Chapter 10
Journey to Crime Estimation

The *Journey to Crime* (Jtc) routine is a distance-based method which makes estimates about the likely residential location of a serial offender. It is an application of *location theory*, a framework for identifying optimal locations from a distribution of markets, supply characteristics, prices, and events. The following discussion gives some background to the technique. Those wishing to skip this part can go to page 10-19 for the specifics of the Jtc routine.

Location Theory

Location theory is concerned with one of the central issues in geography. This theory attempts to find an optimal location for any particular distribution of activities, population, or events over a region (Haggett, Cliff and Frey, 1977; Krueckeberg and Silvers, 1974; Stopher and Meyburg, 1975; Oppenheim, 1980, Ch. 4; Bossard, 1993). In classic location theory, economic resources were allocated in relation to idealized representations (Anselin and Madden, 1990). Thus, von Thünen (1826) analyzed the distribution of agricultural land as a function of the accessibility to a single population center (which would be more expensive towards the center), the value of the product produced (which would vary by crop), and transportation costs (which would be more expensive farther from the center). In order to maximize profit and minimize costs, a distribution of agricultural land uses (or crop areas) emerges flowing out from the population center as a series of concentric rings. Weber (1909) analyzed the distribution of industrial locations as a function of the volume of materials to be shipped, the distance that the goods had to be shipped, and the unit distance cost of shipping; consequently, industries become located in particular concentric zones around a central city. Burgess (1925) analyzed the distribution of urban land uses in Chicago and described concentric zones of both industrial and residential uses. Their theory formed the backdrop for early studies on the ecology of criminal behavior and gangs (Thrasher, 1927; Shaw, 1929).

In more modern use, the location of persons with a certain need or behavior (the 'demand' side) is identified on a spatial plane and places are selected as to maximize value and minimize travel costs. For example, for a consumer faced with two retail shops selling the same product, one being closer but more expensive while the other being farther but less expensive, the consumer has to trade off the value to be gained against the increased travel time required. In designing facilities or places of attraction (the 'supply' side), the distance between each possible facility location and the location of the relevant population is compared to the cost of locating near the facility. For example, given a distribution of consumers and their propensity to spend, such a theory attempts to locate the optimal placement of retail stores, or, given the distribution of patients, the theory attempts to locate the optimal placement of medical facilities.

Predicting Locations from a Distribution

One can also reverse the logic. Given the distribution of demand, the theory could be applied to estimate a central location from which travel distance or time is minimized. One of the earliest uses of this logic was that of John Snow, who was interested in the causes of cholera in the mid-19th century (Cliff and Haggett, 1988). He postulated the theory that water was the major vector transmitting the cholera bacteria. After investigating water sources in the London metropolitan area and concluding that there was a relationship between contaminated water and cholera cases, he was able to confirm his theory by an outbreak of cholera cases in the Soho district. By plotting the distribution of the cases and looking for water sources in the center of the distribution (essentially, the center of minimum distance; see chapter 4), he found a well on Broad Street that was, in fact, contaminated by seepage from nearby sewers. The well was closed and the epidemic in Soho receded. Incidently, in plotting the incidents on a map and looking for the center of the distribution, Snow applied the same logic that had been followed by the London Metropolitan Police Department who had developed the famous 'pin' map in the 1820s.

Theoretically, there is an optimal solution that minimizes the distance between demand and supply (Rushton, 1979). However, computationally, it is an almost impossible task to define, requiring the enumeration of every possible combination. Consequently in practice, approximate, though sub-optimal, solutions are obtained through a variety of methods (Everett, 1974, Ch. 4).

Travel Demand Modeling

A sub-set of location theory models the travel behavior of individuals. It actually is the converse. If location theory attempts to allocate places or sites in relation to both a supply-side and demand-side, travel demand theory attempts to model how individuals travel between places, given a particular constellation of them. One concept that has been frequently used for this purpose is that of the *gravity function*, an application of Newton's fundamental law of attraction (Oppenheim, 1980). In the original Newtonian formulation, the attraction, F, between two bodies of respective masses M_1 and M_2, separated by a distance D, will be equal to

$$F = g \frac{M_1 M_2}{D^2} \tag{10.1}$$

where g is a constant or scaling factor which ensures that the equation is balanced in terms of the measurement units (Oppenheim, 1980). As we all know, of course, g is the gravitational constant in the Newtonian formulation. The numerator of the function is the *attraction* term (or, alternatively, the attraction of M_2 for M_1) while the denominator of the equation, d^2, indicates that the attraction between the two bodies falls off as a function of their *squared* distance. It is an *impedance* term.

Social Applications of the Gravity Concept

The gravity model has been the basis of many applications to human societies and has been applied to social interactions since the 19th century. Ravenstein (1895) and Andersson (1897) applied the concept to the analysis of migration by arguing that the tendency to migrate between regions is inversely proportional to the squared distance between the regions. Reilly's 'law of retail gravitation' (1929) applied the Newtonian gravity model directly and suggested that retail travel between two centers would be proportional to the product of their populations and inversely proportional to the square of the distance separating them:

$$T_{ij} = \alpha \frac{P_i \, P_j}{D_{ij}^2} \qquad (10.2)$$

where T_{ij} is the interaction between centers *i* and *j*, P_i and P_j are the respective populations, D_{ij} is the distance between them raised to the second power and α is a balancing constant. In the model, the initial population, P_i, is called a *production* while the second population, P_j, is called an *attraction*.

Stewart (1950) and Zipf (1949) applied the concept to a wide variety of phenomena (migration, freight traffic, exchange of information) using a simplified form of the gravity equation

$$T_{ij} = \alpha \frac{P_i \, P_j}{D_{ij}} \qquad (10.3)$$

where the terms are as in equation 10.2 but the exponent of distance is only 1. In doing so, they basically linked location theory with travel behavior theory. Given a particular pattern of interaction for any type of goods, service or human activity, an optimal location of facilities should be solvable.

In the Stewart/Zipf framework, the two P's were both population sizes and, therefore, their sums had to be equal. However, in modern use, it's not necessary for the productions and attractions to be identical units (e.g., P_i could be population while P_j could be employment).

The total volume of productions (trips) from a single location, i, is estimated by summing over all destination locations, j:

$$T_i = K \quad P_i \sum_j (P_j / D_{ij}) \qquad (10.4)$$

Over time, the concept has been generalized and applied to many different types of travel behavior. For example, Huff (1963) applied the concept to retail trade between zones in an urban area using the general form of

$$T_{ij} = \alpha \frac{A_j^{\beta}}{D_{ij}^{\lambda}} \qquad (10.5)$$

where T_{ij} is the number of purchases in location *j* by residents of location *i*, A_j is the attractiveness of zone j (e.g., square footage of retail space), D_{ij} is the distance between zones *i* and *j*, β is the exponent of S_j, and λ is the exponent of distance, and α is a constant (Bossard, 1993). $D_{ij}^{-\lambda}$ is sometimes called an *inverse distance* function. This is a *single constraint* model in that only the attractiveness of a commercial zone is constrained, that is the sum of all attractions for j must equal the total attraction in the region.

Again, it can be generalized to all zones by, first, estimating the total trips generated from one zone, i, to another zone, j,

$$T_{ij} = \alpha \frac{P_i^{\rho} A_j^{\beta}}{D_{ij}^{\lambda}} \qquad (10.6)$$

where T_{ij} is the interaction between two locations (or zones), P_i is productions of trips from location/zone i, A_j is the attractiveness of location/zone j, D_{ij} is the distance between zones i and j, β is the exponent of S_j, ρ is the exponent of H_i, λ is the exponent of distance, and α is a constant.

Second, the total number of trips generated by a location, i, to all destinations is obtained by summing over all destination locations, j:

$$T_i = \alpha P_i^{\rho} \sum_j (A_j^{\beta}/D_{ij}^{\lambda}) \qquad (10.7)$$

This differs from the traditional gravity function by allowing the exponents of the production from location i, the attraction from location j, and the distance between zones to vary. Typically, these exponents are calibrated on a known sample before being applied to a forecast sample and the locations are usually measured by zones. Thus, retailers in deciding on the location of a new store can use this type of model to choose a site location to optimize travel behavior of patrons; they will, typically, obtain data on actual shopping trips by customers and then calibrate the model on the data, estimating the exponents of attraction and distance. The model can then be used to predict future shopping trips if a facility is built at a particular location.

This type of function is called a *double constraint* model because the balancing constant, K, has to be constrained by the number of units in both the origin and destination locations; that is, the sum of P_i over all locations must be equal to the total number of productions while the sum of P_2 over all locations must be equal to the total number of attractions. Adjustments are usually required to have the sum of individual productions and attractions equal the totals (usually estimated independently).

The equation can be generalized to other types of trips and different metrics can be substituted for distance, such as travel time, effort, or cost (Isard, 1960). For example, for commuting trips, usually employment is used for attractions, frequently sub-divided into retail and non-retail employment. In addition, for productions, median household income or car ownership percentage is used as an additional production variable. Equation 10.7 can be generalized to include any type of production or attraction variable (10.8 and 10.9):

$$T_{ij} = \alpha_1 P_i^{\rho} \alpha_2 A_j^{\beta}/D_{ij}^{\lambda} \tag{10.8}$$

$$T_i = \alpha_1 P_i^{\rho} \Sigma (\alpha_2 A_j^{\beta}/D_{ij}^{\lambda}) \tag{10.9}$$

where T_{ij} is the number of trips produced by location i that travel to location j, P_i is either a single variable associated with trips produced from a zone or the cross-product of two or more variables associated with trips produced from a zone, A_j is either a single variable associated with trips attracted to a zone or the cross-product of two or more variables associated with trips attracted to a zone, D_{ij} is either the distance between two locations or another variable measuring travel effort (e.g., travel time, travel cost), ρ, β, and λ are exponents of the respective terms, α_1 is a constant associated with the productions to ensure that the sum of trips produced by all zones equals the total number of trips for the region (usually estimated independently), and α_2 is a constant associated with the attractions to ensure that the sum of trips attracted to all zones equals the total number of trips for the region. Without having two constants in the equation, usually conflicting estimates of K will be obtained by balancing the equation against productions or attractions. The summation over all destination locations, j (equation 10.9), produces the total number of trips from zone i.

Intervening Opportunities

Stouffer (1940) modified the simple gravity function by arguing that the attraction between two locations was a function not only of the characteristics of the relative attractions of two locations, but of intervening opportunities between the locations. His hypothesis "..assumes that there is no necessary relationship between mobility and distance... that the number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities"(Stouffer, 1940, p. 846). This model was used in the 1940s to explain interstate and intercounty migration (Bright and Thomas, 1941; Isbell, 1944; Isard, 1979). Using the gravity type formulation, we can write this as:

$$T_{ji} = \alpha \frac{A_j^{\beta}}{\Sigma(A_k^{\xi})\ D_{ij}^{\lambda}} \tag{10.10}$$

where T_{ji} is the attraction of location j by residents of location i, A_j is the attractiveness of zone j, A_k is the attractiveness of all other locations that are *intermediate* in distance between locations i and j, D_{ij} is the distance between zones *i* and *j*, β is the exponent of S_j, ξ is the exponent of S_k, λ is the exponent of distance, and α is a constant. While the intervening opportunities are implicit in equation 10.5 in the exponents, β and λ, and coefficient, K, equation 10.10 makes the intervening opportunities explicit. The importance of the concept is that the interaction between two locations becomes a complex function of the spatial environment of nearby areas and not just of the two locations.

Urban Transportation Modeling

This type of model is incorporated as a formal step in the urban transportation planning process, implemented by most regional planning organizations in the United States and elsewhere (Stopher and Meyburg, 1975; Krueckeberg and Silvers, 1974; Field and MacGregor, 1987). The step, called *trip distribution*, is linked to a five step model. First, data are obtained on travel behavior for a variety of trip purposes. This is usually done by sampling households and asking each member to keep a travel diary documenting all their trips over a two or three day period. Trips are aggregated by individuals and by households. Frequently, trips by different purposes are separated. Second, the volume of trips produced by and attracted to zones (called traffic analysis zones) is estimated, usually on the basis of the number of households in the zone and some indicator of income or private vehicle ownership. Third, trips produced by each zone are distributed to every other zone usually using a gravity-type function (equation 10.9). That is, the number of trips produced by each origin zone and ending in each destination zone is estimated by a gravity model. The distribution is based on trip productions, trip attractions, and travel 'resistance' (measured by travel distance or travel time). Fourth, zone-to-zone trips are allocated by mode of travel (car, bus, walking, etc); and, fifth, trips are assigned to particular routes by travel mode (i.e., bus trips follow different routes than private vehicle trips). The advantage of this process is that trips are allocated according to origins, destinations, distances (or travel times), modes of travel and routes. Since all zones are modeled simultaneously, all intermediate destinations (i.e., intervening opportunities) are incorporated into the model. Chapters 11-17 present a crime travel demand model.

Alternative Distance Decay Functions

One of the problems with the traditional gravity formulation is in the measurement of travel resistance, either distance or time. For locations separated by sizeable distances in space, the gravity formulation can work properly. However, as the distance between locations decreases, the denominator approaches infinity. Consequently, an alternative expression for the interaction has been proposed which uses the negative exponential function (Hägerstrand, 1957; Wilson, 1970).

$$A_{ji} = S_j^{\beta} e^{(-\alpha D_{ij})} \qquad (10.11)$$

where A_{ji} is the attraction of location j for residents of location i, S_j is the attractiveness of location j, D_{ij} is the distance between locations *i* and *j*, β is the exponent of S_j, *e* is the base of the natural logarithm (i.e., 2.7183...), and α is an empirically-derived exponent. Sometimes known as *entropy maximization*, the latter part of the equation includes a negative exponential function which has a maximum value of 1 (i.e., $e^{-0} = 1$). This has the advantage of making the equation more stable for interactions between locations that are close together. For example, Cliff and Haggett (1988) used a negative exponential gravity-type model to describe the diffusion of measles into the United States from Canada and Mexico. It has also been argued that the negative exponential function generally gives a better fit to urban travel patterns, particularly by automobile (Foot, 1981; Bossard, 1993; NCHRP, 1995).

Other functions have also be used to describe the distance decay - negative linear, normal distribution, lognormal distribution, quadratic, Pareto function, square root exponential, and so forth (Haggett and Arnold, 1965; Taylor, 1970; Eldridge and Jones, 1991). Later in the chapter, we will explore several different mathematical formulations for describing the distance decay. One, in fact, does not need to use a mathematical function at all, but could empirically describe the distance decay from a large data set and utilize the described values for predictions. The use of mathematical functions has evolved out of both the Newtonian tradition of gravity as well as various location theories which used the gravity function. A mathematical function makes sense under two conditions: 1) if travel is uniform in all directions; and 2) as an approximation if there is inadequate data from which to calibrate an empirical function. The first assumption is usually wrong since physical geography (i.e., oceans, rivers, mountains) as well as asymmetrical street networks make travel easier in some directions than others. As we shall see below, the distance decay is quite irregular for journey to crime trips and would be better described by an empirical, rather than mathematical function.

In short, there is a long history of research on both the location of places as well as the likelihood of interaction between these places, whether the interaction is freight movement, land prices or individual travel behavior. The gravity model and variations on it have been used to describe the interactions between these locations.

Travel Behavior of Criminals

Journey to Crime Trips

The application of travel behavior theory to crime has a sizeable history as well. The analysis of distance for journey to crime trips was applied in the 1930s by White (1932), who noted that property crime offenders generally traveled farther distances than offenders committing crimes against people, and by Lottier (1938), who analyzed the ratio of chain store burglaries to the number of chain stores by zone in Detroit. Turner (1969) analyzed delinquency behavior by a distance decay travel function showing how more crime trips tend to be close to the offender's home with the frequency dropping off with distance. Phillips (1980) is, apparently, the first to use the term *journey to crime* is describing the travel distances that offenders make though Harries (1980) noted that the average distance traveled has evolved by that time into an analogy with the journey to work statistic.

Rhodes and Conly (1981) expanded on the concept of a *criminal commute* and showed how robbery, burglary and rape patterns in the District of Columbia followed a distance decay pattern. LeBeau (1987a) analyzed travel distances of rape offenders in San Diego by victim-offender relationships and by method of approach. Boggs (1965) applied the intervening opportunities model in analyzing the distribution of crimes by area in relation to the distribution of offenders. Other empirical descriptions of journey to crime distances and other travel behavior parameters have been studied by Blumin (1973), Curtis (1974), Repetto (1974), Pyle (1974), Capone and Nichols 1975), Rengert (1975), Smith (1976), LeBeau (1987b), and Canter and Larkin (1993). It has generally been accepted that property crime trips are longer than personal crime trips (LeBeau, 1987a), though exceptions have been noted (Turner, 1969). Also, it would be expected that average trip distances will vary by a number of factors: crime type; method of operation; time of day; and, even, the value of the property realized (Capone and Nichols, 1975).

Modeling the Offender Search Area

Conceptual work on the type of model have been made by Brantingham and Brantingham (1981) who analyzed the *geometry of crime* and conceptualized a criminal search area, a geographical area modified by the spatial distribution of potential offenders and potential targets, the awareness spaces of potential offenders, and the exchange of information between potential offenders. In this sense, their formulation is similar to that of Stouffer (1940), who described intervening opportunities, though their's is a behavioral framework. An important concept developed by the Brantingham's is that of decreased criminal activity near to an offender's home base, a sort of a safety area around their near neighborhood. Presumably, offenders, particularly those committing property crimes, go a little way from their home base so as to decrease the likelihood that they will get caught. This was noted by Turner (1969) in his study of delinquency in Philadelphia. Thus, the Brantingham's postulated that there would be a small safety area (or 'buffer' zone) of relatively little offender activity near to the offender's base location; beyond that zone, however, they postulated that the number of crime trips would decrease according to a distance decay model (the exact mathematical form was never specified, however).

Crime trips may not even begin at an offender's residence. Routine activity theory (Cohen and Felson, 1979; 1981) suggests that crime opportunities appear in the activities of everyday life. The routine patterns of work, shopping, and leisure affect the convergence in time and place of would be offenders, suitable targets, and absence of guardians. Many crimes may occur while an offender is traveling from one activity to another. Thus, modeling crime trips as if they are referenced relative to a residence is not necessarily going to lead to better prediction.

The mathematics of journey to crime has been modeled by Rengert (1981) using a modified general opportunities model:

$$P_{ij} = K\ U_i\ V_j\ f(D_{ij}) \qquad (10.12)$$

where P_{ij} is the probability of an offender in location (or zone) i committing an offense at location j, U_i is a measure of the number of crime trips produced at location i (what Rengert called *emissiveness*), V_j is a measure of the number of crime targets (attractiveness) at location j, and $f(D_{ij})$ is an unspecified function of the cost or effort expended in traveling from location i to location j (distance, time, cost). He did not try to operationalize either the production side or the attraction side. Nevertheless, conceptually, a crime trip would be expected to involve both elements as well as the cost of the trip.

In short, there has been a great deal of research on the travel behavior of criminals in committing acts as well as a number of statistical formulations.

Predicting the Location of Serial Offenders

The journey to crime formulation, as in equation 10.9, has been used to estimate the origin location of a serial offender based on the distribution of crime incidents. The logic is to plot the distribution of the incidents and then use a property of that distribution to estimate a likely origin location for the offender. Inspecting a pattern of crimes for a central location is an intuitive idea that police departments have used for a long time. The distribution of incidents describes an activity area by an offender, who lives somewhere in the center of the distribution. It is a *sample* from the offender's activity space. Using the Brantingham's terminology, there is a search area by an offender within which the crimes are committed; most likely, the offender also lives within the search area.

For example, Canter (1994) shows how the area defined by the distribution of the 'Jack the Ripper' murders in the east end of London in the 1880s included the key suspects in the case (though the case was never solved). Kind (1987) analyzed the incident locations of the 'Yorkshire Ripper' who committed thirteen murders and seven attempted murders in northeast England in the late 1970s and early 1980s. Kind applied two different geographical criteria to estimate the residential location of the offender. First, he estimated the center of minimum distance. Second, on the assumption that the locations of the murders and attempted murders that were committed late at night were closer to the offender's residence, he graphed the time of the offense on the Y axis against the month of the year (taken as a proxy for length of day) on the X axis and plotted a trend line through the data to account for seasonality. Both the center of minimum distance and the murders committed at a later time than the trend line pointed towards the Leeds/Bradford area, very close to where the offender actually lived (in Bradford).

Rossmo Model

Rossmo (1993; 1995) has adapted location theory, particularly travel behavior modeling, to serial offenders. In a series of papers (Rossmo, 1993a; 1993b; 1995; 1997) he outlined a mathematical approach to identifying the home base location of a serial offender, given the distribution of the incidents. The mathematics represent a formulation of the Brantingham and Brantingham (1981) search area model, discussed above in which the search behavior of an offender is seen as following a distance decay function with decreased activity near the offender's home base. He has produced examples showing how the model can be applied to serial offenders (Rossmo, 1993a; 1993b; 1997).

The model has four steps (what he called *criminal geographic targeting*):

1. First, a rectangular study area is defined that extends beyond the area of the incidents committed by the serial offender. The average distance between points is taken in both the Y and X direction. Half the Y inter-point distance is added to the maximum Y value and subtracted from the minimum Y value. Half the X inter-point distance is added to the maximum X value and subtracted from the minimum X value. These are based on projected coordinates; presumably, the directions would have to be adjusted if spherical coordinates were used. The rectangular study defines a grid from which columns and rows can be defined.

2. For each grid cell, the Manhattan distance to each incident location is taken (see chapter 3 for definition).

3. For each Manhattan distance from a grid cell to an incident location, MD_{ij}, one of two functions is evaluated:

   A. If the Manhattan distance, MD_{ij}, is less than a specified buffer zone radius, B, then

$$P_{ij} = \prod_{c=1}^{T} \{k[\,(1-\phi)(B^{g-f}) \,/\, (2B - |\,x_i - x_c| \;+ |\,y_i - y_c|\,)^g]\,\} \qquad (10.13)$$

   where P_{ij} is the resultant of offender interaction for grid cell, i; c is the incident number, summing to T; $\phi = 0$; k is an empirically determined constant; g is an empirically determined exponent; and f is an empirically determined exponent.

   The Greek letter, Π, is the product sign, indicating that the results for each grid cell-incident distance, MD_{ij}, are *multiplied* together across all incidents, c. This equation reduces to

$$P_{ij} = \prod_{c=1}^{T} \{k(1-0)(B^{g-f}) \,/\, (2B - |\,x_i - x_c| \;+ |\,y_i - y_c|\,)^g\,\} \qquad (10.14)$$

$$P_{ij} = \prod_{c=1}^{T} \frac{KB^{g-f}}{(2B - |\,x_i - x_c| \;+ |\,y_i - y_c|\,)^g} \qquad (10.15)$$

   Within the buffer region, the function is the ratio of a constant, k, times the radius of the buffer, B, raised to another constant (g-f), divided by the difference between the diameter of the circle (2B) and the Manhattan distance, MD_{ij}, raised to a constant, g. This is a *non-linear* function.

B. If the Manhattan distance, MD_{ij}, is greater than a specified buffer zone radius, B, then

$$P_{ij} = \prod_{c=1}^{T} \{k [\phi / (| x_i - x_c| + | y_i - y_c|)^f \} \qquad (10.16)$$

where P_{ij} is the resultant of offender interaction for grid cell, i, and incident location, j; *c* is the incident number, summing to T; $\phi = 1$; *k* is an empirically determined constant (the same as in equation 10.15 above); and *f* is an empirically determined exponent (the same as in equation 10.15 above).

Again, the Greek letter, Π, indicates that the results for each grid cell-incident distance, MD_{ij}, are multiplied together across all incidents, c. This equation reduces to

$$P_{ij} = \prod_{c=1}^{T} \{k [1/(| x_i - x_c| + | y_i - y_c|)^f \} \qquad (10.17)$$

$$P_{ij} = \prod_{c=1}^{T} \left\{ \frac{k}{(| x_i - x_c| + | y_i - y_c|)^f} \right\} \qquad (10.18)$$

Outside of the buffer region, the function is a constant, k, divided by the distance, MD_{ij}, raised to an exponent, f. It is an inverse distance function and drops off rapidly with distance

4. Finally, for each grid cell, i, the functions evaluated in step 3 above are summed over all incidents.

For both the 'within buffer zone' (near to home base) and 'outside buffer zone' (far from home base) functions, the coefficient, k, and exponents, f and g, are empirically determined. Though he doesn't discuss how these are calculated, they are presumably estimated from a sample of known offender locations where the distance to each incident is known (e.g., arrest records).

The result is a surface model indicating a likelihood of the offender residing at that location. He describes it as a probability surface, but it is actually a *density* surface. Since the probability of interaction between any one grid cell, i, and any one incident, j, cannot be greater than 1, the surface actually indicates the product of individual likelihoods that the offender uses that location as the home base. To be an actual probability function, it would have to be re-scaled so that the sum of the grid cells was equal to 1.

The second function - 'outside the buffer zone' (equation 10.16) is a classic gravity function, similar to equation 10.5 except there is no attraction definition. It is the distance decay part of the gravity function. The first function, equation 10.13, is an increasing curvilinear function designed to model the area of decreased activity near the offender's home base.

Strengths and weaknesses of the Rossmo model

The Rossmo model has both strengths and weaknesses. First, the model has some theoretical basis utilizing the Brantingham and Brantingham (1981) framework for an offender search area as well as the mathematics of the gravity model and distinguishes two types of travel behavior - near to home and farther from home. Second, the model does represent a systematic approach towards identifying a likely home base location for an offender. By evaluating each grid cell in the study area, an independent estimate of the likelihood is obtained, which can then be integrated into a continuous surface with an interpolation graphics routine.

There are problems with the particular formulation, however. First, the exclusive use of Manhattan distances is questionable. Unless the study area has a street network that follows a uniform grid, measuring distances horizontally and vertically can lead to overestimation of travel distances; further, the more the layout differs from a north-south and east-west orientation, the greater the distortion. Since many urban areas do not have a uniform grid street layout, the method will necessarily lead to overestimation of travel distances in places where there are diagonal or irregular streets.[1]

Second, the use of a product term, Π, complicates the mathematics. That is, the technique evaluates the distance from a particular grid cell, i, to a particular incident location, j. It then *multiplies* this result by all other results. Since the P values are actually densities, which can be greater than 1.0, the process, if strictly applied, would be a compounding of probabilities with overestimation of the likelihood for grid cells close to incident locations and underestimation of the likelihood for grid cells farther away. In the description of the method, however, Rossmo actually mentions summing the terms. Thus, the substitution of a summation sign, Σ, for the product sign would help the mathematics.

A third problem is in the distance decay function (equation 10.16). The use of an inverse distance term has problems as the distance between the grid cell location, i, and the incident location, j, decreases. For some types of crimes, there will be little or no buffer zone around the offender's home base (e.g., rapes by acquaintances). Consequently, the buffer zone radius, B, would approach 0. However, this would cause the model to become unstable since the inverse distance term will approach infinity.

