activity. Using this temporal component allows us to account for periods of criminal activity that match the routine activities and population dynamics of the study region. We will discuss the inital conditions for the variance estimates in the parameter model.

We use a Markov random field (MRF) construct as the second random effect by assuming that the likelihood of a crime at a specific location is dependent only on its neighbors and its previous temporal state (Zhu, Huang, and Wu 2006). Recent work on point processes uses MRFs as a secondary structure that results from an aggregation process of event counts. For our crime data, we construct the MRF along an irregular lattice structure defined by political and cultural boundaries using the construct provided by Illian et al. (2008). We consider a MRF effect that accounts for the past value at the location s and the second-order neighbors such that $\omega_s \sim N(\omega_{j-1}, \tau_o)$. The index j accounts for the second-order neighbors of location s. The inclusion of the neighborhood spatial effects gives us a method to include criminal repeat information into the feature-space model. Studies on criminal repeats have shown that for short temporal intervals, locations that have experienced crime have an increased likelihood for repeat victimization (Townsley, Homel, and Chaseling 2000).

The third random effect considered for this article is an interaction term. We consider an interaction term $\psi_{s,t} \sim N(0, \tau_p)$. The interaction term is uncorrelated but can identify potential spatial-temporal interactions within the data that are not accounted for in the base feature-space model (Lawson 2009). The final random effect is an uncorrelated error term $v_s \sim N(0, tau_v)$ that accounts for any uncorrelated spatial components of the criminal site-selection process. The research design section outlines the four primary models considered for this article using different combinations of these random effects.

Finally, we specify the ***parameter models*** by establishing the initial distributions for the parameters. As seen in Figure 1, the $\boldsymbol{\beta}$ vector appears in the process model. However, we provide initial estimates for the individual βs within the parameter model. Estimating the $\boldsymbol{\beta}$ values increases the complexity of the parameter model, since for both the long term and short term data study, we initially estimate each β for each feature during the model fitting phase. In order to reduce the computational requirements, we substitute a feature-space prior calculated from linear model regression (Lunn et al. 2000). The initial assumptions for the parameter model follow:

$$\begin{gathered}\beta \sim N(\hat{\beta}, \tau_b) \\ \tau_b \sim N(0, svb), svb \sim U(0, 10) \\ \tau_u \sim N(0, svu), svu \sim U(0, 10) \\ \tau_g \sim N(0, svg), svg \sim U(0, 10) \\ \tau_o \sim N(0, svo), svo \sim U(0, 10) \\ \tau_p \sim N(0, svp), svp \sim U(0, 10)\end{gathered} \tag{3}$$

The parameter model sets the initial conditions for the simulation methods used to estimate the process and data model and completes the model hierarchy (Wickle 2003). More details on the simulation methods can be found in (Lawson 2009; Kery 2010).

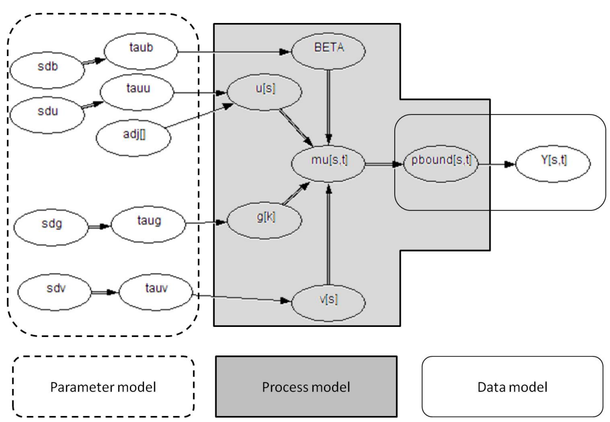

Figure 1: Directed acylic graph for Bayesian hierarchical feature-space model.

