The Ambient Population and Crime Analysis

Article *in* The Professional Geographher · March 2011

DOI: 10.1080/00330124.2010.547151

The Ambient Population and Crime Analysis

Martin A. Andresen
Institute for Canadian Urban Research Studies, Simon Fraser University

This article uses an alternative measure of the population at risk, the ambient population (provided by Oak Ridge National Laboratory), in crime rate calculations. It is shown through a variety of statistical analyses at two different scales of aggregation that this alternatively calculated crime rate is not always related to the conventionally calculated crime rate. The implications of this finding are that past theoretical testing and policy formation might have been based on spurious results, showing the importance of remaining current with the developments of geographic information science technologies and data availability when undertaking a spatial analysis of crime. **Key Words: ambient population, crime rates, population at risk.**

本文在犯罪案率计算中使用一个高危人群的替代措施，周围环境人群（由橡树岭国家实验室提供）。它通过在两个不同尺度的聚集统计分析表明，这种替代计算的犯罪案率并不总是与传统方法计算的犯罪案率相关。这一发现的意义是，过去的理论测试和政策的形成可能是基于不可靠的结果，显示了在进行犯罪空间分析时，与地理信息科学技术和数据的可用性保持相当步调的重要性。**关键词：环境人口，犯罪率，高危人群。**

Este artículo utiliza una medida alternativa a la de población en riesgo, la población ambiente (suministrada por el Laboratorio Nacional de Oak Ridge), para el cálculo de las tasas de criminalidad. A través de una variedad de análisis estadísticos para dos diferentes escalas de agregación, se muestra que esta tasa de criminalidad calculada alternativamente no siempre está concuerda con la tasa de crimen calculada de manera convencional. Lo que este hallazgo implica es que las anteriores pruebas teóricas y formación de políticas podrían haberse basado en resultados espurios, mostrando con ello lo importante que es mantenerse actualizado sobre los desarrollos de las tecnologías de la ciencia de información geográfica y la disponibilidad de datos cuando se emprenda un análisis espacial del crimen. **Palabras clave: población ambiente, tasas de criminalidad, población en riesgo.**

Environmental criminology begins its analysis of the criminal event by asking where and when a crime has occurred (Brantingham and Brantingham 1991). Through this explicitly spatial-temporal approach to studying crime (broadly encompassing routine activity, rational choice, opportunity, and pattern theories), significant insight is garnered: Human activity patterns and, hence, crime patterns follow both spatial and temporal regularities. The development of geographic information science (GIScience) and the availability of spatially referenced crime data allow the spatial dimension of these theories to flourish. Most often, this occurs through aggregation of criminal event data to census boundary units and analyses in relation to census variables. These types of analyses, however, can be problematic. To calculate crime rates for census boundary units, the residential population is most often used. A potential problem arises because people leave their census boundary units during the day, a spatial and temporal regularity. Consequently, crime rate calculations based on the residential population might not provide accurate representations of risk.

This potential problem in crime rate calculations has been known in the criminological literature for many years (Boggs 1965; Harries 1991)—see Andresen and Jenion (2010) for a discussion of these issues. There are those, however, who state that the conventionally calculated crime rate in the geography of crime literature is better than alternatively calculated crime rates that might come at a high cost in terms of money, time, or both (see

The Professional Geographer, 63(2) 2011, pages 193–212
Initial submission, December 2009; final acceptance, April 2010.
Published by Taylor & Francis Group, LLC.

Cohen, Kaufman, and Gottfredson 1985). The conventionally calculated crime rates use the resident population rather than a population at risk, such as the presence of people for violent crime or the number of residential units for burglary. The proponents of alternatively calculated crime rates do not deny this claim for some crimes but do not believe this claim to be universal. The difficulty with alternatively calculated crime rates is that they are not readily available. As such, the status quo has been the conventionally calculated crime rate. The recent availability of a new data set alleviates past concerns of monetary and time costs for crimes that have a population at risk tied to the presence of people.

This article uses an alternative measure of the population at risk, the ambient population (a twenty-four-hour average estimate of the population present in a spatial unit made available through Oak Ridge National Laboratory), in violent crime rate calculations. The ambient population has been used in the crime analysis literature previously to test criminological theory (Andresen 2006, 2007), to show its general utility in crime prevention (Andresen and Jenion 2008), and in crime mapping (Andresen and Jenion 2010). However, these analyses have not systematically assessed the impact the ambient population has on a variety of spatial analysis techniques. This article shows the utility of the ambient population in crime analysis through the use of global indicators of spatial association, local indicators of spatial association, and spatial statistics at two different scales of aggregation. It is shown that alternatively calculated crime rates are not always related to conventionally calculated crime rates. This is important because the ambient population, as discussed later, is argued to be a better representation of the population at risk for some crime classifications. This shows the importance of remaining current with the developments of GIScience technologies and data availability when undertaking a spatial analysis of crime. The following section presents the data and methodology employed in the analysis.

Data and Methodology

All data described here are for 2001 in the city of Vancouver, British Columbia, Canada.[1] The Vancouver Census Metropolitan Area (CMA) is the third largest metropolitan area in Canada, based on population (approximately 2 million people), and the largest metropolitan area in western Canada. In 2001, Vancouver had a population of 546,000. In recent years, Vancouver experienced substantial growth in its resident population: 431,000 in 1986, 472,000 in 1991, and 514,000 in 1996. This high rate of growth is often attributed to the 1986 World Exposition on Transportation and Communication that garnered Vancouver tremendous world attention. This attention was expected to continue because of the 2010 Winter Olympics. With an area of approximately 115 square kilometers, the City of Vancouver has 110 census tracts (CTs) and 990 dissemination areas (DAs), defined by Statistics Canada (2003).

Although Vancouver's crime rate trend decreased from 1991 to 2005, its crime rate remains substantially higher than the national average. In fact, the Vancouver CMA had the highest crime rates among the three largest metropolitan areas in Canada, at 11,367 criminal code offenses per 100,000 persons in 2001, more than doubling the rate found in Toronto (5,381 per 100,000 persons) and almost doubling that in Montreal (6,979 per 100,000 persons). The same relative standing held for the 2001 violent crime rate in the Vancouver CMA (1,058 per 100,000 persons) in comparison to the Toronto CMA (882 per 100,000 persons) and the Montreal CMA (886 per 100,000) but to a lesser degree. The differences in crime rates between these three cities have been decreasing in recent years (Kong 1997; Savoie 2002; Wallace 2003).

Crime Data

The discipline of criminology groups crime into three classifications: property, violent, and nuisance or other crimes (Ellis and Walsh 2000). The utility of the ambient population in crime analysis is best shown using violent crime. Although, in essence, the target in any crime is a person (property and nuisance crimes affect people), violent crimes specifically target people. Given that the ambient population measures the presence of people within a given area, the natural point of departure for considering the ambient population and criminal activity is violent crime. Violent crimes include assault, fighting, hold-ups, homicide, robbery, sexual assault, and stabbing.

The violent crime data used here come from the Vancouver Police Department's Calls for Service Database (VPD-CFS database) generated by its computer-aided dispatch system. The VPD-CFS database is the set of requests for police service made directly to the VPD or through the 911 emergency service and allocated to the VPD. The VPD-CFS database contains information on both the location and the complaint code or description for each call. For each call, there are two codes: the initial complaint code and a complaint code filed by the officer on the scene. The code provided by the officer is always taken to be correct. Although the VPD-CFS database is actually a proxy for actual crime data because not all calls for service represent actual crimes, the primary advantage of the VPD-CFS database is this raw form—these data are not dependent on a criminal charge. It should be noted, however, that few calls for service are subsequently unfounded by the VPD.

Geocoding the point locations of any phenomenon provides the potential for error. Because of the limitations of computer algorithms, geocoding algorithms have issues regarding their accuracy (see Ratcliffe 2001). Even when these limitations are not of particular concern, geocoding algorithms are not always able to locate specific point locations: addresses might be recorded incorrectly in the field, only a 100-block is given, or the street network might be out of date. In an effort to quantify the reliability of geocoded spatial point data, Ratcliffe (2004) ran simulations to determine the minimum acceptable hit rate (perfect match). Above this minimum acceptable hit rate, the researcher should have little concern for bias: This hit, or success, rate is deemed to be 85 percent. A 93 percent hit rate was achieved in the geocoding procedure used in this analysis—11,629 violent crimes were geocoded from an original 12,504 violent crimes. Given this high hit rate in the geocoding, all analyses were performed with little concern for any bias.[2]

The Ambient Population

The ambient population data, LandScan Global Population Database, were provided by Oak Ridge National Laboratory. These data provide a twenty-four-hour estimate of the expected population present at a spatial scale, considered a "quantum leap" in precision from previous databases (Dobson 2003, 163): 30 arc seconds by 30 arc seconds (latitude and longitude). This is approximately one square kilometer depending on the location's distance from the equator. The expected population, being a twenty-four-hour estimate, is an average for the day regardless of the time of year because the calculations incorporate both diurnal and seasonal population movements.