Fourth, the use of a mathematical function to describe the distance decay, while easy to define, probably oversimplifies actual travel behavior. A mathematical function to describe distance decay is an approximation to actual travel behavior. It assumes that travel is equally likely in each direction, that travel distance is uniformly easy (or difficult) in each direction, and that, similarly, opportunities are uniformly distributed. For most urban areas, these conditions would not be true. Few cities form a perfect grid (Salt Lake City is, of course, an exception), though most cities have sections that are grided. Both physical geography limit travel in certain directions as does the historical street structure, which is often derived from earlier communities. A mathematical function does not consider this structure, but rather assumes that the 'impedance' in all directions is uniform.

This latter criticism, of course, would be true for all mathematical formulations of travel distance. There are corrections that can be made to adjust for this. For example, in the urban travel demand type model, trip distribution between locations is estimated by a gravity model, but then the distributed trips are constrained by, first, the total number of trips in the region (estimated separately), second, by mode of travel (bus v. single driver v. drivers plus passengers v. walk, etc.), and, third, by the route structure upon which the trips are eventually assigned (Krueckeberg and Silvers, 1974; Stopher and Meyburg, 1975; Field and MacGregor, 1987). Calibration at all stages against known data sets ensures that the coefficients and exponents fit 'real world' data as closely as possible. It would take these types of modifications to make the travel distribution type of model postulated by Rossmo and others be more realistic.

Fifth, the model imposes mathematical rigidity on the data. While there are two different functions that could vary from place to place, the particular type of distance decay function might also vary. Specifying a strict form for the two equations limits the flexibility of applying the model to different types of crime or to places where the distance decay does not follow the form specified by Rossmo.

A sixth problem is that opportunities for committing crimes - the attractiveness of locations, are never measured. That is, there is no enumeration of the opportunities that would exist for an offender nor is there an attempt to measure the strength of this attraction. Instead, the search area is inferred strictly from the distribution of incidents. Because the distribution of offender opportunities would be expected to vary from place to place, the model would need to be re-calibrated at each location. In this sense, both the Canter model and my journey to crime model (both described below) also share this weakness. It is understandable in that victim/target opportunities are difficult to define *a priori* since they can be interpreted differently by individuals. Nevertheless, a more complete theory of journey to crime behavior would have to incorporate some measure of opportunities, a point that both Brantingham and Brantingham (1981) and Rengert (1981) have made.

Finally, the 'buffer zone' concept is but one interpretation of the tendency of many crimes not to be committed close to the home location. There are other interpretations that are applicable. For example, the distribution of crime opportunities is often not close to the home location, either. Many crimes occur in commercial areas. In most American cities, residential areas are not located in commercial areas. Thus, there will usually be a distance between a residential location and a nearby crime opportunity. This does not imply anything about a 'safety zone' for the offender but, instead, may illustrate the distribution of the opportunities. If we could map the travel distance of, say, shopping trips, we would probably find a similar distribution to that seen in most of journey to crime studies (and illustrated below).

The concept of a 'buffer zone' is a hypothesis, not a certainty. The language of it is so appealing that many people believe it to be true. But, to demonstrate the existence of a 'buffer zone' would require interviewing offenders (or offenders who have been arrested) and demonstrating that they did not commit crimes near their residence even though there were opportunities (i.e., they valued safety over opportunity). To my knowledge, there has not been a study that demonstrated this yet. Otherwise, one cannot distinguish between the 'buffer zone' hypothesis and the distribution of available opportunities. They may very well be the same thing.

Canter Model

Canter's group in Liverpool (Canter and Tagg, 1975; Canter and Larkin, 1993; Canter and Snook, 1999; Canter, Coffey and Huntley, 2000) have modified the distance decay function for journey to crime trips by using a negative exponential term, instead of the inverse distance. Their *Dragnet* program uses the negative exponential function

$$Y = \alpha\, e^{(-\beta\, D_{ij}/\, P)} \tag{10.19}$$

where Y is the likelihood of an offender traveling a certain distance to commit a crime,, D_{ij} is the distance (from a home base location to an incident site), α is an arbitrary constant, β is the coefficient of the distance (and, hence, an exponent of *e*), P is a normalization constant, and *e* is the base of the natural logarithm. The model is similar to equation 10.11 except, like Rossmo, it does not include the attractiveness of the location.

Using the logic that most crimes are committed near the offender's home base, Canter, Coffey and Huntley (2000) use a five step process to estimate a search strategy:

1. The study area is defined by a rectangle that is 20% larger in area than that defined by the minimum and maximum X/Y points. A grid cell structure of 13, 300 cells is imposed over the rectangle. Each grid cell is a reference location, i.

2. A decay coefficient is selected. In equation 10.19, this would be the coefficient, β, for the distance term, D_{ij}, both of which are exponents of e. Unlike Rossmo, Canter uses a series of decay coefficients from 0.1 to 10 to estimate the sensitivity of the model. The equation indicates the likelihood with which any location is likely to be the home base of the offender based on one incident.

3. Because different offenders have different search areas, the measured distances for each cell are divided by a normalization coefficient, P, that adjusts all offenses to a comparable range. Canter uses two different types of normalization function: 1) mean inter-point distance between all offenses (across a group of offenders); and 2) the QRange, which is an index that takes into account asymmetry in the orientation of the incidents.

4. For each reference cell, i, the distance between each grid cell and each incident location is evaluated with the function and the standardized likelihoods are summed to yield an estimate of location potential.

5. A *search cost* index is defined by the proportion of the study area that has to be searched to find the offender. By calibrating the model against known cases, an estimate of search efficiency is obtained.

Additional modifications can be added to the functions to make them more flexible (Canter, Coffey and Huntley, 2000). For example, 'steps' are distances near to home where offenders are not likely to act while 'plateaus' are constant distances near to home where there is the highest likelihood of acting. For example, Canter and Larkin (1993) found an area around serial offenders' homes of about 0.61 mile in radius within which they were less likely to commit crimes.

Canter and Snook (1999) provide estimates of the search cost (or efficiency) associated with various distance coefficients. For example, with the known home base locations of 32 burglars, a β of 1.0 yielded a mean search cost of 18.06%; that is, on average, only 18.06% of the study area had to be searched to find the location of 32 burglars in the calibration sample. Clearly, for some of them, a larger area had to be searched while for others a smaller area; the average was 18.06%. Conversely, the mean search cost index for 24 rapists was 21.10% and for 37 murderers 28.28%. They further explored the marginal increase in locating offenders by increasing the percentage of the study area that had to be searched. They found for their three samples (burglary, rape, homicide) that more than half the offenders could be located within 15% of the area searched.

The Canter model is different from the Rossmo model is that it suggests a search strategy by the police for a serial offender rather than a particular location. The strength of it is to indicate how narrow an area the police should concentrate on in order to optimize finding an offender. Clearly, in most cases, only a small area needs be searched.

Strengths and weaknesses of the Canter model

The model has both strengths and weaknesses. First, the model provides a search strategy for law enforcement. By examining what type of function best fits a certain type of crime, police can target their search efforts more efficiently. The model is relatively easy to implement and is practical. Second, the mathematical formulation is stable. Unlike the inverse distance function in the Rossmo model, equation 10.19 will not have problems associated with distances that are close to 0. Further, the model does provide a search strategy for identifying an offender. It is a useful tool for law enforcement officers, particularly as they frame a search for a serial offender.

There are also weaknesses to the model. First, it lacks a theoretical basis. Canter's research has provided a great deal in terms of understanding the activity spaces of serial offenders (Canter and Larkin, 1993; Canter and Gregory, 1994; Canter, 1994; Hodge and Canter, 2000). However, the empirical model used is strictly pragmatic. Second, mathematically, it imposes the negative exponential function without considering other distance decay models. In the *Dragnet* program, the decay function is a string of 20 numbers so that, in theory, any function can be explored. However, the default is a negative exponential. The negative exponential has been used in many travel behavior studies (Foot, 1981; Bossard, 1993), but it does not always produce the best fit. Later on, I'll show examples of travel behavior which show a distinctly non-monotonic function, even beyond a home base 'buffer zone'. While the model can be adapted to be more flexible by different exponents and including steps and plateaus, for example, it is still tied to the negative exponential form. Thus, the model might work in some locations, but may fail in others; a user can't easily adjust the model to make it fit new data.

Third, the coefficient of the negative exponential, α, is defined arbitrarily. In the *Dragnet* program, it is usually set as 0.5. While this ensures that the result never exceed 1.0 for any one incident, there is a limit on the location potential summation since the total potential is a function of the number of incidents (i.e., it will be higher for more incidents). Thus, the use of α ends up being arbitrary. It would have been better if the coefficient were calibrated against a known sample.

Fourth, and finally, also similar to the Rossmo model (and to my Jtc model below), criminal opportunities (or attractions) are never measured, but are inferred from the pattern of crime incidents. As a pragmatic tool for informing a police search, one could argue that this is not important. However, in a different location, the distance coefficient is liable to differ as is the search cost index. It would need to be re-calibrated each time.

Nevertheless, the Canter model is a useful tool for police department and can help shape a search strategy. It is different from the other location models in that it is not focused so much on the best prediction for a location of an offender (though the summation discussed above in step 4 can yield that) as it does in defining where the search should be optimized.

Geographic Profiling

Journey to crime estimation should be distinguished from *geographical profiling*. Geographical profiling involves understanding the geographical search pattern of criminals in relation to the spatial distribution of potential offenders and potential targets, the awareness spaces of potential offenders including the labeling of 'good' targets and crime areas, and the interchange of information between potential offenders who may modify their awareness space (Brantingham and Brantingham, 1981). According to Rossmo:

"...Geographic profiling focuses on the probable spatial behaviour of the offender within the context of the locations of, and the spatial relationships between, the various crime sites. A psychological profile provides insights into an offender's likely motivation, behaviour and lifestyle, and is therefore directly connected to his/her spatial activity. Psychological and geographic profiles thus act in tandem to help investigators develop a picture of the person responsible for the crimes in question" (Rossmo, 1997).

In other words, geographic profiling is a framework for understanding how an offender traverses an area in searching for victims or targets; this, of necessity, involves understanding the social environment of an area, the way that the offender understands this environment (the 'cognitive map') as well as the offender's motives.

On the other hand, journey to crime estimation follows a much simpler logic involving the distance dimension of the spatial patterning of a criminal. It is a method aimed at estimating the distance that serial offenders will travel to commit a crime and, by implication, the likely location from which they started their crime 'trip'. In short, it is a strictly statistical approach to estimating the residential whereabouts of an offender compared to understanding the dynamics of serial offenders.

It remains an empirical question whether a conceptual framework, such as geographic profiling, can predict better than a strictly statistical framework. Understanding of a phenomena, such as serial murders, serial rapists, and so forth, is an important research area. We seek more than just statistical prediction in building a knowledge base. However, it doesn't necessarily follow that understanding produces better predictions. In many areas of human activity, strictly statistical models are better in predicting than explanatory models. I will return to this point later in the section.

The *CrimeStat* Journey to Crime Routine

The journey to crime (Jtc) routine is a diagnostic designed to aid police departments in their investigations of serial offenders. The aim is to estimate the likelihood that a serial offender lives at any particular location. Using the location of incidents committed by the serial offender, the program makes statistical guesses at where the offender is liable to live, based on the similarity in travel patterns to a known sample of serial offenders for the same type of crime. The *Jtc* routine builds on the Rossmo (1993a; 1993b; 1995) framework, but extends its modeling capability.

1. A grid is overlaid on top of the study area. This grid can be either imported or can be generated by *CrimeStat* (see chapter 2). The grid represents the entire study area. Unlike Rossmo or Canter and Snook, there is no optimal study area. The technique will model that which is defined. Thus, the user has to select an area intelligently.

2. The routine calculates the distance between each incident location committed by a serial offender (or group of offenders working together) and

each cell, defined by the centroid of the cell. Rossmo (1993a; 1995) used indirect (Manhattan) distances. However, this would be appropriate only when a city falls on a uniform grid. The *Jtc* routine allows both direct and indirect distances. In most cases, direct distances would be the most appropriate choice as a police department would normally locate origin and destination locations rather than particular routes that are taken (see below).

3. A distance decay function is applied to each grid cell-incident pair and sums the values over all incidents. The user has a choice whether to model the travel distance by a mathematical function or an empirically-derived function.

4. The resultant of the distance decay function for each grid cell-incident pair are summed over all incidents to produce a likelihood (or density) estimate for each grid cell.
5. In both cases, the program outputs the two results: 1) the grid cell which has the peak likelihood estimate; and 2) the likelihood estimate for every cell. The latter output can be saved as a *Surfer® for Windows* 'dat', *ArcView Spatial Analyst©* 'asc', ASCII 'grd', *ArcView®* '.shp', *MapInfo®* '.mif', *Atlas*GIS™* '.bna' file or as an Ascii grid 'grd' file which can be read by many GIS packages (e.g., *ARC/INFO®*, *Vertical Mapper©*). These files can also be read by other GIS packages (e.g., *Maptitude*).

Figure 10.1 shows the logic of the routine and figure 10.2 shows the Journey to Crime (Jtc) screen. There are two parts to the routine. First, there is a calibration model which is used in the empirically-derived distance function. Second, there is the Journey to Crime (Jtc) model itself in which the user can select either the already-calibrated distance function or the mathematical function. The empirically-derived function is, by far, the easiest to use and is, consequently, the default choice in *CrimeStat*. The discussion of it is on p. 35. However, the mathematical function can be used if there is inadequate data to construct an empirical distance decay function or if a particular form is desired.

Distance Modeling Using Mathematical Functions

We'll start by illustrating the use of the mathematical functions because this has been the traditional way that distance decay has been examined. The *CrimeStat* Jtc routine allows the user to define distance decay by a mathematical function.

Probability Distance Functions

The user selects one of five probability density distributions to define a likelihood that the offender has traveled a particular distance to commit a crime. The advantage of having five functions, as opposed to only one, is that it provides more flexibility in describing travel behavior. The travel distance distribution followed will vary by crime type, time of day, method of

Figure 10.1: # Journey to Crime Interpolation Routine

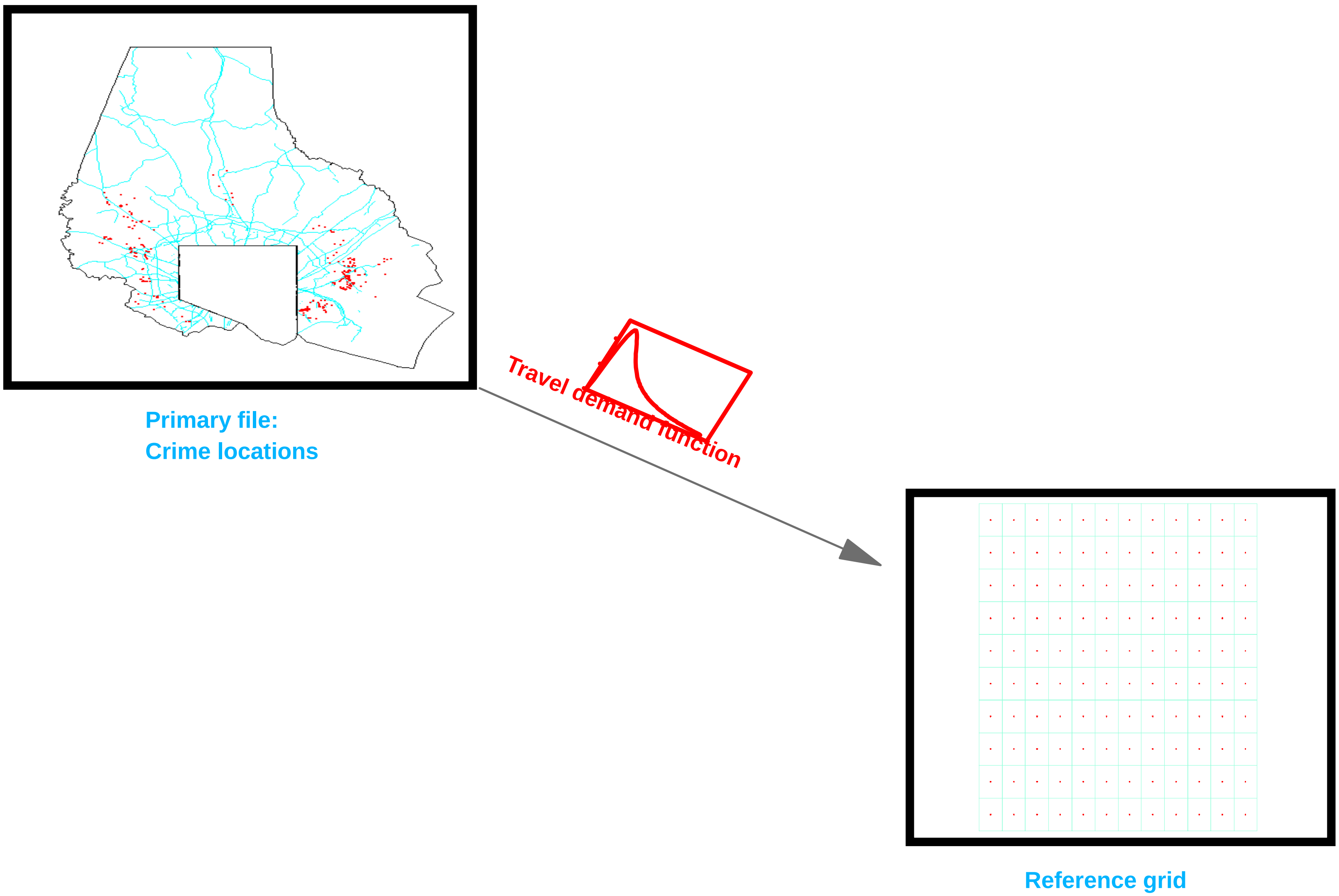
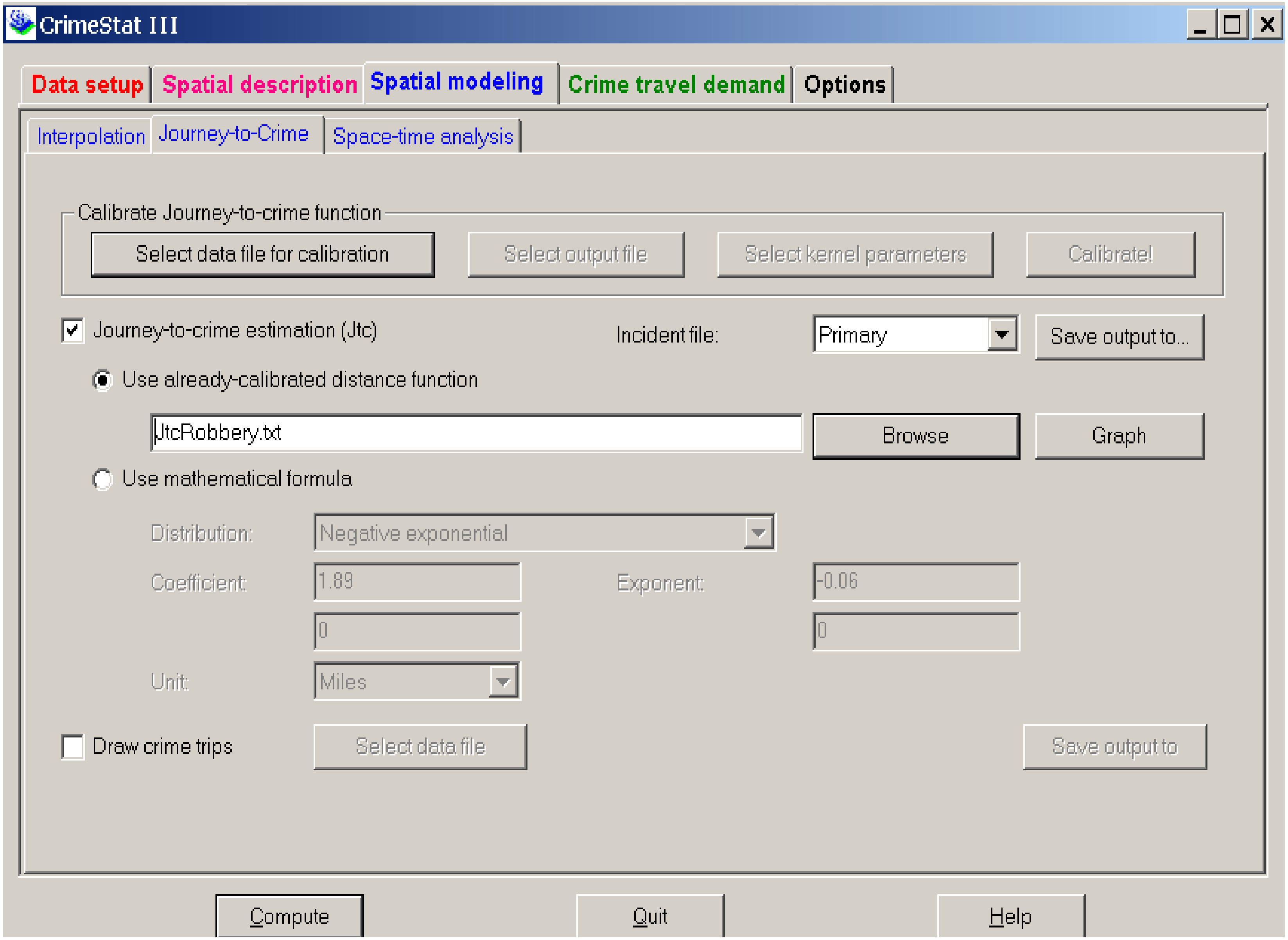


Figure 10.2: Journey to Crime Screen



operation, and numerous other variables. The five functions allow an approach that can simulate more accurately travel behavior under different conditions. Each of these has parameters that can be modified, allowing a very large number of possibilities for describing travel behavior of a criminal.

Figure 10.3 illustrates the five types.[2] Default values based on Baltimore County have been provided for each. The user, however, can change these as needed.

Briefly, the five functions are:

Linear

The simplest type of distance model is a linear function. This model postulates that the likelihood of committing a crime at any particular location declines by a constant amount with distance from the offender's home. It is highest near the offender's home but drops off by a constant amount for each unit of distance until it falls to zero. The form of the linear equation is:

$$f(d_{ij}) = A + B*d_{ij} \qquad (10.20)$$

where $f(d_{ij})$ is the likelihood that the offender will commit a crime at a particular location, i, defined here as the center of a grid cell, d_{ij} is the distance between the offender's residence and location i, A is a slope coefficient which defines the fall off in distance, and B is a constant. It would be expected that the coefficient B would have a negative sign since the likelihood should decline with distance. The user must provide values for A and B. The default for A is 1.9 and for B is -0.06. This function assumes no buffer zone around the offender's residence. When the function reaches 0 (the X axis), the routine automatically substitutes a 0 for the function.

Negative Exponential

A slightly more complex function is the negative exponential. In this type of model, the likelihood is also highest near the offenders home and drops off with distance. However, the decline is at a constant *rate* of decline, thus dropping quickly near the offender's home until is approaches zero likelihood. The mathematical form of the negative exponential is

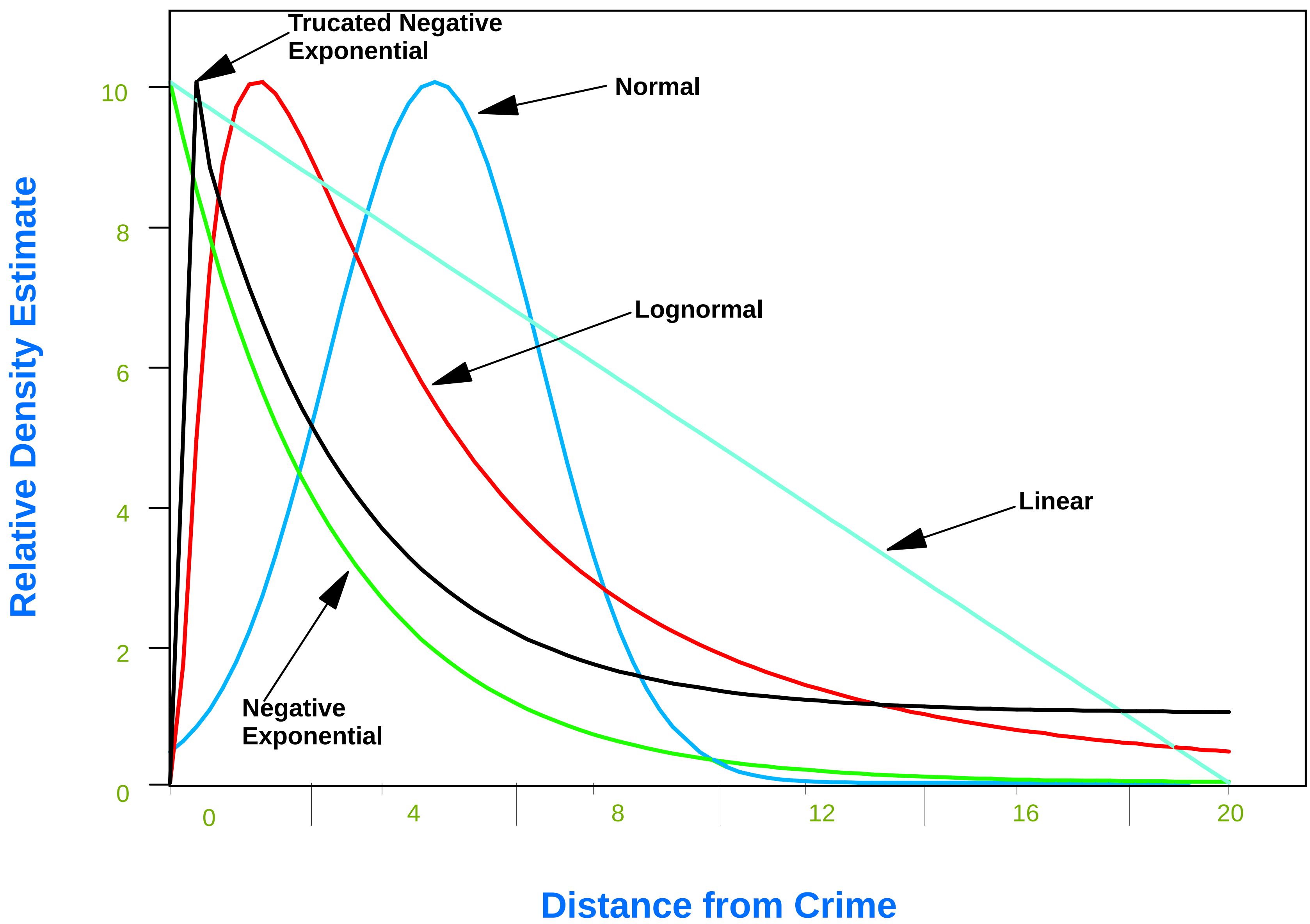
$$f(d_{ij}) = A*e^{-B*d_{ij}} \qquad (10.21)$$

where $f(d_{ij})$ is the likelihood that the offender will commit a crime at a particular location, i, defined here as the center of a grid cell, d_{ij} is the distance between each reference location

Figure 10.3:

Journey to Crime Travel Demand Functions

Five Mathematical Functions



and each crime location, e is the base of the natural logarithm, A is the coefficient and B is an exponent of e. The user inputs values for A - the coefficient, and B - the exponent. The default for A is 1.89 and for B is -0.06. This function is similar to the Canter model (equation 10.19) except that the coefficient is calibrated. Also, like the linear function, it assumes no buffer zone around the offender's residence.