Bayesian methods provide a means to calculate the posterior distribution from our three stage-hierarchical model. Using the example from Wickle (2003), our posterior distribution is proportional to our data model conditioned upon the process and parameter models times the process model conditioned upon the parameters:

$$\begin{gathered}[process, parameters|data] \propto \\ [data|process, parameters] \times \\ [process|parametersl][parameters]\end{gathered} \tag{4}$$

Since our goals in modeling criminal site-selection problems include both predictive inference and parameter understanding, we desire to solve for the left hand side of Equation 4. However, the complexity of the posterior distribution makes obtaining a closed form solution almost, if not completely, unobtainable. Using simulation methods, built upon empirical knowledge from the data and expert knowledge on the prior distributions, we obtain samples that provide estimates of our target variables (Lawson 2009).

Research design

The Bayesian hierarchical feature-space model (BHFSM) is a limited feature-space logistic regression model with an auto-regression on the state of the neighboring locations across an irregular lattice at discrete temporal intervals. Following work from disease mapping and geostatistics, we examine four models of random effects for our variable $\theta_{s,t}$. The models considered provide several methods for including other random effects (Lawson 2009). The four models considered for random effects include:

- A time-varying trend g_t plus an uncorrelated error v_s
- A Markov random field ω_s accounting for the sum of the neighboring effects at a previous time plus v_s
- g_t plus ω_s plus v_s
- g_t plus ω_s plus v_s and an interaction term $\psi_{s,t}$

Figure 1 displays a graphical representation of the third model developed for this article without an interaction term.

Study dataset

The primary source of data for this article is an incident database for the city of Charlottesville, Virginia. We sample the complete dataset to develop a subset that contains a time horizon spanning four years with over 2,000 incidents. We restrict the crime types analyzed for this article to assaults, both simple and aggravated. We drape an irregular lattice over the study area and aggregate the criminal incidents at the daily level. Although the aggregation introduces some level of discreteness, we treat the temporal intervals as continuous points along the temporal horizon. The irregular lattice structure is based on the thirty-seven US Census block-groups for the city. Using the this lattice structure facilities inclusion of demographic information at the block-group level. We use the census information as proxies for complex factors that actually affect criminals. We are not claiming that a criminal actually considers the percent of houses in area that are rentals when deciding to execute a crime. However, the percentage of rental houses in an area might correlate with other factors that are part of the criminal site-selection process. Figure 2 depicts the study region draped with the irregular lattice and shows spatial-temporal patterns of assaults over four distinct temporal intervals. The analysis that follows uses a second-order neighbor structure over the irregular lattice depicted in Figure 2.

We set $Y_{i,t} = 1$ if a criminal assault occurs within the specified block-group $i = 1, ..., 37$ during one of the days $t = 1, ..., 1095$ of the study horizon. The block-group and daily aggregation results in a 37×365 matrix for a total of 40,515 observations in space-time. Figure 3 depicts a one year snapshot of criminal events across the entire spatial region.

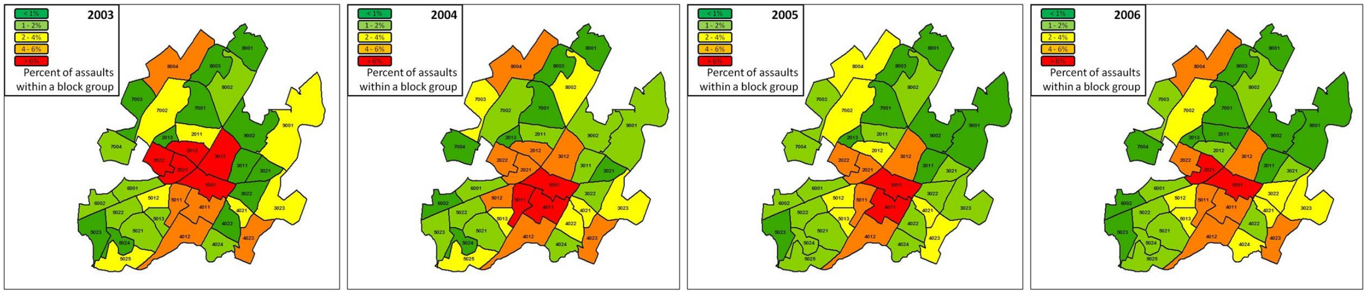

Figure 2: Evolution of spatial patterns over continuous temporal horizon. Since we identify changes in the map over time, we hypothesize that we have spatial and temporal effects within the criminal site-selection process.