Because of peoples' daily and seasonal routine activities, the population in any city moves from home to work, shop, and engage in recreational activities. The LandScan database considers the relative attractiveness of each square kilometer cell (an urban area attracts more people throughout the day than a rural area, for example) and assigns a value to each cell based on this relative attractiveness. This value is referred to as a *probability coefficient*. The probability coefficient is calculated considering the following: road proximity, slope, land cover, and nighttime lights. Road proximity and density[3] are good predictors of ambient populations, both of which are positively associated with population density. Slope is important because people tend to settle on relatively flat terrain.[4] Because of its suitability for human settlement, land cover (desert, water, wetlands, ice, urban, etc.) is included in the calculation (Dobson et al. 2000). Finally, "[n]ighttime lights are the best available global indicator of where people live, work, and play, and the amount of light emitted is roughly proportional to the number of people" (Dobson 2004, 577).

The probability coefficients are used to redistribute national or subnational census populations—the subnational population, province or state, is preferred (Dobson et al. 2003). As such, the LandScan Global Population Database is an estimate of the population at the one kilometer by one kilometer scale that "integrates diurnal movements and collective travel habits into a single measure" and its "purpose is to distribute populations based on their likely ambient locations integrated over a 24-hour period for typical days, weeks, and seasons" (Dobson et al. 2000, 849–50).

The general methodology for these calculations is the same for all regions of the world, but because some variables do represent one place differently from another there are regional differences in calculations. The presence of nighttime lights, for example, in energy-rich nations needs to be interpreted differently than for energy-poor nations. Although Oak Ridge National Laboratory does undertake verification and validation studies in all regions of the world on a routine basis, and intensely for some regions within the United States (Dobson et al. 2000), at this time there are no known independent verifications of these data.[5]

At times, the differences between ambient and census populations are striking. In a case study of the downtown region of Houston, Texas, the ambient population was estimated as being twenty-six times the value of the census population: 180,000 versus 7,000 (Gold 2003). With such differences in some areas, crime rate calculations could be radically different when using different populations at risk. These extreme differences are not present in Vancouver, but the utility of the ambient population is apparent in what follows.

It should be noted that this population estimate is not without its weaknesses. Two are immediately apparent. First, although Oak Ridge National Laboratory conducts verification studies, without local knowledge in each (urban) area, light might be misinterpreted. This will result in placing a high ambient population count in an area with few actual individuals. Such a phenomenon does occur in Vancouver near the shipping yards that have many bright lights. Second, the ambient population is a twenty-four-hour average. Consequently, some places have population underestimated and others overestimated. Therefore, this population at risk measure still suffers from measurement error. The error is simply less than census resident population counts.

There is one final issue with the use of this ambient population measure in the current analysis: The ambient population data are in a grid format and the remaining data are in, or converted to, census boundary units, CTs, and DAs. To convert the ambient data grids into census boundary units, the grid format is converted into an Environmental Systems Research Institute shapefile. Through the use of an identity overlay with the census boundary units, this shapefile is "cut" into many smaller pieces that are aggregated to the census boundary units using a unique identifier. There is one source of error in this conversion that is common to all area-based analyses: It is assumed that the ambient population is evenly distributed within each grid. This assumption is necessary to avoid double (or triple or more) counting when the grids are cut based on the census boundary units. For example, if a grid is cut into three equal-sized polygons, each of those polygons are assigned one-third of the grid's initial ambient population count.

Census Variables

The independent variables used in the inferential analysis, discussed later, come from Statistics Canada's 2001 Census of Population. Although some of the data used in this analysis are obtained from the "short form" of the Census filled out by all households in Canada, much of the detailed socioeconomic data used in this analysis come from the "long form" of the Census that is completed by only 20 percent of Canadian households. All of these data exclude the institutional population, those living in hospitals, nursing homes, prisons, and other institutions.

The purpose of this analysis is to show differences that arise when using a different population at risk in spatial crime analysis, not to test any particular theory. However, because there are two theories commonly used in the geography of crime literature, social disorganization theory and routine activity theory, variable selection is made with these theories in mind.[6] These variables are organized as follows: population characteristics, socioeconomic status, and dwelling characteristics.[7]

The population characteristic variables are the presence of young populations (males, ages fifteen to twenty-four), never-married persons, lone-parent families, recent immigrants (1991–2001), visible minorities, ethnic heterogeneity,[8] and people who have recently moved. The socioeconomic variables include the population receiving government transfer payments,[9] the population aged twenty years and older without a secondary school diploma, the population aged twenty and older who have obtained a postsecondary education,[10] persons designated as low income,[11] the unemployment rate for those fifteen and older participating in the labor force, and the average household income in thousands of dollars. The dwelling characteristic variables include dwellings constructed before 1961,[12] dwellings in need of major repairs, households spending more than 30 percent of total household income on shelter,[13] owner-occupied dwellings, and average dwelling value. The interested reader is referred to the following previous studies for more information regarding these variables and their theoretical connections: Ackerman (1998), Andresen (2006, 2007), Cahill and Mulligan (2003), Cohen and Cantor (1980), Cohen and Felson (1979), Cohen, Kluegel, and Land (1981), Harries (1974, 1995), Hirschi and Gottfredson (1983), Kelling and Coles (1998), Morenoff, Sampson, and Raudenbush (2001), Sampson (1997), Sampson, Raudenbush, and Earls (1997), Shaw and McKay (1942), and Tseloni et al. (2002).

All of the variables used in the following analysis, with the exception of ethnic heterogeneity, are transformed into natural logarithms; ethnic heterogeneity is an index generated from other variables, so it is not transformed. This transformation is undertaken to ease interpretation of results and the comparison of results at different scales. The resulting estimated parameters in the following statistical analysis are then interpreted as elasticities (see Kennedy 2003). Elasticities measure the degree of sensitivity between the dependent and independent variables: A 1 percent increase in an independent variable leads to a β-percent increase (or decrease if the estimated parameter is negative) in the dependent variable, also transformed into a natural logarithm.

Empirical Methodology

The purpose of this analysis is to show the impact of a different population at risk in crime rate calculations on the spatial analysis of crime. As such, a number of empirical approaches are taken here to show any similarities and differences in different contexts. These different

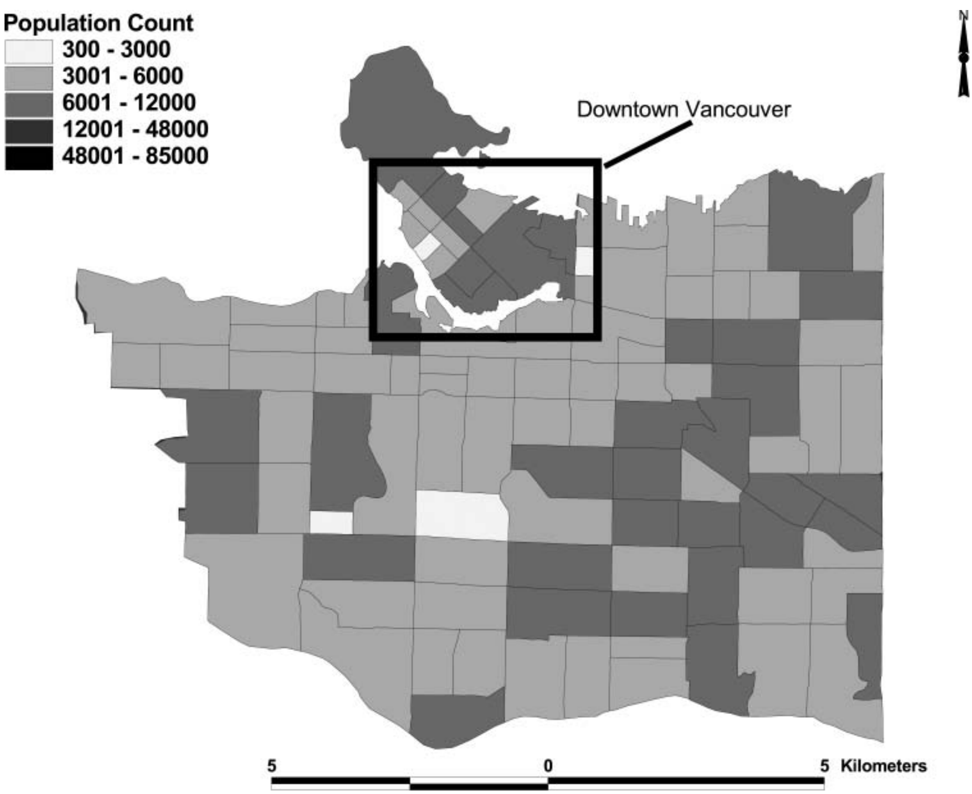

Figure 1 *Resident population, census tracts. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

contexts are a descriptive global analysis, a descriptive local analysis, and a spatial regression analysis.