Normal

A normal distribution assumes the peak likelihood is at some optimal distance from the offender's home base. Thus, the function rises to that distance and then declines. The rate of increase prior to the optimal distance and the rate of decrease from that distance is symmetrical in both directions. The mathematical form is:

$$Z_{ij} = \frac{(d_{ij} - MeanD)}{S_d} \tag{10.22}$$

$$f(d_{ij}) = A * \frac{1}{S_d * SQRT(2\pi)} * e^{-0.5*Z_{ij}^2} \tag{10.23}$$

where $f(d_{ij})$ is the likelihood that the offender will commit a crime at a particular location, i (defined here as the center of a grid cell), d_{ij} is the distance between each reference location and each crime location, MeanD is the mean distance input by the user, S_d is the standard deviation of distances, **e** is the base of the natural logarithm, and A is a coefficient. The user inputs values for MeanD, S_d, and A. The default values are 4.2 for the mean distance, MeanD, 4.6 for the standard deviation, S_d, and 29.5 for the coefficient, A.

By carefully scaling the parameters of the model, the normal distribution can be adapted to a distance decay function with an increasing likelihood for near distances and a decreasing likelihood for far distances. For example, by choosing a standard deviation greater than the mean (e.g., MeanD = 1,S_d = 2), the distribution will be skewed to the left because the left tail of the normal distribution is not evaluated. The function becomes similar to the model postulated by Brantingham and Brantingham (1981) in that it is a single function which describes travel behavior.

Lognormal

The lognormal function is similar to the normal except it is more skewed, either to the left or to the right. It has the potential of showing a very rapid increase near the offender's home base with a more gradual decline from a location of peak likelihood (see Figure 10.3). It is also similar to the Brantingham and Brantingham (1981) model. The mathematical form of the function is:

$$f(d_{ij}) = A * \frac{1}{d^2_{ij} * S_d * SQRT(2\pi)} * \mathbf{e}^{-[\ln(d^2_{ij}) - MeanD]^2 / 2 * s_d^2} \tag{10.24}$$

where $f(d_{ij})$ is the likelihood that the offender will commit a crime at a particular location, i, defined here as the center of a grid cell, d_{ij} is the distance between each reference location and each crime location, MeanD is the mean distance input by the user, S_d is the standard deviation of distances, ***e*** is the base of the natural logarithm, and A is a coefficient. The user inputs MeanD, S_d , and A. The default values are 4.2 for the mean distance, MeanD, 4.6 for the standard deviation, S_d, and 8.6 for the coefficient, A. They were calculated from the Baltimore County data (see table 10.3).

Truncated Negative Exponential

The truncated negative exponential is a joined function made up of two distinct mathematical functions - the linear and the negative exponential. For the near distance, a positive linear function is defined, starting at zero likelihood for distance 0 and increasing to d_p, a location of peak likelihood. Thereupon, the function follows a negative exponential, declining quickly with distance. The two mathematical functions making up this spline function are

Linear: $$f(d_{ij}) = 0 + B*d_{ij} = B*d_{ij} \qquad \text{for } d_{ij} \geq 0, d_ij \leq d_p \qquad (10.25)$$

Negative Exponential: $$f(d_{ij}) = A*e^{-C*d_{ij}} \qquad \text{for } X_i > d_p \qquad (10.26)$$

where d_{ij} is the distance from the home base, B is the slope of the linear function and for the negative exponential function A is a coefficient and C is an exponent. Since the negative exponential only starts at a particular distance, d_p, A, is assumed to be the intercept *if* the Y-axis were transposed to that distance. Similarly, the slope of the linear function is estimated from the peak distance, d_p, by a peak likelihood function. The default values are 0.4 for the peak distance, d_p, 13.8 for the peak likelihood, and -0.2 for the exponent, C. Again, these were calculated with Baltimore County data (see table 10.3)

This function is the closest approximation to the Rossmo model (equations 10.13 and 10.16). However, it differs in several mathematical properties. First, the 'near home base' function is linear (equation 10.25), rather than a non-linear function (equation 10.13). It assumes a simple increase in travel likelihoods by distance from the home base, up to the edge of the safety zone.[3] Second, the distance decay part of the function (equation 10.26) is a negative exponential, rather than an inverse distance function (equation 10.13); consequently, it is more stable when distances are very close to zero (e.g., for a crime where there is no 'near home base' offset).

Calibrating an Appropriate Probability Distance Function

The mathematics are relatively straightforward. However, how does one know which distance function to use? The answer is to get some data and calibrate it. It is important to obtain data from a sample of known offenders where both their residence at the time they committed crimes as well as the crime locations are known. This is called the *calibration data set*. Each of the models are then tested against the calibration data set using an approach similar to that explained below. An error analysis is conducted to determine which of the models best fits the data. Finally, the 'best fit' model is used to estimate the likelihood that a particular serial offender lives at any one location. Though the process is tedious, once the parameters are calculated they can be used repeatedly for predictions.

Because every jurisdiction is unique in terms of travel patterns, it is important to calibrate the parameters for the particular jurisdiction. While there may be some similarities between cities (e.g., Eastern "centralized" cities v. Western "automobile" cities), there are always unique travel patterns defined by the population size, historical road pattern, and physical geography. Consequently, it is necessary to calibrate the parameters anew for each new city. Ideally, the sample should be a large enough so that a reliable estimate of the parameters can be obtained. Further, the analyst should check the errors in each of the models to ensure that the best choice is used for the Jtc routine. However, once it has been completed, the parameters can be re-used for many years and only periodically re-checked.

Data Set from Baltimore County

I'll illustrate with data from Baltimore County. The steps in calibrating the *Jtc* parameters were as follows:

1. 49,083 matched arrest and incident records from 1992 through 1997 were obtained in order to provide data on where the offender lived in relation to the crime location for which they were arrested.[4]

2. The data set was checked to ensure that there were X and Y coordinates for both the arrested individual's residence location and the crime incident location for which the individual was being charged. The data were cleaned to eliminate duplicate records or entries for which either the offender's residence or the incident location were missing. The final data set had 41,424 records. There were many multiple records for the same offender since an individual can commit more than one crime. In fact, more than half the records involved individuals who were listed two or more times. The distribution of offenders by the number of offenses for which they were charged is seen in Table 10.1. As would be expected, a small proportion of individuals account for a sizeable proportion of crimes; approximately 30% of the offenders in the database accounted for 56% of the incidents.

3. The data were imported into a spreadsheet, but a database program could equally have been used. For each record, the direct distance between the arrested individual's residence and the crime incident location was calculated. Chapter 2 presented the formulas for calculating direct distances between two locations and are repeated in endnote 5.[5]

Table 10.1

Number of Offenders and Offenses in Baltimore County: 1993-1997
Journey to Crime Database

Number of Offenses	Number of Individuals	Percent of Offenders	Number of Incidents	Percent of Incidents
1	18,174	70.0%	18,174	43.9%
2	4,443	17.1%	8,886	21.5%
3	1,651	6.4%	4,953	12.0%
4	764	2.9%	3,056	7.4%
5	388	1.5%	1,940	4.7%
6-10	482	1.9%	3,383	8.2%
11-15	61	0.2%	757	1.8%
16-20	10	<0.0%	175	0.4%
21-25	3	<0.0%	67	0.2%
26-30	0	<0.0%	0	0.0%
30+	1	<0.0%	33	<0.0%
	25,977		41,424	

4. The records were sorted into sub-groups based on different types of crimes. For the Baltimore County example, eleven categories of crime incident were used. Table 10.2 presents the categories with their respective sample sizes. Of course, other sub-groups could have been identified. Each sub-group was saved as a separate file. The same records can be part of multiple files (e.g., a record could be included in the 'all robberies' file as well as in the 'commercial robberies' file). All records were included in the 'all crimes' file.

5. For each type of crime, the file was grouped into distance intervals of 0.25 miles each. This involved two steps. First, the distance between the offender's residence and the crime location was sorted in ascending order. Second, a frequency distribution was conducted on the distances and grouped into 0.25 mile intervals (often called *bins*). The degree of precision in distance would depend on the size of the data set. For 41,426 records, quarter mile bins were appropriate.

6. For each type of crime, a new file was created which included only the frequency distribution of the distances broken down into quarter mile distance intervals, d_i.

7. In order to compare different types of crimes, which will have different frequency distributions, two new variables were created. First, the frequency in the interval was converted into the percentage of all crimes of in each interval by dividing the frequency by the total number of incidents, N,

and multiplying by 100. Second, the distance interval was adjusted. Since the interval is a range with a starting distance and an ending

Table 10.2

Baltimore County Files Used for Calibration

Crime Type	Sample Size
All crimes	41,426
Homicide	137
Rape	444
Assault	8,045
Robbery (all)	3,787
Commercial robbery	1,193
Bank robbery	176
Burglary	4,694
Motor vehicle theft	2,548
Larceny	19,806
Arson	338

distance but has been identified by spreadsheet program as the beginning distance only, a small fraction, representing the midpoint of the interval, is added to the distance interval. In our case, since each interval is 0.25 miles wide, the adjustment is half of this, 0.125. Each new file, therefore, had four variables: the interval distance, the adjusted interval distance, the frequency of incidents within the interval (the number of cases falling into the interval), and the percentage of all crimes of that type within the interval.

8. Using the regression program in the crime travel demand model (see chapter 12), a series of regression equations was set up to model the frequency (or the percentage) as a function of distance. In this case, I used our routines, but other statistical packages could equally have been used. Again, because comparisons between different types of crimes were of interest, the percentage of crimes (by type) within an interval was used as the dependent variable (and was defined as a percentage, i.e., 11.51% was recorded as 11.51). Five equations testing each of the five models were set up.

Linear

For the linear function, the test was

$$Pct_i = A + Bd_i \tag{10.27}$$

where Pct_i is the percentage of all crimes of that type falling into interval i, d_i is the distance for interval i, A is the intercept, and B is the slope. A and B are estimated directly from the regression equation.

Negative Exponential

For the negative exponential function, the variables have to be transformed to estimate the parameters. The function is

$$Pct_i = A * e^{-B*d_i} \qquad (10.28)$$

A new variable is defined which is the natural logarithm of the percentage of all crimes of that type falling into the interval, *ln(Pct$_i$)*. This term was then regressed against the distance interval, d_i.

$$\ln(Pct_i) = K - B*d_i \qquad (10.29)$$

However, since the original equation has been transformed into a log function, B is the coefficient and A can be calculated directly from

$$\ln(Pct_i) = \ln(A) - B*d_i \qquad (10.30)$$

$$A = e^K \qquad (10.31)$$

If the percentage in any bin was 0 (i.e., $Pct_i = 0$), then a value of -16 was taken since the natural logarithm of 0 cannot be solved (it approximates -16 as the percentage approaches 0.0000001).

Normal

For the normal function, a more complex transformation must be used. The normal function in the model is

$$Pct_i = A * \frac{1}{S_d * SQRT(2\pi)} * e^{-0.5*Zij^2} \qquad (10.32)$$

First, a standardized *Z* variable for the distance, d_i, is created

$$Z_i = \frac{(d_i - MeanD)}{S_d} \qquad (10.33)$$

where MeanD is the mean distance and S_d is the standard deviation of distance. These are calculated from the original data file (*before* creating the file of frequency distributions). Second, a normal transformation of Z is constructed with

$$\text{Normal}(Z_i) = \frac{1}{S_d * \text{SQRT}(2\pi)} * e^{-0.5 * Zij^2} \quad (10.34)$$

Finally, the normalized variable is regressed against the percentage of all crimes of that type falling into the interval, *Pct_i* with *no* constant

$$Pct_i = A * \text{Normal}(Z_i) \quad (10.35)$$

A is estimated by the regression coefficient.

Lognormal

For the lognormal function, another complex transformation must be done. The lognormal function for the percentage of all crimes of a type for a particular distance interval is

$$Pct_i = \quad A * \frac{1}{d^2_{ij} * S_d * \text{SQRT}(2\pi)} * e^{-[\ln(d^2_i) - \text{MeanD}]^2 / 2 * S_d^2} \quad (10.36)$$

The transformation can be created in steps. First, create L

$$L \quad = \quad \ln(d_i^2) \quad (10.37)$$

Second, create M

$$M \; = (l - \text{MeanD})^2 \quad (10.38)$$

Third, create O

$$O = \frac{m}{(2 * S_d^2)} \quad (10.39)$$

Fourth, create P by raising **e** to the O^{th} power.

$$P = e^O \quad (10.40)$$

Fifth, create the lognormal conversion, Lnormal

$$\text{Lnormal}(d_i) = A * \frac{1}{d^2_{ij} * S_d * \text{SQRT}(2\pi)} * P \quad (10.41)$$

Finally, the lognormal variable is regressed against the percentage of all crimes of that type falling into the interval, Pct_i with *no* constant

$$Pct_i = A^* \ Lnormal(d_i) \qquad (10.42)$$

A is estimated with the regression coefficient.

Truncated Negative Exponential

For the truncated negative exponential function, two models were set up. The first applied to the distance range from 0 to the distance at which the percentage (or frequency) is highest, $Maxd_i$. The second applied to all distances greater than this distance

$$\text{Linear:} \quad Pct_i = A + Bd_i \text{ for } d_{ij} \geq 0, d_ij \leq Maxd_{ij} \qquad (10.43)$$

$$\text{Negative Exponential:} \quad Pct_i = A^*e^{-C^*di} \quad \text{for } d_ij > Maxd_{ij} \qquad (10.44)$$

To use this function, the user specifies the distance at which the peak likelihood occurs, d_p (the *peak distance*) and the value for that peak likelihood, P (the *peak likelihood*). For the negative exponential function, the user specifies the exponent, C.

In order to splice the two equations together (the spline), the *CrimeStat* truncated negative exponential routine starts the linear equation at the origin and ends it at the highest value. Thus,

$$A = 0 \qquad (10.45)$$

$$B = P/d_p \qquad (10.46)$$

where P is the peak likelihood and d_p is the peak distance.

The exponent, C, can be estimated by transforming the dependent variable, Pct_i, as in the negative exponential above (equation 10.28) and regressing the natural log of the percentage ($\ln(Pct_i)$ against the distance interval, d_i, *only* for those intervals that are greater than the peak distance. I have found that estimating the transformed equation with a coefficient, A in

$$Pct_i = A * e^{-C^*d_i} \qquad (10.47)$$

$$\ln(Pct_i) = Ln(A) - C^*d_i \qquad (10.48)$$

gives a better fit to the equation. However, the user need only input the exponent, C, in the Jtc routine as the coefficient, A, of the negative exponential is calculated internally to produce a distance value at which the peak likelihood occurs. The formula is:

$$A = e^{\ln(P) + C*(d_p - d_i)} \tag{10.49}$$

where P is the peak likelihood, d_p is the distance for the peak likelihood, C is an exponent (assumed to be positive) and d_i is the distance interval for the histogram.

9. Once the parameters for the five models have been estimated, they can be compared to see which one is best at predicting the travel behavior for a particular type of crime. It is to be expected that different types of crimes will have different optimal models and that the parameters will also vary.

Examples from Baltimore County

Let's illustrate with the Baltimore County data. Figure 10.4 shows the frequency distribution for all types of crime in Baltimore County. As can be seen, at the nearest distance interval (0 to 0.25 miles with an assigned 'adjusted' midpoint of 0.125 miles), about 6.9% of all crimes occur within a quarter mile of the offender's residence (it can be seen on the Y-axis). However, for the next interval (0.25 to 0.50 miles with an assigned midpoint of 0.375 miles), almost 10% of all crimes occur at that distance (9.8%). In subsequent intervals, however, the percentage decreases, a little less than 6% for 0.50 to 0.75 miles (with the midpoint being 0.625 miles), a little more than 4% for 0.75 to 1 mile (the midpoint is 0.875 miles), and so forth.

The best fitting statistical function was the negative exponential. The particular equation is

$$Pct_i = 5.575 * e^{-0.229*d_i} \tag{10.50}$$

This is shown with the solid line. As can be seen, the fit is good for most of the distances, though it underestimates at close to zero distance and overestimates from about a half mile to about four miles. There is only slight evidence of decreased activity near to the location of the offender.

However, the distribution varies by type of crime. With the Baltimore County data, property crimes, in general, occur farther away than personal crimes. The truncated negative exponential generally fit property crimes better, lending support for the Brantingham and Brantingham (1981) framework for these types. For example, larceny offenders have a definite safety zone around their residence (figure 10.5). Fewer than 2% of larceny thefts occur within a quarter mile of the offender's residence. However, the percentage jumps to about 4.5% from a quarter mile to a half. The truncated negative exponential function fits the data reasonably well though it overestimates from about 1 to 3 miles and underestimates from about 4 to12 miles.

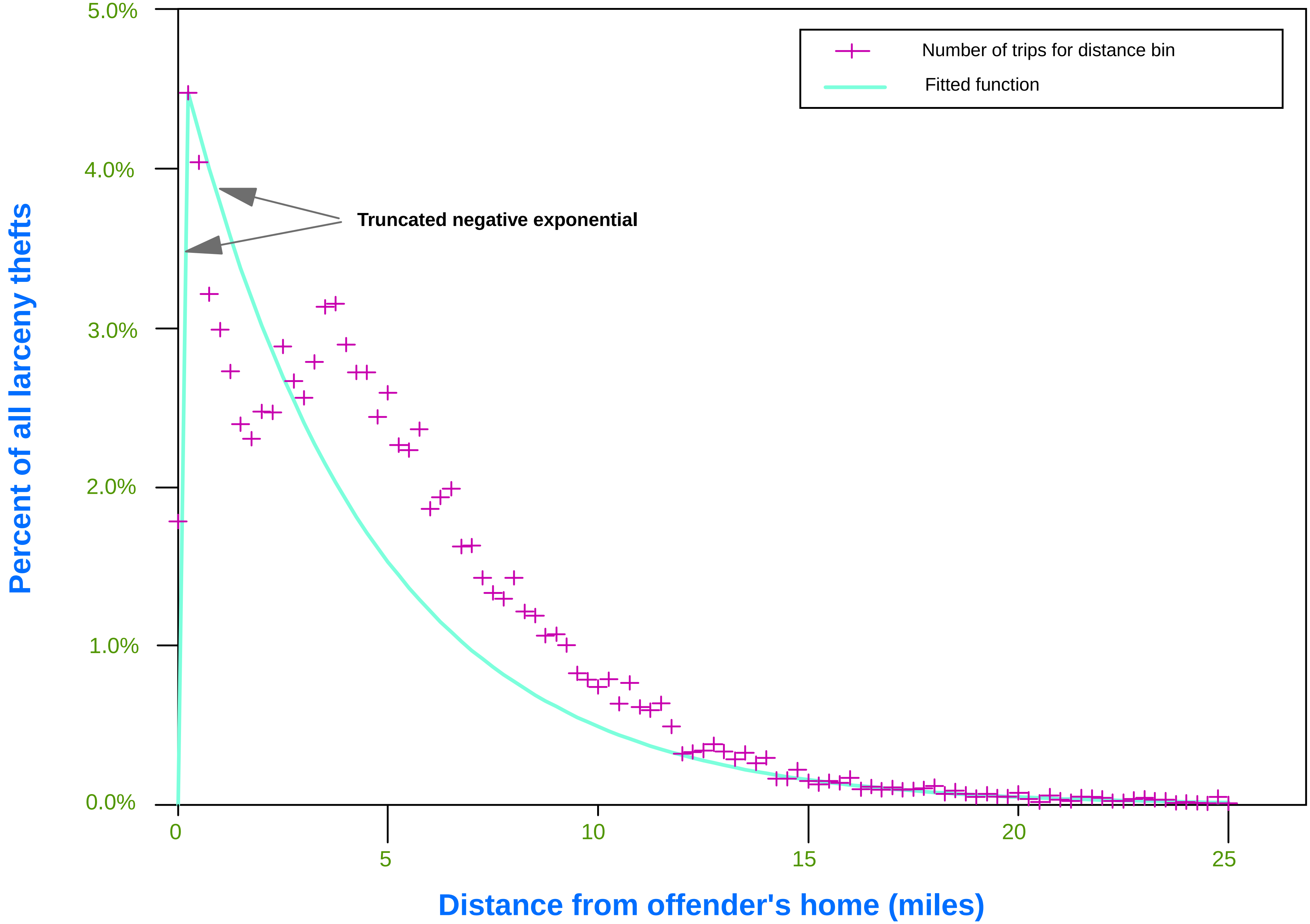
Figure 10.4:

Journey to Crime Distances: All Crimes
Negative Exponential Distribution
Number of trips for distance bin
Fitted function
Negative exponential
Percent of all crimes
10%
8%
6%
4%
2%
0%
0
5
10
15
20
25
Distance from offender's home (miles)

Figure 10.5:

Journey to Crime Distances: Larceny

Truncated Negative Exponential Function



Similarly, motor vehicle thefts show decreased activity near the offender's resident, though it is less pronounced than larceny theft. Figure 10.6 shows the distribution of motor vehicle thefts and the truncated negative exponential function which was fit to the data. As can be seen, the fit is reasonably good though it tends to underestimate middle range distances (approximately 3-12 miles).

Some types of crime, on the other hand, are very difficult to fit. Figure 10.7 shows the distribution of bank robberies. Partly because there were a limited number of cases (N=176) and partly because it's a complex pattern, the truncated negative exponential gave the best fit, but not a particularly good one. As can be seen, the linear ('near home') function underestimates some of the near distance likelihoods while the negative exponential drops off too quickly; in fact, to make this function even plausible, the regression was run only up to 21 miles (otherwise, it underestimated even more).

For some crimes, it was very difficult to fit any single function. Figure 10.8 shows the frequency distribution of 137 homicides with three functions being fitted to the data - the truncated negative exponential, the lognormal, and the normal. As can be seen each function fits only some of the data, but not all of it.

Testing for Residual Errors in the Model

In short, the five mathematical functions allow a user to fit a variety of distance decay distributions. Each of the models will predict some parts of the distribution better than others. Consequently, it is important to conduct an error analysis to determine which model is 'best'. In an error analysis, the residual error is defined as

$$\text{Residual error} = Y_i - E(Y_i) \qquad (10.51)$$

where Y_i is the observed (actual) likelihood for distance i and $E(Y_i)$ is the likelihood predicted by the model. If raw numbers of incidents are used, then the likelihoods are the number of incidents for a particular distance. If the number of incidents are converted into proportions (i.e., probabilities), then the likelihoods are the proportions of incidents for a particular distance.

The choice of 'best model' will depend on what part of the distribution is considered most important. Figure 10.9, for example, shows the residual errors on vehicle theft for the five fitted models. That is, each of the five models was fit to the proportion of vehicle thefts by distance intervals (as explained above). For each distance, the discrepancy between the actual percentage of vehicle thefts in that interval and the predicted percentage was calculated. If there was a perfect fit, then the discrepancy (or residual) was 0%. If the actual percentage was greater than the predicted (i.e., the model underestimated), then the residual was positive; if the actual was smaller than the predicted (i.e., the model overestimated), then the residual was negative.

Figure 10.6:

Journey to Crime Distances: Vehicle Theft

Truncated Negative Exponential Function

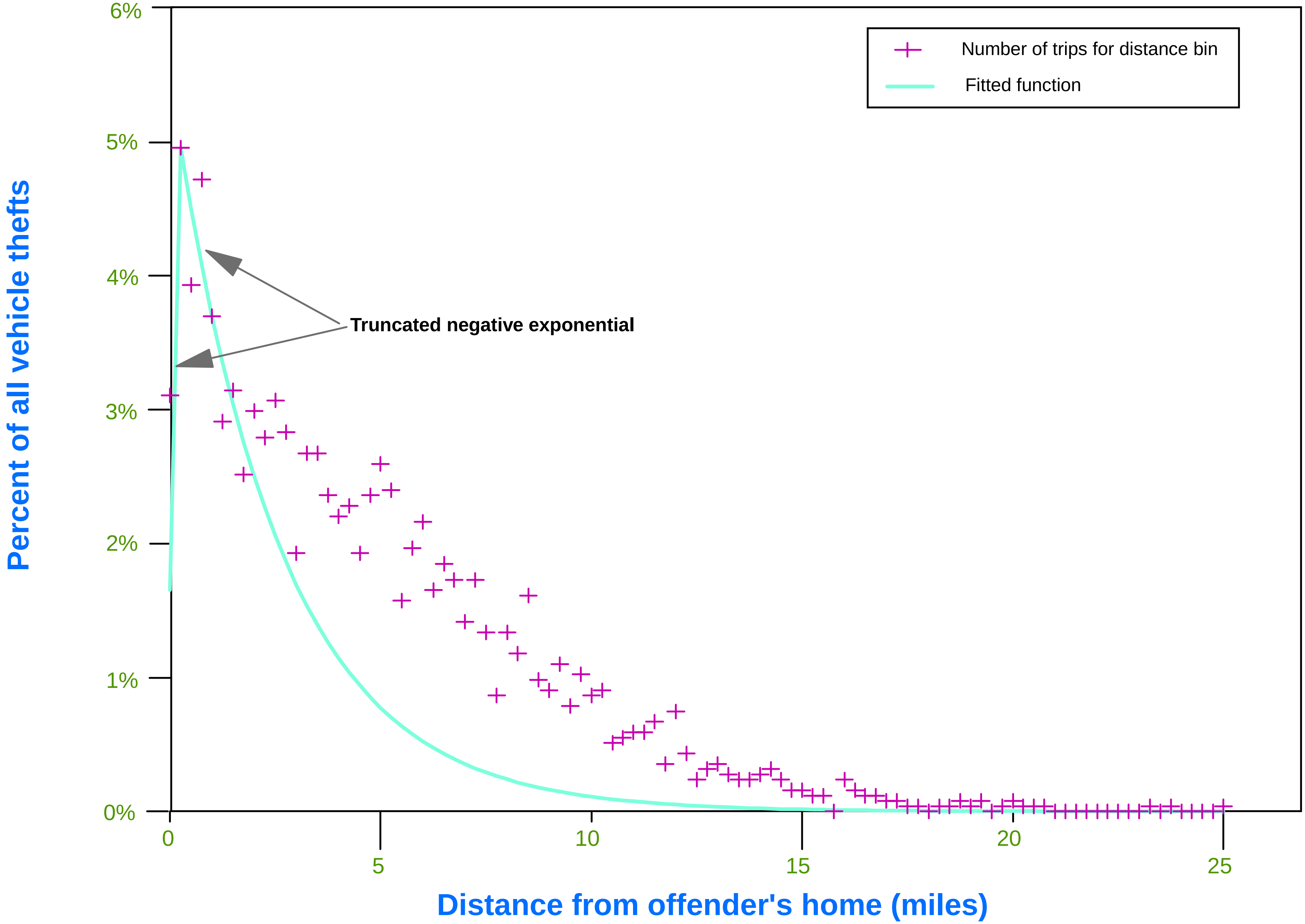


Figure 10.7:

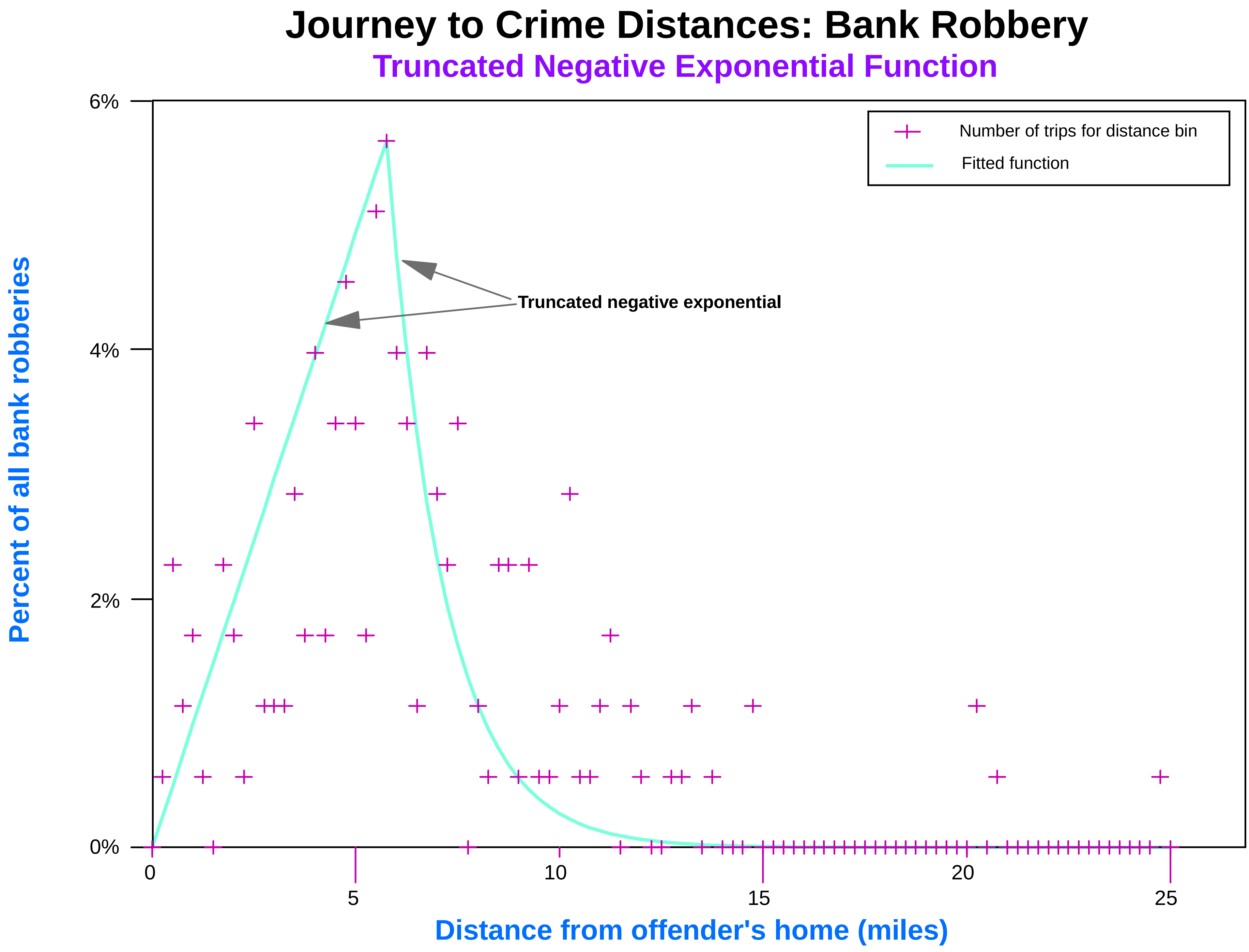

Journey to Crime Distances: Bank Robbery
Truncated Negative Exponential Function
Number of trips for distance bin
Fitted function
Truncated negative exponential
Percent of all bank robberies
6%
4%
2%
0%
0
5
10
15
20
25
Distance from offender's home (miles)

Figure 10.8:

Journey to Crime Distances: Homicide

Normal, Lognormal, and Truncated Negative Exponential Functions

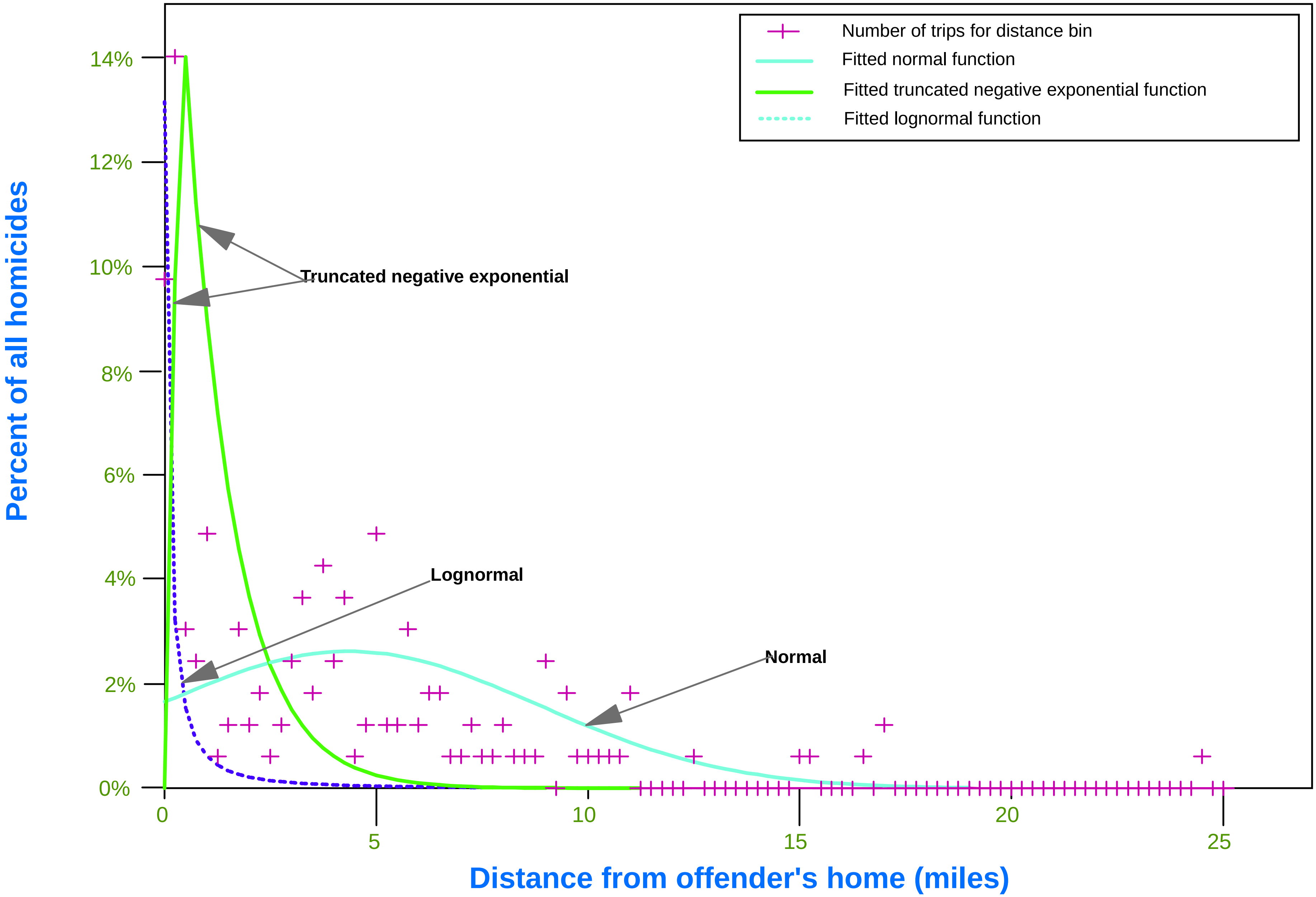
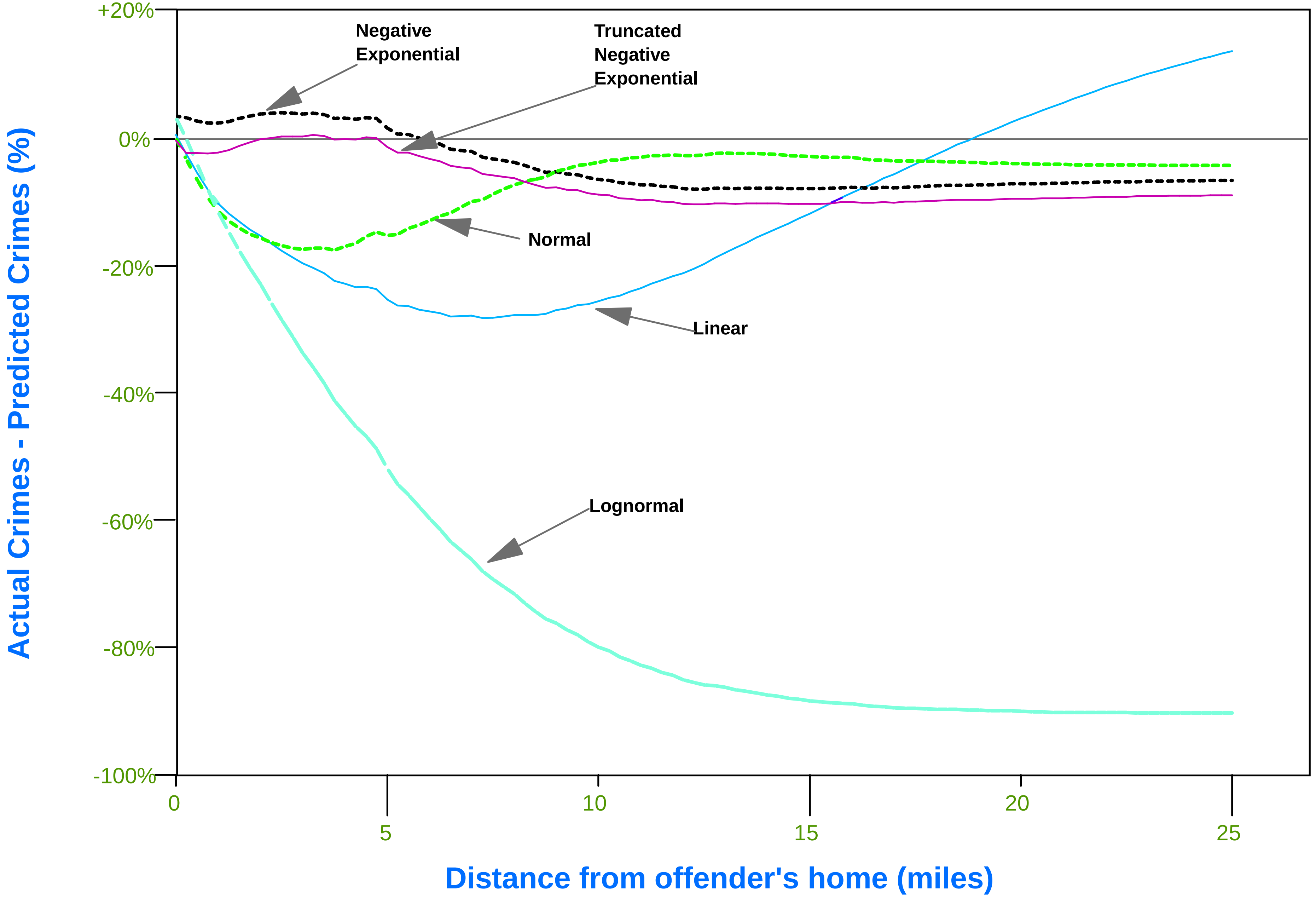


Figure 10.9:

Residual Error for Jtc Mathematical Models

Vehicle Theft



Using *CrimeStat* for Geographic Profiling

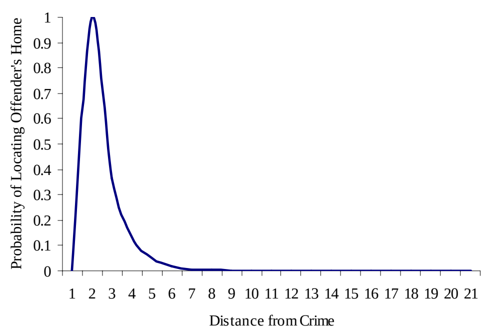
Brent Snook, Memorial University of Newfoundland,
Paul J. Taylor, University of Liverpool, Liverpool
Craig Bennell, Carleton University, Ottawa

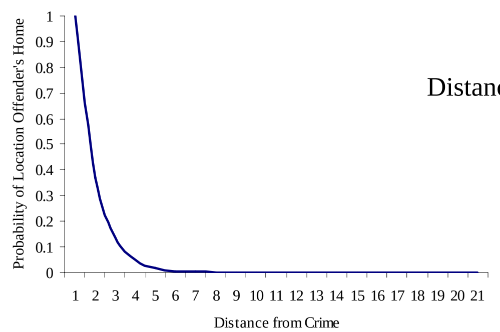
A challenge for researchers providing investigative support is to use information about crime locations to prioritize geographic areas according to how likely they are to contain the offender's residence. One prescient solution to this problem uses *probability distance functions* to assign a likelihood value to the activity space around each crime location. A research goal is to identify the function that assigns the highest likelihood to the offender's actual residence, since this should prove more efficient in future investigations.

CrimeStat was used to test of the effectiveness of two functions for a sample of 68 German serial murder cases, using a measure known as *error distance*. The top figures below illustrate the two functions used and the bottom figures portray the corresponding effectiveness of the functions by plotting the percentage of the sample 'located' by error distance. A steeper effectiveness curve indicates that home locations were closer to the point of highest probability and that, consequently, the probability distance function was more efficient. In this particular test, no difference was found between the two functions in their ability to classify geographic areas.

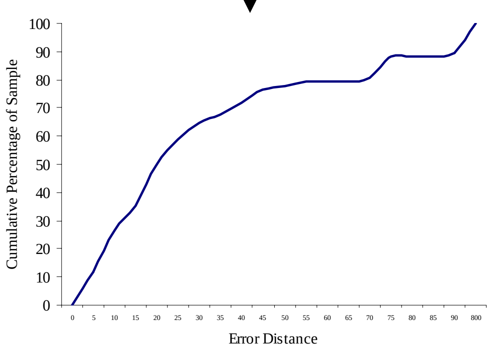
Truncated Negative Exponential

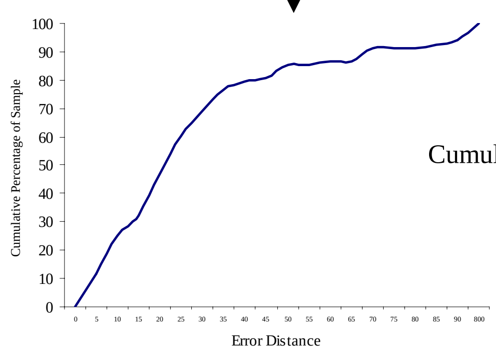
Negative Exponential





Distance Decay Model





Cumulative Error Distance

Original Article: Taylor, P.J., Bennell, C., & Snook B. (2002) *Problems of Classification in Investigative Psychology.* Proceedings of the 8th Conference of the International Federation of Classification Societies, Krakow, Poland

As can be seen in figure 10.9, the truncated negative exponential fit the data well from 0 to about 5 miles, but then became poorer than other models for longer distances. The negative exponential model was not as good as the truncated for distances up to about 5 miles, but was better for distances beyond that point. The normal distribution was good for distances from about 10 miles and farther. The lognormal was not particularly good for any distances other than at 0 miles, nor was the linear.

The degree of predictability varied by type of crime. For some types, particularly property crimes, the fit was reasonably good. I obtained R^2 in the order of 0.86 to 0.96 for burglary, robbery, assault, larceny, and auto theft. For other types of crime, particularly violent crimes, the fit was not very good with R^2 values in the order of 0.53 (rape), 0.41 (arson) and 0.30 (homicide). These R^2 values were for the entire distance range; for any particular distance, however, the predictability varied from very high to very low.

In modeling distance decay with a mathematical function, a user has to decide which part of the distribution is the most important as no simple mathematical function will normally fit all the data (even approximately). In these cases, I assumed that the near distances were more important (up to, say, 5 miles) and, therefore, selected the model which 'best' fit those distances (see table 10.2). However, it was not always clear which model was best, even with that limited criteria.

Problems with Mathematical Distance Decay Functions

There are several reasons that mathematical models of distance decay distributions, such as illustrated in the Jtc routine, do not fit data very well. First, as mentioned earlier, few cities have a completely symmetrical grid structure or even one that is approximately grid-like (there are exceptions, of course). Limitations of physical topography (mountains, oceans, rivers, lakes) as well as different historical development patterns makes travel asymmetrical around most locations.

Second, there is population density. Since most metropolitan areas have much higher intensity of land use in the center (i.e., more activities and facilities), travel tends to be directed towards higher land use intensity than away from them. For origin locations that are not directly in the center, travel is more likely to go towards the center than away from it.

This would be true of an offender as well. If the person were looking for either persons or property as 'targets', then the offender would be more likely to travel towards the metropolitan center than away from it. Since most metropolitan centers have street networks that were laid out much earlier, the street network tends to be irregular. Consequently, trips will vary by location within a metropolitan area. One would expect shorter trips by an offender living close to the metropolitan center than one living farther away; shorter trips for offenders living in more built-up areas than in lower density areas; shorter trips for offenders in mixed use neighborhoods than in strictly residential neighborhoods; and so forth. Thus, the distribution of trips of any sort (in our case, crime trips from a residential location to a crime location), will tend to follow an irregular, distance decay type of distribution. Simple mathematical models will not fit the data very well and will make many errors.

Third, the selection of a best mathematical function is partly dependent on the interval size used for the bins. In the above examples, an interval size of 0.25 miles was used to calculate the frequency distribution. With a different interval size (e.g., 0.5 miles), however, a slightly different distribution is obtained. This effects the mathematical function that is selected as well as the parameters that are estimated. For example, the issue of whether there is a safety zone near the offender's residence from which there is decreased activity or not is partly dependent on the interval size. With a small interval, the zone may be detected whereas with a slightly larger interval the subtle distinction in measured distances may be lost. On the other hand, having a smaller interval may lead to unreliable estimates since there may be few cases in the interval. Having a technique depend on the interval size makes it vulnerable to mis-specification.

Uses of Mathematical Distance Decay Functions

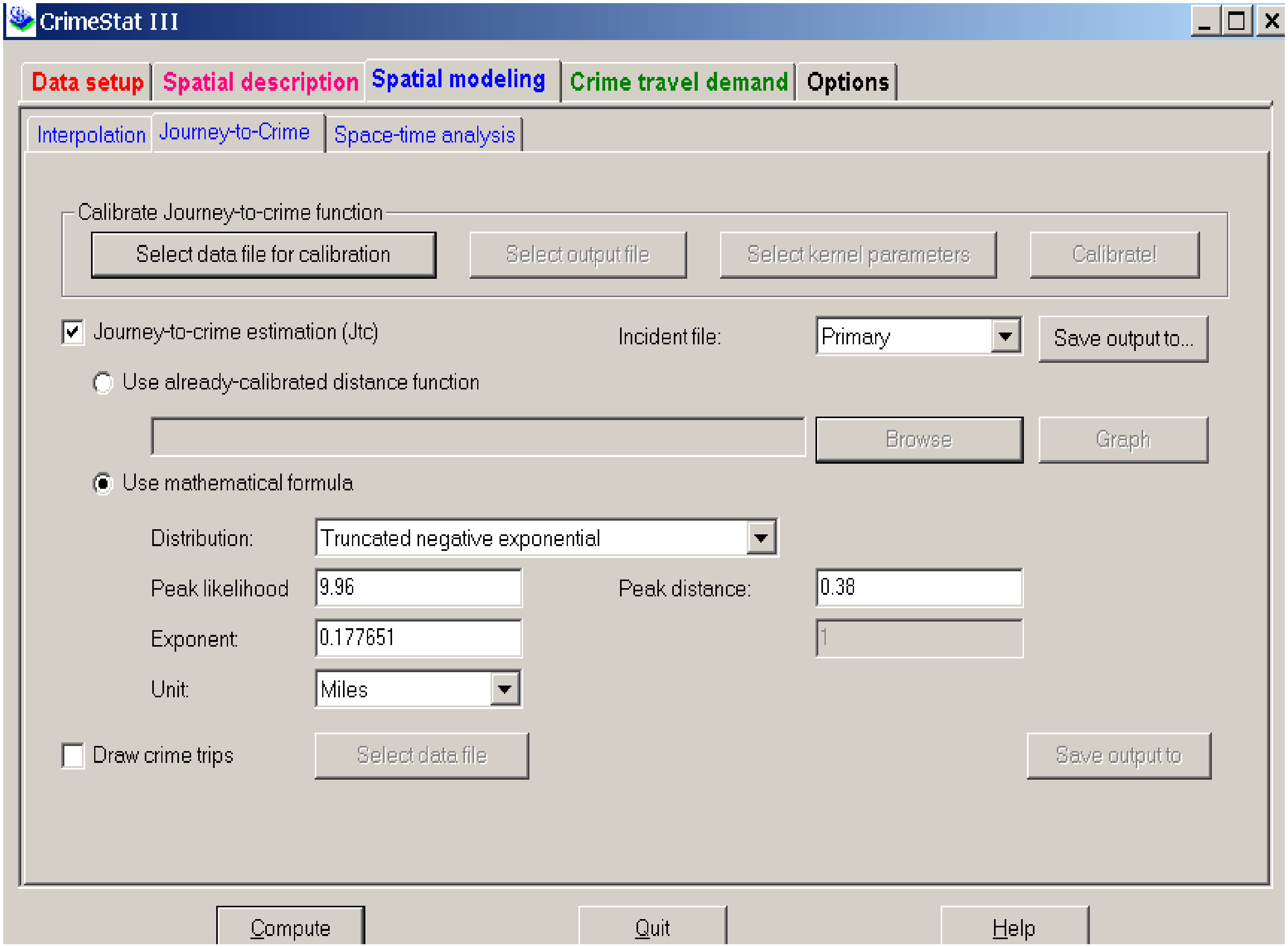
Does this mean that one should not use mathematical distance functions? I would argue that under most circumstances, a mathematical function will give less precision than an empirically-derived one (see below). However, there are two cases when a mathematical model would be appropriate. First, if there is either no data or insufficient data to model the empirical travel distribution, the use of a mathematical model can serve as an approximation. If the user has a good sense of what the distribution looks like, then a mathematical model may be used to approximate the distribution. However, if a poorly defined function is selected, then the selected function may produce many errors.

A second case when mathematical models of distance decay would be appropriate is in theory development or application. Many models of travel behavior, for example, assume a simple distance decay type of function in order simplify the allocation of trips over a region. This is a common procedure in travel demand modeling where trips from each of many zones are assigned to every other zone using a gravity type of function (Stopher and Meyburg, 1975; Field and MacGregor, 1987). Even though the model produces errors because it assumes uniform travel behavior in all directions, the errors are corrected later in the modeling process by adjusting the coefficients for allocating trips to particular roads (traffic assignment). The model provides a simple device and the errors are corrected down the line. Still, I would argue that an empirically-derived distribution will produce fewer errors in allocation and, thus, require less adjustment later on. Errors can never help a model and its better to get it more correct initially to have to adjust it later on; the adjustment may be inadequate. Nevertheless, this is common practice in transportation planning.

The Journey to Crime Routine Using a Mathematical Formula

The Jtc routine which allows mathematical modeling is simple to use. Figure 10.10 illustrates how the user specifies a mathematical function. The routine requires the use of a grid which is defined on the reference file tab of the program (see chapter 3). Then, the

Figure 10.10: **Jtc Mathematical Distance Decay Function**



user must specify the mathematical function and the parameters. In the figure, the truncated negative exponential is being defined. The user must input values for the peak likelihood, the peak distance, and the exponent (see equations 10.43 and 10.44 above). In the figure, since the serial offenses were a series of 18 robberies, the parameters for robbery have been entered into the program screen. The peak likelihood was 9.96% (entered as a whole number - i.e., 9.96); the distance at which this peak likelihood occurred was the second distance interval 0.25-0.50 miles (with a mid-point of 0.38 miles); and the estimated exponent was 0.177651. As mentioned above, the coefficient for the negative exponential part of the equation is estimated internally.

Table 10.3 gives the parameters for the 'best' models which fit the data for the 11 types of crime in Baltimore County. For several of these (e.g., bank robberies), two or more functions gave approximately equally good fits. Note that these parameters were estimated with the Baltimore County data. They will not fit any other jurisdiction. If a user wishes to apply this logic, then the parameters should be estimated anew from existing data. Nevertheless, once they have been calibrated, they can be used for predictions.

The routine can be output to *ArcView*, *MapInfo*, *Atlas*GIS*, *Surfer for Windows*, *Spatial Analyst*, and as an Ascii grid file which can be read by many other GIS packages. All but *Surfer for Windows* require that the reference grid be created by *CrimeStat*.

Distance Modeling Using an Empirically Determined Function

An alternative to mathematical modeling of distance decay is to empirically describe the journey to crime distribution and then use this empirical function to estimate the residence location. *CrimeStat* has a two-dimensional kernel density routine that can calibrate the distance function if provided data on trip origins and destinations. The logic of kernel density estimation was described in chapter 8, and won't be repeated here. Essentially, a symmetrical function (the 'kernel') is placed over each point in a distribution. The distribution is then referenced relative to a scale (an equally-spaced line for two-dimensional kernels and a grid for three-dimensional kernels) and the values for each kernel are summed at each reference location. See chapter 8 for details.

Calibrate Kernel Density Function

The *CrimeStat* calibration routine allows a user to describe the distance distribution for a sample of journey to crime trips. The requirements are that:

1. The data set must have the coordinates of *both* an origin location and a destination location; and

2. The records of all origin and destination locations have been populated with legitimate coordinate values (i.e., no unmatched records are allowed).

Table 10.3

Journey to Crime Mathematical Models for Baltimore County
Parameter Estimates for Percentage Distribution
(Sample Sizes in Parentheses)

ALL CRIMES		
Negative Exponential:	Coefficient:	5.575107
	Exponent:	0.229466
HOMICIDE		
Truncated Negative Exponential:	Peak likelihood	14.02%
	Peak distance	0.38 miles
	Exponent	0.064481
RAPE		
Lognormal:	Mean	3.144959
	Standard Deviation	4.546872
	Coefficient	0.062791
ASSAULT		
Truncated Negative Exponential:	Peak likelihood	27.40%
	Peak distance	0.38 miles
	Exponent	0.181738
ROBBERY		
Truncated Negative Exponential:	Peak likelihood	9.96%
	Peak distance	0.38 miles
	Exponent	0.177651
COMMERCIAL ROBBERY		
Truncated Negative Exponential:	Peak likelihood	4.9455%
	Peak distance	0.625 miles
	Exponent	0.151319

Table 10.3 (continued)

BANK ROBBERY

Truncated Negative Exponential:	Peak likelihood	9.96%
	Peak distance	5.75 miles
	Exponent	0.139536

BURGLARY

Truncated Negative Exponential:	Peak likelihood	20.55%
	Peak distance	0.38 miles
	Exponent	0.162907

AUTO THEFT

Truncated Negative Exponential:	Peak likelihood	4.81%
	Peak distance	0.63 miles
	Exponent	0.212508

LARCENY

Truncated Negative Exponential:	Peak likelihood	4.76%
	Peak distance	0.38 miles
	Exponent	0.193015

ARSON

Truncated Negative Exponential:	Peak likelihood	38.99%
	Peak distance	0.38 miles
	Exponent	0.093469

Data Set Definition

The steps are relatively easy. First, the user defines a calibration data set with both origin and destination locations. Figure 10.11 illustrates this process. As with the primary and secondary files, the routine reads *ArcView* 'shp', *dBase* 'dbf', Ascii 'txt', and *MapInfo* 'dat' files. For both the origin location (e.g., the home residence of the offender) and the destination location (i.e., the crime location), the names of the variables for the X and Y coordinates must be identified as well as the type of coordinate system and data unit (see chapter 3). In the example, the origin locations has variable names of HomeX and HomeY and the destination locations has variable names of IncidentX and IncidentY for the X and Y coordinates of the two locations respectively. However, any name is acceptable as long as the two locations are distinguished.