Model comparison

For this article, we compare each model's predictive performance against a test set from the dataset. For the long term study, we use a 365 day temporal window for model fitting and then evaluate against a ninety day test. For the short term study, we use a thirty day temporal window surrounding special events in Charlottesville for model fitting and then evaluate against the thirty day temporal window surrounding the same special event in the following year.

Prior to comparing predictive performance, we use a goodness of fit measure to evaluate each model. Borrowing from conventional generalized linear modeling, we use deviance as a measure of how well the model fits the data. In the software used for this article, we can expect the deviance to decrease by 1 for each predictor added to the model (Gelman and Hill 2007).

As an additional method for comparing goodness of fit, we use the mean squared predictive error (MSPE). Given our known spatial-temporal dataset from the test period, $\boldsymbol{Y}$, our estimated spatial-temporal dataset, $\hat{\boldsymbol{Y}}$, and a number of observations m from a simulation sample of G, we use Lawson's (2009) formulation such that:

$$MSPE = \frac{|\boldsymbol{Y} - \hat{\boldsymbol{Y}}|^2}{(G \times m)} \tag{5}$$

One of the challenges for spatial-temporal data is selecting an appropriate statistical measure for examining model performance. Originally used to assess radar performance in World War II, the receiver operating characteristic (ROC) curve are particularly useful for evaluating the ability of a model to predict the occurrence of an event accurately while minimizing the number of false positive predictions (Bradley 1997; Swets, Dawes, and Monahan 2000). Similiar to the ROC curve, the surveillance plot provides a method for evaluating model performance in spatial-temporal classification problems. The surveillance plot gives the analyst a method for monitoring the amount of area within the study region that needed to be observed in order to identify the highest percentage of crimes (Huddleston and Brown 2009; Kewley and Evangelista 2007). Using a contingency table, or decision matrix, similar to Table 1, we record the possible outcomes of prediction estimated with the model being considered against the true conditions observed in the test set.

Table 1: **Contingency Table**

	True Condition		
Test Result	Positive	Negative	Measures
Positive	TP	FP	TP + FP
Negative	FN	TN	FN + TN
Measures	TP + FN	FP + TN	

We build the surveillance plot by plotting the rate of accurate crime predictions against the rate of crime incidents predicted where crimes did not occur. Although the surveillance plot provides a measure for comparing model performance visually, translating the surveillance plot into a numerical measure provides a method for comparing the performance of multiple models against a common test set. A model with high accuracy — predicting all the crime locations perfectly — would have a ratio of all true positives versus zero false positives while a model with an equal ratio of true positives and false positives is basically guessing (Bradley 1997; Swets, Dawes, and Monahan 2000).

$$PLR = \frac{n \times TP}{(TP + FN) \times (TP + FP)} \tag{6}$$

The performance limit ratio (PLR) measures the model's trade-off in accuracy and precision by focusing on the model's better-than-chance ratio (Gorr 2009b) of correctly predicting crimes within a test set of size n. A model that is more accurate in predicting crimes across the space-time surface will have a higher PLR. Rather than focusing on the entire area under the curve, we reduce the focus to the first 20% of the space-time surface observed while discounting the area under the curve that accounts for random guessing.

Long Term Study Results

For the long term study, the block-group and daily aggregation results in a 37×365 matrix for a total of 13,505 observations in space-time. We use a second-order neighbor model

