Modeling Local-Scale Violent Crime Rate: A Comparison of Eigenvector Spatial Filtering Models and Conventional Spatial Regression Models

Yeran Sun, Shaohua Wang, Jing Xie & Xuke Hu

Published online: 29 Jan 2021.

Submit your article to this journal ↗

Article views: 86

View related articles ↗

View Crossmark data ↗

Modeling Local-Scale Violent Crime Rate: A Comparison of Eigenvector Spatial Filtering Models and Conventional Spatial Regression Models

Yeran Sun
Swansea University

Shaohua Wang
University of Illinois at Urbana–Champaign

Jing Xie
The University of Hong Kong

Xuke Hu
German Aerospace Center (DLR)

Environmental factors have both direct and indirect impacts on crime behavior decision making. This study aimed to examine to what degree the occurrences of violent crimes can be affected by social and built environment over space. Although a few studies have attempted to model crime rate using spatial regression models, there is a lack of comparison of spatial regression models. Particularly, the eigenvector spatial filtering type of spatial regression models has reportedly been effective in urban and regional studies, but it has not been widely applied to crime data. In this study, we aimed to examine whether the spatial filtering type of spatial regression models outperforms conventional types of spatial regression models in modeling violent crime rates over space. Moreover, we aimed to investigate the impacts of land use mix and street connectivity on the occurrences of violent crimes as the routine activity theory explained. In empirical studies, two types of spatial regression models (i.e., spatial error model and eigenvector spatial filtering model) were selected and estimated successfully to model local-scale violent crime rates across New York City. The eigenvector spatial filtering models outperform the spatial error models as well as the nonspatial models. Model estimation results show that occurrences of violent crimes (i.e., assaults and robberies) can be well determined by socioeconomic and built environment factors and thereby environmental factors can affect the occurrences of violent crimes. The contributions of socioeconomic and built environment factors to violent crime can offer insights on urban planning and policymaking toward violent crime prevention. Particularly, this study offers new evidence on the routine activity theory that increasing land use mix and street connectivity can enhance street activity, thereby reducing occurrences of violent crimes. Policymakers and urban planners should continue to enhance street activity through increasing land use mix and street connectivity. In addition, eigenvector spatial filtering models are advocated for use in crime or other applications in urban and regional studies. **Key Words: eigenvector spatial filtering, land use mix, spatial regression model, street connectivity, violent crime.**

Crimes are found to exhibit spatially clustered distribution over space rather than random distribution. Although crimes are found to be influenced by biological or psychological factors at the individual level, they are mainly influenced by environmental factors at the aggregate level. Environmental factors have both direct and indirect impacts on crime behavior decision making by affecting psychological factors. On the one hand, clustering (hotspots) of crime is found to be strongly associated with socioeconomic factors such as poverty, education level, ethnicity, and so forth (Sherman, Gartin, and Buerger 1989; Loukaitou-Sideris 1999; Weisburd et al. 2004; Uittenbogaard and Ceccato 2012; Helbich and Arsanjani 2015; Pope and Song 2015). On the other hand, the built environment is found to influence crime behavior by affecting perception of crime committers (Bernasco and Block 2011; Ceccato and Uittenbogaard 2014). Typically, according to routine activity theory, increasing land use mix or street connectivity can enhance street activity, thereby reducing neighborhood crime due to the increase in natural surveillance (Jacobs 1961; Gilderbloom, Riggs, and Meares 2015; Phillips and Sandler 2015). Little empirical evidence on the contributions of increasing land use mix or street connectivity on the prevention of crimes has been offered by empirical studies, however (Sohn 2016). Therefore, more empirical studies are needed to investigate the association of land use mix, street connectivity, and crime rate.

Furthermore, because environmental effects are likely to differ between violent crime and property crime (Speer et al. 1998; Smith, Frazee, and Davison 2000; Grinols and Mustard 2006; Wolfe and Pyrooz 2014; Sohn 2016), this study separates violent crimes from property crimes. Therefore, this study focuses on violent crimes as safety perceptions of residents are influenced more by violent crime occurrences than by property crime occurrences. Specifically, we examined to what degree occurrences of violent crimes can be affected by social and built environment at the aggregate level. In particular, whether and how land use mix or street connectivity affects the occurrence of violent crimes over space is of high interest. Consequently, models were established to examine how environment factors contribute to local-scale occurrences of violent crimes over a city. A better understanding of environmental effects of violent crime would offer new policy implications for crime reduction and prevention efforts.

Because geospatial data are used in this study, spatial regression models are recommended to replace nonspatial regression models (e.g., ordinary least squares [OLS] models). Residuals might be spatially autocorrelated when nonspatial regression models are applied to geospatial data. In this case, the existence of residual spatial autocorrelations will undermine the assumption of independence in regression models. To address this issue, spatial regression models are developed for geospatial data. Because there are two basic types of spatial regression models (i.e., spatial lag and spatial error models), these two types of models were used in the vast majority of previous studies to estimate crime rate over space. In recent years, a new type of spatial regression model, eigenvector spatial filtering models, has been used in a few studies (Chun 2014; Helbich and Arsanjani 2015). Previous studies used either conventional types or the eigenvector spatial filtering type of spatial regression models to estimate spatial distribution of crime rate (e.g., Gorman et al. 2001; Andresen 2006; Takagi, Ikeda, and Kawachi 2012; Chun 2014; Helbich and Arsanjani 2015; Troy, Nunery, and Grove 2016). A direct comparison between conventional types and the eigenvector spatial filtering type of spatial regression models has not been done; therefore, this study aims to compare them directly in estimates of crime rates across a city.