The user should specify whether there are any missing values for these four fields (X and Y coordinates for both origin and destination locations). By default, *CrimeStat* will ignore records with blank values in any of the eligible fields or records with non-numeric values (e.g.,alphanumeric characters, #, *). Blanks will always be excluded unless the user selects <none>. There are 8 possible options:

1. <blank> fields are automatically excluded. This is the default
2. <none> indicates that no records will be excluded. If there is a blank field, *CrimeStat* will treat it as a 0
3. 0 is excluded
4. –1 is excluded
5. 0 and –1 indicates that both 0 and -1 will be excluded
6. 0, -1 and 9999 indicates that all three values (0, -1, 9999) will be excluded

Any other numerical value can be treated as a missing value by typing it (e.g., 99)Multiple numerical values can be treated as missing values by typing them, separating each by commas (e.g., 0, -1, 99, 9999, -99).

The program will calculate the distance between the origin location and the destination location for each record. If the units are spherical (i.e., lat/lon), then the calculations use spherical geometry; if the units are projected (either meters or feet), then the calculations are Euclidean (see chapter 3 for details).

Kernel Parameters

Next, the user must define the kernel parameters for calibration. There are five choices that have to be made (Figure 10.12):

1. The method of interpolation. As with the two-dimensional kernel technique described in chapter 8, there are five possible kernel functions:

Figure 10.11: **Jtc Calibration Data Input**

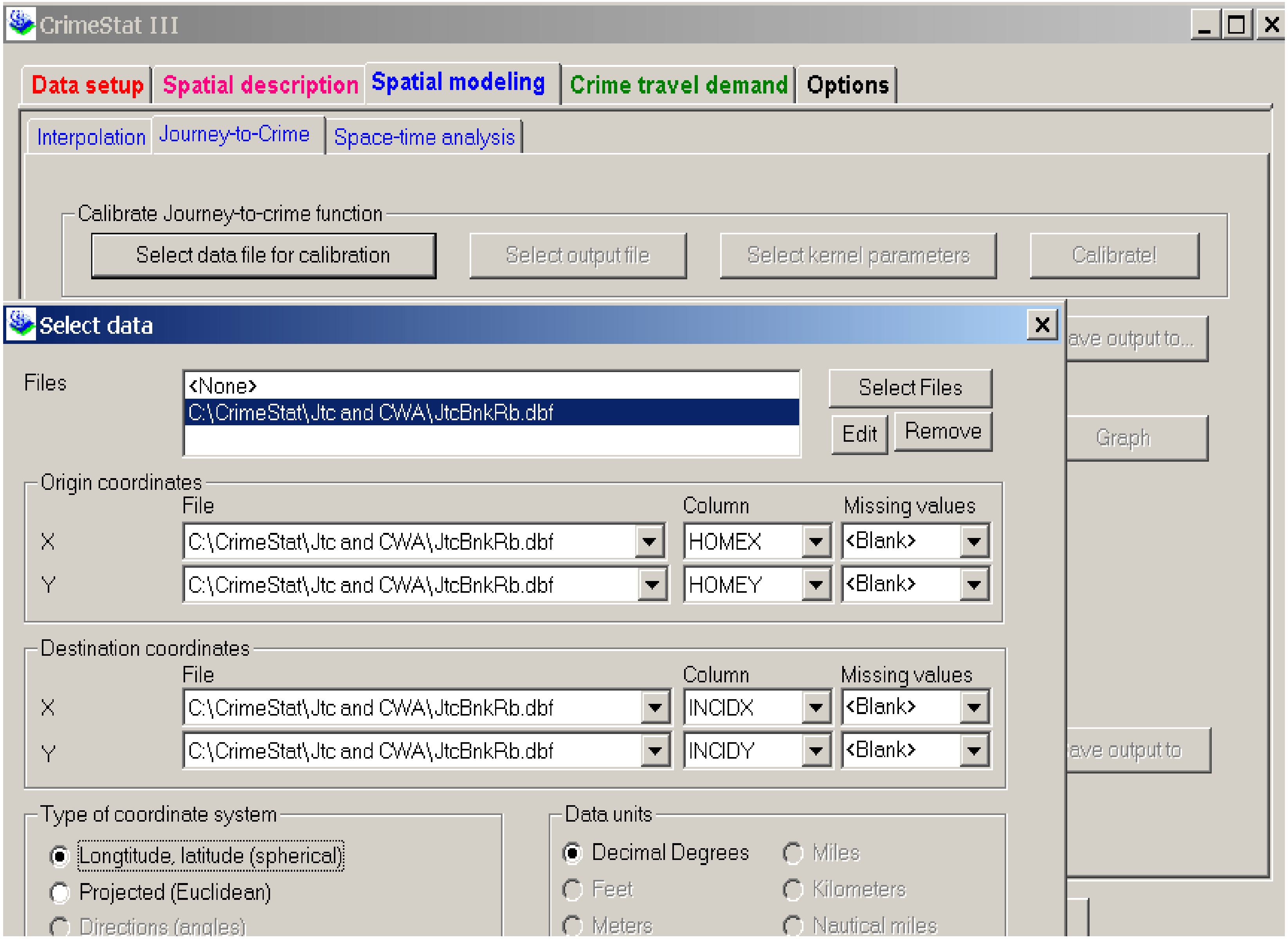
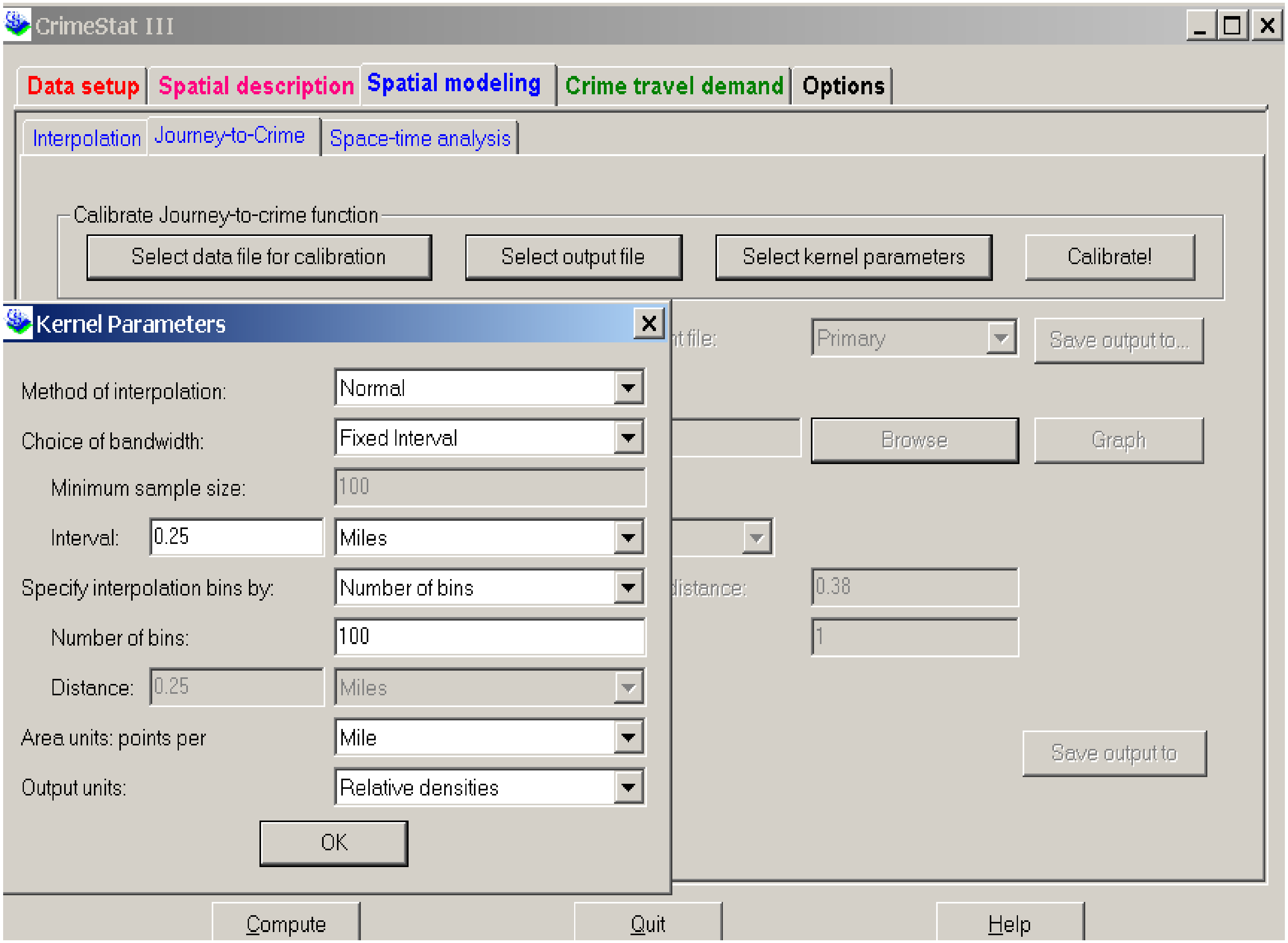


Figure 10.12: Jtc Calibration Kernel Parameters



A. Normal (the default);
B. Quartic;
C. Triangular (conical);
D. A negative exponential (peaked); and
E. A uniform (flat) distribution.

2. Choice of bandwidth. The bandwidth is the width of the kernel function. For a normal kernel, it is the standard deviation of the normal distribution whereas for the other four kernels (quartic, triangular, negative exponential, and uniform), it is the radius of the circle defined by the kernel. As with the two-dimension kernel technique, the bandwidth can be fixed in length or adaptive (variable in length). However, for the one-dimensional kernel, the fixed bandwidth is the default since an even estimate over an equal number of intervals (bins) is desirable. If the fixed bandwidth is selected, the interval size must be specified and the units (in miles, kilometers, feet, meters, and nautical miles). The default is 0.25 mile intervals. If the adaptive bandwidth is selected, the user must identify the minimum sample size that the bandwidth should incorporate; in this case, the bandwidth is widened until the specified sample size is counted.

3. The number of interpolation bins. The bins are the intervals along the distance scale (from 0 up to the maximum distance for a journey to crime trip) and are used to estimate the density function. There are two choices. First, the user can specify the number of intervals (the default choice with 100 intervals). In this case, the routine calculates the maximum distance (or longest trip) between the origin location and the destination location and divides it by the specified number of intervals (e.g., 100 equal-sized intervals). The interval size is dependent on the longest trip distance measured. Second, the user can specify the distance between bins (or the interval size). The default choice is 0.25 miles, but another value can be entered. In this case, the routine counts out intervals of the specified size until it reaches the maximum trip distance.

4. The output units. The user specifies the units for the density estimate (in units per mile, kilometer, feet, meters, and nautical miles).

5. The output calculations. The user specifies whether the output results are in probabilities (the default) or in densities. For probabilities, the sum of all kernel estimates will equal 1.0. For densities, the sum of all kernel estimates will equal the sample size.

Saved Calibration File

Third, the user must define an output file to save the empirically determined function. The function is then used in estimating the likely home residence of a particular function. The choices are to save the file as a 'dbf' or Ascii text file. The saved file then can be used in the Jtc routine. Figure 10.13 illustrates the output file format.

Calibrate

Fourth, the calibrate button runs the routine. A calibration window appears and indicates the progress of the calculations. When it is finished, the user can view a graph illustrating the estimated distance decay function (Figure 10.14). The purpose is to provide quick diagnostics to the user on the function and selection of the kernel parameters. While the graph can be printed, it is not a high quality print. If a high quality graph is needed, the output calibration file should be imported into a graphics program.

Examples from Baltimore County

Let's illustrate this method by showing the results for the same data sets that were calculated above in the mathematical section (figures 10.4-10.8). In all cases, the normal kernel function was used. The bandwidth was 0.25 miles except for the bank robbery data set, which had only 176 cases, and the homicide data set, which only had 137 cases; because of the small sample sizes, a bandwidth of 0.50 miles was used for these two data sets. The interval width selected was a distance of 0.25 miles between bins (0.5 miles for bank robberies and homicides) and probabilities were output.

Figure 10.15 shows the kernel estimate for all crimes (41,426 trips). A frequency distribution was calculated for the same number of intervals and is overlaid on the graph. It was selected to be comparable to the mathematical function (see figure 10.4). Note how closely the kernel estimate fits the data compared to the negative exponential mathematical function. The fit is good for every value but the peak value; that is because the kernel averages several intervals together to produce an estimate.

Figure 10.16 shows the kernel estimate for larceny thefts. Again, the kernel method produces a much closer fit as a comparison with figure 10.5 will show. Figure 10.17 shows the kernel estimate for vehicle thefts. Figure 10.18 shows the kernel estimate for bank robberies and figure 10.19 shows the kernel estimate for homicides. An inspection of these graphs shows how well the kernel function fits the data, compared to the mathematical function, even when the data are irregularly spaced (in vehicle thefts, bank robberies, and homicides). Figure 10.20 compares the distance decay functions for homicides committed against strangers compared to homicides committed against known victims.

In short, the Jtc calibration routine allows a much closer fit to the data than any of the simpler mathematical functions. While it's possible to produce a complex mathematical function that will fit the data more closely (e.g., higher order polynomials), the kernel method is much simpler to use and gives a good approximation to the data.

Figure 10.3:

Journey to Crime Travel Demand Functions

Five Mathematical Functions

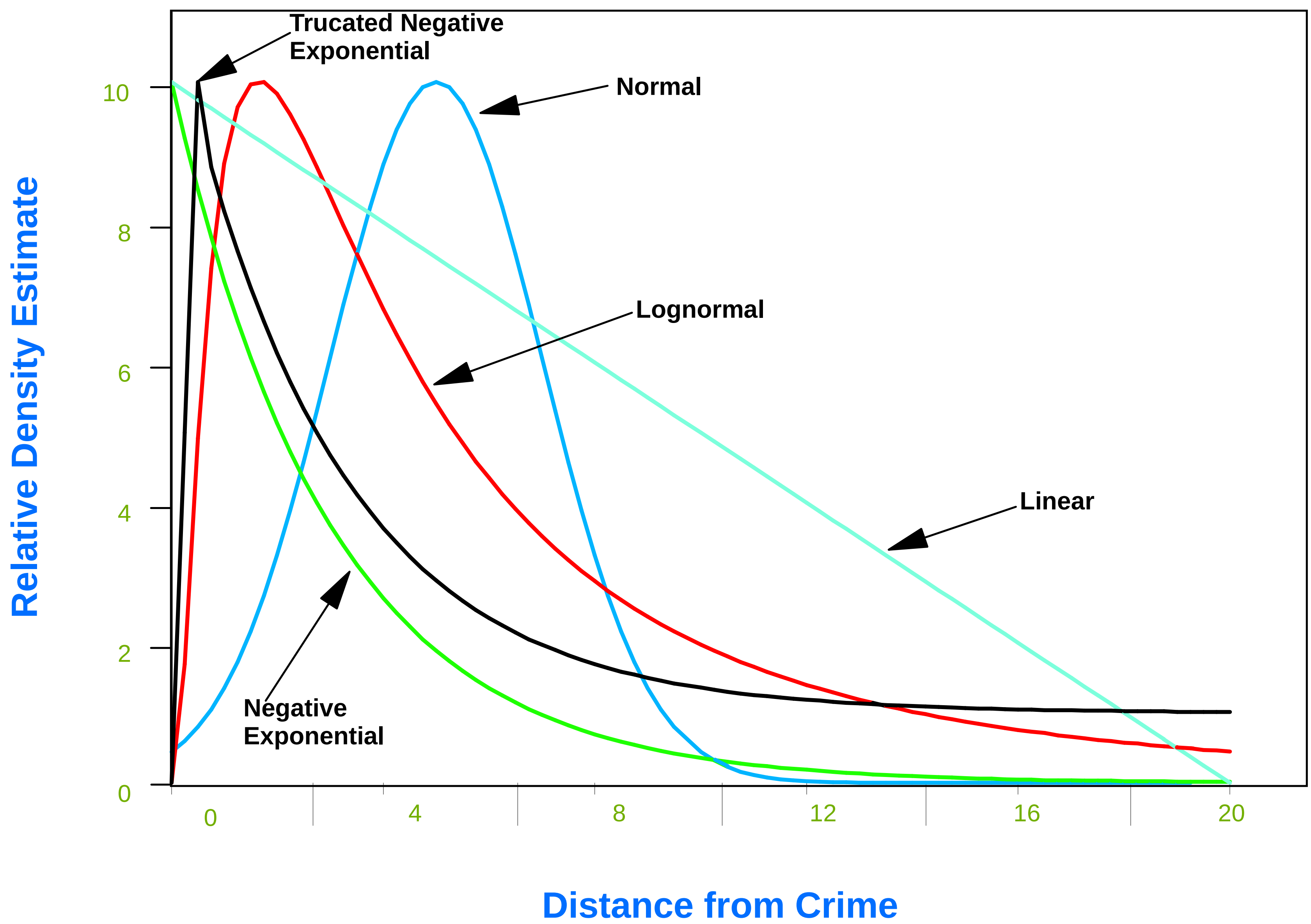


Figure 10.14: Jtc Calibration Graphic Output

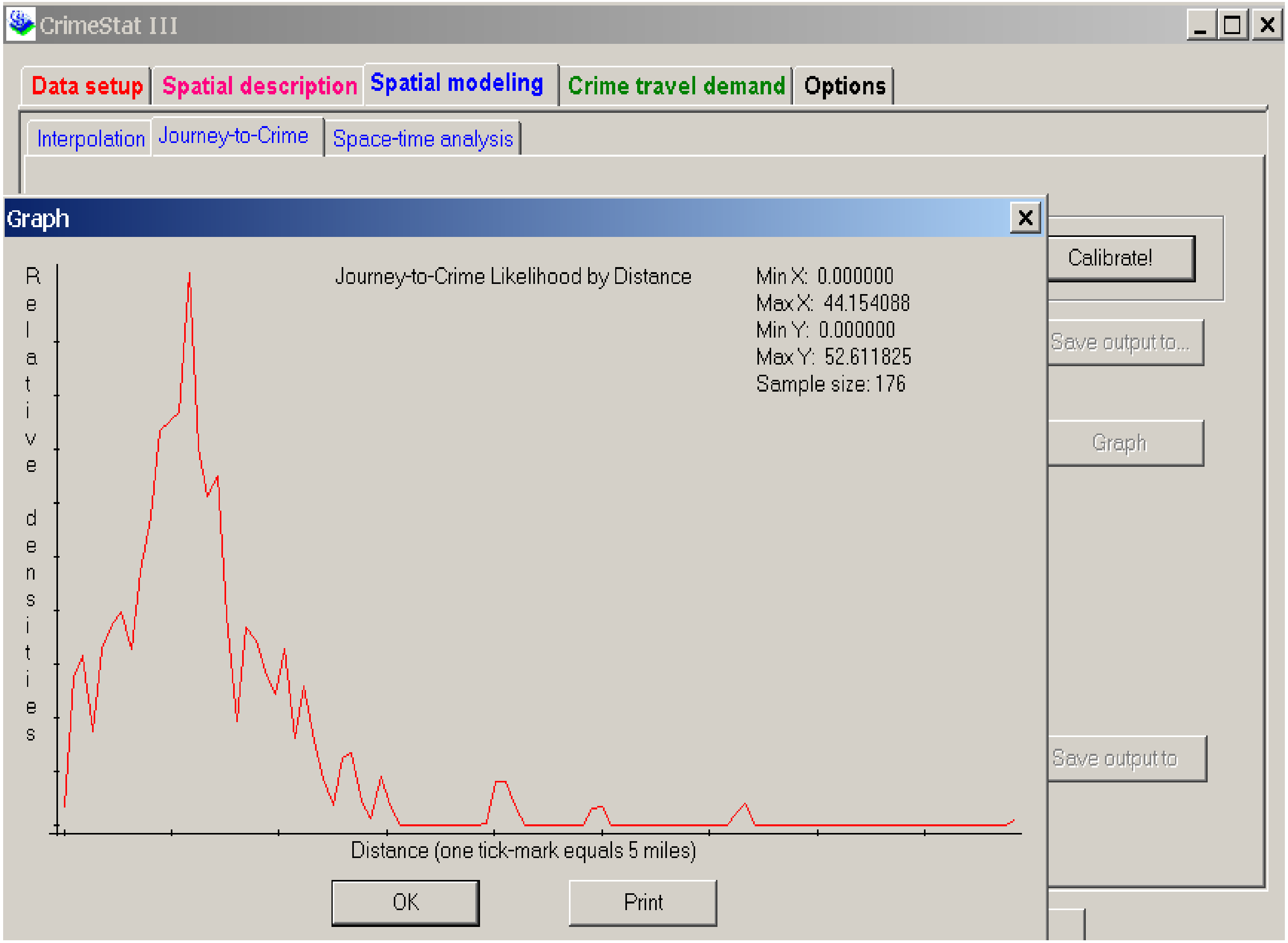


Figure 10.15:

Journey to Crime Distances: All Crimes

Frequencies and Kernel Density Estimate

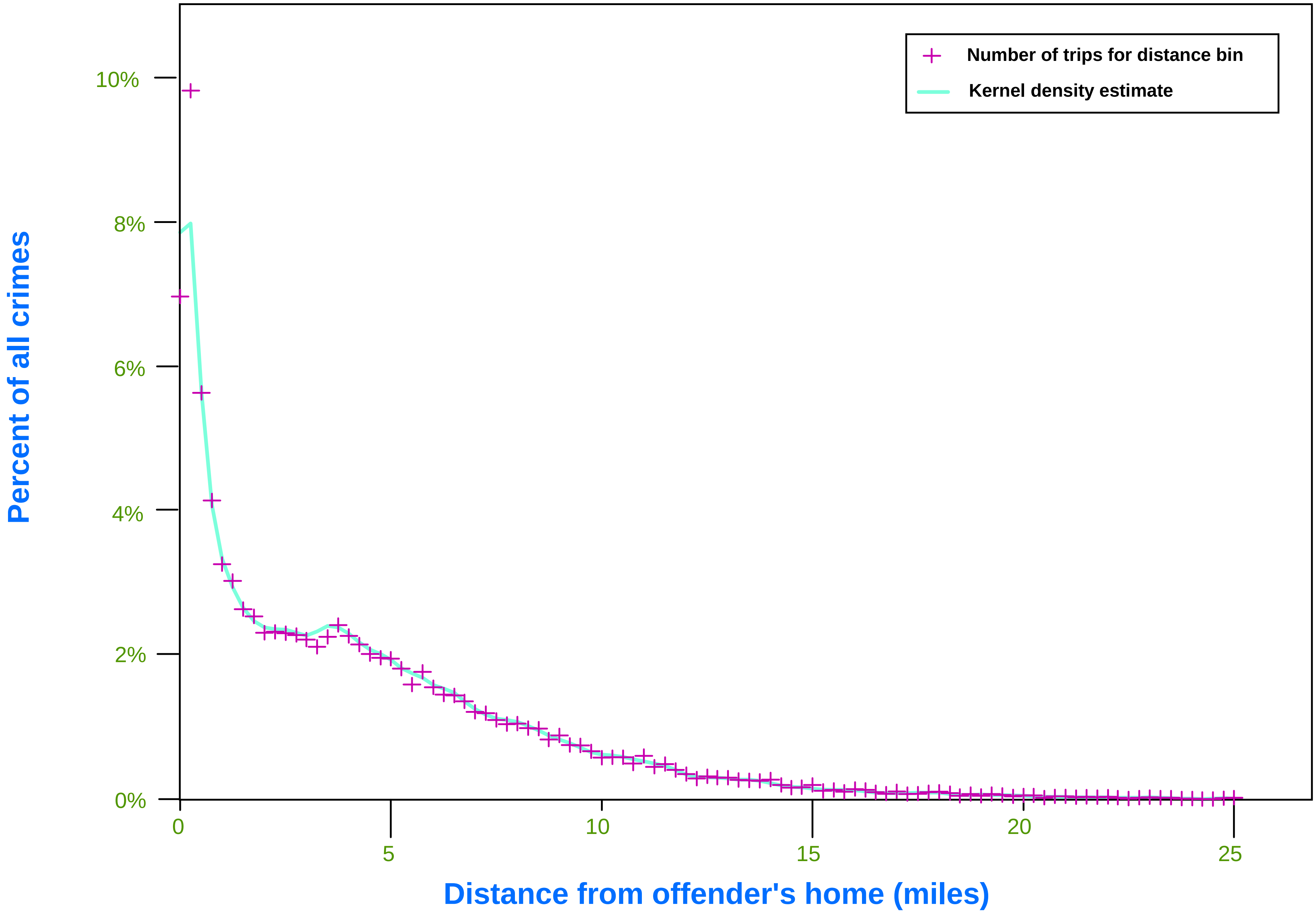


Figure 10.16:

Journey to Crime Distances: Larceny

Frequencies and Kernel Density Estimate

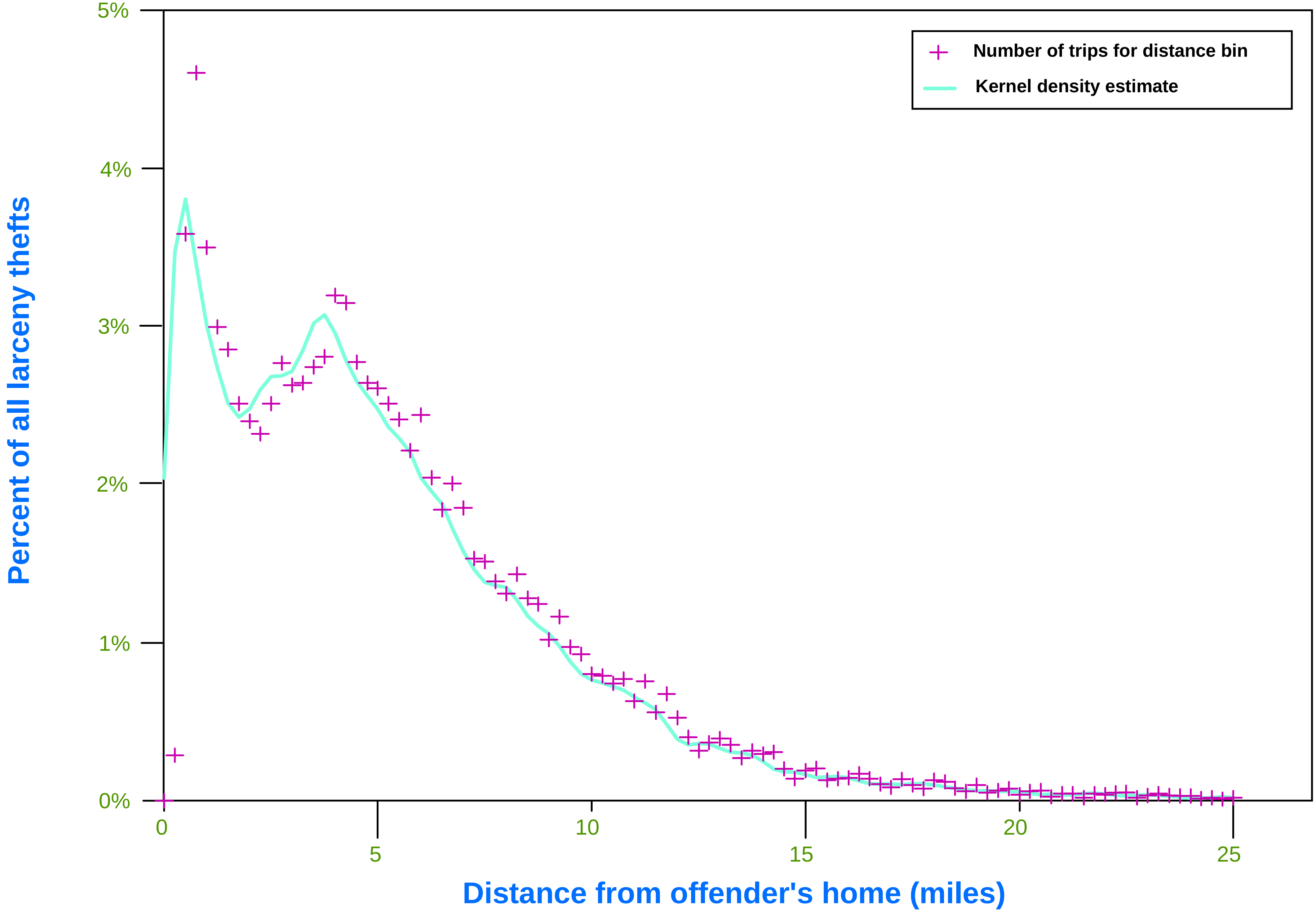


Figure 10.17:

