Chapter 28:

Crime Trip Distribution

Ned Levine
Ned Levine & Associates
Houston, TX

Richard Block
Loyola University
Chicago, IL

Dan Helms
Scytale Consulting
Reston, VA

Phil Canter
Towson University
Towson, MD

Table of Contents

Table of Contents (continued)

Chapter 28:
Crime Trip Distribution

In this chapter, the mechanics of the second crime travel demand modeling stage -trip distribution, is explained. *Trip distribution* is a model of the number of trips that occur between each origin zone and each destination zone. It uses the predicted number of trips originating in each origin zone (trip production model) and the predicted number of trips ending in each destination zone (trip attraction model). Thus, trip distribution is a model of travel between zones - trips or links. The modeled trip distribution can then be compared to the actual distribution to see whether the model produced a reasonable approximation.

Theoretical Background

The theoretical background behind the trip distribution module is presented first. Next, the specific procedures and tests are discussed with the model being illustrated with data from Baltimore County.

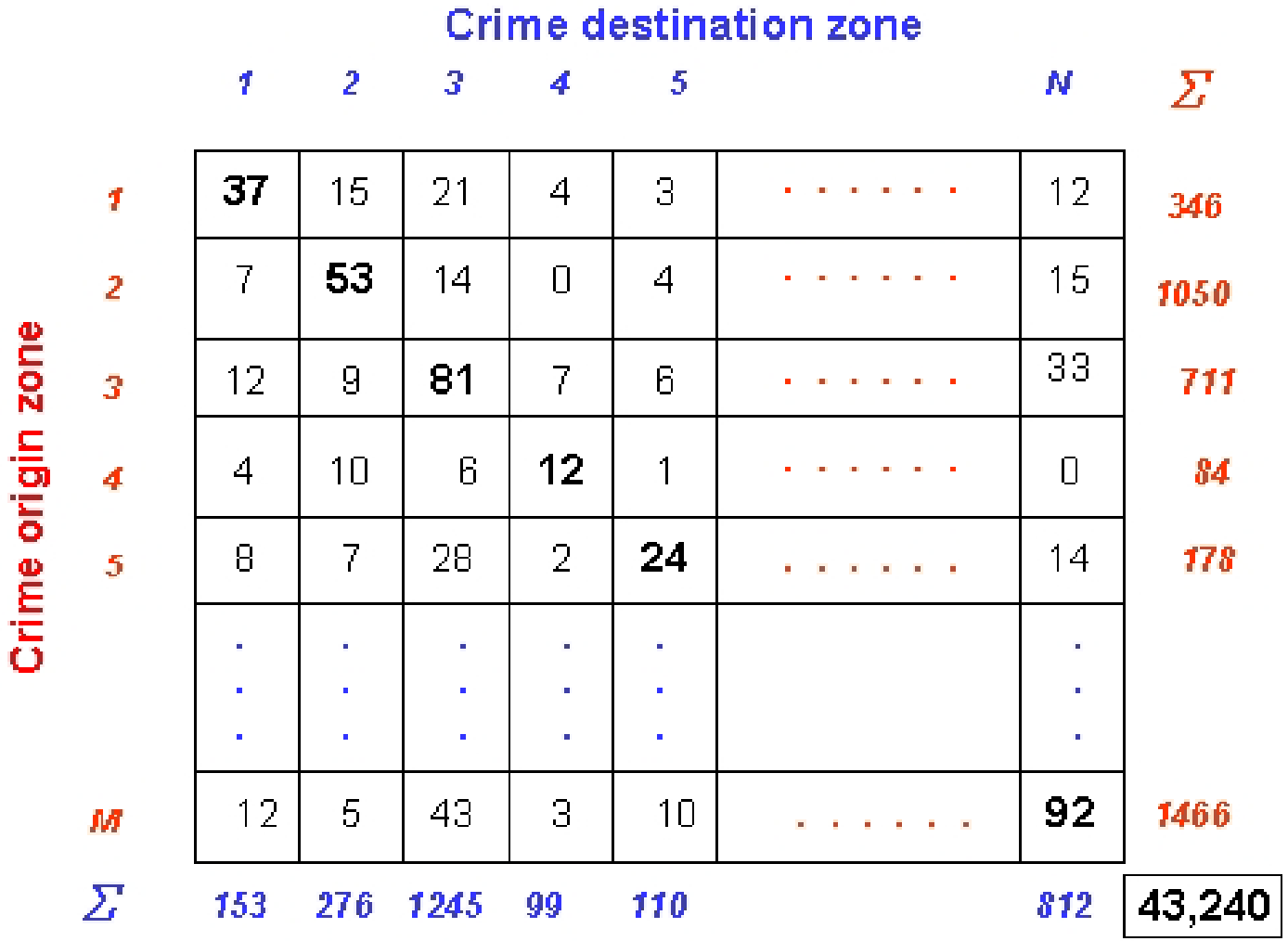
Logic of the Model

Trip distribution usually occurs through an allocation model that splits trips from each origin zone into distinct destinations. That is, there is a matrix which relates the number of trips originating in each zone to the number of trips ending in each zone. Figure 28.1 illustrates a typical arrangement. In this matrix, there are a number of origin zones, M, and a number of destination zones, N. The origin zones include ***all*** the destination zones but may also include additional ones. The reasons that there would be different numbers of zones for the origin and destination models are that crime data for other jurisdictions are not available but that many crimes that occurred in the study jurisdiction were committed by individuals who lived in other jurisdictions.

For example, with crimes that occurred in Baltimore County, approximately 35% were committed by offenders who lived in the City of Baltimore. Thus, it is important to include the City of Baltimore as an originating area for Baltimore County crimes. Hence, there are 325 destinations zones for Baltimore County while the origin zones include both the 325 in Baltimore County and 207 more from the adjacent City of Baltimore. If it were possible to obtain crime data for the City of Baltimore, then it would be possible to have the same number of zones for both the origin file and the destination file. As Chapter 26 pointed out, the study area should extend beyond the modeling area until the origins of at least 95% of all trips ending in the study area are counted.

Figure 28.1:

Example Crime Origin-Destination Matrix



Each cell in the matrix indicates the number of *trips* that go from each origin zone to each destination zone. To use the example in Figure 28.1, there were 15 trips from zone 1 to zone 2, 21 trips from zone 1 to zone 3, and so forth. Note that the trips are asymmetrical; that is, trips in one direction are different than trips in the opposite direction. To use the table, there were 15 trips from zone 1 to zone 2, but only 7 trips from zone 2 to zone 1.