Specifically, in this study, we aimed to model local-scale violent crime rates according to socioeconomic and built environment factors over New York City. We replaced OLS models with spatial regression models because residual spatial autocorrelation exists in the OLS model established. Consequently, we discuss the performance of the three models (i.e., one nonspatial and two spatial models). Furthermore, we discuss the contributions of socioeconomic and built environment factors to the violent crime rate, thereby offering advice on urban planning and policymaking. The remainder of this article is organized as follows. We first introduce the data and methods used in this article. We then carry out an empirical model and, finally, we present a conclusion and offer suggestions for future work.

Materials and Methods

In this section, data on crime, socioeconomic, and built environment factors are introduced. Subsequently, the socioeconomic and built environment factors are presented. Finally, the two conventional types and the eigenvector spatial filtering type of spatial regression models used are briefly presented.

Data

The neighborhood tabulation area (NTA) was used as the areal unit due to its size and availability of socioeconomic data. Boundaries of NTAs were created by the New York City Department of City Planning based on census tracts. In total, there are 195 NTAs within New York City. Of them, three NTAs have no residents and thus they are excluded from this study (see gray polygons in Figure 1).

Crime data were downloaded from the NYC OpenData database (NYC OpenData 2017). This data set includes the violent crimes reported to the New York City Police Department from 2006 onward. The violent crime categories include arson, assault, homicide, murder, kidnapping, rape, and robbery. In this study, we focus on two categories of violent crimes: assault and robbery. We selected geo-referenced assaults and robberies from 2012 to 2016 and then aggregated them from the street level to the NTA level. We divided the number of five-year assaults and robberies by population to represent the NTA-level assault rate and robbery rate (cases/1,000 persons), respectively. Figure 1 maps the NTA-level five-year assault rate and robbery rate across New York City, and the gray areas represent the three unpopulated NTAs. In addition, Figure 2 shows box plots of NTA-level annual violent crime rate across New York City in different years. As Figure 2 shows, interannual variability in assault rate or robbery rate is not high across New York City.

The NYC OpenData database also offers socioeconomic data at the NTA level (NYC OpenData 2017). The date set extracted from the American Community Survey contains demographic, economic, housing, and sociological profiles between 2012 and 2016. Road network data were downloaded from NYC OpenData as well, and land use data were downloaded from the New York City Department of City Planning (New York City 2017). We overlapped the street network
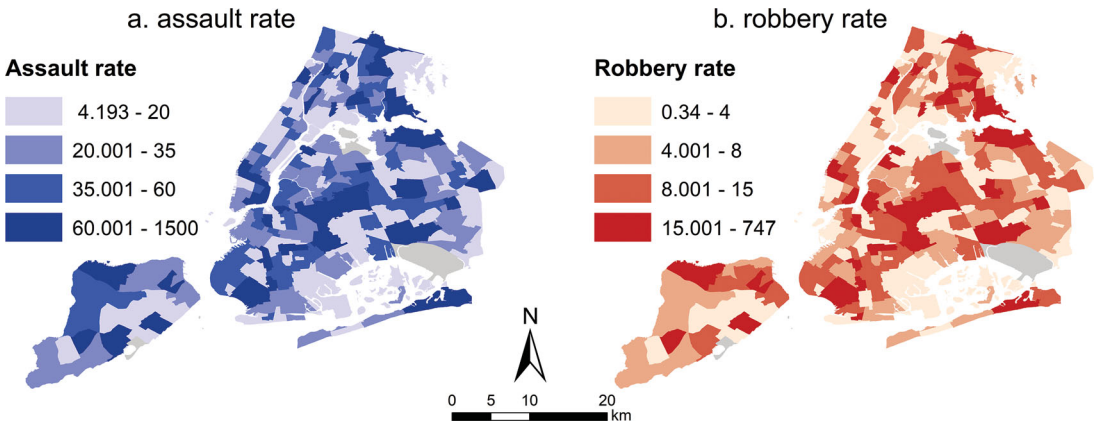
a. assault rate

b. robbery rate

Figure 1 *Five-year violent crime rate across New York City (cases/1,000 persons).*

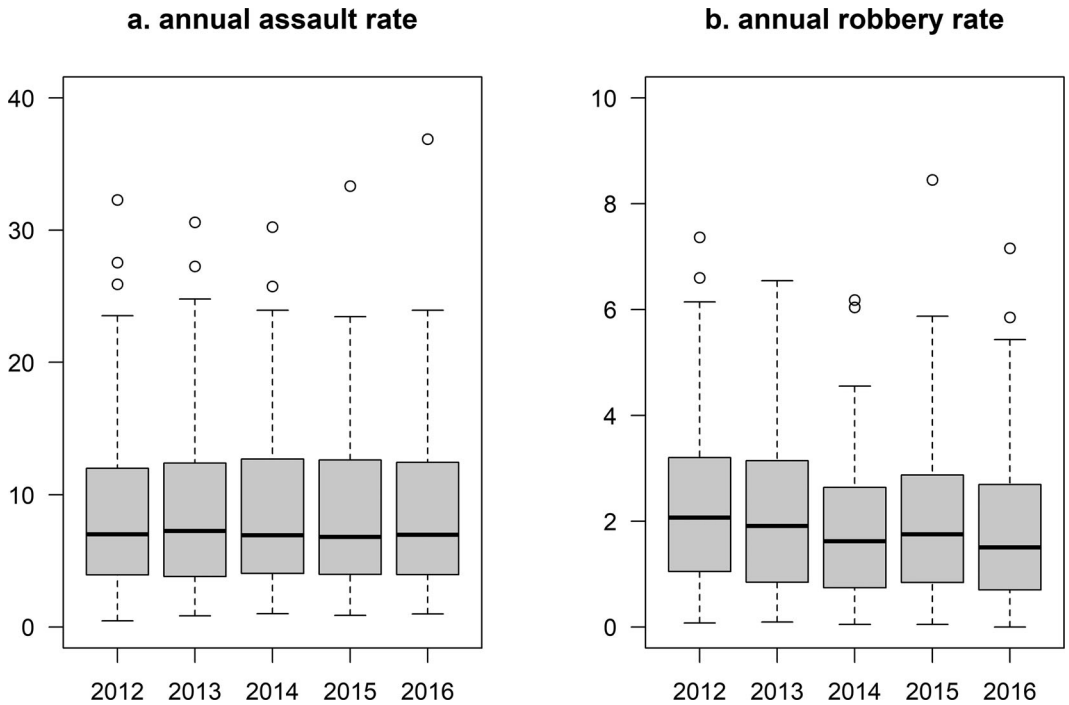## a. annual assault rate

b. annual robbery rate

Figure 2 *Box plot of annual violent crime rate across New York City in different years (cases/1,000 persons).*

data and the land use data onto NTA boundaries to calculate road density and proportions of distinct land use categories.