The descriptive global analysis is in two forms: First, the simple correlations between the different calculated crime rates are reported and discussed; second, the global Moran's I statistics are reported and discussed. This global analysis is followed by the descriptive local analysis using the local Moran's I of Anselin (1995). Because global measures of spatial association might mask local spatial relationships (see Anselin 1995; Getis and Ord 1992, 1996; Ord and Getis 1995), the use of local indicators of spatial association (LISA) might show that the nature of the clustering for each of the calculated crime rates is qualitatively different. For example, at the global level two different crime rates might exhibit the same (or similar) degree of clustering according to Moran's I, but each crime rate might have that clustering occurring in entirely different regions within the study space. Additionally, local indicators of spatial association show where and what type of spatial concentrations are occurring—Anselin, Syabri, and Kho (2006) discussed four categories of local clusters that are mapped and discussed here. Both the global and local Moran's I need to have their spatial neighbors defined. This is done using queen's contiguity.[14] This means that any census unit sharing any boundary, even touching at a corner, with the census unit of concern is considered to be contiguous.

The spatial regression procedure employed is the spatial error model. The spatial error model filters out the spatial effect within both the dependent and independent variables. This specification usually makes it easier to

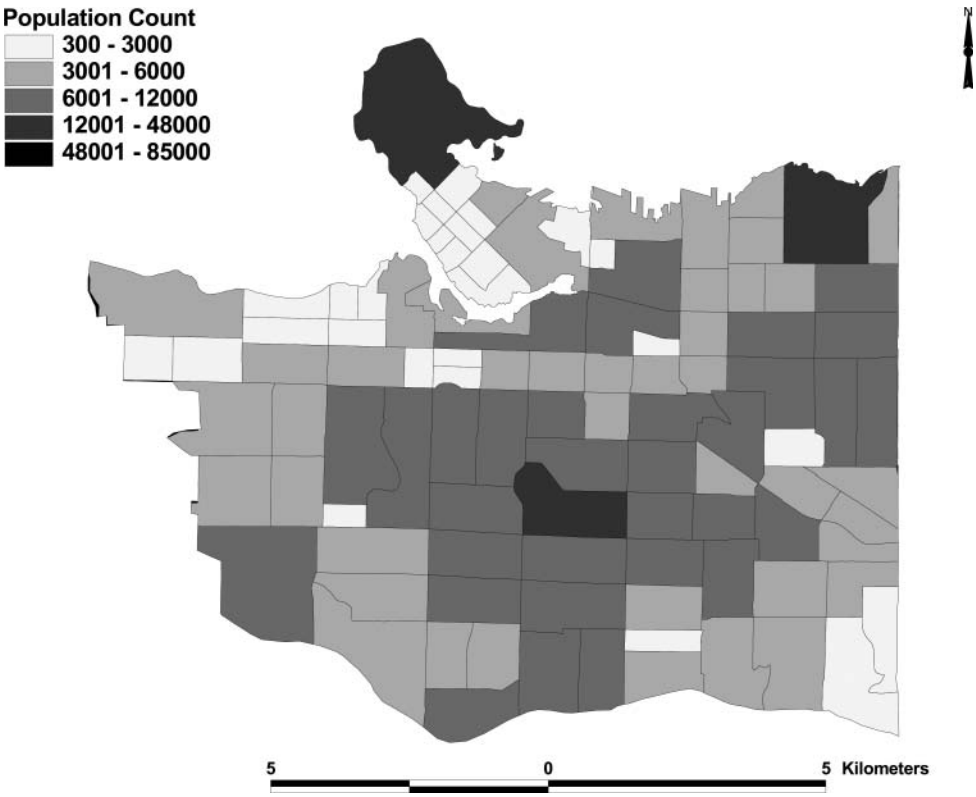

Figure 2 *Ambient population, census tracts. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

reject spatial autocorrelation in the residuals for valid statistical inference (versus the spatial lag model) because the spatial effect is filtered from both sides of the equation. Hence, the spatial error model is used here instead of the spatial lag model. The general functional form of the spatial error model is as follows:

$$y = X\beta + \rho W\varepsilon + u \tag{1}$$

where y is the crime rate, $X\beta$ is the matrix of independent variables and its estimated parameters, W is the spatial weights matrix that captures the spatial association between the different census units, ρ measures the strength of spatial association, ε is shorthand for $y-X\beta$, and u is the independent and identically distributed error.

Queen's contiguity is also used to define the spatial weights matrices used in the spatial regression analysis. However, the "order" of the queen's contiguity varies depending on the presence of spatial autocorrelation. For example, if two census units (a and b) share a boundary, the order of contiguity is 1. With the introduction of a third census unit, c, that borders census unit b, a, and c are then considered contiguous of order 2 because there are two boundaries separating census units a and c. For the global and local Moran's I queen's contiguity is always set to order 1. The final model selection is based on a general-to-specific method: the initial inclusion of all independent variables, removing statistically insignificant variables one at a time. Additionally, joint significance tests are performed to prevent the removal of an important, but highly

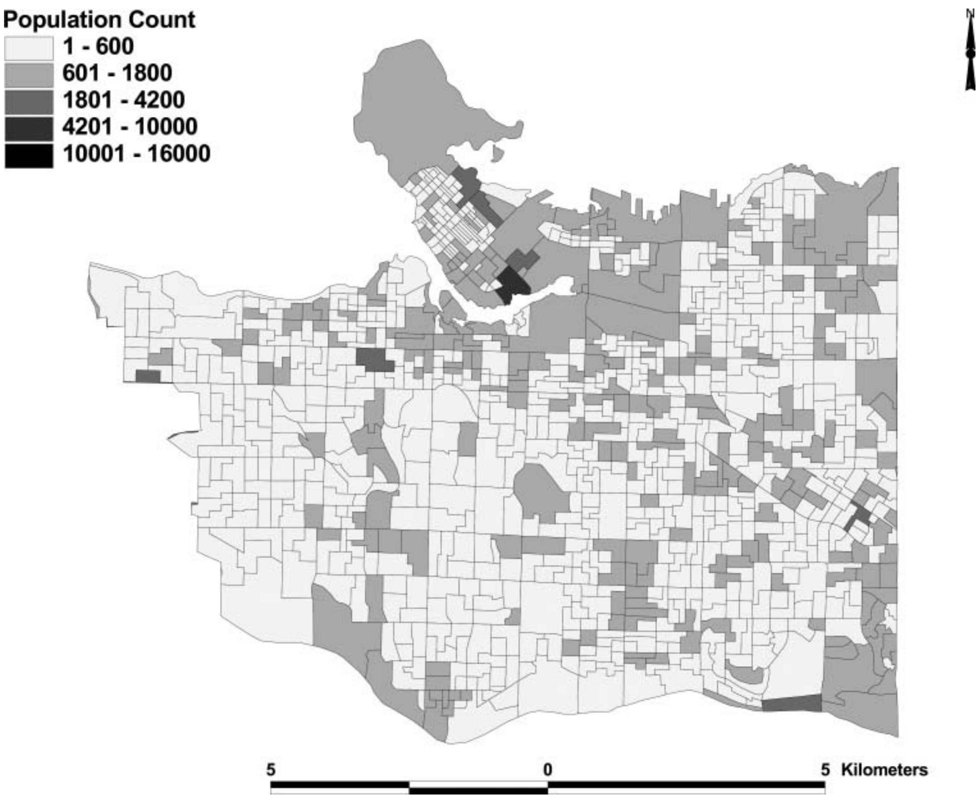

Figure 3 *Resident population, dissemination areas. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

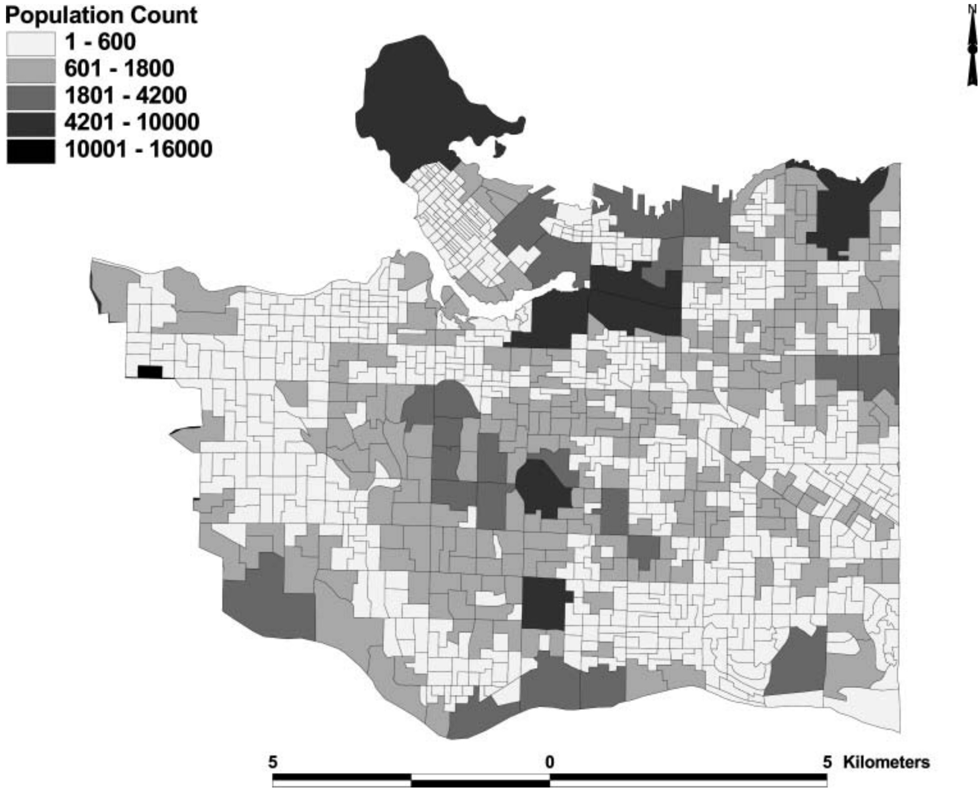

Figure 4 *Ambient population, dissemination areas. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

collinear, variable. For each crime rate model, the analysis begins with all independent variables, estimated parameters are tested for statistical significance (p value $= 0.10$),[15] insignificant variables are removed, and the final models for each crime rate only show the remaining statistically significant independent variables. All estimation is performed using GeoDa 0.9.5i (http://geoda.uiuc.edu), a spatial statistical freeware package developed by Luc Anselin and his coworkers in the Department of Geography at the University of Illinois, Urbana–Champaign; they are now located at Arizona State University.