The trips on the diagonal are *intra-zonal* trips, trips that originate and end in the same zone. Again, to use the example above, there were 37 trips that both originated and ended in zone 1, 53 trips that both originated and ended in zone 2.

In such a model, constancy is maintained in that the number of trips originating from all origins zones *must equal* the number of trips ending in all destination zones. This is the fundamental balancing equation for a trip distribution. In equation form, it is expressed as:

$$\sum_{i=1}^{M} O_i = \sum_{j=1}^{N} D_j \qquad (28.1)$$

where the origins, O_i, are summed over M origin zones while the destinations, D_j, are summed over N destination zones. To use the example in Figure 28.1, the total number of origins is equal to the total number of destinations, and is equal to 43,240.

The balancing equation is implemented in a series of steps that include modeling the number of crimes originating in each zone, adding in trips originating from outside the study area (external trips), and statistically balancing the origins and destinations so that equation 28.1 holds. This was done in the trip generation stage. But, it is essential that the step should have been completed for the trip distribution to be implemented.

Observed and Predicted Distributions

There are two trip distribution matrices that need to be distinguished. The first is the *observed* (or empirical) distribution. This is the actual number of trips that are observed traveling between each origin zone and each destination zone. In general, with crime data, such an empirical distribution would be obtained from an arrest record where the residence (or arrest) location of each offender is listed for each crime that the offender was charged with. In this case, the residence/arrest location would be considered the origin while the crime location would be considered the destination.

In Chapter 26, it was mentioned that there is always uncertainty as to the true origin location of a crime incident, whether the offender actually traveled from the residence location to the crime location or even whether the offender was actually living at the residence location.

But absent any alternative evidence, a meaningful distribution can still be obtained by simply treating the residence location as an approximate origin.

The observed distribution is calculated by simply enumerating the number of trips by each origin-destination combination. This is sometimes called a *trip link* (or trip pair). The second distribution, however, is a *model* of the trip distribution matrix. This is usually called the *predicted* distribution. In this case, a simple model is used to approximate the actual empirical distribution. The trips originating in each origin zone are allocated to destination zones usually on the basis of being directly proportional to attractions and inversely proportional to costs (or impedance).