Journey to Crime Distances: Vehicle Theft

Frequencies and Kernel Density Estimate

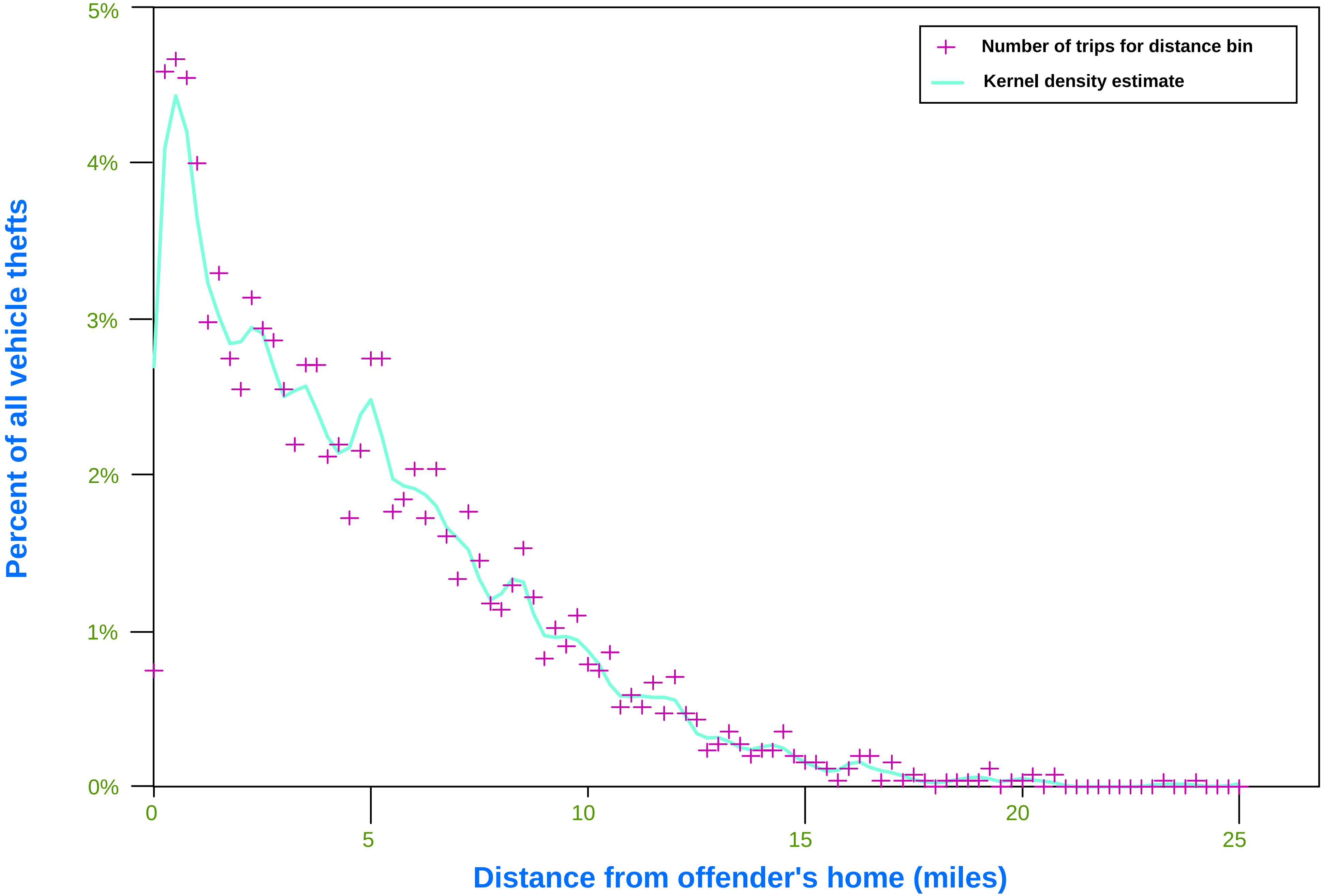
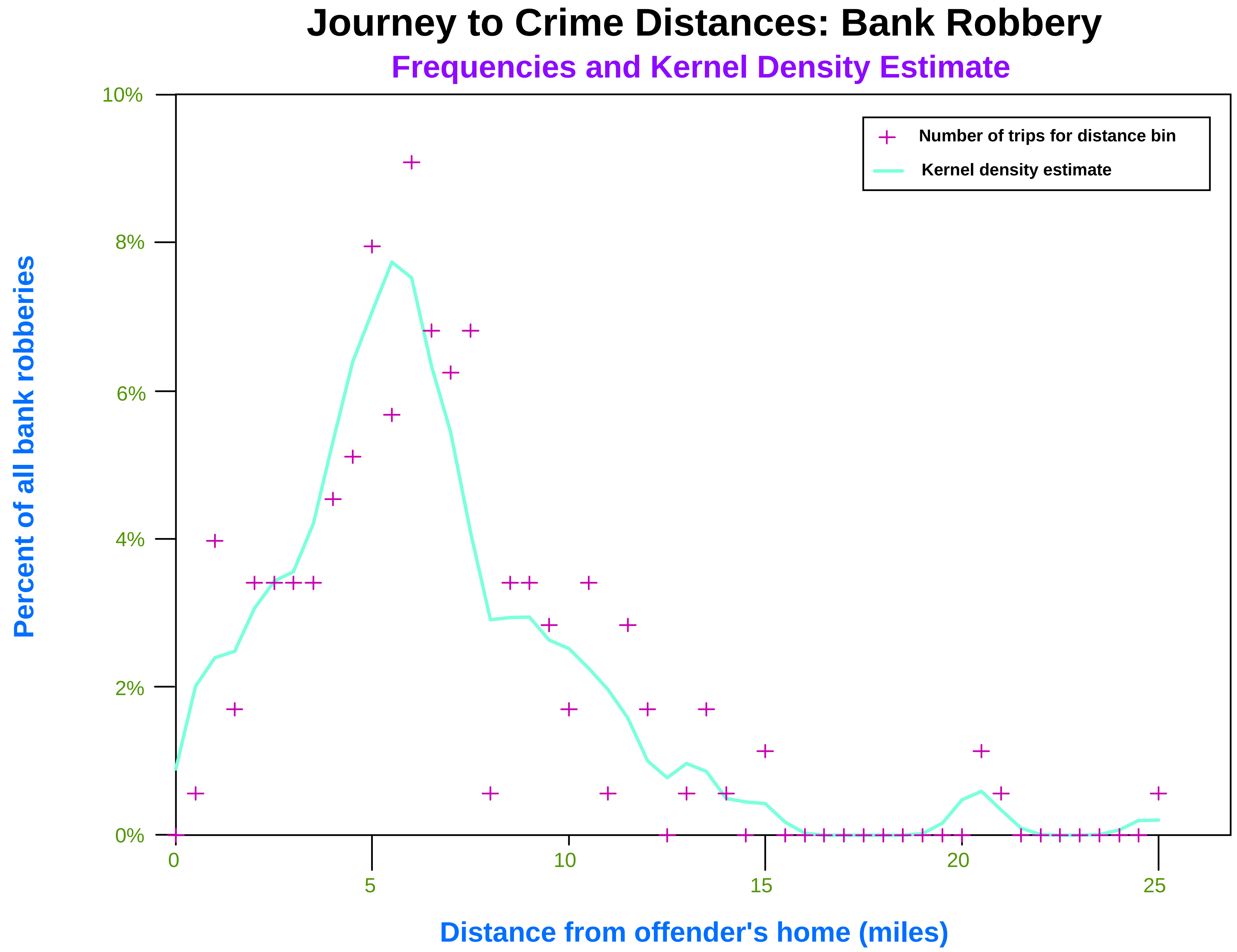


Figure 10.18:



Journey to Crime Distances: Bank Robbery
Frequencies and Kernel Density Estimate
Number of trips for distance bin
Kernel density estimate
Percent of all bank robberies
10%
8%
6%
4%
2%
0%
0
5
10
15
20
25
Distance from offender's home (miles)

Figure 10.19:

Journey to Crime Distances: Homicide

Frequencies and Kernel Density Estimate

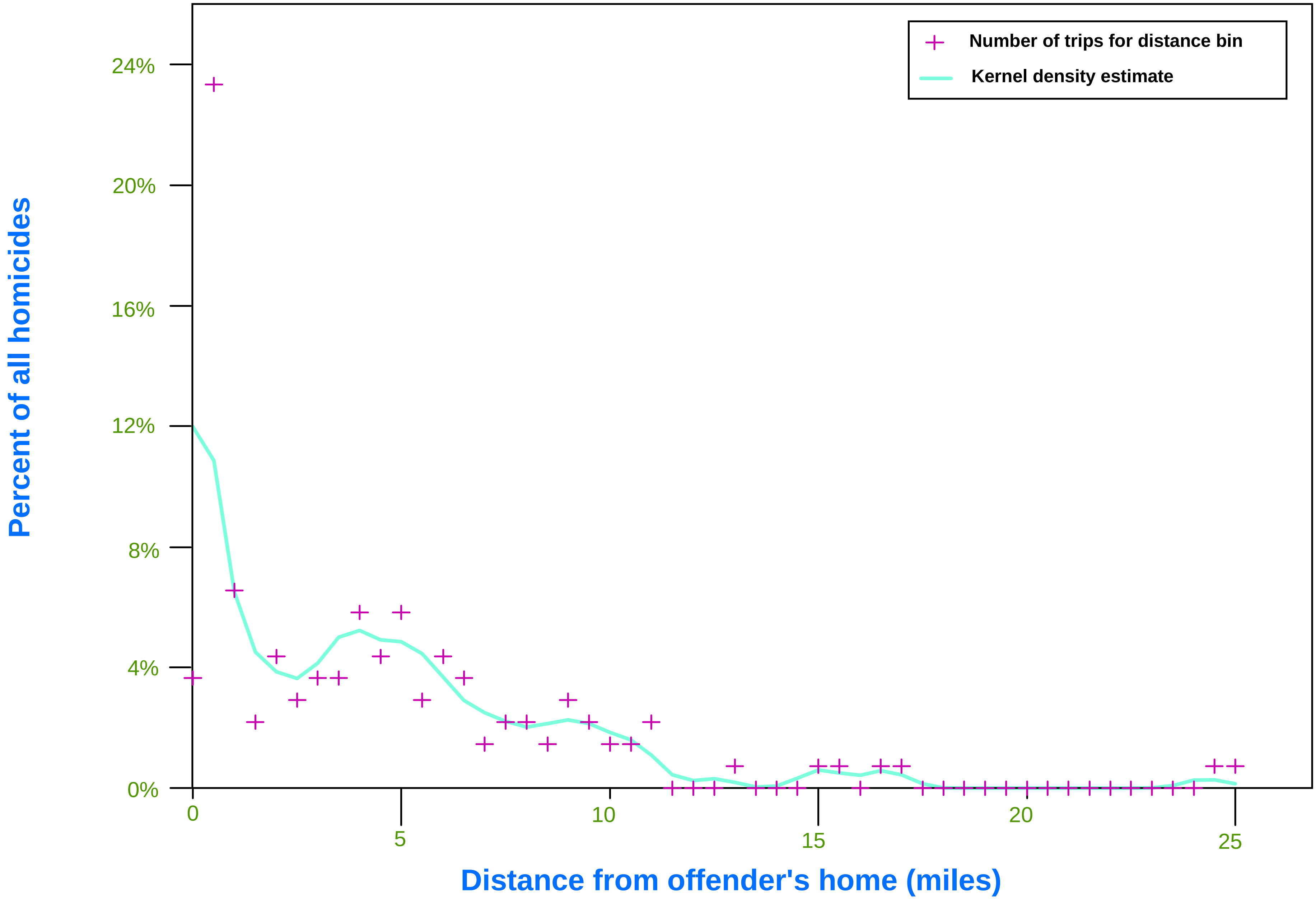
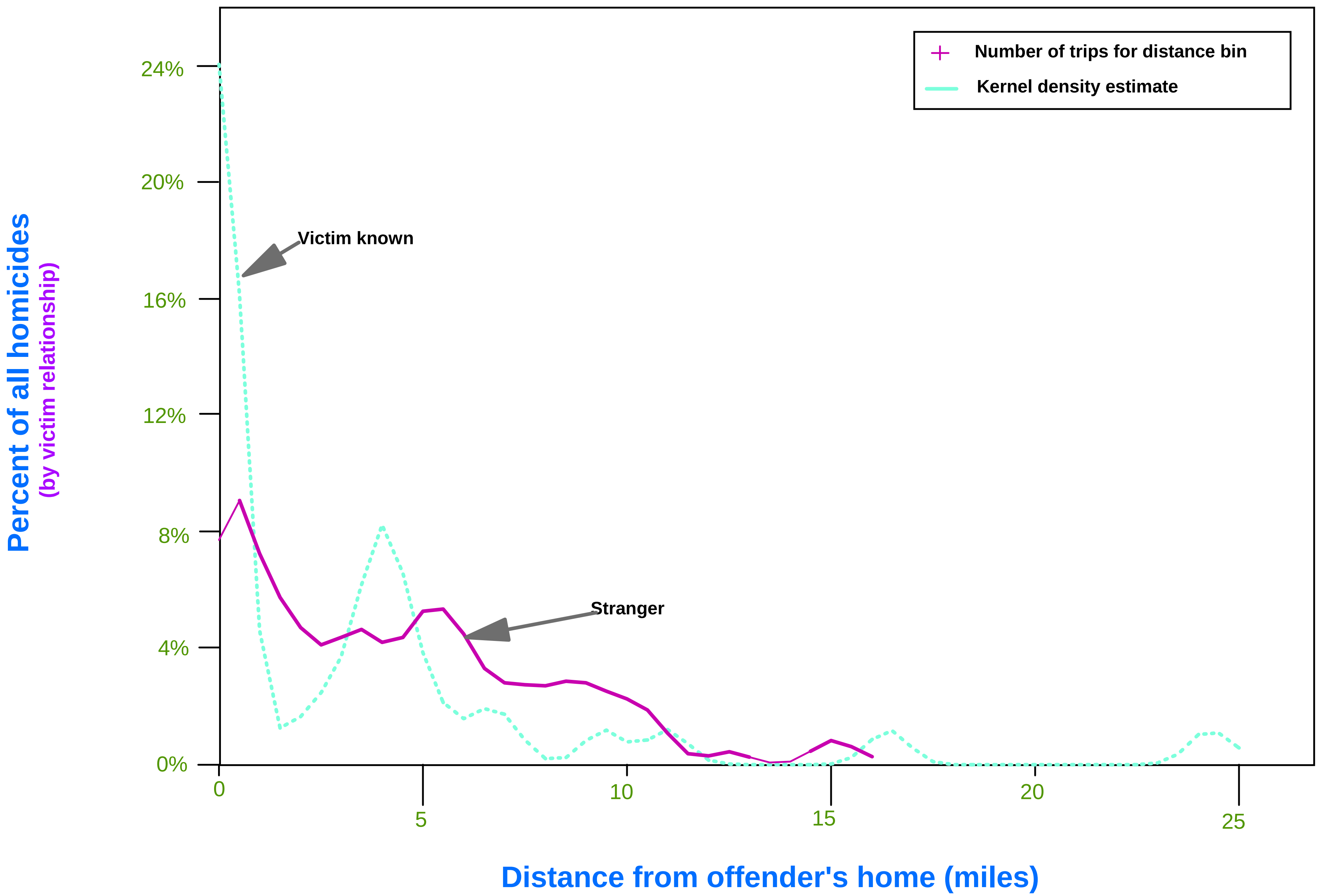


Figure 10.20:

Journey to Crime Distances: Homicide by Victim Relationship

Frequencies and Kernel Density Estimate

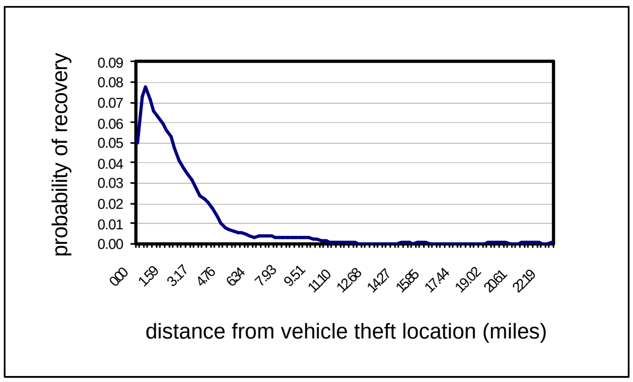


Using Journey-To-Crime Routine for Journey-After-Crime Analysis

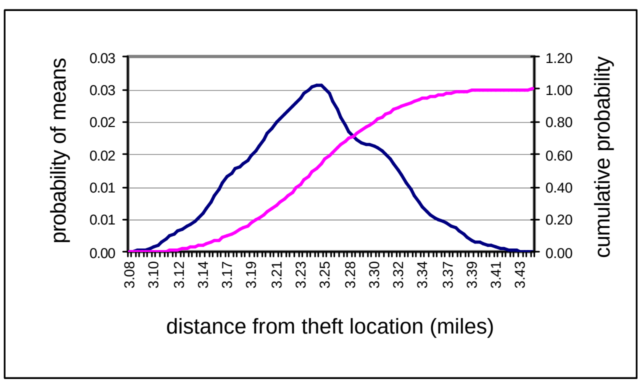
Yongmei Lu
Department of Geography
Southwest Texas State University
San Marcos, TX

The study of vehicle theft recovery locations can fill a gap in the knowledge about criminal travel patterns. Although the journey-to-crime routine of *CrimeStat* was designed to analyze the distance between offense location and offender's residential location, it can be used to describe the distance between vehicle theft location and the corresponding recovery location.

There were more than 3000 vehicle thefts in the City of Buffalo in 1998. Matching the offenses with vehicle recoveries in the same year, 1600 location pairs were identified for a journey-after-vehicle-theft analysis. To evaluate the randomness of the distances, 1000 groups of simulations were conducted. Every group contains 1600 simulated trips of journey-after-vehicle-theft. The results indicate that 1) short distances dominate journey-after-vehicle-theft, and 2) the observed trips are significantly shorter than the random trips given the distribution of possible vehicle theft and recovery locations.



Probability of recovering a stolen vehicle by distance from vehicle theft location



Distribution of mean distances of simulated vehicle theft-recovery location pairs.

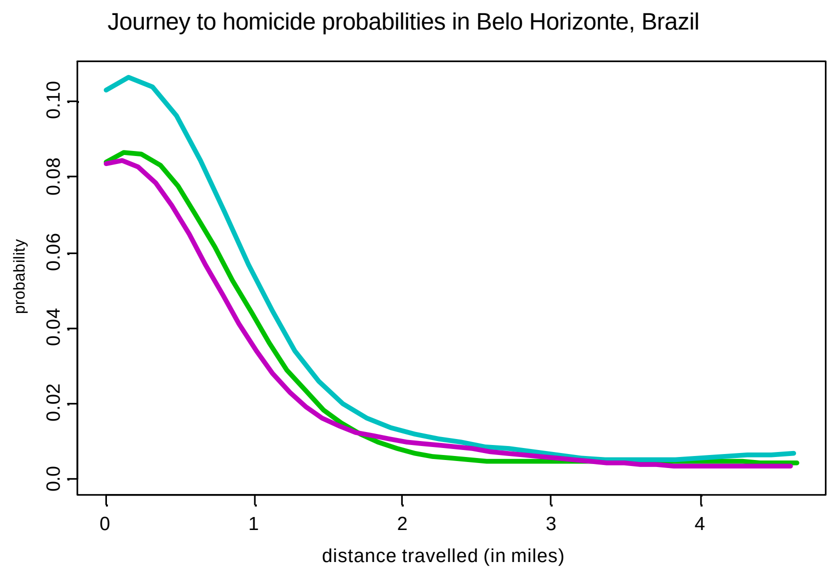
Using Journey to Crime for Different Age Groups of Offenders

Renato Assunção, Cláudio Beato, Bráulio Silva
CRISP, Universidade Federal de Minas Gerais , Brazil

CrimeStat offers a method for analysing the distance between the crime scene and the residence of the offender using the journey to crime routine within the spatial modeling module. We analysed homicide incidents in Belo Horizonte, a Brazilian city of 2 million inhabitants, for the period January 1996 – December 2000. We used 496 homicide cases for which the police identified an offender who was living in Belo Horizonte, and for which both the crime location and offender residence could be identified. The cases were divided into three groups according to the offender's age: 1) 14 to 24 (N=201); 2) 25 to 34 (N=176); and 3) 35 or older (N=119). The journey to crime calibration routine was used to produce a probability curve *P(d)* that gives the approximate chance of an offender travelling approximately distance *d* to commit the crime.

We used the normal kernel, a fixed bandwidth of 1000 meters, 100 output bins, and the probability (or proportion of all points) option, rather than densities. This is to allow comparisons between the three age groups since they have different number of homicides. We tested for each age group separately and directed the output to a text file to analyse the three groups simultaneously.

The green, blue, and purple curves are associated with the 14-24, 25-34, 35+ year olds respectively. There are more similarities than differences between the groups. Most homicides are committed near to the residence of the offenders with between 60% t o 70% closer than one mile from their home. However, the curve does not vanish totally even for large distances because there are around 15% of offenders, of any age group, travelling longer than 3 miles to commit the crime. The oldest offenders travel longer distances, on average, followed by the youngest group, with the 25-34 year olds travelling the shortest distances.



The Journey to Crime Routine Using the Calibrated File

After the distance decay function has been calibrated and saved as a file, the file can be used to calculate the likelihood surface for a serial offender. The user specifies the name of the already-calibrated distance function (as a 'dbf' or an Ascii text file) and the output format. As with the mathematical routine, the output can be to *ArcView*, *MapInfo*, *Atlas*GIS*, *Surfer for Windows*, *Spatial Analyst*, and as an Ascii grid file which can be read by many other GIS packages. All but *Surfer for Windows* require that the reference grid be created by *CrimeStat*.

The result is produced in three steps:

1. The routine calculates the distance between each reference cell of the grid and each incident location;

2. For each distance measured, the routine looks up the calculated value from the saved calibration file; and

3. For each reference grid cell, it sums the values of all the incidents to produce a single likelihood estimate.

Application of the Routine

To illustrate the techniques, the results of the two methods on a single case are compared. The case has been selected because the routines accurately estimate the offender's residence. This was done to demonstrate how the techniques work. In the next section, I'll ask the question about how accurate these methods are in general.

The case involved a man who had committed 24 offenses. These included 13 thefts, 5 burglaries, 5 assaults, and one rape. The spatial distribution was varied; many of the offenses were clustered but some were scattered. Since there were multiple types of crimes committed by this individual, a decision had to be made over which model to use to estimate the individual's residence. In this case, the theft (larceny) model was selected since that was the dominant type of crime for this individual.

For the mathematical function, the truncated negative exponential was chosen from table 10.3 with the parameters being:

Peak likelihood	4.76%
Peak distance	0.38 miles
Exponent	0.193015

For the kernel density model, the calibrated function for larceny was selected (figure 10.16).

Figure 10.21 shows the results of the estimation for the two methods. The output is from *Surfer for Windows* (Golden Software, 1994). The left pane shows the results of the mathematical function while the right pane shows the results for the kernel density function. The incident locations are shown as circles while the actual residence location of the offender is shown as a square. Since this is a surface model, the highest location has the highest predicted likelihood.

In both cases, the models predicted quite accurately. The discrepancy (error) between the predicted peak location and the actual residence location was 0.66 miles for the mathematical function and 0.36 miles for the kernel density function. For the mathematical model, the actual residence location (square) is seen as slightly off from the peak of the surface whereas for the kernel density model the discrepancy from the peak cannot be seen.

Nevertheless, the differences in the two surfaces show distinctions. The mathematical model has a smooth decline from the peak likelihood location, almost like a cone. The kernel density model, on the other hand, shows a more irregular distribution with a peak location followed by a surrounding 'trough' followed a peak 'rim'. This is due to the irregular distance decay function calibrated for larceny (see figure 10.16). But, in both cases, they more or less identify the actual residence location of the offender.

Choice of Calibration Sample

The calibration sample is critical for either method. Each method assumes that the distribution of the serial offender will be similar to a sample of 'like' offenders. Obviously, distinctions can be made to make the calibration sample more or less similar to the particular case. For example, if a distance decay function of all crimes is selected, then a model (of either the mathematical or kernel density form) will have less differentiation than for a distance decay function from a specific type of crime. Similarly, breaking down the type of crime by, say, mode of operation or time of day will produce better differentiation than by grouping all offenders of the same type together. This process can be taken on indefinitely until there is too little data to make a reliable estimate. An analyst should try to find as close a calibration sample to the actual as is possible, given the limitations of the data.