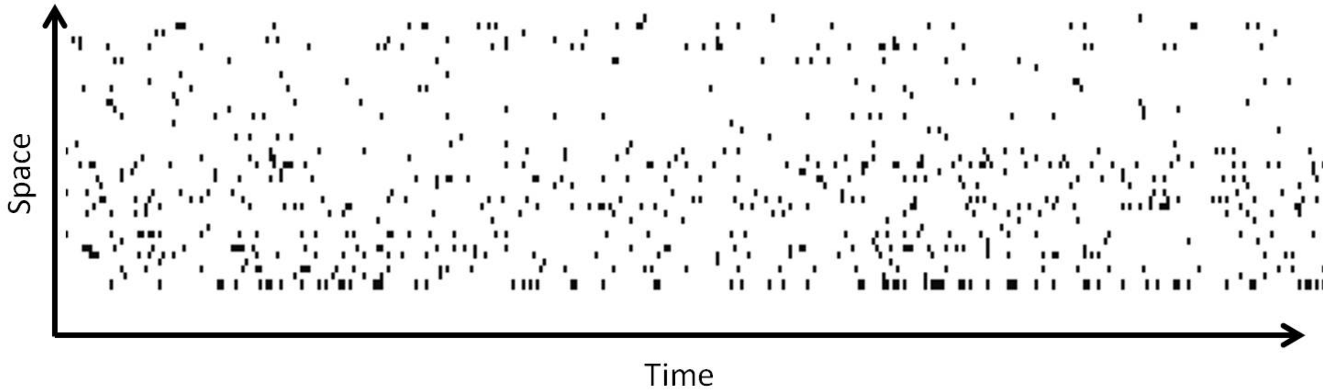

Figure 3: Space-time lattice for study domain. The arrangement of spatial regions along the y axis might falsely identify spatial clusters, however, the temporal horizon along the x axis does allow for visual identification of temporal clusters for assaults within the study region. When using the surveillance plot for measuring model performance, we iteratively evaluate all spatial locations within each temporal interval.

to account for all the criminal activity in all the surrounding census blocks. Table 2 outlines the specific models examined using both the demographic features and a feature-space prior obtained from a generalized linear regression similar to the work of (Liu and Brown 2003). As discussed in the research design section, we consider four alternatives to model the random effects using our variable $\theta_{s,t}$ in the process model: 1) a time-varying trend; 2) a Markov random field accounting for the sum of the neighboring effects at a previous time; 3) a time-varying trend with a Markov random field; 4) a time-varying trend with a Markov random field and an interaction term. For every alternative we include a space-time independent noise term. For the first four alternatives, we attempt to account for the criminal site-selection preference by modeling $\boldsymbol{\beta}$ as seen in Figure 1. After model fitting, we evaluate performance using the MSPE discussed above.

Although the predictive performance of the BHFSM is not significantly better than the base feature-space model, we were expecting to see significant lift in the parameter estimation related to identifying criminal preferences for certain spatial features. In fact, even with all four models converging, the only feature-space variable with significantly better estimation was the preference for areas with high percent vacancy. However, (Lawson 2009) shows that the combination of spatially-referenced explanatory variables within a Markov random field construct often yields poor estimates of the regression coefficients and produces computational challenges related to multi-collinearity. Both of our approaches to reduce the impact of correlation created additional challenges. First, removing the features that are spatially dependent limits our insight into the criminal site-selection process for identifying feature-space preferences. Second, introducing new variables that have a stationary spatial attribute but are non-stationary temporally limits our ability to identify how the criminal's feature-space preferences evolve over time. Overall, the Bayesian approach offers promise for reducing uncertainty in the predictive surfaces. However, as discussed in (Withers 2002; Zhu et al. 2008), the computational time required for sampling from the posterior distribution for Bayesian inference for criminal site-selection problems is a major drawback. We discuss an alternative approach in the conclusion that offers computational advantages while remaining sufficiently accurate for prediction. In the next section, we scale down the horizon of the study period as an additional step in examining the BHFSM.

Short Term Study Results

Although applying the Bayesian framework to the long term study data did not result in significant gains in predictive performance, the initial disappointment was not entirely unexpected. Previous research shows that spatial-temporal analysis focused on criminal site-selection requires focused efforts on periods of temporal transition and local knowledge of the environment (Kerchner 2000; Bernasco and Block 2009). A more appropriate methodology for including temporal information into the BHFSM reduces the scope of the temporal horizon to those intervals with the greatest variance in crime rates. Research has also shown that spatial regions experience great variance in crime rates for certain locations depending on the temporal proximity to special events (Cohen, Gorr, and Olligschlaeger 2007). Reducing the temporal horizon to a smaller scale — such as a thirty day window before and after large spikes in crime rates — makes it easier to examine the impact of these special events on the criminal site-selection process. More importantly, including additional data from local law enforcement personnel takes advantage of their local knowledge of the temporal environment (Cressie and Wikle 2011).