Results

Before the results from the different analyses are presented, it is instructive to view the different maps representing the resident and ambient population for both census units, as well as the corresponding violent crime rates. As shown in Figures 1 and 3, the resident population has a relatively uniform distribution across the city. This is because census boundaries are designed to have samples and corresponding populations to be as similar as possible. Such a design improves the validity of measurements for comparisons across space. The ambient population, Figures 2 and 4, shows far more clustering in both the CTs and DAs. This is particularly evident, and expected, in the downtown area of Vancouver (the northern peninsula and surrounding areas) and the center of the city, which contains the largest shopping areas outside of the downtown area. This is expected because these are the areas of the city with the greatest population draws during the day.

At the municipal level, the actual resident and ambient population counts are not that different, at 546,000 and 598,000, respectively.

This 10 percent increase in Vancouver's population is representative of its drawing working populations during the day. At the CT and DA levels, however, the differences are far greater. For CTs, the average population does not differ much between resident (5,700) and ambient (6,000) populations. The range in population counts is, however, rather different: 2,600 to 10,000 for the resident population and 350 to 85,000 for the ambient population. This is similar for DAs: Resident and ambient averages are similar, 565 and 600, respectively, but the ranges are vastly different: 225 to 2,500 for the resident population and 1 to 15,000 for the ambient population. Quite clearly, there are significant spatial shifts in Vancouver's population throughout the day when considering these census boundary units.

The resident- and ambient-based violent crime rates, shown in Figures 5 through 8, do not exhibit such systematic changes in their maps. There are, however, noticeable spatial shifts in the distribution of crime. One notable change is that the crime rates in the center of the city and the downtown area fall in the ambient-based maps because of the higher ambient population.

Global Analysis

The first results to note from the global analysis are the bivariate correlations between the resident- and ambient-based violent crime rates for the CTs and DAs, $r = 0.823$ and 0.801, respectively. With such high degrees of correlation, one would expect that there would not be much difference between the results using the resident- or ambient-based violent crime rates. As shown here, however, significant differences are present. It should be noted that

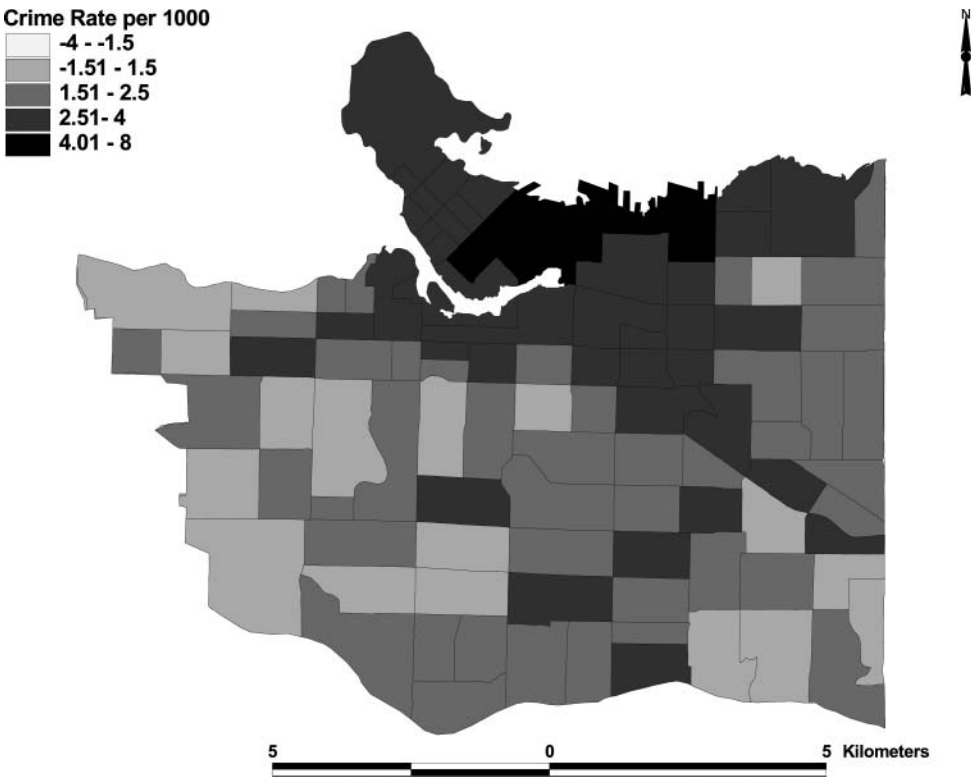

Figure 5 *Resident-based violent crime rate, natural logarithm, census tracts. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

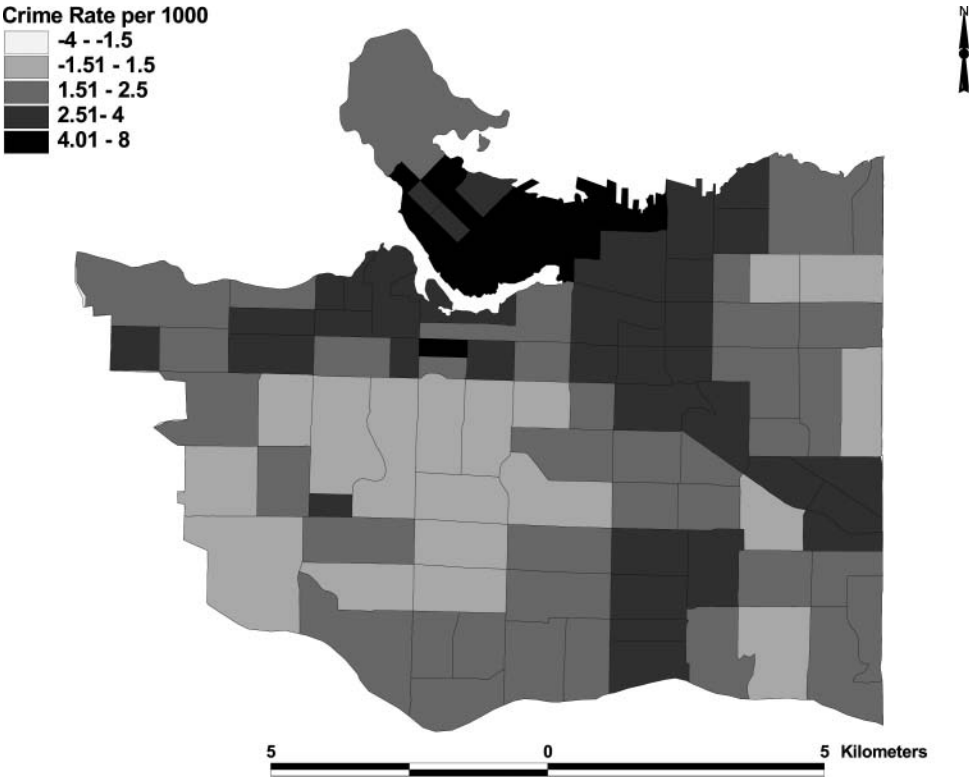

Figure 6 *Ambient-based violent crime rate, natural logarithm, census tracts. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

using Statistics Canada's1996 census boundaries and data that the correlations for CTs and enumeration areas[16] are $r = 0.737$ and $r = 0.095$, respectively. The correlations between the resident- and ambient-based violent crime rates at the CT level are not too dissimilar, but the correlations between the resident- and ambient-based violent crime rates for the enumeration areas or DAs are radically different. Although both correlations are statistically significant at the 1 percent level, the magnitude for 2001 indicates that the gains from using an alternative crime rate might be small (as would be argued by Cohen, Kaufman, and Gottfredson 1985). The magnitude for 1996 indicates that using an alternative crime rate might change the results significantly. Because Vancouver's urban landscape changed very little in those five years (the spatial distribution of the population will not have changed very much), this change in the magnitude of the correlation between the resident- and ambient-based violent crime rates is the result of subtle changes in the census unit boundaries. The purpose of this discussion is not to invoke the modifiable areal unit problem (see Openshaw 1984) but to show that the correlation between the resident- and ambient-based violent crime rates is not necessarily high.