Variables

Although some studies suggested replacing population-standardized measures with area-standardized measures may produce a more biased estimate owing to the possible presence of ecological fallacy or the modifiable areal unit problem (Sohn 2016), this study selected that population-standardized crime rate, because most of the independent variables are population-standardized measures

as well. Moreover, we replace the violent crime rate measures (i.e., assault rate and robbery rate) with their logarithmic forms (i.e., ln(assault rate) and ln(robbery rate)) as the dependent variables (responses) because the violent crime rate measures are not normally distributed. In this study, we took account of socioeconomic factors, land use patterns, and connectivity factors as the independent variables (predictors). Table 1 lists the explanatory variables as potential affecting factors and Table 2 lists the descriptive statistics of variables. The socioeconomic, land use, and connectivity factors are all measured at the NTA level. The entropy-based land use mix is calculated as

Table 1 *Responses and explanatory variables (predictors) considered in this study*

Category	Variable	Full name
Responses	AR	Assault rate
	RR	Robbery rate
Socioeconomic factors	FemP	Percentage of female population
	PopU18P	Percentage of population under 18 years old
	PopO65P	Percentage of population 65 years and older
	NonWtNHP	Percentage of nonwhite population
	CvLFUEm1P	Percentage of unemployed working population
	FamBwPvP	Percentage of families below poverty
	VacHUP	Percentage of vacant housing units
	HOVacRtE	Homeowner vacancy rate
	RntVacRtE	Rental vacancy rate
	MrdFamP	Percentage of married-couple families
	LgOEnLEP1P	Percentage of population speaking English less than "very well"
Built environment factors	LUX	Land use mix (entropy-based measures)
	Road_Den	Road density (km/km^2)

Table 2 *Descriptive statistics of variables in this study*

Category	Variables	*M*	*SD*
Responses	ln(AR)	3.57	0.84
	ln(RR)	2.09	0.96
Explanatory variables	FemP	52.09	3.94
	PopU18P	21.1	6.04
	PopO65P	13.38	4.72
	NonWtNHP	66.04	27.79
	CvLFUEm1P	5.44	2.1
	FamBwPvP	16.3	10.76
	VacHUP	8.12	3.92
	HOVacRtE	2.44	2.54
	RntVacRtE	3.53	3.01
	MrdFamP	38.19	13.72
	LgOEnLEP1P	22.23	14.15
	LUX	2.28	0.59
	Road_Den	18.56	3.63

$$LUX = -\sum_{i=1}^{N} p_i log_2(p_i), \qquad (1)$$

where the vector p_i is the proportion of ith land use category in the areal unit (NTA), and N is equal to 11 because there are eleven land use categories. The higher the *LUX* value, the higher the land use mix level is.

Spatial Regression Model

Spatial regression models are recommended if residual spatial autocorrelation is present in nonspatial regression models. As the two conventional types of spatial regression models, the spatial lag (SL) model and spatial error (SE) model are widely used to fit linear models with spatially dependent errors and spatially lagged dependent variable, respectively (Larch and Walde 2008; LeSage and Pace 2009). Moran's *I* tests and Lagrange multiplier (LM) tests had been employed to decide whether and in which form spatial correlation exists in regression models, respectively (Larch and Walde 2008). According to the results of LM tests, appropriate conventional spatial models will be selected from SL or SE models. Specifically, there are two basic types of LM tests: a simple LM test for spatial lag dependence (LMlag) and a simple LM test for spatial error dependence (LMerr). The results of LM tests are interpreted as follows:

1. If either LMlag or LMerr is statistically significant, either SL dependence or SE dependence is found to exist in the regression models and thereby either an SL model or SE model should be selected.
2. If both LMlag and LMerr are statistically significant, both SL dependence and SE dependence are found to exist in the regression models and thereby both the SL model and SE model can be selected.
3. If neither LMlag nor LMerr is statistically significant, neither SL dependence nor SE dependence is found to exist in the regression models and thereby neither an SL model nor SE model should be selected. In this case, a nonspatial regression model should be selected.

Conventional Types of Spatial Regression Models (Spatial Lag or Spatial Error Models)

In the standard linear regression model, spatial dependence can be incorporated in two distinct ways: as an additional regressor in the form of a spatially lagged dependent variable (Wy) or in the error structure (E[ε$_i$ε$_j$] ≠ 0; Anselin 2001). Specifically, the two conventional types of spatial regression models are presented as follows (Anselin 2001).

Formally, a spatial lag model is expressed as

$$\mathbf{y} = \rho\mathbf{Wy} + \mathbf{X\beta} + \varepsilon, \qquad (2)$$

where ρ is a spatial autoregressive coefficient, $\mathbf{W}$ is a $n \times n$ spatial weights matrix, $\mathbf{X}$ is a $n \times K$ matrix of the explanatory variables, $\boldsymbol{\beta}$ is a vector of coefficients, and ε is a vector of error terms. Unlike what holds for the time series counterpart of this model, the spatial lag term Wy is correlated with the disturbances, even when the latter are independent and

identically distributed. Therefore, this can be seen from the reduced form of (2),

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\varepsilon, \qquad (3)$$

in which each inverse can be expanded into an infinite series, including both the explanatory variables and the error terms at all locations (the spatial multiplier). Consequently, the spatial lag term must be treated as an endogenous variable and proper estimation methods must account for this endogeneity (OLS will be biased and inconsistent due to the simultaneity bias).