Thus, a model of the trip distribution is produced that approximates the actual, empirical distribution. There are a number of reasons why this would be useful - to be able to apply the model to a different data set from which it was calibrated, to use the model for evaluating a policy intervention, or to use the model for forecasting future crime trip distribution. But, whatever the reason, it has to be realized that the model is not the observed distribution. There will always be a difference between the observed distribution from which a model is constructed and the resulting predicted distribution of the model. It is useful to compare the observed and predicted model because this allows a test of the validity of the impedance function. But, rarely, if ever, will the predicted distribution be identical to the empirical distribution.

Another way to think of this is that the actual distribution of crime trips is complex, representing a large number of different decisions on the part of offenders who do not necessarily use the same decision logic. The model, on the other hand, is a simple allocation on the basis of three or, sometimes, four variables. Almost by definition, it will be much simpler than the real distribution. Still, the simple model can often capture the most important characteristics of the actual distribution. Hence, modeling can be an extremely useful analytical exercise that allows other types of questions to be asked that are not possible with just the observed distribution.

The Gravity Model

A model that is usually used for trip distribution is that of the *gravity function*, an application of Newton's fundamental law of attraction (Oppenheim, 1980; Field & MacGregor, 1987; Ortuzar & Willumsen, 2001). Much of the discussion below is also repeated in Chapter 13 on journey-to-crime modeling since there is a common theoretical basis. 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

$$F = g\frac{M_1 M_2}{d^2} \tag{28.2}$$

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* (or resistance) 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:

$$I_{ij} = \alpha\frac{P_1 P}{d_{ij}^2} \tag{28.3}$$

where I_{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 variety of phenomena (migration, freight traffic, information) using a simplified form of the gravity equation:

$$I_{ij} = K\frac{P_1 P}{d_{ij}} \tag{28.4}$$

where the terms are as in equation 28.3 but the exponent of distance is only 1. 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. However, in modern use, it is not necessary for the productions and attractions to be identical units (e.g., P_i could be population while P_j could be employment).

Trips as Interactions

It should be obvious that this interaction equation can be applied to trips from one area (zone) to another. Changing the symbols slightly, the total volume of trips from a particular origin zone, i, to a single location, j, is directly proportional to the product of the productions at *i* and the attractions at *j*, and inversely proportion to the impedance (or cost) of travel between the two zones:

$$T_{ij} = \frac{\alpha P_1 \beta P}{d_{ij}} \tag{28.5}$$

where P_i are the productions for zone I, A_j are the attractions zone j, α is a production constant, β is an attraction constant, and d_{ij} is the impedance (cost) of travel between zone ii and zone j.

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:

$$A_{ij} = \alpha \frac{S_j}{d_{ij}^{\rho}} \tag{28.6}$$

where A_{ij} is the number of purchases in location *j* by residents of location$i$, S_j is the attractiveness of zone j (e.g., square footage of retail space), d_{ij} is the distance between zones i *and* j, α is a constant, λ is the exponent of S_j, and ρ is the exponent of distance (Bossard, 1993). $D_{ij}^{-\rho}$ is sometimes called an *inverse distance* function. 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.

Equation 28.6 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*,

$$T_{ij} = \alpha \frac{P_i \, A_j^{\tau}}{d_{ij}^{\rho}} \tag{28.7}$$

where T_{ij} is the interaction between two locations (or zones), P_i is productions of trips from zone i, A_j is the attractiveness of zone j, d_{ij} is the distance between zones i and$j$$i$, λ is the exponent of P_i, τ is the exponent of A_j, ρ is the exponent of distance, and α is a constant.

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

$$T_i = \alpha P_i^{\lambda} \sum_{j=1}^{N} \frac{A_j^{\tau}}{d_{ij}^{\rho}} \tag{28.8}$$

and generalizing this to all zones, we get:

$$T_{ij} = \frac{\alpha P_i \ \beta A_j^{\tau}}{d_{ij}^{\rho}} \tag{28.9}$$

where α is a constant for the productions, P_i^{λ} and β is a constant for the attractions, A_j^{τ}. This type of function is called a *double constraint* model because the equation 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 A_j 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).

Negative Exponential Distance Function

One of the problems with the traditional gravity formulation is in the measurement of travel impedance (or cost). 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 uses the negative exponential function (Hägerstrand, 1957; Wilson, 1970).