Turning to the global spatial autocorrelation results, Table 1 shows that positive spatial autocorrelation is present for all of the variables. The Moran's I values are statistically significant for all variables except for the resident population of the CTs. Evident from these Moran's I statistics, as discussed earlier, is that the ambient variables have a stronger degree of clustering than the resident variables, particularly for the DAs. It should be noted, however, that the test statistics are low in magnitude. This result should not come as a surprise because commercial land uses in any city (work and shopping) that attract populations during the

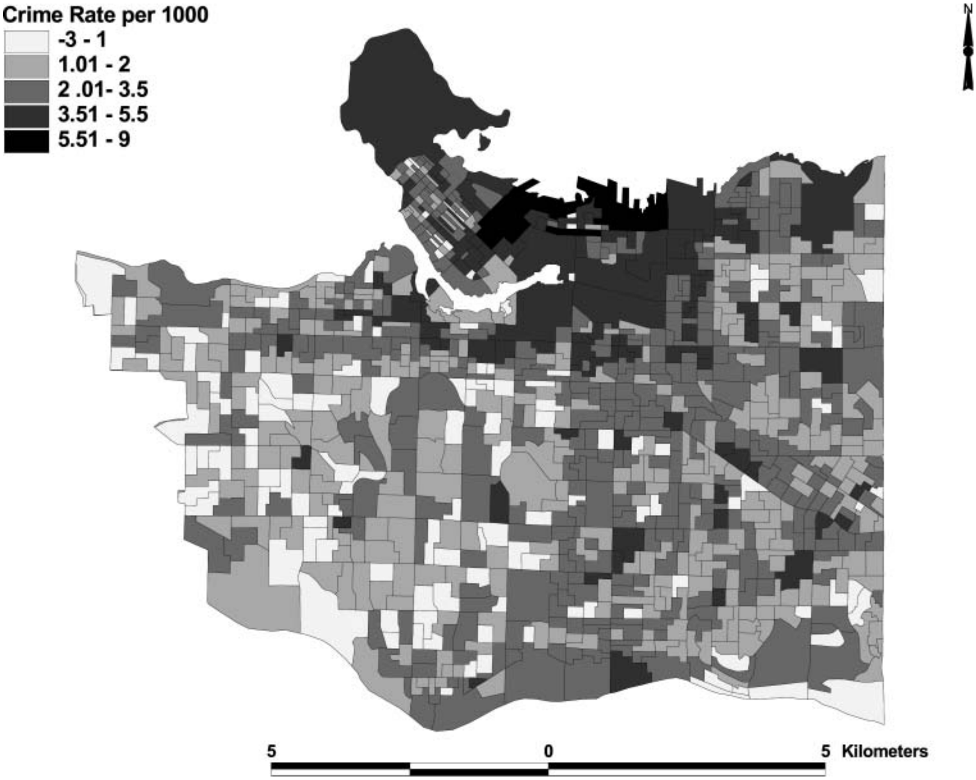

Figure 7 *Resident-based violent crime rate, natural logarithm, dissemination areas. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

day tend to be far more spatially concentrated than the residential land uses. Of course, the resident population is "concentrated" in places such as suburbs, but commercial land use in a city (square footage, for example) is neither randomly nor uniformly distributed across the urban landscape. Given that the DAs exhibit the greatest differences between the resident- and ambient-based violent crime rates, it is expected that the DAs will also exhibit the greatest differences in the analyses that follow.

Table 1 *Global Moran's I*

	Census tracts	Dissemination areas
Resident-based violent crime rate	0.489**	0.341**
Ambient-based violent crime rate	0.515**	0.567**
Residential population	0.001	0.164**
Ambient population	0.054*	0.211**

Note: Crime rates are natural logarithms.
$^{*}p < 0.10$.
$^{**}p < 0.01$.

Local Analysis

The LISA output is mapped and shown for the CTs (Figures 9 and 10) and DAs (Figures 11 and 12). The areas labeled as High–High and Low–Low are census units that have high crime rates that are surrounded by other census units with high crime rates and census units that have low crime rates that are surrounded by other census units with low crime rates, respectively. The areas labeled as High–Low are census units with high crime rates surrounded by other census units with low crime rates, and vice versa for Low–High.

In all the figures there are three areas where clustering occurs: the downtown and

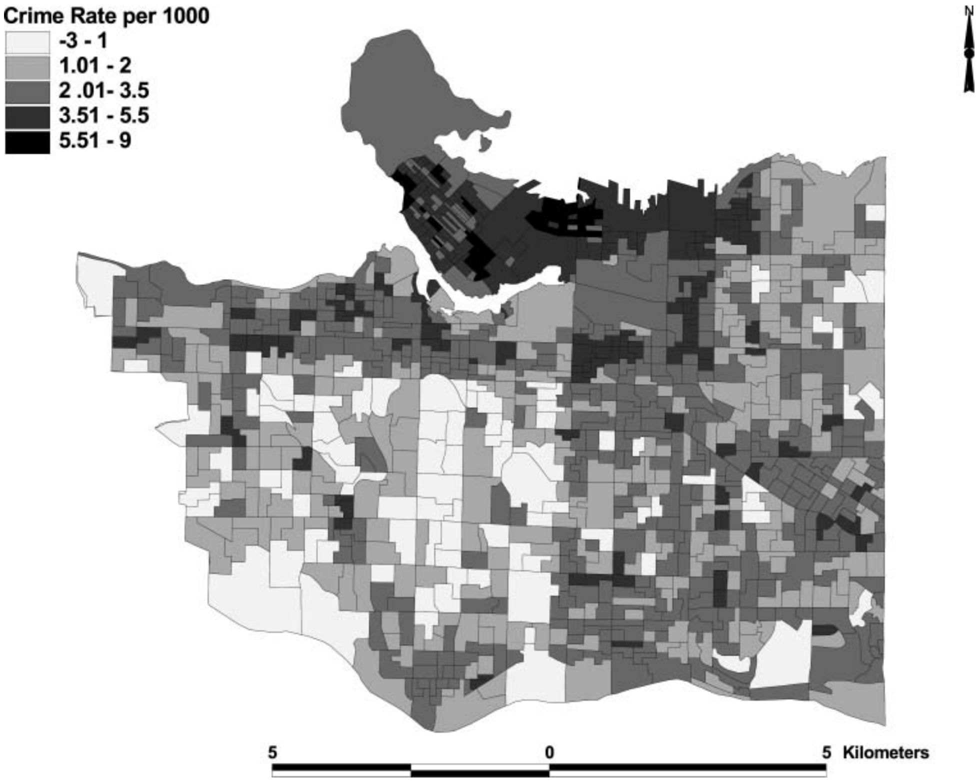

Figure 8 *Ambient-based violent crime rate, natural logarithm, dissemination areas. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

surrounding areas, the western border of the city, and the eastern border of the city. For the CTs in Figures 9 and 10, there is not much change that results from using the resident-based violent crime rate versus the ambient-based violent crime rate. The cluster in the downtown and surrounding areas shifts more into the peninsula, the cluster on the western border moves east, and the eastern cluster is essentially unchanged. The only notable change is the eastward movement of the Low–Low cluster at the western border of the city. The significance of this movement is only because crime is expected to increase as one moves east in Vancouver. Although global statistics might mask local spatial relationships, the similarity of the global Moran's *I* statistics for the CTs is mirrored here with the local analysis.

The LISA maps for the DAs, Figures 11 and 12, show the same general shifting, but the magnitude of the shift is far greater, similar to the global Moran's *I* results. The cluster in the downtown and surrounding areas for the resident-based LISA map shifts almost entirely into the peninsula of downtown and then trails east out of the peninsula. The resident-based LISA map does not exhibit any surprises in that High–High clusters extend as far south as they do, but the ambient-based LISA is more representative of the high crime area of Vancouver. The cluster on the western border of the city shifts eastward (similar to the CT LISA map), but this cluster's shift is further east and grows significantly in size. Although there is some overlap of the Low–Low areas in the western portion of the city, far more areas of Vancouver are considered to be Low–Low using the ambient-based violent crime rate. Referring back to Figures 3 and 4, this should not be a surprise because of the population draw in the central areas of Vancouver. Rather than being areas that are intrinsically high in crime,