A spatial error model is a special case of a regression with a nonspherical error term, in which the off-diagonal elements of the covariance matrix express the structure of spatial dependence. Consequently, OLS remains unbiased, but it is no longer efficient and the classical estimators for standard errors will be biased. The spatial structure can be specified in a number of different ways and (except for the nonparametric approaches) results in an error variance–covariance matrix of the form

$$\mathrm{E}[\varepsilon\varepsilon'] = \Omega(\theta), \qquad (4)$$

where θ is a vector of parameters, such as the coefficients in a simultaneous spatial autoregressive error process.

Eigenvector Spatial Filtering Model

As a new type of spatial regression model, spatial regression models with eigenvector spatial filtering are also applied to geospatial data (Griffith 2003; Tiefelsdorf and Griffith 2007). Moran's eigenvector-based spatial regression approach is called eigenvector spatial filtering (ESF; Griffith 2003) in regional science, and ESF with a small number of eigenvectors (i.e., small L) can greatly reduce model misspecification errors and increase model accuracy (Murakami and Griffith 2019). The ESF model is presented as follows (Chun et al. 2016).

ESF utilizes the spectral decomposition of a transformed spatial weights matrix, $\mathbf{C}$. The spectral decomposition of matrix $\mathbf{MCM}$ (where $\mathbf{M} = (\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}})/n$ and $\mathbf{1}$ is a vector of ones) produces a set of n eigenvalues and their corresponding eigenvectors:

$$\mathbf{MCM} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^{-1} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^{\mathrm{T}}, \qquad (5)$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal elements are the n eigenvalues $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ ordered from the largest value to the smallest value, and $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ represents the n corresponding eigenvectors. As an output of the spectral decomposition, the eigenvectors are mutually orthogonal and uncorrelated and $(n-1)$ have a zero mean, and one is proportional to the vector $\mathbf{1}$. Each of these eigenvectors represents a distinct nature and degree of spatial autocorrelation. ESF introduces a subset of the eigenvectors as control variables

in a regression model specification in order to capture its spatial stochastic component.

In linear regression, an ESF model specification can be expressed as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}_k\boldsymbol{\beta}_{\mathrm{E}} + \varepsilon, \qquad (6)$$

where $\mathbf{Y}$ denotes the dependent variable, $\mathbf{X}$ denotes a matrix of independent variables, $\mathbf{E}_k$ denotes a selected set of k eigenvectors selected from the n eigenvectors $\mathbf{E}$, $(\boldsymbol{\beta}, \boldsymbol{\beta}_{\mathrm{E}})$ denote parameters, and ε denotes random noise that is distributed $N(\mathbf{0}, \mathbf{I}\sigma^2)$. In this specification, $\mathbf{E}_k\boldsymbol{\beta}_{\mathrm{E}}$ captures the spatial stochastic component in the dependent variable $\mathbf{Y}$. Hence, the regression model does not suffer from complications attributable to spatial autocorrelation, which is likely to be observed in its residuals if the spatial stochastic component is not explicitly addressed.

The identification of $\mathbf{E}_k$ can be achieved through a stepwise procedure. Specifically, eigenvectors that minimize the level of spatial autocorrelation at each step can be selected. Intuitively, although this minimizing residual spatial autocorrelation criterion adheres to the notion of isolating spatial autocorrelation, it becomes computationally demanding as n increases. That is, in order to evaluate whether the addition of an eigenvector reduces spatial autocorrelation in residuals, the expected value and variance of Moran's I for the residuals needs to be recalculated repeatedly, which involves the inversion of large matrices.

This identification procedure can be assisted further by excluding irrelevant eigenvectors. The stepwise procedure can be conducted from a noticeably smaller set (i.e., a candidate set) instead of the entire set of eigenvectors, $\mathbf{E}$. A candidate set can be demarcated based upon several criteria. First, eigenvectors that do not explain much spatial variation can be excluded. Second, eigenvectors that represent negative spatial autocorrelation can be excluded when variable $\mathbf{Y}$ displays positive spatial autocorrelation and vice versa. This exclusion procedure can be assisted by the eigenvalues λ, because λ_i is proportional to Moran's I value of a map that is a portrayal of $\mathbf{E}_i$ on the spatial tessellation from which $\mathbf{C}$ is created; Moran's $I = \lambda_i \, n/\mathbf{1}^{\mathrm{T}}\mathbf{C}\mathbf{1}$. Hence, a candidate set is often constructed with a threshold minimum Moran's I value of 0.25, which is related to approximately 5 percent of the variation in a response variable being attributable to spatial autocorrelation.

In addition, the selection and estimates of the spatial regression models can be implemented in R. Specifically, the package 'spdep' allows users to conduct Moran's I tests and LM tests for spatial dependence (see https://cran.r-project.org/web/packages/spdep/index.html). The package 'spatialreg' was developed to support estimates of SL or SE models (see https://cran.r-project.org/web/packages/spatialreg/index.html). The package 'spmoran' was developed to implement estimates of ESF models (see https://cran.r-project.org/web/packages/spmoran/index.html).

Table 3 *Moran's I tests for the selection of models*

Response	Moran's I
ln(AR)	0.096**
ln(RR)	0.2309***

**p < 0.01.
***p < 0.001.

Results and Discussion

This section demonstrates the model selection and estimation results and discusses the contributions of explanatory variables, specifically the empirical evidence on the associations of land use mix, street connectivity, and violent crime rate.