For example, in our calibration data set, there were 4,694 burglary incidents where both the offender's home residence and the incident location were known. The approximate time of the offense for 2,620 of the burglaries was known and, of these, 1,531 occurred at night between 6 pm and 6 am. Thus, if a particular serial burglar for whom the police are interested in catching tends to commit most of his burglaries at night, then choosing a calibration sample of nighttime burglars will generally produce a better estimate than by grouping all burglars together. Similarly, of the 1,531 nighttime burglaries, 409 were committed by individuals who had a prior relationship with the victim. Again, if the analysts suspect that the burglar is robbing homes of people he knows or is acquainted with, then selecting the subset of nighttime burglaries committed against a known victim

Figure 10.21:

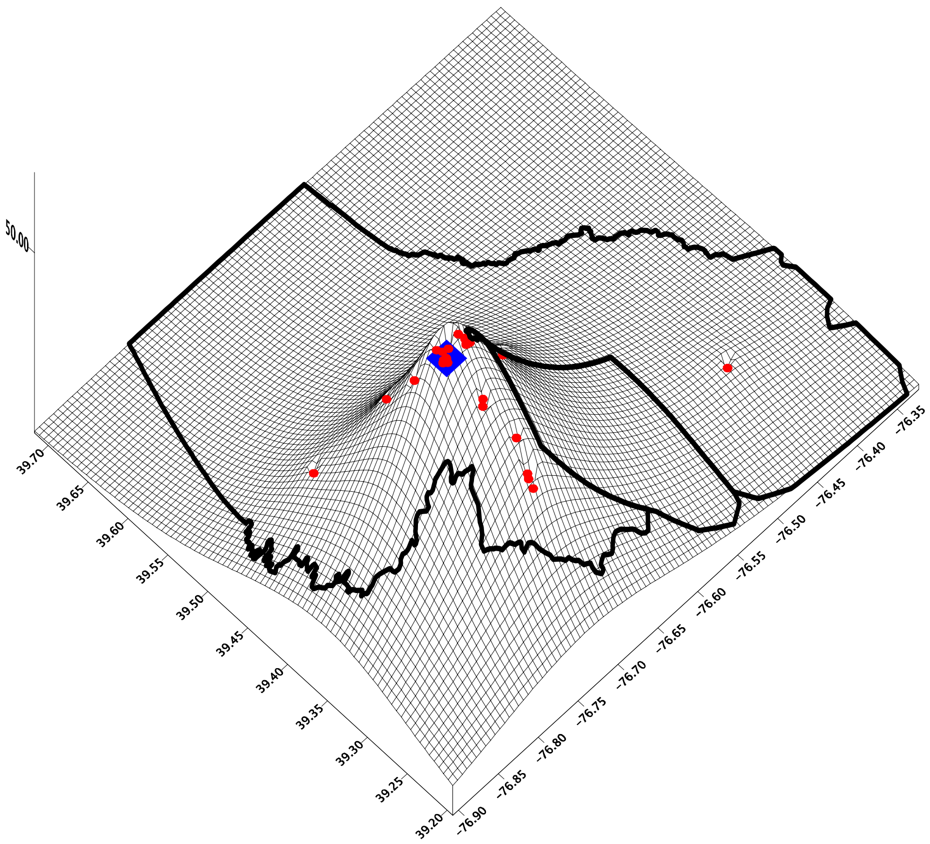
Predicted and Actual Location of Serial Thief

Man Charged with 24 Offenses in Baltimore County

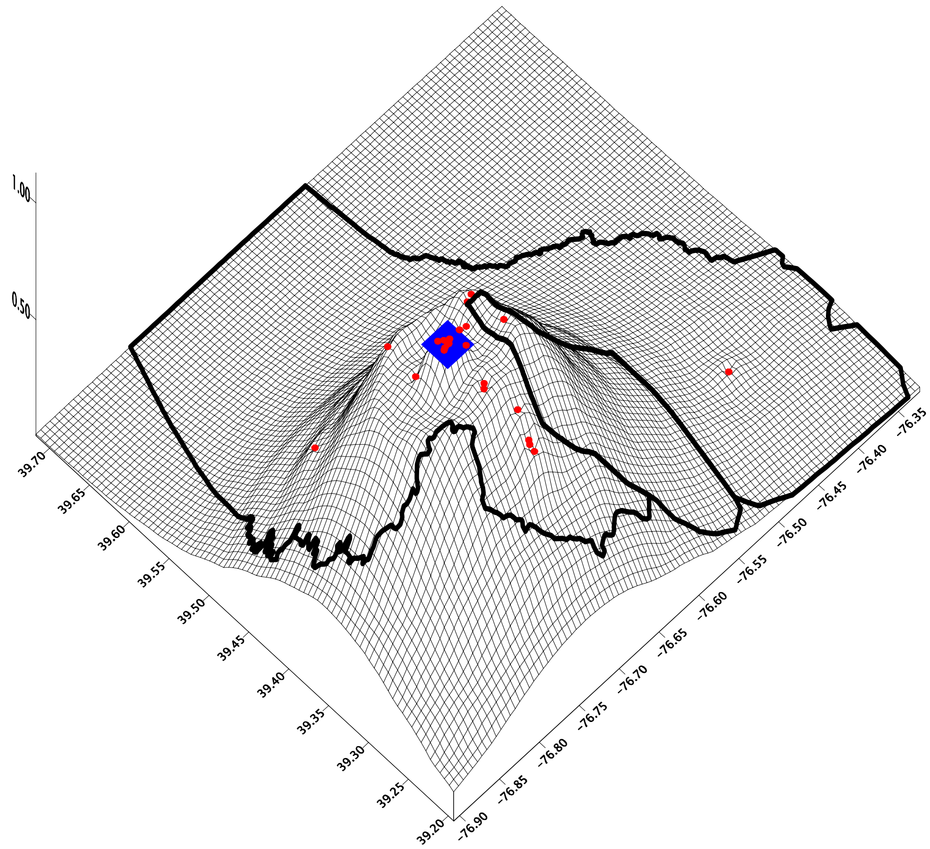
Predicted with Mathematical and Kernel Density Models for Larceny

Residence Location = Square

Crime Locations = Circles



Mathematical Model:
Truncated Negative Exponential



Kernel Density Model

would produce even better differentiation in the model than taking all nighttime burglars. However, eventually, with further sub-groupings there will be insufficient data.

This point has been raised in a recent debate. Van Koppen and De Keijser (1997) argued that a distance decay function that combined multiple incidents committed by the same individuals could distort the estimated relationship compared to selecting incidents committed by different individuals.[6] Rengert, Piquero and Jones (1999) argued that such a distribution is nevertheless meaningful. In our language, these are two different sub-groups - persons committing multiple offenses compared to persons committing only one offense. Combining these two sub-groups into a single calibration data set will only mean that the result will have less differentiation in prediction than if the sub-groups were separated out.

Actually, there is not much difference, at least in Baltimore County. From the 41,426 cases, 18,174 were committed by persons who were only listed once in the database while 23,251 offenses were committed by persons who were listed two or more times (7,802 individuals). Categorizing the 18,174 crimes as committed by 'single incident offenders and the 23,251 crimes as committed by 'multiple incident offenders', the density distance decays functions were calculated using the kernel density method (Figure 10.22).

The distributions are remarkably similar. There are some subtle differences. The average journey to crime trip distance made by a single incident offender is longer than for multiple incident offenders (4.6 miles compared to 4.0 miles, on average); the difference is highly significant ($p \leq .0001$), partly because of the very large sample sizes. However, a visual inspection of the distance decay functions shows they are similar. The single incident offenders tend to have slightly more trips near their home, slightly fewer for distances between about a mile up to three miles, and slightly more longer trips. But, the differences are not very large.

There are several reasons for the similarity. First, some of the 'single incident offenders' are actually multiple incident offenders who have not been charged with other incidents. Second, some of the single incident offenders are in the process of becoming multiple incident offenders so their behavior is probably similar. Third, there may not be a major difference in travel patterns by the number of offenses an individual commits, certainly compared to the major differences by type of crime (see graphs above). In other words, the distinction between a single offender crime trip and a multiple offender crime trip is just another sub-group comparison and, apparently, not that important. Nevertheless, it is important to choose an appropriate sample from which to estimate a likely home base location for a serial offender. The method depends on a similar sample of offenders for comparison.

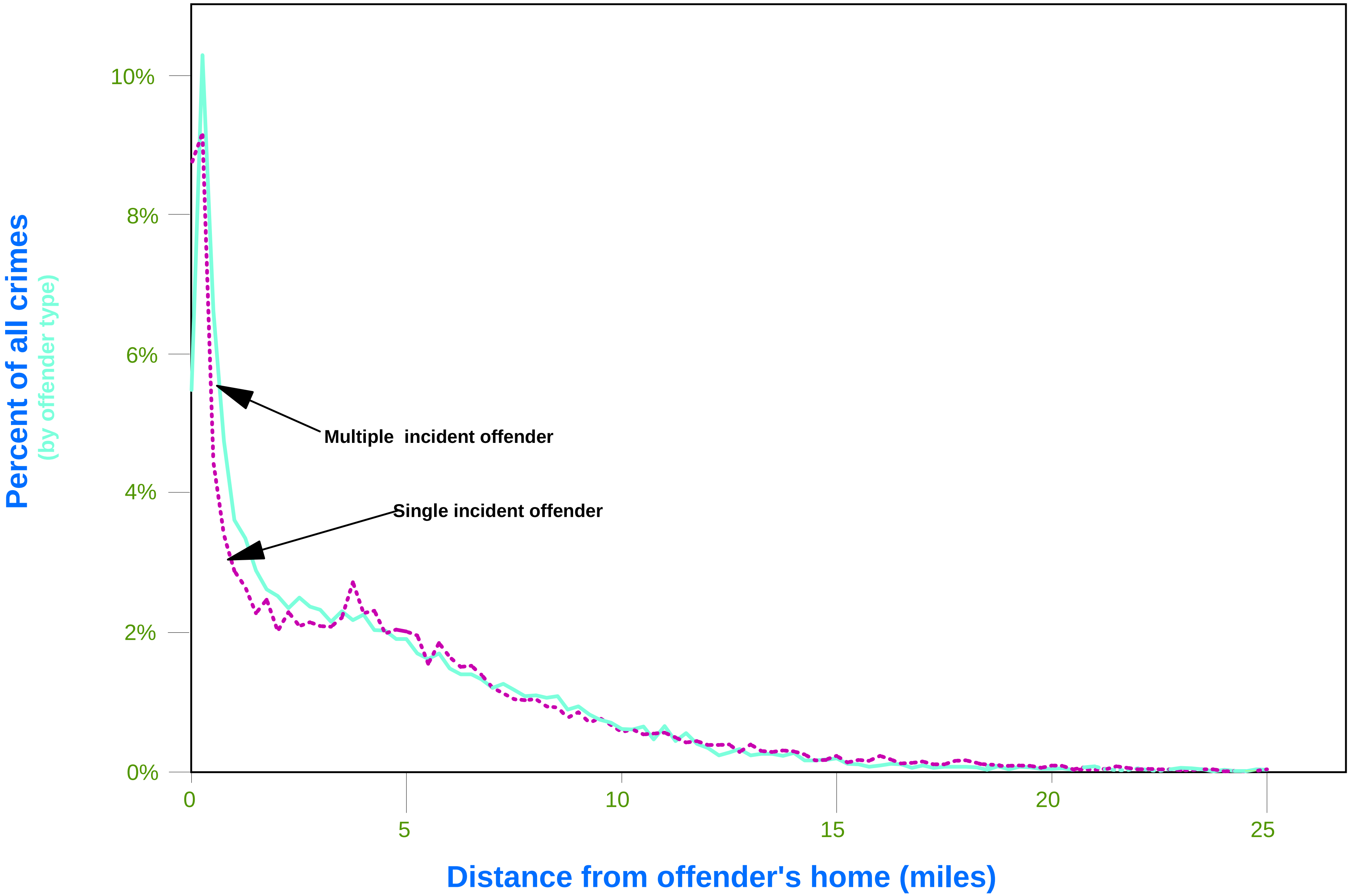
Sample Data Sets for Journey to Crime Routines

Three sample data sets from Baltimore County have been provided for the journey to crime routine. The data sets are simulated and do not represent real data. The first file - JtcTest1.dbf, are 2000 simulated robberies while the second file - JtcTest2.dbf, are 2500

Figure 10.22:

Journey to Crime Distances

Kernel Density Estimate of Single and Multiple Incident Offenders

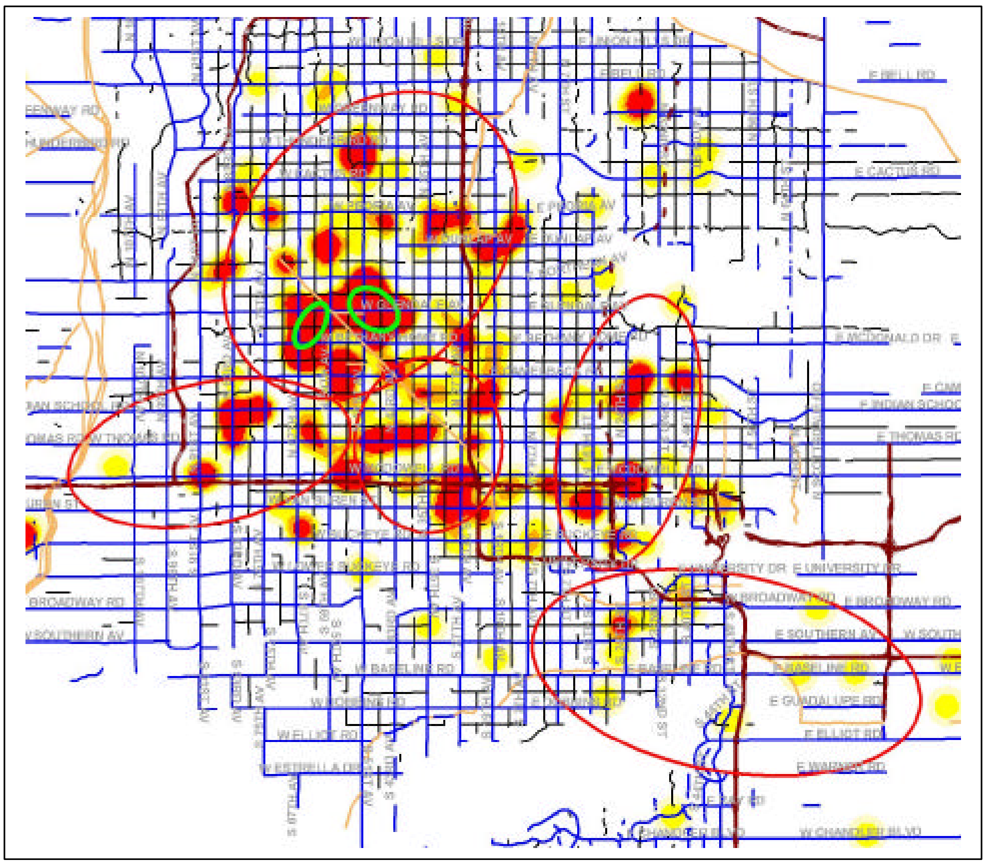


Hot Spot Verification in Auto Theft Recoveries

Bryan Hill
Glendale Police Department
Glendale, AZ

We use *CrimeStat* as a verification tool to help isolate clusters of activity when one application or method does not appear to completely identify a problem. The following example utilizes several *CrimeStat* statistical functions to verify a recovery pattern for auto thefts in the City of Glendale (AZ). The recovery data included recovery locations for the past 6 months in the City of Glendale which were geocoded with a county-wide street centerline file using *ArcView*.

First, a spatial density "grid" was created using *Spatial Analyst* with a grid cell size of 300 feet and a search radius of 0.75 miles for the 307 recovery locations. We then created a graduated color legend, using standard deviation as the classification type and the value for the legend being the *CrimeStat* "Z" field that is calculated.



In the map, the K-means (red ellipses), Nnh (green ellipses) and *Spatial Analyst* grid (red-yellow grid cells) all showed that the area was a high density or clustering of stolen vehicle recoveries. Although this information was not new, it did help verify our conclusion and aided in organizing a response

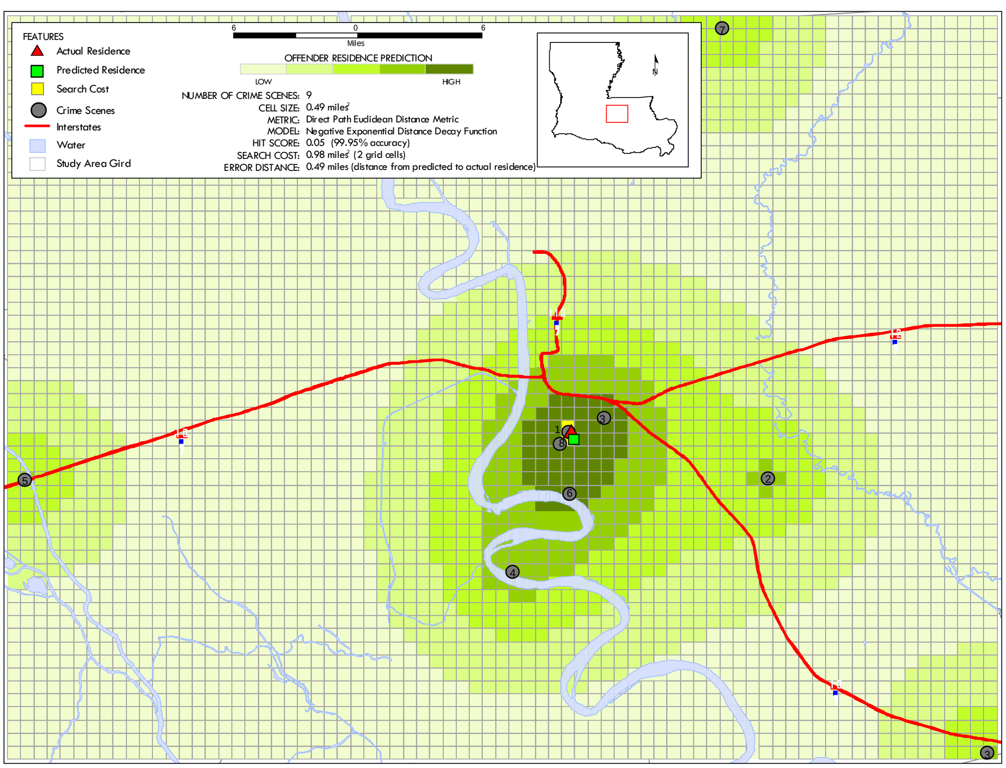
Constructing Geographic Profiles Using the *CrimeStat* Journey-To-Crime Routine

Josh Kent,
Michael Leitner,
Louisiana State University
Baton Rouge, LA

The map below shows a geographic profile constructed from nine crime sites associated with a Baton Rouge serial killer, Sean Vincent Gillis, who was apprehended on April 29, 2004 at his residence in Baton Rouge. Eight of the nine are body dump sites and the ninth is a point of fatal encounter. All crime sites were located in the City of Baton Rouge and surrounding parishes. Gillis's hunting style can best be described as that of a typical 'localized marauder'.

The Journey-to-crime routine, implemented in *CrimeStat*, was applied to simulate the travel characteristics of Gillis to and from the known crime sites. Gillis's travel behavior was calibrated with different mathematical functions that were derived from the known travel patterns of 301 homicide cases in Baton Rouge.

The profile was estimated using Euclidean distance and the negative exponential distance decay function. It predicts the actual residence of Gillis extremely accurately. The straight-line error distance between the predicted and the actual residence is only 0.49 miles. The proportion of the entire study area that must be searched in order to successfully identify the serial offender's residence is 0.05% (approximately 0.98 square miles out of a 2094.75 square miles study area).



simulated burglaries. Both files have coordinates for an origin location (HomeX, HomeY) and a destination location (IncidentX, IncidentY). Users can use the calibration routine to calculate the travel distances between the origins and the destinations. A third data set - Serial1.dbf, are simulated incident locations for a serial offender. Users can use the Jtc estimation routine to identify the likely residence location for this individual. In running this routine, a reference grid needs to be overlaid (see chapter 3). For Baltimore County, appropriate coordinates for the lower-left corner are -76.91^0 longitude and 39.19^0 latitude and for the upper-right corner are -76.32^0 longitude and 39.72^0 latitude.

Draw Crime Trips

The Journey to Crime module includes one utility that can help visualize the pattern before selecting a particular estimation model. This is a Draw Crime Trips routine that simply draws lines between the origin and destination of individual crime trips. The X and Y coordinates of an origin and destination location are input and the routine draws a line in *ArcView* 'shp', *MapInfo* 'mif', *Atlas*GIS* 'bna' or Ascii format.

Figure 10.23 illustrates the drawing of the known travel distances for 444 rape cases for which the residence location of the rapist was known. Of the 444 cases, 113 (or 25.5%) occurred in the residence of the rapist. However, for the remaining 331 cases, the rape location was not the residence location. As seen, many of the trips are of quite long distances. This would suggest the use of an journey to crime function that has many trips at zero distance but with a more gradual decay function.

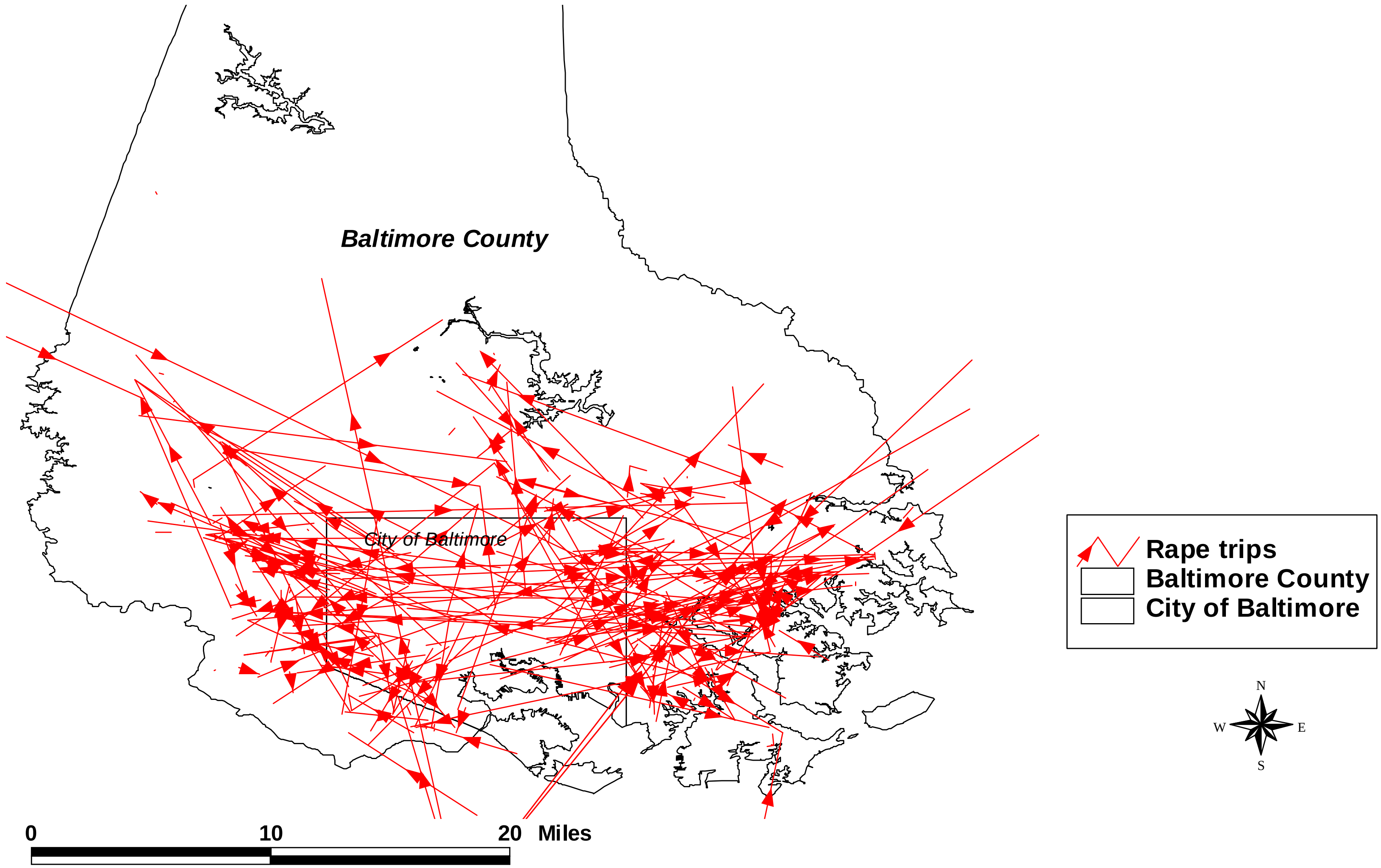
How Accurate are the Methods?

A critical question is how accurate are these methods? The journey to crime model is just that, a model. Whether it involves using a mathematical function or an empirically-derived one, the assumption in the Jtc routine is that the distribution of incidents will provide information about the home base location of the offender. In this sense, it's not unlike the way most crime analysts will work when they are trying to find a serial offender. A typical approach will be to plot the distribution of incidents and routinely search a geographic area in and around a serial crime pattern, noting offenders who have an arrest history matching case attributes (MO, type weapon, suspect description, etc.). Because a high proportion of offenses are committed within a short distance of offender residence's, the method can frequently lead to their apprehension. But, in doing this method, the analysts are not using a sophisticated statistical model.

Test Sample of Serial Offenders

To explore the accuracy of the approach, a small sample of 50 serial offenders was isolated from the database and used as a target sample to test the accuracy of the methods. The 50 offenders accounted for 520 individual crime incidents in the database. To test the Jtc method systematically, the following distribution was selected (table 10.4). The sample was not random, but was selected to produce a balance in the number of incidents

Figure 10.23:
Journey to Rape Trips
Distance from Residence to Rape Location for 444 Known Offenders



committed by each individual and to, roughly, approximate the distribution of incidents by serial offenders. Each of the 50 offenders was isolated as a separate file so that each could be analyzed in *CrimeStat*.

Identifying the Crime Type

Each of the 50 offenders was categorized by a crime type. Only two of the offenders committed the same crime for all their offenses; most committed two or more different types of crimes. Arbitrarily, each offender was typed according to the crime type that he/she most frequently committed; in the two cases where there was a tie between two crime types, the most severe was selected (i.e., personal crime over property crime). While I recognize that there is arbitrariness in the approach, it seemed a practical solution. Any error in categorizing an offender would be applicable to all the methods. The crime types for the 50 offenders approximately mirrored the distribution of incidents: larceny (29); vehicle theft (7); burglary (5); robbery (5); assault (2); bank robbery (1); and arson (1).

Identifying the Home Base and Incident Locations

In the database, each of the offenders was listed as having a residence location. For the analysis, this was taken as the *origin* location of the journey to crime trip. Similarly, the incident location was taken as the *destination* for the trip. Operationally, the crime trip is taken as the distance from the origin location to the destination location. However, it is very possible that some crime trips actually started from other locations. Further, many of these individuals have moved their residences over time; we only have the last known residence in the database. Unfortunately, there was no other information in the digital database to allow more accurate identification of the home location. In other words, there may be, and probably are, numerous errors in the estimation of the journey to crime trip. However, these errors would be similar across all methods and should not affect their relative accuracy.

Evaluated Methods

Eleven methods were compared in estimating the likely residence location of the offenders. Four of the methods used the Jtc routines and seven were simple spatial distribution methods (table 10.5).

The mean center and center of minimum distance are discussed in chapter 4. The center of minimum distance, in particular, is more or less the geographic center of distribution in that it ignores the values of particular locations; thus, locations that are far away from the cluster (extreme values) have no effect on the result. When the center of minimum distance is calculated on a road network in which each segment is weighted by travel time or speed, the result is the center of minimum travel time, the point at which travel time to each of the incidents is minimized. The directional mean, triangulated mean, geometric and harmonic means are discussed in chapter 4.