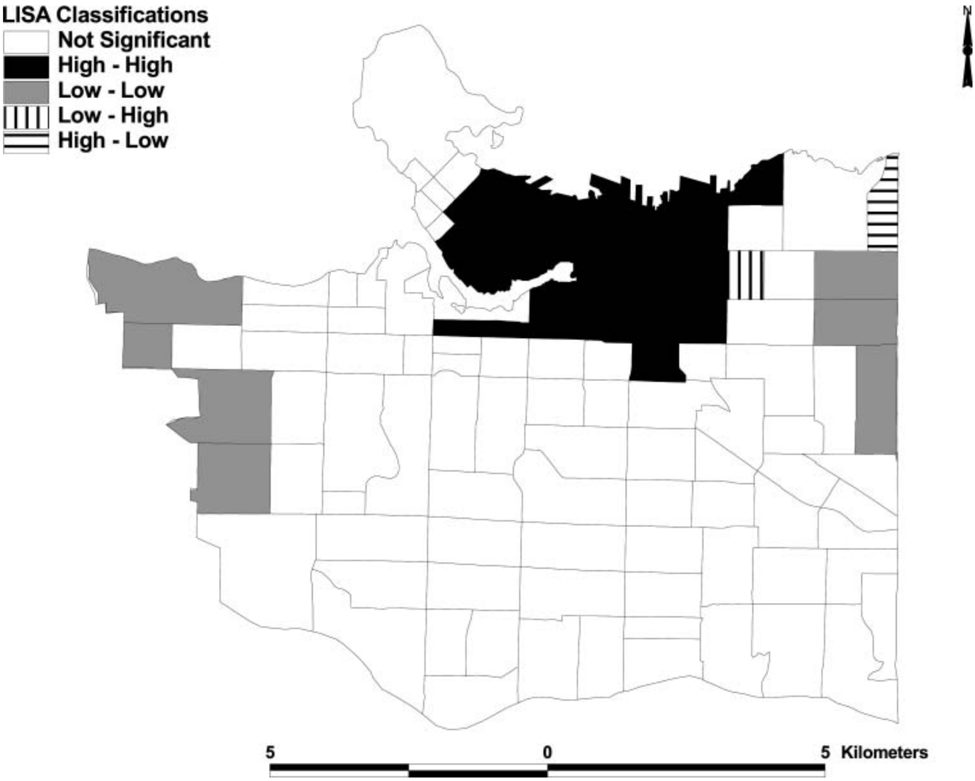

Figure 9 *Resident-based local indicators of spatial association (LISA) map, natural logarithm, census tracts. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

they are areas of many opportunities for crime because of the population attraction, particularly during the day. Needless to say, the alternatively calculated violent crime rate has had a significant effect in the analysis of DAs.

Spatial Regression Analysis

Thus far, the results presented are descriptive in nature. It is, however, important to test whether resident and ambient measures of the population are different in how they relate to sets of conditions representative of those in the criminological literature. Such an inferential analysis is done next using spatial regression. The spatial regression results, shown in Tables 2 and 3, are similar to the previous results in that the CTs are less sensitive to which violent crime rate is being used in the analysis. This does not mean, however, that there are no changes in the results.

As shown in Table 2, the goodness of fit, measured using a pseudo R^2, is notably stronger in the ambient-based output that retains one less variable than the resident-based output, 0.86 versus 0.77. Although a higher pseudo R^2 does not necessarily mean a better model,[17] the ambient-based violent crime rate is better able to fit the data than the resident-based violent crime rate. Curiously, with such a difference in the pseudo R^2, the signs and magnitudes of the variables retained in both models are almost identical—they are identical in three cases. Consequently, the results for the CTs do not indicate a strong need for the use of an alternatively calculated violent crime rate, unless one is more satisfied with a higher pseudo R^2 value. This is not the case for the DAs.

The results for the DA spatial regression model, shown in Table 3, present a much stronger case for use of an alternatively

Table 2 *Spatial error model results, census tracts* Text

	Resident-based violent crime rate	Ambient-based violent crime rate
Resident population density	−1.99***	−0.97***
Ambient population	−1.57***	−1.55***
Ambient population density	1.86***	0.84***
Population turnover	0.76***	0.63**
Ethnic heterogeneity	0.01	0.01**
Government assistance	0.05***	0.06***
Unemployment	0.65***	0.75***
Average family income		0.75*
Low income	−0.74**	−0.56**
Housing affordability	2.04***	1.66***
Average dwelling value	−0.39	−0.71***
Rentals	−0.66***	−0.75***
Major repair	−0.22	−0.27**
Apartments		0.29**
Pseudo R^2	0.78	0.88
Moran's I, residuals	–0.011, p value = 0.42	–0.014, p value = 0.30
Queen's contiguity order	1	1

Note: All estimated coefficients are elasticities, except ethnic heterogeneity.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

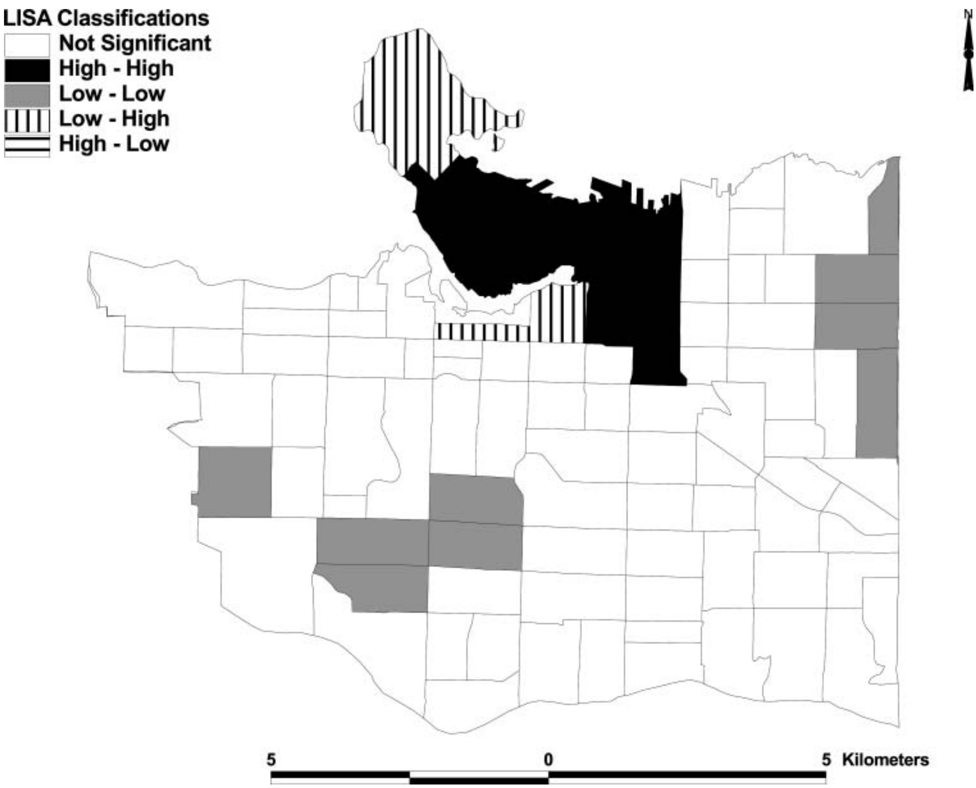

Figure 10 *Ambient-based local indicators of spatial association (LISA) map, natural logarithm, census tracts. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

Table 3 *Spatial error model results, dissemination areas*

	Resident-based violent crime rate	Ambient-based violent crime rate
Resident population	−0.83**	
Resident population density	−0.58**	−0.35**
Ambient population		−0.77**
Ambient population density	0.24*	
Never married	1.03**	1.01**
Postsecondary	−0.46**	−0.47**
Average family income	−0.37**	−0.37**
Low income	−0.14*	−0.15*
Housing affordability	0.21*	0.21*
Average dwelling value	−0.38**	−0.38**
Apartments	0.13*	0.13**
Pseudo R^2	0.38	0.60
Moran's I, residuals	–0.001, p value = 0.21	–0.001, p value = 0.22
Queen's contiguity order	1	1

Note: All estimated coefficients are elasticities.
$^{*}p < 0.05$.
$^{**}p < 0.01$.

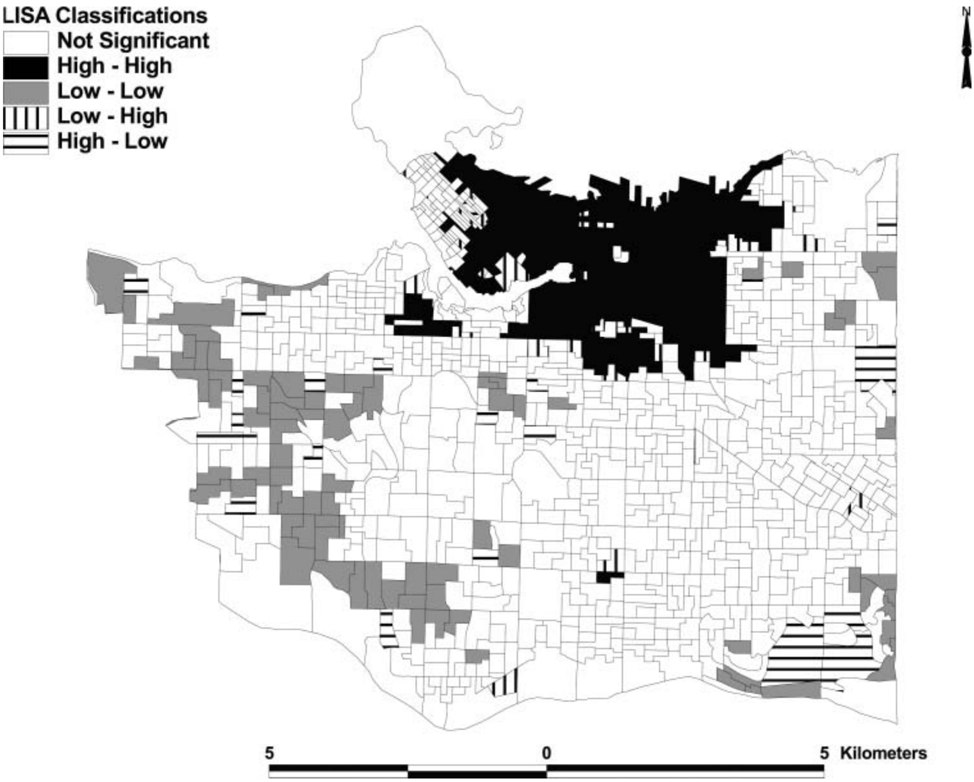

Figure 11 *Resident-based local indicators of spatial association (LISA) map, natural logarithm, dissemination areas. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