Model Selection Results

To select a more appropriate model, two nonspatial regression models (OLS models) were initially estimated based on 192 samples (192 NTAs) in this study. Subsequently, Moran's *I* tests and LM tests were used to decide whether and in which form spatial correlation is present in the nonspatial regression models (the OLS models). Tables 3 and 4 list the Moran's *I* tests and LM tests for the selection of regression models. In Table 3, Moran's *I* tests for residuals are statistically significant in both the OLS models for ln(AR) and ln(RR). This indicates the presence of residual spatial dependence in the two nonspatial regression models (OLS models). Table 4 shows the values of the LM tests as well as the *p* values to measure the statistical significance of the LM tests. In Table 4, simple LM tests for SE dependence are statistically significant in both the OLS models for ln(AR) and ln(RR), because their *p* values are all below 0.05. This indicates the presence of SE dependence in the two nonspatial regression models (OLS models). Therefore, replacing the nonspatial regression models (the OLS models) with the spatial regression model is necessary. More specifically, among the two conventional types of spatial regression models, the SE model is more appropriate than the SL model owing to the presence of SE dependence other than the presence of SL dependence. Moreover, the ESF model is recommended due to the presence of residual spatial dependence in the nonspatial models initially estimated.

Model Estimate Results

Consequently, the two types of spatial regression models were estimated based on 192 samples (192 NTAs) for ln(AR) and ln(RR). In this study, because the geodata used are areal data, we selected the contiguity-based spatial weights matrix. The SE models were estimated by the maximum likelihood method. In the estimation of ESF models, the eigenvectors were selected by a stepwise method. As mentioned previously, the threshold for the eigenvalues is set to 0.25. As a result, 35 of 192 eigen-pairs were extracted.

Table 5 lists the estimation results for the spatial regression models (SE and ESF models) and nonspatial models (OLS models) estimated. The ESF models have the lowest Akaike's information criterion (AIC) values and the highest R^2 values. This indicates that the ESF models outperform SE and OLS models. The R^2 values of ESF models estimated are more than 0.8, indicating that the response variable (i.e., assault rate and robbery rate) can be determined by the explanatory variables (socioeconomic and built environment factors) in the ESF model estimated (see Table 5). In addition, Moran's *I* tests for residuals of the models estimated show that the presence of residual spatial autocorrelation is not statistically significant in the spatial models (i.e., SE and ESF models). This indicates that replacing nonspatial models with spatial models can address the presence of residual spatial dependence. In addition, variance inflation factor values for all of the explanatory variables are below ten, indicating that multicollinearity is not problematic in the models estimated. This means that all of the explanatory variable (predictors) are not highly correlated with each other.

Moreover, Table 5 displays more details of the model estimation results. In both the SE and ESF models estimated, some factors such as FemP, NonWtNHP, CvLFUEm1P, FamBwPvP, RntVacRtE, MrdFamP, LUX, and Road_Den exhibit substantial effects on the dependent variables (responses), because their corresponding *p* values are below 0.05 (see Table 5). More specifically, FemP, MrdFamP, LUX, and Road_Den exhibit negative effects, whereas NonWtNHP, CvLFUEm1P, FamBwPvP, VacHUP, RntVacRtE, and LgOEnLEP1P exhibit positive effects. Because the coefficients of predictors are similar in the three types of models estimated, the effects of predictors exhibiting statistically significant effects are similar in the three types of models. The exception is CvLFUEm1P, which is statistically significant in the OLS and SE models but not in the ESF models.

Additionally, we modeled annual violent crime rate (i.e., assault rate and robbery rate) in different years by ESF models to examine interannual temporal variability in the models estimated. Table 6 shows estimation results for ESF models in different years. Generally, interannual temporal variability in the models estimated is not high for assault rate and robbery rate as variable coefficients, AIC values, and R^2 values are similar between different years.

Environmental Effects on Violent Crime

Because the effects of predictors exhibiting statistically significant effects are similar in the three types of models and ESF models outperform the other

Table 4 *LM tests for the selection of models*

Response	Simple LM test for spatial lag dependence (LMlag)	Simple LM test for spatial error dependence (LMerr)
ln(AR)	0.1469	4.4054*
ln(RR)	15.787***	25.456***

Notes: LM = Lagrange multiplier.
*p < 0.05.
***p < 0.001.

Table 5 *Estimation results for nonspatial and spatial regression models*

	Model		
Coefficient **Response: ln(RR)**	**OLS**	**SE**	**ESF**
Response: ln(AR)			
Intercept	7.5799***	7.6592***	7.9432***
FemP	−0.0525***	−0.0523***	−0.0577***
PopU181P	0.0054	0.0012	−0.0071
Pop65pl1P	−0.0159+	−0.0154+	−0.012
NonWtNHP	0.0064***	0.0066***	0.0077***
CvLFUEm1P	0.0519*	0.0476*	0.0241
FamBwPvP	0.0161*	0.0184*	0.0254***
VacHUP	0.032***	0.0343***	0.0485***
HOVacRtE	0.0076	0.0087	0.0058
RntVacRtE	0.0674***	0.0707***	0.0612***
MrdFamP	−0.0294***	−0.0294***	−0.0295***
LgOEnLEP1P	0.0002	0.0005	−0.0025
LUX	−0.244***	−0.2761***	−0.2871***
Road_Den	−0.0531***	−0.0543***	−0.0485***
AIC	194.7	192.42	154.24
Adjusted R^2	0.7883		0.8412
Pseudo R^2		0.8071	
Moran's *I* test for residuals	0.096**	−0.0067	−0.1123
Intercept	5.656***	5.6572***	5.76***
FemP	−0.0433***	−0.0407***	−0.049***
PopU181P	0.0106	−0.0063	−0.0126
Pop65pl1P	−0.021+	−0.0158+	0.0024
NonWtNHP	0.0067**	0.008***	0.0099***
CvLFUEm1P	0.0822**	0.0495*	0.0242
FamBwPvP	0.0075	0.0153*	0.021*
VacHUP	0.038**	0.04***	0.0438s***
HOVacRtE	−0.0123	0.0093	0.0078
RntVacRtE	0.0715***	0.0877***	0.0704***
MrdFamP	−0.0377***	−0.0316***	−0.0293***
LgOEnLEP1P	0.0098*	0.0009	0.0091+
LUX	−0.3326***	−0.3588***	−0.282***
Road_Den	−0.0445***	−0.0576***	−0.0579***
AIC	316.74	294.28	236.78
Adjusted R^2	0.6958		0.8165
Pseudo R^2		0.7504	
Moran's *I* test for residuals	0.2309***	−0.0346	−0.1131