As with the long term study, we consider all four alternatives to model the random effects using our variable $\theta_{s,t}$ in the process model. Table 3 outlines the specific models ex-

Table 2: Bayesian Hierarchical Feature-Space Model Development for Long Term Data Study

Model	Predictors	Time	Deviance	MSPE
Spatial Choice and Trend	35	3049	4632	0.0430
Spatial Choice and MRF	35	2587	4622	0.0429
Spatial Choice and MRF and Trend	36	4694	4625	0.0429
Spatial Choice and MRF and Trend and Interaction	43	19481	4609	0.0428
Feature-Space Prior and Trend	31	2273	4785	0.0435
Feature-Space Prior and MRF	30	2314	4632	0.0430
Feature-Space Prior and MRF and Trend	38	7219	4614	0.0429
Feature-Space Prior and MRF and Trend and Interaction	40	9515	4619	0.0429

Table 3: Bayesian Hierarchical Feature-Space Model Development for Short Term Data Study

Model	Predictors	Time	Deviance	MSPE	PLR
Feature-Space Model	7	5	423	0.0479	0.46
Spatial Choice and Trend	19	182	423	0.0479	0.52
Spatial Choice and MRF	19	303	423	0.0479	0.53
Spatial Choice and MRF and Trend	22	332	422	0.0478	0.53
Spatial Choice and MRF and Trend and Interaction	23	900	421	0.0477	0.53
Feature-Space Prior and Trend	8	180	420	0.0477	0.49
Feature-Space Prior and MRF	8	138	420	0.0477	0.47
Feature-Space Prior and MRF and Trend	10	371	419	0.0476	0.48
Feature-Space Prior and MRF and Trend and Interaction	12	1022	417	0.0475	0.50

amined using both the demographic features and a feature-space prior obtained from a generalized linear regression similar to the work of (Liu and Brown 2003) and as seen in our visual graph from Figure 1. Again, the only feature-space variable with significantly better estimation was the preference for areas with high percent vacancy. After model fitting, we evaluate performance using the PLR discussed above. While each BHFSM performs better than the base feature-space model, the computational time required for sampling from the posterior distribution for Bayesian inference is still several orders of magnitude greater than the time required for using generalized linear regression on the base feature-space model.

Conclusions

For city-wide, or regional-level, crime monitoring, the BHFSM offers a methodology for modeling criminal activity across continuous time. For this article, we added a Bayesian framework to the base feature-space model to include variables that account for both spatial *and* temporal patterns within the criminal site-selection process. We applied this methodology to both a long term and short term data study for criminal events in a small US city. Using data aggregated at the census block-group level for a medium temporal resolution, the BHFSM allowed us to model an actor's spatial-temporal preferences within a limited temporal period. Incorporating elements of the feature-space methodology into the Bayesian construct allowed us to blend the benefits gained from understanding multiple covariates within the actor's spatial-temporal decision process with the basic elements of geographic recency and spatial dependence found in hotspot modeling. Although the overall predictive performance is not significantly improved, by reducing the variance on estimates for a criminal's feature-space preferences, we gain understanding into the temporal variations of the criminal site-selection process. Enhanced understanding of the criminal site-selection process allows law enforcement personnel to adjust resource allocation strategies to better mitigate short term changes in the criminal site-selection process.

Several challenges remain for further consideration of the Bayesian framework for feature-space modeling of the criminal's site-selection process. The methodology examined in this article is computationally intensive. Although the BHFSM did provide improvement in predictive performance over the base feature-space model for the short term data study, the increased computational requirements hinder the application of the BHFSM for "just-in-time" crime modeling. Extending the Bayesian framework for modeling data at either a finer temporal or spatial resolution would increase the computational complexity since the size of the spatial-temporal event matrix is a multiple of the temporal intervals and the spatial dimensions. Future work will attempt to reduce this computational complexity by adding temporal and neighborhood indicator functions to the base feature-space model (Diggle, Tawn, and Moyeed 1998). Using indicator functions allows for faster sampling from the data while still accounting for temporal preferences in the criminal's site-selection process.