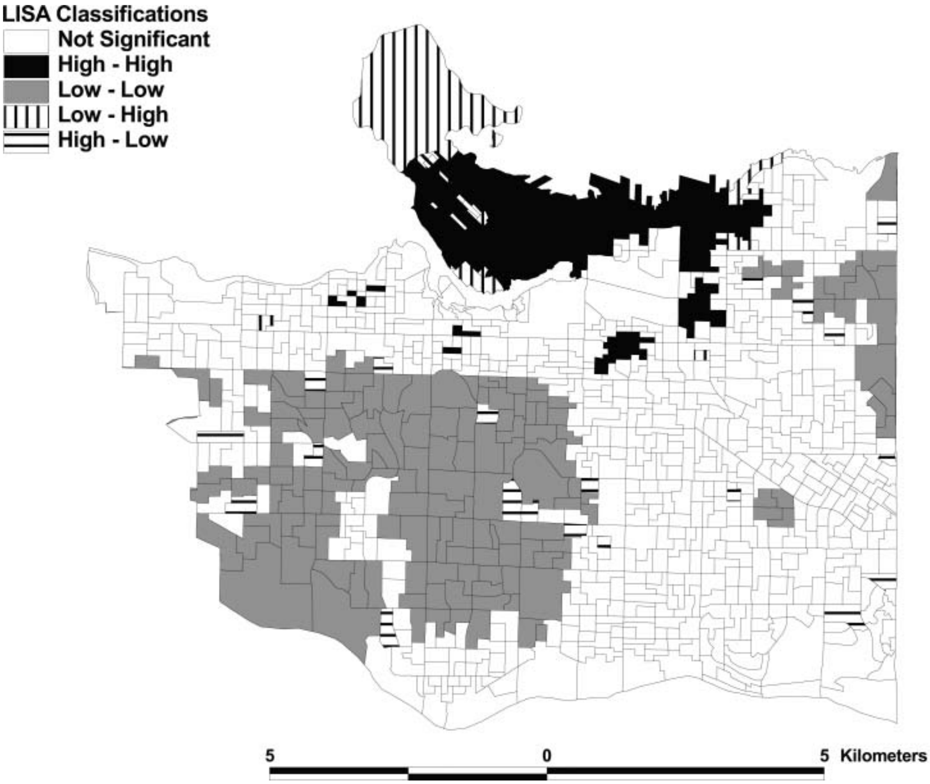

Figure 12 *Ambient-based local indicators of spatial association (LISA) map, natural logarithm, dissemination areas. Source: Oak Ridge National Laboratory (2003) and Statistics Canada (2003).*

calculated violent crime rate. Similar to the CT results, when a variable is present in both outputs its sign is always the same and its magnitude is similar. The primary difference is variable retention and, again, the pseudo R^2. The pseudo R^2 is much stronger in the ambient-based results (0.60 vs. 0.38). Although a pseudo R^2 value of 0.38 is acceptable for a spatial cross-section analysis, it is low enough to be noticed, whereas a pseudo R^2 value of 0.60 is not.

The most notable result from the DA spatial regression output is the consistency with theoretical expectations. As shown in Table 2, there are a number of inconsistencies with theory present in the CT results: Areas with higher income have more crime, areas with low income and more rentals have less crime, and areas with more homes under major repair have less crime. However, these inconsistencies in the results have been resolved for DAs in Table 3, aside from low income. These changing results point to the importance of choosing the appropriate units of analysis in the study of crime. As argued in the "crime at places" literature, smaller units of analysis are critical for understanding the spatial distribution of crime because many "dangerous neighborhoods" only have a few "places" that are truly problematic—there are many safe places in dangerous neighborhoods (Sherman, Gartin, and Buerger 1989; Taylor 1997; Weisburd et al. 2004).

Conclusion

The geography of crime literature asks the following question: Where and why do crimes occur? As initially stated by Boggs (1965) over forty years ago, a great deal of attention is given to the spatial referencing of the criminal events, but far less attention is given to the population at risk. Although at that time and for many years afterward, constraints regarding time and money impeded the development of widely available populations at risk appropriate for different crimes, a data set is now available to serve this purpose for some, but not all, crime classifications—the ambient population produced by Oak Ridge National Laboratory. This article investigates the utility of this new data set in the context of violent crime at two different scales of aggregation and multiple methods of analysis.

It was shown earlier that the use of the ambient population in crime rate calculations has an impact on the results relative to the conventionally calculated violent crime rate using the resident population. This impact, however, is greater for the analysis involving the DAs, although still present for the CTs. In all analyses (global, local, and spatial regression) the general result is the same, but the specifics change: Positive spatial autocorrelation is always present, but the magnitude is different; each local analysis produced three clusters of census units, but those clusters shifted in both space and size; and the spatial regression results are similar when variables are present in the different models, but the goodness of fit and variable retention varied depending on whether the analysis was resident or ambient based. The most surprising factor in the context of these results is that the resident- and ambient-based violent crime rates had high bivariate correlations, all $r > 0.80$.

However, because of the similarity in all of these spatial regression results, one might question whether it is worthwhile to incorporate the ambient population into an inferential analysis. Indeed, Harries (1991) found that although the use of alternative denominators in crime rate calculations did improve measurements, the improvements were most often not worth the costs in both time and money to obtain the alternative denominators. Regardless, there are at least three reasons why the ambient population needs to be considered in spite of these similar spatial regression results.

First, as just stated, other analyses did show qualitatively different results when using the ambient population in crime rate calculations. This is particularly true for the local analysis at the DA level. As such, the utility of the ambient population might be more important in particular contexts. Second, the spatial regression results presented here are for one municipality in one year using particular spatial boundaries. As noted earlier, using 1996 census boundary units alters the statistical relationship between the resident- and ambient-based violent crime rates significantly. If the relevant census variables were available,[18] the differences between the resident- and ambient-based spatial regression outputs would no doubt be far greater than presented here. Finally, the costs in both time and money for incorporating the ambient population data into an analysis are negligible. The data are free for noncommercial use and are easily incorporated into an analysis with a moderate level of training in geographic information system (GIS) software. As such, past concerns regarding resources are now overcome.

These data, however, are not without limitations. The ambient population might not always be the appropriate population at risk. Is the presence of people the appropriate population at risk for homicide, assault, rape, and robbery? Perhaps not, but it is still better than the alternative of the census resident population. This is the key point: Even if the ambient population is not best, it is likely better than the resident population when analyzing violent crime. This reduces measurement error. A further limitation is that the ambient population cannot account for varying differences in risk at different times of the day. This limitation, however, is somewhat mediated with LandScan USA, which includes a daytime population estimate. Unfortunately, as the name suggests, these data are not available globally.

The primary implication of these results is that the choice of the population at risk alters empirical results. It is unlikely that any particular theory will be discarded through the use of an alternative population at risk, but the theoretical and policy variables of interest might be altered (insignificant variables become significant, and vice versa), possibly altering theoretical interpretations and changing the focus of criminal justice policy. This analysis has then shown the importance of being aware of the latest technologies and available data sets. This is particularly true with the spatial analysis of crime and the developments within GIScience technologies.

A further implication for these results is that smaller units of analysis in crime data are in order. Evident in the maps is significant variation in relatively small areas: DAs exhibit much variation within their respective CTs. Although not at such a small scale, this result is consistent with the "crime at places" literature (Sherman, Gartin, and Buerger 1989; Taylor 1997; Weisburd et al. 2004): smaller units of analysis are critical for understanding crime patterns. ■

Notes

[1] The ambient population data, described later, are updated every year, but data for previous years are not made available.

[2] Of course, there is always potential for error arising from converting point data into area data: the modifiable areal unit problem (Openshaw 1984). In an attempt to mediate this error, analyses are undertaken at two levels of areal aggregation for a sensitivity analysis.

[3] "Roads" are interpreted loosely to include roads, rail, water, and air transportation networks.

[4] Flat terrain is also attractive for other land uses such as commercial and industrial. For a small percentage of residences, sloping land is attractive for its views and corresponding property value.

[5] These data are free for noncommercial use, but the algorithm used in the calculation is proprietary and, hence, not available.

[6] These named theories are to be interpreted loosely, not necessarily representing a particular component of social disorganization or routine activity theory. For example, some of the dwelling characteristics might be interpreted as representing broken windows theory, but the author considers broken windows to be an extension of social disorganization theory. Also, because of the nature of the Canadian census, some of the variables are defined differently than in U.S. studies.