Notes: OLS = ordinary least squares; SE = spatial error; ESF = eigenvector spatial filtering; AIC = Akaike's information criterion.
+p < 0.1.
*p < 0.05.
**p < 0.01.
***p < 0.001.

two types, we focus more on the estimate results of the ESF models to further discuss the environmental effects. Specifically, we discuss the effects of independent variables (predictors) as shown in Table 5. Of personal socioeconomic factors, FemP (percentage of female population) exhibits negative effects on both assault rate and robbery rate, and NonWtNHP (percentage of non-white population) exhibits a positive effect on both assault rate and

robbery rate. Of household socioeconomic factors, FamBwPvP (percentage of families below poverty), VacHUP (percentage of vacant housing units), and RntVacRtE (rental vacancy rate) exhibit positive effects on both assault rate and robbery rate and MrdFamP (percentage of married-couple families) exhibits a negative effect on both assault rate and robbery rate. Of built environment factors, LUX (land use mix) and Road_Den (road density) exhibit

Table 6 *Estimation results for eigenvector spatial filtering models in different years*

Coefficient Response: ln(RR)	Year				
	2012	2013	2014	2015	2016
Response: ln(AR)					
Intercept	6.6302***	5.6791***	6.436***	6.163***	6.2841***
FemP	−0.0632***	−0.0501***	−0.0607***	−0.0543***	−0.0564***
PopU181P	−0.0052	−0.0027	−0.007	−0.0034	−0.0061
Pop65pl1P	−0.0098	−0.0084	−0.0092	−0.0121	−0.0075
NonWtNHP	0.0078***	0.0089***	0.0087***	0.0082***	0.0077***
CvLFUEm1P	0.0175	0.0389+	0.0239	0.0156	0.0208
FamBwPvP	0.0261**	0.0223**	0.0247**	0.0239**	0.0275***
VacHUP	0.0508***	0.0463***	0.0438***	0.0453***	0.0469***
HOVacRtE	0.0077	0.0125	0.0055	0.0083	0.0045
RntVacRtE	0.0627***	0.0604***	0.0604***	0.0616***	0.06316***
MrdFamP	−0.028***	−0.0258***	−0.0303***	−0.03101***	−0.03001***
LgOEnLEP1P	−0.0042	−0.0017	−0.0021	−0.001	−0.0025
LUX	−0.2647***	−0.3135***	−0.3245***	−0.2817***	−0.2813***
Road_Den	−0.0563***	−0.0538***	−0.0426***	−0.0498***	−0.0535***
AIC	164.4687	164.5858	145.7855	184.4764	177.2286
Adjusted R^2	0.8311	0.8366	0.8485	0.8195	0.8285
Intercept	3.2011***	3.7184***	4.3045***	4.8535***	4.3596**
FemP	−0.0232+	−0.0451***	−0.0527***	−0.0544***	−0.0594*
PopU181P	−0.0233	−0.0125	−0.0146	−0.0179	0.0132
Pop65pl1P	−0.0173	0.0044	0.0064	0.0029	0.0042
NonWtNHP	0.0068*	0.0109***	0.0097***	0.0081**	0.0084
CvLFUEm1P	0.034	0.0262	0.0287	0.0051	0.0145
FamBwPvP	0.0243*	0.0135	0.0263*	0.0292**	−0.0136
VacHUP	0.045***	0.0579***	0.0536***	0.0401**	0.0497+
HOVacRtE	−0.0024	0.0141	0.0122	0.0057	0.001
RntVacRtE	0.0869***	0.0621***	0.0562***	0.0578***	0.1354***
MrdFamP	−0.0257***	−0.0283***	−0.0291***	−0.0299***	−0.0654***
LgOEnLEP1P	0.0108	0.0152**	0.0055	0.0024	0.0289
LUX	−0.2982**	−0.2462**	−0.3405***	−0.2338*	−0.0493
Road_Den	−0.0506***	−0.0613***	−0.0593***	−0.0648***	−0.0356
AIC	271.9586	259.4286	279.5344	294.0505	564.5207
Adjusted R^2	0.7766	0.8022	0.7964	0.7730	0.5260

Note: AIC = Akaike's information criterion.
+$p < 0.1$.
*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

negative effects on both assault rate and robbery rate.

Discussions and Implications for Policies

In this study, replacing the OLS model with ESF models can reduce the effects of residual spatial dependence. Moreover, although spatial regression models have been used in a variety of applications, the majority of those studies selected the conventional types of spatial regression models (i.e., SL or SE models) but neglected to compare them with other types (e.g., ESF). This study shows the necessity of selecting ESF models to enhance model estimation.