Structural vector autoregressive models (SVARs) show promise for forecasting employment rates given spatially based economic indicators (Rickman, Miller, and McKenzie 2009). Using an SVAR construct for modeling criminal site-selection might improve predictive ability if temporal changes in other features affect a criminal's temporal considerations for certain sites. However, the computational requirements for SVARs, like the Bayesian construct, are still rather demanding (Petris, Petrone, and Campagnoli 2009).

The social sciences offer another approach for reducing the computational demands of criminal site-selection modeling. Spatial-temporal designs for environmental research often include panel methods for monitoring and detecting temporal patterns and spatial relationships (Dobbie, Henderson, and Stevens 2008). We are not designing a method for collecting criminal event data, but rather examining historical collections of crime data. And as mentioned above, studies at fine temporal and spatial resolutions require a large spatial-temporal event matrix. Using a variation of stratified sampling (Gilbert 1987; Dobbie, Henderson, and Stevens 2008) on the spatial-temporal event matrix might reduce the computational time while retaining comparable predictive performance.

References

Bannatyne, J. C., and Edwards, H. P. 2003. A Bayesian explorations of the relationship between crime and unemployment in New Zealand for the time period: 1986-2002. In *International Workshop on Bayesian Data Analysis*.

Bernasco, W., and Block, R. 2009. Where offenders choose to attack: A discrete choice model of robberies in chicago. *Criminology* 93–130.

Bradley, A. P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30:1145–1159. Retrieved July 2010.

Brantingham, P. J., and Brantingham, P. L. 1984. *Patterns in Crime*. New York: Macmillan Publishing Company.

Brown, D. E.; Dalton, J.; and Holye, H. 2004. Spatial forecast methods for terrorist events in urban environments. In *Proceedings of the Second NSF/NIJ Symposium on Intelligence and Security Informatics*. Heidelberg: Springer-Verlag.

Brown, D. E.; Liu, H.; and Xue, Y. 2001. Mining preferences from spatial-temporal data. In *Proceedings of the SIAM Conference*. Chicago: Society for Industrial and Applied Mathematics.

Burnett, P. 1976. Behavioral geography and the philosophy of mind. In Golledge, R., and Rushton, G., eds., *Spatial Choice and Spatial Behavior*. Columbus: Ohio State University Press. pp. 23–50.

Clark, R. V. 1980. Situational crime prevention: Theory and practice. *British Journal of Criminology* 136–147.

Cohen, J.; Gorr, W.; and Olligschlaeger, A. 2007. Leading indicators and spatial interactions: a crime forecasting model for proactive police deployment. *Geographical Analysis* 39:105–127.

Cressie, N. A., and Wikle, C. K. 2011. *Statistics for Spatial-Temporal Data*. Hoboken, New Jersey: John Wiley and Sons.

Diggle, P. J.; Tawn, A. J.; and Moyeed, R. A. 1998. Model-based geostatistics. *Applied Statistics* 299–350.

Dobbie, M. J.; Henderson, B. L.; and Stevens, D. L. 2008. Sparse sampling: Spatial design for monitoring stream networks. *Statistics Surveys* 2:113–153.

Eck, J. E.; Chainey, S.; Cameron, J. G.; Leitner, M.; and Wilson, R. E. 2005. Mapping crime: Understanding hot spots. Technical report, National Institute of Justice.

Ewing, G. O. 1976. Environmental and spatial preferences of interstate migrants in the United States. In Golledge, R. G., and Rushton, G., eds., *Spatial Choice and Spatial Behavior*. Columbus: Ohio State University Press. pp. 250–270.

Gelman, A., and Hill, J. 2007. *Data Analysis using Regression and Multilevel / Hierarchical Models*. Cambridge: Cambridge University Press.

Gilbert, R. O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. New York: John Wiley and Sons.

Gorr, W.; Olligschlaeger, A.; and Thompson, Y. 2003. Short term forecasting of crime. *International Journal of Forecasting* 19:579–594.

Gorr, W. L. 2009a. Cloudy with a chance of theft. *Wired*.

Gorr, W. L. 2009b. Forecast accuracy measures for exception reporting using receiver operating characteristic curves. *International Journal of Forecasting* 25(1):48–61.