[7] In the interest of brevity, descriptive statistics and correlations of the independent variables are not reported here. These tables, however, are available to the interested reader from the author.

[8] Ethnic heterogeneity is measured using an index that ranges in value from zero to one hundred, with zero representing no ethnic mix and one hundred representing a perfectly even mix of ethnic groups: $-\frac{\sum_{i=1}^{n} p_i \, In \, p_i}{In \, n}$, where p_i is the proportion of ethnic group i and n is the number of ethnic groups.

[9] These include employment insurance benefits, old age security benefits, net federal supplements, Canada and Quebec pension plan benefits, the Canada Child Tax benefit, provincial government family allowances, the goods and services tax credit, workers' compensation benefits, social assistance, and provincial or territorial refundable tax credits.

[10] A completed certificate, diploma, or degree.

[11] This is defined as individuals spending 20 percent or more of disposable income compared to the average private household on food, shelter, and clothing.

[12] This represents the perception of physical disorder and decay.

[13] This is a measure of housing affordability, an indicator of socioeconomic disadvantage.

[14] The use of rook's contiguity does not alter the results in any meaningful way. Most often, rook's contiguity only necessitates a higher degree of contiguity to control for spatial autocorrelation. Also, in the present context, areas that might only be considered spatial neighbors at one point (a corner) are still expected to be similar because of the nature of urban landscapes; as such, queen's contiguity is appropriate.

[15] Some might consider a 10 percent significance level too high and prefer a 5 percent significance level. However, for the purposes of this comparative analysis, a more liberal testing methodology is used to err on the side of variable inclusion. If a 5 percent significance level was used in the following analysis, the qualitative results do not change, but the case for using the ambient population in spatial crime analysis becomes even stronger.

[16] *Enumeration area* is the previous name used for the current DA.

[17] It is a relatively easy task to inflate the pseudo R^2 by including more variables, such as a lagged dependent variable. This is easily justified in a temporal analysis because of path dependency and a desire to predict crime rates.

[18] Many of the variables are not available because of low population counts and the corresponding issue of confidentiality.

Literature Cited

Ackerman, W. V. 1998. Socioeconomic correlates of increasing crime rates in smaller communities. *The Professional Geographer* 50:372–87.

Andresen, M. A. 2006. Crime measures and the spatial analysis of criminal activity. *British Journal of Criminology* 46:258–85.

———. 2007. Location quotients, ambient populations, and the spatial analysis of crime in Vancouver, Canada. *Environment and Planning A* 39: 2423–44.

Andresen, M. A., and G. W. Jenion. 2008. Crime prevention and the science of where people are. *Criminal Justice Policy Review* 19:164–180.

———. 2010. Ambient populations and the calculation of crime rates and risk. *Security Journal* 23:114–33.

Anselin, L. 1995. Local indicators of spatial association—LISA. *Geographical Analysis* 27:93–115.

Anselin, L., I. Syabri, and Y. Kho. 2006. GeoDa: An introduction to spatial data analysis. *Geographical Analysis* 38:5–22

Boggs, S. L. 1965. Urban crime patterns. *American Sociological Review* 30:899–908.

Brantingham, P. J., and P. L. Brantingham. 1991. Introduction: The dimensions of crime. In *Environmental criminology*, ed. P. J. Brantingham and P. L. Brantingham, 7–26. Prospect Heights, IL: Waveland Press.

Cahill, M. E., and G. F. Mulligan. 2003. The determinants of crime in Tucson, Arizona. *Urban Geography* 24:582–610.

Cohen, L. E., and D. Cantor. 1980. The determinants of larceny: An empirical and theoretical study. *Journal of Research in Crime and Delinquency* 17:146–59.

Cohen, L. E., and M. Felson. 1979. Social change and crime rates. *American Sociological Review* 44:588–608.

Cohen, L. E., R. L. Kaufman, and M. R. Gottfredson. 1985. Risk-based crime statistics: A forecasting comparison for burglary and auto theft. *Journal of Criminal Justice* 13:445–57.

Cohen, L. E., J. R. Kluegel, and K. C. Land. 1981. Social inequality and predatory criminal victimization: An exposition and test of a formal theory. *American Sociological Review* 46:505–24.

Dobson, J. E. 2003. Estimating populations at risk. In *The geographical dimensions of terrorism*, ed. S. L. Cutter, D. B. Richardson, and T. J. Wilbanks, 161–67. London and New York: Routledge.

———. 2004. The GIS revolution in science and society. In *Geography and technology*, ed. S. D. Brunn, S. L. Cutter, and J. W. Harrington Jr., 573–87. Dordrecht, The Netherlands: Kluwer Academic.

Dobson, J. E., E. A. Bright, P. R. Coleman, and B. L. Bhaduri. 2003. LandScan: A global population database for estimating populations at risk. In *Remotely sensed cities*, ed. V. Mesev, 267–79. London and New York: Taylor and Francis.

Dobson, J. E., E. A. Bright, P. R. Coleman, R. C. Durfee, and B. A. Worley. 2000. LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing* 66:849–57.

Ellis, L., and A. Walsh. 2000. *Criminology: A global perspective.* Boston: Allyn & Bacon.

Getis, A., and J. K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24:189–206.

———. 1996. Spatial analysis: Modelling in a GIS environment. In *Local spatial statistics: An overview*, ed. P. Longley and M. Batty, 261–77. London: Geoinformation International.

Gold, S. S. 2003. Watch us move: Homeland security requires a new kind of population map. *Popular Science* January:40.

Harries, K. D. 1974. *The geography of crime and justice.* New York: McGraw-Hill.

———. 1991. Alternative denominators in conventional crime rates. In *Environmental criminology*, ed. P. J. Brantingham and P. L. Brantingham, 147–65. Prospect Heights, IL: Waveland.

———. 1995. The ecology of homicide and assault: Baltimore City and County, 1989–1991. *Studies in Crime and Crime Prevention* 4:44–60.

Hirschi, T., and M. Gottfredson. 1983. Age and the explanation of crime. *American Journal of Sociology* 89:552–84.

Kelling, G., and C. Coles. 1998. *Fixing broken windows: Restoring order and reducing crime in our communities.* New York: Touchstone.

Kennedy, P. 2003. *A guide to econometrics.* Fifth ed. Malden, MA: Blackwell.

Kong, R. 1997. *Canadian crime statistics, 1996.* Ottawa, ON: Statistics Canada, Canadian Centre for Justice Statistics.

Morenoff, J. D., R. J. Sampson, and S. W. Raudenbush. 2001. Neighborhood inequality, collective efficacy and the spatial dynamics of urban violence. Research Report No. 00-451, Population Studies Centre, Institute for Social Research. University of Michigan, Ann Arbor, MI.

Oak Ridge National Laboratory. 2003. LandScan global population database. Oak Ridge, TN: Oak Ridge National Laboratory. http://www.ornl.gov/sci/landscan/ (last accessed 21 May 2006).

Openshaw, S. 1984. *The modifiable areal unit problem.* Concepts and Techniques in Modern Geography 38. Norwich, UK: Geo Books.

Ord, J. K., and A. Getis. 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis* 27:286–306.

Ratcliffe, J. H. 2001. On the accuracy of TIGER type geocoded address data in relation to cadastral and census areal units. *International Journal of Geographical Information Science* 15:473–85.

———. 2004. Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science* 18:61–72.

Sampson, R. J. 1997. The embeddedness of child and adolescent development: A community-level perspective on urban violence. In *Violence and childhood in the inner city*, ed. J. McCord, 31–77. Cambridge, UK: Cambridge University Press.

Sampson, R. J., S. W. Raudenbush, and F. Earls. 1997. Neighbourhoods and violent crime: A multilevel study of collective efficacy. *Science* 277:918–24.

Savoie, J. 2002. *Crime statistics in Canada, 2001.* Ottawa. ON: Statistics Canada, Canadian Centre for Justice Statistics.

Shaw, C. R., and H. D. McKay. 1942. *Juvenile delinquency and urban areas: A study of rates of delinquency in relation to differential characteristics of local communities in American cities.* Chicago: University of Chicago Press.

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

Statistics Canada. 2003. *2001 census of population*. Ottawa, ON: Statistics Canada.

Taylor, R. B. 1997. Social order and disorder of street blocks and neighborhoods: Ecology, microecology, and the systemic model of social disorganization. *Journal of Research in Crime and Delinquency* 34:113–55.

Tseloni, A., D. R. Osborn, A. Trickett, and K. Rose. 2002. Modelling property crime using the British Crime Survey: What have we learnt? *British Journal of Criminology* 42:109–28.

Wallace, M. 2003. *Crime statistics in Canada, 2002.* Ottawa, ON: Statistics Canada, Canadian Centre for Justice Statistics.

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

MARTIN A. ANDRESEN is an Assistant Professor in the School of Criminology and Member in the Institute for Canadian Urban Research Studies at Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6 Canada. E-mail: andresen@sfu.ca. His research interests include environmental criminology, the geography of crime, and spatial crime analysis.