We further discuss the effects of socioeconomic and built environmental factors on violent crime. First, violent crimes are likely to occur in the areas with a high percentage of families below poverty. Second, violent crimes are likely to occur in the areas with a high percentage of non-white people, but they are unlikely to occur in the areas with a high percentage of female population or aging population. Third, violent crimes are likely to occur in areas with a high percentage of vacant housing units or of high rental vacancy rate, but they are unlikely to occur in the areas with a high percentage of married-couple families. Fourth, violent crimes are unlikely to occur in areas with highly mixed land uses or a high level of street connectivity. This finding is consistent with Jacobs' (1961) theory that mixed land use or street connectivity is beneficial to prevention of residential crime, thereby offering new empirical evidence on routine activity theory (Jacobs 1961). According to Jacobs' (1961) explanations on the contributions of mixed land use or high levels of street connectivity to crime reduction and prevention, mixed-use or highly connected neighborhoods encourage more residents onto streets (Browning and Jackson 2013). The more pedestrians there are on the streets, the more "eyes on the street." Consequently, increasing land use mix or street connectivity can promote natural surveillance through increasing street activity. This study empirically supports the contributions of increasing land use mix or street connectivity to preventing and reducing violent crime, as increasing street activity can enhance

natural surveillance on streets. Likewise, according to Jacobs' (1961) theory, a high frequency of violent crimes in the areas with a high percentage of vacant housing units or a high rental vacancy rate might also be attributable to lack of natural surveillance from residents nearby, because fewer residents are likely to live in those areas.

After discussing the effects of socioeconomic and built environment factors on the occurrence of violent crimes, we offer implications for policymaking and urban planning toward crime prevention and reduction. First, policymakers and urban planners should continue to enhance street activity through increasing land use mix and street connectivity. High street density and high land use mix can increase the number of pedestrians and vehicles, increasing natural surveillance. Enhancement of street activity could further enhance natural surveillance to reduce violent crime. Second, more attention could be given to an increase in surveillance and monitoring by means of police patrol or closed-circuit television in areas with a high percentage of vacant housing units or a high rental vacancy rate, where natural surveillance from residents or travelers is weak.

Conclusions

In this study, we aimed to model violent crime rate according to socioeconomic and built environment factors. Different types of spatial regression models were established to model local-scale violent crime rates across New York City. Specifically, the spatial regression models (i.e., the SE and ESF models) outperform the nonspatial models (OLS models), and the ESF models outperform the SE models. The responses (i.e., assault rate and robbery rate) can be determined by the explanatory variables (socioeconomic and built environment factors) in the ESF models estimated ($R^2 > 0.8$). ESF models are strongly recommended for use in modeling crime or other applications of urban and regional studies. Moreover, the results of model estimation show that environmental factors can affect the occurrence of violent crimes. In addition, the contributions of socioeconomic and built environment factors to violent crime offer implications for urban planning and policymaking. Specifically, violent crimes are unlikely to occur in the areas with highly mixed land uses or a high level of street connectivity. Urban regeneration and planning policies should continue to enhance street activity through increasing land use mix and street connectivity to lower the occurrences of violent crimes.

There are some limitations of this study. First, in this study, we have not discussed the temporal variation of violent crime. Violent crime might vary over time, and there might be a shift from one season to another in the concentration of violent crime (Ceccato and Uittenbogaard 2014). Second, some other environmental factors (e.g., quality of natural environment) were not considered due to the lack of data. For instance, some recent studies indicate that high levels of air pollution or high lead levels are likely to increase criminal activity (Boutwell et al. 2017; Lu et al. 2018). Third, exposure to alcohol or drugs is not considered due to the lack of data. Exposure to alcohol is thought to trigger violent crime behavior (Carpenter and Nevin 2010).

In the future, we will improve on this study by addressing those limitations. First, time-varying environmental factors like temperature, wind speed, and brightness could be incorporated into further studies. Second, natural environment factors like air quality measures (e.g., $PM_{2.5}$, ozone, CO, and SO_2) and lead exposure (blood lead levels) could be incorporated into further studies. Third, we would attempt to assess exposure to alcohol or drugs by mapping locations of alcohol or drug stores or delivering questionnaire-based surveys. Finally, we would take account of electronic surveillance coverage if the locations of cameras are available in the future. ∎

Literature Cited

Andresen, M. A. 2006. A spatial analysis of crime in Vancouver, British Columbia: A synthesis of social disorganization and routine activity theory. *The Canadian Geographer/Le Géographe Canadien* 50 (4):487–502.

Anselin, L. 2001. Spatial econometrics. In *A companion to theoretical econometrics*, ed. B. H. Baltagi, 310–30. Oxford, UK: Blackwell.

Bernasco, W., and R. Block. 2011. Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime and Delinquency* 48 (1):33–57.

Boutwell, B. B., E. J. Nelson, Z. Qian, M. G. Vaughn, J. P. Wright, K. M. Beaver, J. C. Barnes, M. Petkovsek, R. Lewis, M. Schootman, et al. 2017. Aggregate-level lead exposure, gun violence, homicide, and rape. *PLoS ONE* 12 (11). doi: 10.1371/journal.pone.0187953.

Browning, C. R., and A. L. Jackson. 2013. The social ecology of public space: Active streets and violent crime in urban neighborhoods. *Criminology: An Interdisciplinary Journal* 51 (4):1009–43. doi: 10.1111/1745-9125.12026.

Carpenter, D. O., and R. Nevin. 2010. Environmental causes of violence. *Physiology & Behavior* 99 (2):260–68. doi: 10.1016/j.physbeh.2009.09.001.

Ceccato, V., and A. C. Uittenbogaard. 2014. Space–time dynamics of crime in transport nodes. *Annals of the Association of American Geographers* 104 (1):131–50.

Chun, Y. 2014. Analyzing space–time crime incidents using eigenvector spatial filtering: An application to vehicle burglary. *Geographical Analysis* 46 (2):165–84.

Chun, Y., D. A. Griffith, M. Lee, and P. Sinha. 2016. Eigenvector selection with stepwise regression techniques to construct eigenvector spatial filters. *Journal of Geographical Systems* 18 (1):67–85.