Groff, E. R., and LaVigne, N. G. 2002. Forecasting the future of predictive crime mapping. In Tilley, N., ed., *Analysis for Crime Prevention*, volume 13 of *Crime Prevention Series*. Monsey, NY: Lynne Rienner Publishers. 29–57.

Huddleston, S., and Brown, D. E. 2009. A statistical threat assessment. *Systems, Man, and Cybernetics, Part A*. 39:1307–1315.

Illian, J.; Penttinen, A.; Stoyan, H.; and Stoyan, D. 2008. *Statistical Analysis and Modeling of Spatial Point Patterns*. West Sussex: John Wiley and Sons Ltd.

Kerchner, S. H. 2000. Spatial-temporal event prediction. Master's thesis, University of Virginia, Charlottesville.

Kery, M. 2010. *Introduction to WinBUGS for Ecologists*. Amsterdam: Elsevier.

Kewley, R. H., and Evangelista, P. 2007. Evaluating machine learning methods for geospatial prediction problems. being prepared for submission to IEEE.

Lawson, A. B. 2009. *Bayesian Disease Mapping*. Boca Raton: CRC Press.

Liu, H., and Brown, D. E. 1998. Spatial-temporal event prediction: A new model. In *Proceedings of the 1998 IEEE International Conference on Systems, Man, and Cybernetics*, volume 3, 2933–2937. San Diego: IEEE.

Liu, H., and Brown, D. E. 2003. Criminal incident prediction using a point-pattern based density model. *International Journal of Forecasting* 19:603–622.

Lunn, D.; Thomas, A.; Best, N.; and Spiegelhalter, D. 2000. Winbugs – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 10:325–337.

McFadden, D. 1986. The choice theory approach to market research. *Marketing Science* 5:275–297.

Petris, G.; Petrone, S.; and Campagnoli, P. 2009. *Dynamic Linear Models with R*. Dordrecht: Springer.

Porter, M. D. 2006. *Detecting Space Time Anomalies in Point Process Models of Intelligent Site Selection*. Ph.D. Dissertation, University of Virginia, Charlottesville.

Ratcliffe, J. H. 2004. The hotspot matrix: A framework for the spatio-temporal targeting of crime reduction. *Police Practice and Research* 5:5–23.

Rengert, G. F. 1997. Auto theft in central philadelphia. In Homel, R., ed., *Policing for Prevention: Reducing Crime, Public Intoxication and Injury*, volume 7 of *Crime Prevention Series*. Monsey, NY: Lynne Rienner Publishers. 199–219.

Rickman, D. S.; Miller, S. R.; and McKenzie, R. 2009. Spatial and sectoral linkages in regional models: A bayesian vector autoregression forecast evaluation. *Papers in Regional Science* 88(1):29–41.

Rossmo, K.; Laverty, I.; and Moore, B. 2005. Geographic profiling for serial crime investigation. In Wang, F., ed., *Geographic Information Systems and Crime Analysis*. Hershey: Idea Group Publishing. 137–152.

Smith, M. A., and Brown, D. E. 2004. Hierarchical choice modeling of terror attack site selection. *Decision Support Systems*.

Swets, J. A.; Dawes, R. M.; and Monahan, J. 2000. Better decisions through science. *Scientific American* 283(4):82–87.

Tobler, W. 1970. A computer movie simulating urban growth in the detroit region. *Economic Geography* 46(2):234–240.

Townsley, M.; Homel, R.; and Chaseling, J. 2000. Repeat burglary victimisation: Spatial and temporal patterns. *Australian and New Zealand Journal of Criminology* 33(1):37–63.

Wickle, C. K. 2003. Hierarchical bayesian models for predicting the spread of ecological processes. *Ecology* 84(6):1382–1394.

Withers, S. D. 2002. Quantitative methods: Bayesian inference, bayesian thinking. *Progress in Human Geography* 26:553–566.

Zhu, J.; Zheng, Y.; Carroll, A. L.; and Aukema, B. H. 2008. Autologistic regression analysis of spatial-temporal binary data via monte carlo maximum likelihood. *Journal of Agricultural, Biological and Environmental Statistics* 13:84–98.

Zhu, J.; Huang, H. C.; and Wu, J. 2006. Modeling spatial-temporal binary data using markov random fields. Department of Statistics.