Table 10.4

Serial Offenders Used in Accuracy Evaluation

Number of Offenders	Number of Crimes Committed by Each Person
4	3
4	4
4	5
4	6
4	7
4	8
3	9
3	10
3	11
2	12
2	13
2	14
2	15
1	16
1	17
1	18
1	19
1	20
1	21
1	22
1	24
1	33
50	520

Table 10.5

Comparison Methods for Estimating the Home Base of a Serial Offender

Journey to Crime Methods

Mathematical model for all crimes

Mathematical model for specific crime type

Kernel density model for all crimes

Kernel density model for specific crime type

Spatial Distribution Methods

Mean center

Center of minimum distance

Center of minimum travel time
(calculated on road network weighted by travel time)

Directional mean (weighted) calculated with 'lower left corner' as origin

Triangulated mean

Geometric mean

Harmonic mean

The Test

Each of these eleven methods were run against each of the files created for the serial offenders. For the seven 'means' (mean center, geometric mean, harmonic mean, directional mean, triangulated mean, center of minimum distance, center of minimum travel time), the mean was itself the best guess for the likely residence location of the offender. For the four journey to crime functions, the grid cell with the highest likelihood estimate was the best guess for the likely residence location of the offender.

Measurement of Error

For each of the 50 offenders, error was defined as the distance in miles between the 'best guess' and the actual location. For each offender, the distance between the estimated home base (the 'best guess') and the actual residence location was calculated using direct distances. Table 10.6 presents the results. The data show the error by method for each of the 50 offenders. The three right columns show the average error of all methods and the minimum error and maximum errors obtained by a method. The method with the minimum error is boldfaced; for some cases, two methods are tied for the minimum. The bottom three rows show the median error, the average error and the standard deviation of the errors for each method across all 50 offenders.

There are a number of conclusions from the results. First, the degree of precision for any of these methods varies considerably. The precision of the estimates vary from a low of 0.0466 miles (about 246 feet) to a high of 75.7 miles. The overall precision of the methods is not very high and is highly variable. There are a number of possible reasons for this, some of which have been discussed above. Each of the methods produces a single parameter from what is, essentially, a probability distribution whereas the distribution of many of these incidents are widely dispersed. Few of the offenders had such a concentrated pattern that only a single location was possible. Since these are probability distributions, not everyone follows the 'central tendency'. Also, some of these offenders may have moved during the period indicated by the incidents, thereby shifting the spatial pattern of incidents and making it difficult to identify the last residence.

A second conclusion is that, for any one offender, the methods produce similar results. For many of the offenders the difference between the best estimate (the minimum error) and the worst estimate (the maximum error) is not great. Thus, the simple methods are generally as good (or bad) as the more sophisticated methods.

Third, across all methods, the center of minimum travel time, which is calculated on a road network (see chapters 3 and 16), and it's distance-based 'cousin' - the center of minimum travel time, had the lowest average error. Thus, the approximate geographic center of the distribution where travel time to each of the incidents was minimal produced as good an estimate as the more sophisticated methods. However, it wasn't particularly close (3.8415 miles, on average). The worst method was the triangulated mean; it had an average error of 7.6472 miles. The triangulated mean is produced by vector geometry and will not necessarily capture the center of the distribution. Other than this, there were not great differences. This reinforces the point above that the methods are all, more or less, describing the central tendency of the distribution. For offenders that don't live in the center of their distribution, the error of a method will necessary be high.

Looking at each of the 50 offenders, the methods vary in their efficacy. For example, the Jtc kernel function for all crimes was the best or tied for best for 17 of the offenders, but was also the worst or tied for worst for 9. Similarly, the Jtc kernel function for the specific crimes was best or tied for best for 8 of the offenders, but worse for 4. Even the most consistent was best for 4 offenders, but also worst for one. On the other hand, the triangulated mean, which had the worst overall error, produced the best estimate for 9 of the individuals while it produced the worst estimate for 25 of the individuals. Thus, the triangulated mean tends to be very accurate or very inaccurate; it had the highest variance, by far.

Table 10.6

Accuracy of Methods for Estimating Serial Offender Residences

(N= 50 Serial Offenders)

Dataset	Number of Crimes	Primary Crime Type	Mean Center Error (miles)	Center of Minimum Distance Error (miles)	Center of Minimum Travel Time Error (miles)	Triangulated Mean Error (miles)	Geometric Mean Error (miles)	Harmonic Mean Error (miles)	Jtc Kernel: All Crimes Error (miles)	Jtc Kernel: Crime Type Error (miles)	Jtc Math: All Crimes Error (miles)	Jtc Math: Crime Type Error (miles)	Average Error	All Methods Minimum Error	Maximum Error
3A	3	Larceny	**31.5991**	32.4477	32.2975	32.4109	31.5995	31.6000	32.7824	32.7880	32.7824	32.7880	32.3095	31.5991	32.7880
3B	3	Larceny	13.2303	12.1683	12.1207	24.1531	13.2311	13.2319	**10.7526**	14.4929	**10.7526**	11.2501	13.5384	10.7526	24.1531
3C	3	Bank robbery	2.8348	0.9137	0.9588	2.7767	2.8335	2.8322	**0.6775**	5.8416	**0.6775**	6.0946	2.6441	0.6775	6.0946
3D	3	Burglary	2.9733	3.2603	**2.6907**	6.1013	2.9728	2.9724	4.6038	3.3883	3.3882	3.7931	3.6144	2.6907	6.1013
4A	4	Vehicle theft	4.2436	4.2670	4.3341	**3.8217**	4.2436	4.2436	4.2527	4.2364	4.2527	4.2590	4.2154	3.8217	4.3341
4B	4	Larceny	1.9618	0.3100	**0.1158**	2.0563	1.9621	1.9623	0.3125	0.2018	0.3125	0.2784	0.9473	0.1158	2.0563
4C	4	Larceny	4.4733	4.4733	4.5096	4.6789	4.4733	4.4733	4.9681	4.3563	**4.2637**	4.3563	4.5026	4.2637	4.9681
4D	4	Assault	0.2925	0.1905	**0.0084**	0.0466	0.2925	0.2926	0.0703	0.0703	0.0703	0.4560	0.1790	0.0084	0.4560
5A	5	Larceny	17.3308	16.6459	17.0832	17.8985	17.3292	17.3276	**15.9738**	17.8655	15.9739	16.4526	16.9881	15.9738	17.8985
5B	5	Larceny	1.3609	0.2481	**0.0565**	1.7733	1.3586	1.3564	0.2068	0.6974	0.5140	0.6974	0.8269	0.0565	1.7733
5C	5	Larceny	2.2458	2.6832	2.7443	16.4518	2.2450	**2.2442**	2.7886	2.4205	2.7886	3.0922	3.9704	2.2442	16.4518
5D	5	Larceny	0.9169	0.2250	0.8021	0.2371	0.9171	0.9174	**0.1577**	0.4267	**0.1577**	0.4267	0.5184	0.1577	0.9174
6A	6	Larceny	5.1837	5.2081	5.0644	7.9621	5.1837	5.1837	5.1271	**4.8554**	4.9393	5.2256	5.3933	4.8554	7.9621
6B	6	Vehicle theft	1.3720	1.1869	1.1535	**0.9625**	1.3710	1.3700	3.1126	2.3800	1.3566	2.0831	1.6348	0.9625	3.1126
6C	6	Larceny	1.3199	0.3157	**0.0051**	1.7928	1.3192	1.3184	0.2580	0.5272	0.2580	0.5272	0.7641	0.0051	1.7928
6D	6	Larceny	3.2458	2.3324	3.2838	6.5209	3.2431	3.2405	**1.2506**	2.6253	1.9718	1.9718	2.9686	1.2506	6.5209
7A	7	Larceny	3.9023	3.4185	3.1998	**2.3176**	3.9022	3.9021	2.7419	3.0532	3.1364	3.0532	3.2627	2.3176	3.9023
7B	7	Larceny	12.4100	9.2973	9.9031	14.8293	12.4107	12.4115	**8.5357**	8.6148	**8.5357**	8.8275	10.5776	8.5357	14.8293
7C	7	Burglary	5.0501	7.1477	6.4354	10.8567	5.0481	**5.0460**	7.9975	7.9975	7.9975	7.6274	7.1204	5.0460	10.8567
7D	7	Larceny	2.2686	0.7733	0.3223	75.7424	2.2684	2.2682	**0.0892**	0.7191	**0.0892**	0.7191	8.5260	0.0892	75.7424
8A	8	Larceny	6.0298	**6.0165**	6.3167	6.2653	6.0264	6.0229	8.4210	6.2962	6.2022	6.1166	6.3714	6.0165	8.4210
8B	8	Larceny	**1.0041**	1.1437	1.1458	2.1776	1.0042	1.0042	1.7475	1.3510	1.5298	1.3510	1.3459	1.0041	2.1776
8C	8	Larceny	1.3059	1.6944	1.6203	1.3684	1.3043	1.3027	2.1513	**1.2020**	2.1513	1.8707	1.5971	1.2020	2.1513
8D	8	Vehicle theft	3.5794	2.3780	4.0475	5.5915	3.5809	3.5825	**0.5900**	1.3340	1.9133	1.3340	2.7931	0.5900	5.5915
9A	9	Robbery	5.2527	5.7156	5.4565	**4.8574**	5.2529	5.2532	7.8257	7.1961	6.2520	5.9265	5.8989	4.8574	7.8257
9B	9	Larceny	8.1923	10.6555	9.9787	**6.9916**	8.1886	8.1850	12.4578	10.3957	12.4578	12.0514	9.9554	6.9916	12.4578
9C	9	Robbery	3.7778	3.8454	**3.5670**	11.0042	3.7758	3.7738	4.9015	5.1862	4.6206	4.3445	4.8797	3.5670	11.0042
10A	10	Larceny	0.9358	0.5159	0.4822	1.1003	0.9355	0.9353	**0.0606**	0.3720	0.2601	0.7172	0.6315	0.0606	1.1003
10B	10	Larceny	2.8581	3.4940	4.8179	14.2219	2.8536	**2.8491**	6.4051	6.5709	10.3095	6.4758	6.0856	2.8491	14.2219
10C	10	Larceny	0.8052	**0.7251**	0.7451	5.5938	0.8050	0.8049	0.9059	0.8404	0.9060	1.2786	1.3410	0.7251	5.5938
11A	11	Vehicle theft	**2.9127**	3.2715	3.4493	3.1192	2.9130	2.9134	3.6936	3.4335	3.4282	3.2087	3.2343	2.9127	3.6936
11B	11	Robbery	0.3250	0.3250	0.2709	0.2513	0.3250	0.3250	0.4235	**0.2263**	0.4235	0.7011	0.3596	0.2263	0.7011
11C	11	Vehicle theft	1.2689	1.7157	1.4115	1.4750	1.2709	1.2729	2.8945	**0.6984**	2.8945	2.2049	1.7107	0.6984	2.8945
12A	12	Larceny	3.3881	4.2334	4.2640	10.9241	3.3867	3.3852	6.4050	**3.2639**	5.5843	5.2132	5.0048	3.2639	10.9241
12B	12	Larceny	0.5562	0.5361	**0.4973**	2.8003	0.5562	0.5562	0.7897	0.6709	0.7897	0.9631	0.8716	0.4973	2.8003
13A	13	Larceny	6.3282	7.2857	6.8066	**6.0244**	6.3248	6.3213	7.6438	7.4607	7.6438	7.9915	6.9831	6.0244	7.9915
13B	13	Assault	**1.4943**	**1.4943**	1.4572	1.5279	1.4944	1.4944	1.6501	1.5954	1.6501	2.0824	1.5940	1.4572	2.0824
14A	14	Larceny	1.9363	0.8706	0.6681	1.4498	1.9365	1.9368	0.3434	0.6058	**0.2596**	0.7631	1.0770	0.2596	1.9368
14B	14	Arson	0.6898	0.3727	**0.0251**	0.8086	0.6899	0.6900	0.3359	0.3359	0.3359	0.6213	0.4905	0.0251	0.8086
15A	15	Vehicle theft	0.7282	0.7189	0.8741	**0.3362**	0.7277	0.7271	0.8155	0.4855	0.8155	1.5128	0.7741	0.3362	1.5128
15B	15	Robbery	0.4914	0.4914	**0.3422**	0.8254	0.4914	0.4914	0.6468	0.5693	0.6468	0.6546	0.5651	0.3422	0.8254
16A	16	Vehicle theft	2.1107	2.0995	2.0555	8.2311	2.1107	2.1107	**1.5957**	1.6404	2.5911	2.4033	2.6949	1.5957	8.2311
17A	17	Burglary	1.6484	0.3093	**0.1000**	1.0227	1.6461	1.6438	0.2879	0.2879	0.2879	0.5268	0.7761	0.1000	1.6484
18A	18	Larceny	0.6308	0.4196	**0.0417**	1.0876	0.6329	0.6349	0.2132	0.3383	0.2132	0.6985	0.4911	0.0417	1.0876
19A	19	Larceny	**8.6462**	9.4195	9.3665	8.6772	8.6486	8.6511	10.2869	9.2708	9.7022	9.5548	9.2224	8.6462	10.2869
20A	20	Burglary	6.3520	5.7969	7.4256	28.3094	6.3486	6.3452	**0.5934**	0.8673	**0.5934**	0.7945	6.3426	0.5934	28.3094
21A	21	Burglary	1.2396	0.8861	1.0564	1.2776	1.2393	1.2390	0.5243	0.5243	1.0253	**0.4965**	0.9509	0.4965	1.2776
22A	22	Larceny	3.6828	2.6232	2.3597	**2.0949**	3.6803	3.6777	2.4937	2.8944	2.4937	2.8944	2.8895	2.0949	3.6828
24A	24	Larceny	1.7959	0.5892	0.9322	2.3033	1.7975	1.7991	**0.2658**	0.3574	0.4222	0.6587	1.0921	0.2658	2.3033
33A	33	Robbery	3.9901	5.0481	**3.4056**	7.2505	3.9940	3.9979	7.9485	7.6939	8.1907	7.9439	5.9463	3.4056	8.1907
		Median Error =	2.5517	2.2159	2.2076	3.4704	2.5509	2.5502	**1.9494**	2.0102	2.0615	2.1440			
		Mean Error =	4.0434	3.8441	3.8415	7.6472	4.0429	4.0424	4.0395	4.0305	4.0163	4.1467			
		SD Error =	5.2696	5.4392	5.4619	12.1867	5.2696	5.2695	5.6244	5.6807	5.5961	5.4727			

Fourth, the median error is smaller than the average error. That is, the median is the point at which 50% of the cases had a smaller error and 50% had a larger error. Overall, most of the cases were found within a shorter distance than the average would indicate. This indicates that several cases had very large errors whereas most had smaller errors; that is, they were *outliers*. Over all methods, the Jtc kernel approach for all crimes had the lowest median error (1.95 miles). In fact, all four Jtc methods had smaller median errors than the simple centrographic methods. In other words, they are more accurate than the centrographic methods most of the time. The problem in applying this logic in practice, however, is that one would not know if the case being studied is typical of most cases (in which case, the error would be relatively small) or whether it was an outlier. In other words, the median would define a search area that captured about 50% of the cases, but would be very wrong in the other 50%. If we could somehow develop a method for identifying when a case is 'typical' and when it isn't, increased accuracy will emerge from the Jtc methods. But, until then, the simple center of minimum travel time will be the most accurate method.

Fifth, the amount of error varies by the number of incidents. Table 10.7 below shows the average error for each method as a function of three size classes: 1-5 incidents; 6-9 incidents; and 10 or more incidents. As can be seen, for each of the ten methods, the error decreases with increasing number of incidents. In this sense, the measured error is responsive to the sample size from which it is based. It is, perhaps, not surprising that with only a handful of incidents no method can be very precise.

Sixth, the relative accuracy of each of these methods varies by sample size. The method or methods with the minimum error are boldfaced. For a limited number of incidents (1-5), the Jtc mathematical function for all crimes (i.e., the negative exponential with the parameters from table 10.5) produced the estimate with the least error, followed by the Jtc kernel function for all crimes; the was the third best. The differences in error between these were not very great. For the middle category (6-9 incidents), the center of minimum distance produced the least error followed by the Jtc mathematical function for the specific crime type. For those offenders who had committed ten or more crimes, the Jtc kernel function for the specific crime type produced the best estimate, followed by the center of minimum distance. The two mathematical functions produced the least accuracy for this sub-group, though again the differences in error are not very big (2.2 miles for the best compared to 2.7 miles for the worst). In other words, only with a sizeable number of incidents does the Jtc kernel density approach for specific crimes produce a good estimate. It is better than the other approaches, but only slightly better than the simple measure of the center of minimum distance.

Search Area

A number of researchers have been interested in the concept of a search area for the police (Rossmo, 2000; Canter, 2003). The concept is that the journey to crime method can define a small search area within which there is a higher probability of finding the offender. The average or median error discussed above can be used to define such a search area if treated as a radius of a circle. While intuitive, I'm not sure whether this represent

Table 10.7

Method Estimation Error and Sample Size:
Average Error of Method by Number of Incidents (miles)

Number of Incidents	Mean Center	Center of Minimum Distance	Center of Minimum Travel Time	Triangulated Mean	Geometric Mean	Harmonic Mean	Jtc Kernel: All	Jtc Kernel: Crime types	Jtc Math: All	Jtc Math: Crime types	All Methods Average Error	Minimum Error
3-5	6.9553	6.4861	6.4768	9.3672	6.9160	6.9545	6.4622	7.2321	**6.3278**	6.9954	**7.0173**	**6.3278**
6-9	4.2596	**4.0753**	4.1000	10.6160	4.3331	4.2576	4.4805	4.2489	4.2274	4.2020	**4.8800**	**4.0753**
10+	2.3832	2.3149	2.2980	4.8136	2.4575	2.3827	2.4880	**2.2176**	2.6725	2.6243	**2.6652**	**2.2176**

a meaningful statistic. For example, taking the average error of the center of minimum distance (3.84 miles) would produce a search area of 46.4 square miles, not exactly a small area in which to find a serial offender. Even if we take the median error of 1.94 miles from the Jtc kernel approach for all crimes (1.94 miles) will still produce a search area of 11.9 square miles, and it would be correct only half the time

In other words, these methods are still very imprecise. Further, the error is liable to increase over time, rather than decrease. With about 50% of the U.S. population living in suburbia (Demographia, 1998) and with 90% of American households owning at least one motor vehicle (U.S. Census Bureau, 2000), the average distances traveled by offenders has probably been increasing over time since most types of trips have also shown increases in travel over time. This means that unless police can find a way to narrow down the search area considerably, the methods don't really help beyond what police intuitively do anyway, namely look near the distribution of the incidents committed by serial offenders.

Cautionary Notes

Of course, this is a limited test. It was a small sample (only 50 cases) from a single jurisdiction (Baltimore County). The sample wasn't even randomly selected, but chosen to examine the accuracy by a range of sample sizes. Thus, the conclusions are only tentative and must be seen as hypotheses for further work. Clearly, more research is needed.

Nevertheless, there are certain cautions that must be considered in using either of these journey to crime methods (the mathematical or the empirical). First, a simple technique, such as the center of minimum distance, may be as good as a more sophisticated technique. It doesn't always follow that a sophisticated method will produce any more accuracy than a simple one. For the time being, I would advise crime analysts who are trying to detect a pattern in the distribution of the incidents of a serial offender to do exactly what they have been doing, basically looking at the data and making a subjective guess about where the offender may be residing. The kernel density Jtc routine needs an adequate amount of information (i.e, at least 10 incidents) to produce somewhat precise estimates. These techniques should be seen for now as research tools rather than as diagnostics for identifying the whereabouts of an offender. They are just too imprecise and unreliable to depend on, at least until more definitive results are obtained.

Second, there are other limitations to the technique. The model must be calibrated for each individual jurisdiction. Further, it must be periodically re-calibrated to account for changes in crime patterns. For example, in using the mathematical model, one cannot take the parameters estimated for Baltimore County (Table 10.3) and apply them to another city or if using the kernel density method take the results found at one time period and assume that they will remain indefinitely. The model is a probability model, not a guarantee of certainty. It provides guesses based on the similarity to other offenders of the same type of crime. In this sense, a particular serial offender may not be typical and the model could actually orient police wrongly if the offender is different from the calibration sample. It will take insight by the investigating officers to know whether the pattern is typical or not.

Third, as a theoretical model, the journey to crime approach is quite simple. It is based on a distribution of incidents and an assumed travel distance decay function. From the perspective of modeling the travel behavior of offenders, it is limited. As mentioned above, the method does not utilize information on the distribution of target opportunities nor does it utilize information on the travel mode and route that an offender takes. It is purely a statistical model. The research area of geographic profiling attempts to go beyond statistical description and understand the cognitive maps that offenders use as well as how these interact with their motives. This is good and should clearly guide future research. But it has to be understood that the theory of offender travel behavior is not very well developed, certainly compared to other types of travel behavior. Further, some types of crime trips may not even start from an offender's residence, but may be referenced from another location, such as vehicle thefts occurring near disposal locations. Routine activity theory would suggest multiple origins for crimes (Cohen and Felson, 1979).

The existing models of travel demand used by transportation planners (which have themselves been criticized for being too simple) measure a variety of factors that have only been marginally included in the crime travel literature - the availability of opportunities, the concentration of offender types in certain areas, the mode of travel (i.e., auto, bus, walk), the specific routes that are taken, the interaction between travel time and travel route, and other factors. It will be important to incorporate these elements into the understanding of journey to crime trips to build a much more comprehensive theory of how offenders operate. Travel behavior is very complicated and we need more than a statistical distance model to adequately understand it. The next seven chapters look at an application of travel demand theory to crime travel.

Also, it's not clear whether knowing an offender's 'cognitive map' will help in prediction. There have been no evaluations that have compared a strictly statistical approach with an approach that utilizes information about the offender as he or she understands the environment. It cannot be assumed that integrating information about the perception of the environment will aid prediction. In most travel demand forecasts that transportation engineers and planners make, cognitive information about the environment is not utilized except in the definition of trip purpose (i.e., what the purpose of the trip was). The models use the actual trips by origin and destination as the basis for formulating predictions, not the understanding of the trip by the individual. Understanding is important from the viewpoint of developing theory or for ways to communicate with people. But, it is not necessarily useful for prediction. In short, understanding and prediction are not the same thing.

On the other hand, the journey to crime routine, particularly the kernel density approach, can be useful for police departments *if* used carefully. If there are sufficient cases to build an estimate (i.e., 10 or more incidents), it can provide additional information to officers investigating a serial offender by reducing the number of possible suspects that might be linked to a series of crimes. It can also provide some direction in orienting the deployment of officers and detectives investigating what appear to be serial offenses. It provides guesses about where the offender might be living, but based on similarities with previous offenders for the same type of crime. It's not going to give an exact estimate of where an offender is living, but will provide some insights into which areas the individual might be located. The Jtc model should be seen as a supplement to other techniques, not a complete solution. Like all the statistical tools in *CrimeStat*, it must be used carefully and intelligently. The philosophy of crime analysis must always be to use a technique with thought and with a systematic procedure.

Endnotes for Chapter 10

1. It should also be pointed out that the use of direct distances will underestimate travel distances particularly if the street network follows a grid.

2. There are, of course, many other types of mathematical functions that can be used to describe a declining likelihood with distance. In fact, there are an infinite number of such functions. However, the five types of functions presented here are commonly used. We avoided the inverse distance function because of its potential to distort the likelihood relationship.

$$f(d) = \frac{1}{d_{ij}^{k}}$$

where *k* is a power (e.g., 1, 2, 2.5). For large distances, this function can be a useful approximation of the lessening travel interaction with distance. However, for short distances, it doesn't work. As the distance between the reference cell location and an incident location becomes very small, approaching zero, then the likelihood estimate becomes very large, approaching infinity. In fact, for $d_{ij} = 0$, the function is unsolvable. Since many distances between reference cells and incidents will be zero or close to zero, the function becomes unusable.

3. It is actually the inverse of the inverse distance function. If a distance decay function drops off proportional to the inverse of the distance,

$$Y_{ij} = A^* \ 1/d_{ij}$$

where Y_{ij} is the travel likelihood, A is coefficient, and d_{ij} is the distance from the home base, then the opposite - a distance increase is just the inverse of this function

$$Z_{ij} = \frac{1}{A^* \ 1/d_{ij}} = \frac{d_{ij}}{A} = B^* \ d_{ij}$$

4. There are several sources of error associated with the data set. First, these records were arrest records prior to a trial. Undoubtedly, some of the individuals were incorrectly arrested. Second, there are multiple offenses. In fact, more than half the records were for individuals who were listed two or more times in the database. The travel pattern of repeat offenders may be slightly different than for apparent first-time offenders (see figure 10.19). Third, many of these individuals have lived in multiple locations. Considering that many are young and that most are socially not well adjusted, it would be expected that these individuals would have multiple homes. Thus, the distribution of incidents could reflect multiple home bases, rather than one. Unfortunately, the data we have only gives a single residential location, the place at which they were living when arrested.

5. If the coordinate system is projected with the distance units in feet, meters or miles, then the distance between two points is the hypotenuse of a right triangle using Euclidean geometry.

$$d_{AB} = \sqrt{(X_A - X_B)^2 + (Y_A - Y_B)^2} \qquad (3.1)$$

repeat

where each location is defined by an X and Y coordinate in feet, meters, or miles.

If the coordinate system is spherical with units in latitudes and longitudes, then the distance between two points is the Great Circle distance. All latitudes and longitudes are converted into radians using

$$\text{Radians}(\phi) = \frac{2\pi\,\phi}{360} \qquad (3.2)$$

repeat

$$\text{Radians}(\lambda) = \frac{2\pi\,\lambda}{360} \qquad (3.3)$$

repeat

Then, the distance between the two points is determined from

$$d_{AB} = 2 * \text{Arcsin}\ \{\text{Sin}^2[(\phi_B - \phi_A)/2] + \text{Cos}\ \phi_A * \text{Cos}\phi_B * \text{Sin}^2[(\lambda_B - \lambda_A)/2]^{1/2}\ \} \qquad (3.4)$$

repeat

with all angles being defined in radians (Snyder, 1987, p. 30, 5-3a).

6. They also argued that the combination of incidents - which they called 'aggregation', would distort the relationship between distance and incidence likelihood because of the ecological fallacy. To my mind, they are incorrect on this point. Data on a distribution of incidents by distance traveled is an individual characteristic and is not 'ecological' in any way. An ecological inference occurs when data are aggregated with a *grouping* variable (e.g., state, county, city, census tract; see Langbein and Lichtman, 1978). A frequency distribution of individual crime trip distances is an individual probability distribution, similar, for example, to a distribution of individuals by height, weight, income or any other characteristic. Of course, there are sub-sets of the data that have been aggregated (similar to heights of men v. heights of women, for example). Clearly, identifying sub-groups can make better distinctions in a distribution. But, it is still an individual probability distribution. This doesn't produce bias in estimating a parameter, only variability. For example if a particular distance decay function implies that 70% of the offenders live within, say, 5 miles of their committed incidents, then 30% don't live within 5 miles. In other words, because the data are individual level, then a distance decay function, whether estimated by a mathematical or a kernel density model, is an individual probability model (i.e., an attempt to describe the underlying distribution of individual travel distances for journey to crime trips).