Gilderbloom, J. I., W. W. Riggs, and W. L. Meares. 2015. Does walkability matter? An examination of walkability's impact on housing values, foreclosures and crime. *Cities* 42:13–24. doi: 10.1016/j.cities.2014.08.001.

Gorman, D. M., P. W. Speer, P. J. Gruenewald, and E. W. Labouvie. 2001. Spatial dynamics of alcohol availability, neighborhood structure and violent crime. *Journal of Studies on Alcohol* 62 (5):628–36. doi: 10.15288/jsa.2001.62.628.

Griffith, D. A. 2003. *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization*. Berlin: Springer Science & Business Media.

Grinols, E. L., and D. B. Mustard. 2006. Casinos, crime, and community costs. *The Review of Economics and Statistics* 88 (1):28–45.

Helbich, M., and J. J. Arsanjani. 2015. Spatial eigenvector filtering for spatiotemporal crime mapping and spatial crime analysis. *Cartography and Geographic Information Science* 42 (2):134–48.

Jacobs, J. 1961. *The death and life of great American cities*. New York: Vintage Books.

Larch, M., and J. Walde. 2008. Lag or error? Detecting the nature of spatial correlation. In *Data analysis, machine learning and applications*, ed. C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker, 301–8. Berlin: Springer.

LeSage, J. P., and R. K. Pace. 2009. *Introduction to spatial econometrics*. Boca Raton, FL: CRC Press.

Loukaitou-Sideris, A. 1999. Hot spots of bus stop crime: The importance of environmental attributes. *Journal of the American Planning Association* 65 (4):395–411. doi: 10.1080/01944369908976070.

Lu, J. G., J. J. Lee, F. Gino, and A. D. Galinsky. 2018. Polluted morality: Air pollution predicts criminal activity and unethical behavior. *Psychological Science* 29 (3):340–55. doi: 10.1177/0956797617735807.

Murakami, D., and D. A. Griffith. 2019. Eigenvector spatial filtering for large data sets: Fixed and random effects approaches. *Geographical Analysis* 51 (1):23–49.

New York City. 2017. The official Website of the City of New York. Accessed October 28, 2016. https://www1.nyc.gov/

NYC OpenData. 2017. Open data for all New Yorkers. Accessed July 31, 2016. https://opendata.cityofnewyork.us/.

Phillips, D. C., and D. Sandler. 2015. Does public transit spread crime? Evidence from temporary rail station closures. *Regional Science and Urban Economics* 52:13–26.

Pope, M., and W. Song. 2015. Spatial relationship and colocation of crimes in Jefferson County, Kentucky. *Papers in Applied Geography* 1 (3):243–50.

Sherman, L. W., P. R. Gartin, and M. E. Buerger. 1989. Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology* 27 (1):27–55.

Smith, W. R., S. G. Frazee, and E. L. Davison. 2000. Furthering the integration of routine activity theory and social disorganization theories: Small units of analysis and the study of street robbery as a diffusion process. *Criminology* 38 (2):489–523.

Sohn, D. 2016. Residential crimes and neighbourhood built environment: Assessing the effectiveness of crime prevention through environmental design (CPTED). *Cities* 52:86–93.

Speer, P. W., D. M. Gorman, W. Labouvie, and M. Ontkush. 1998. Violent crime and alcohol availability: Relationships in an urban community. *Journal of Public Health Policy* 19 (3):303–38.

Takagi, D., K. I. Ikeda, and I. Kawachi. 2012. Neighborhood social capital and crime victimization: Comparison of spatial regression analysis and hierarchical regression analysis. *Social Science & Medicine* 75 (10):1895–902. doi: 10.1016/j.socscimed.2012.07.039.

Tiefelsdorf, M., and D. A. Griffith. 2007. Semiparametric filtering of spatial autocorrelation: The eigenvector approach. *Environment and Planning A: Economy and Space* 39 (5):1193–221.

Troy, A., A. Nunery, and J. M. Grove. 2016. The relationship between residential yard management and neighborhood crime: An analysis from Baltimore City and County. *Landscape and Urban Planning* 147:78–87.

Uittenbogaard, A. C., and V. Ceccato. 2012. Space–time clusters of crime in Stockholm. *Review of European Studies* 4 (5):148–56.

Weisburd, D., S. Bushway, C. Lum, and S. M. Yang. 2004. Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle. *Criminology* 42 (2):283–321.

Wolfe, S., and D. C. Pyrooz. 2014. Rolling back prices and raising crime rates? The Walmart effect on crime in the United States. *British Journal of Criminology* 54 (2):199–221.

YERAN SUN is an Assistant Professor at the Department of Geography, College of Science, Swansea University, Swansea SA2 8PP, Wales, UK. E-mail: yeran.sun@swansea.ac.uk. He is a GIScientist and human geographer with interests in urban informatics, urban big data, and sustainable cities.

SHAOHUA WANG is a Postdoctoral Researcher at the CyberGIS Center for Advanced Digital and Spatial Studies, University of Illinois at Urbana–Champaign, Urbana, IL 61801. E-mail: shaohua@illinois.edu. He specializes in spatial optimization, spatial analysis, geoinformatics, and geographic information systems.

JING XIE is a Postdoctoral Fellow at the Faculty of Architecture, The University of Hong Kong, China. E-mail: xiej412@hku.hk. His research priorities include Earth observation, urbanization, vegetation ecology, and GIScience, specifically focusing on urban local climate zone, phenology and productivity of vegetation, mountainous ecosystems, land use and land cover, and climate change.

XUKE HU is a Postdoctoral Researcher at the Data Science Institute of German Aerospace Center (DLR), Mälzerstraβe 3, 07745 Jena, Germany. E-mail: xuke.hu@dlr.de. His research interests are in the fields of indoor positioning, mapping, and navigation; volunteered geographic information; and social media data analysis. In particular, he is interested in using annotation-free deep learning techniques to solve the problems faced in Earth observation and disaster management.