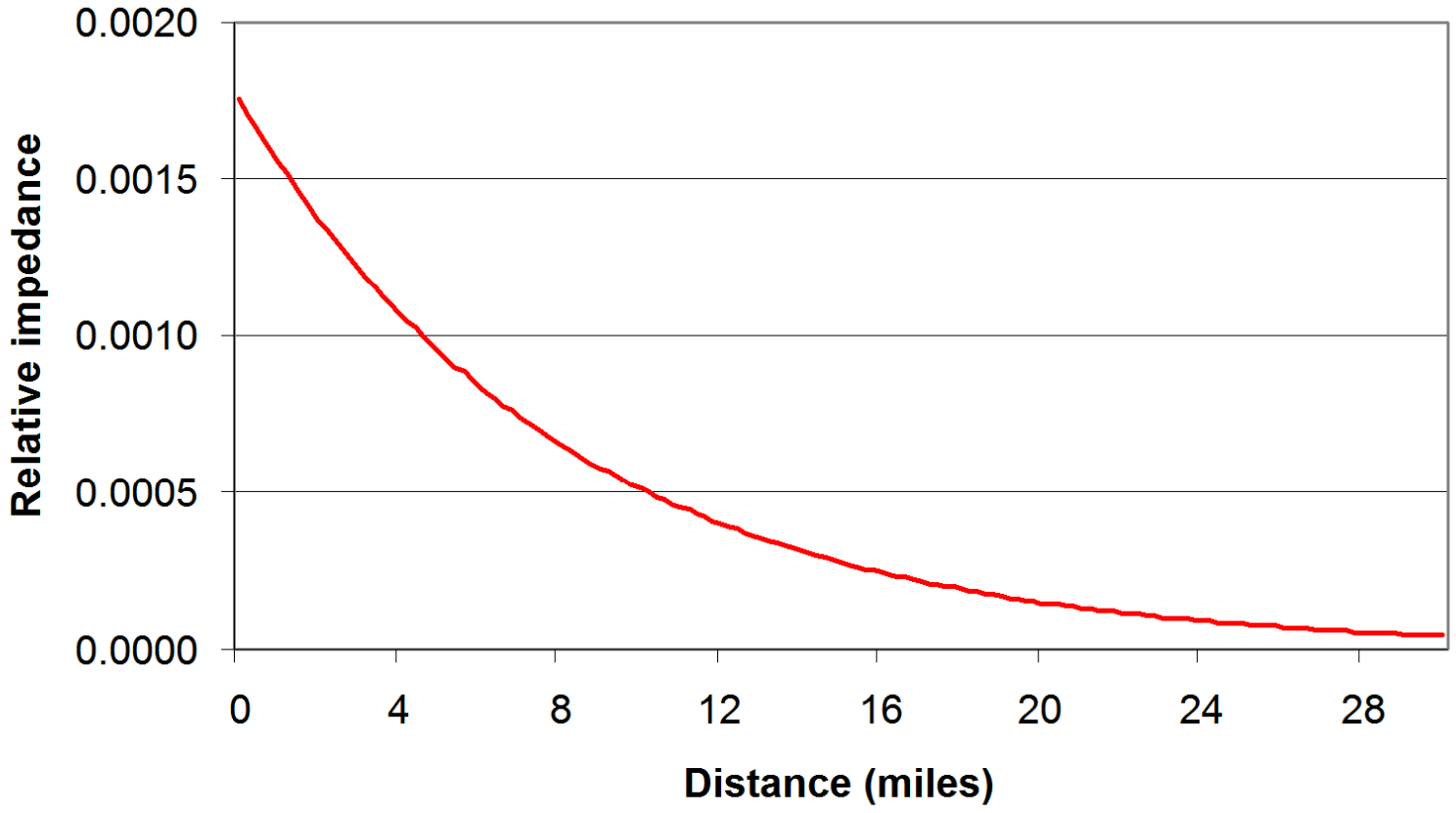
$$T_{ij} = \beta A_j^{\lambda} e^{-\alpha d_{ij}} \tag{28.10}$$

where T_{ji} is the attraction of location j for residents of location I, A_j is the attractiveness of location j, d_{ij} is the distance between locations i and *j*, β is the exponent of A_j, α is a coefficient of d_{ij} (and, also, an exponent) and *e* is the base of the natural logarithm (i.e., 2.7183...). Derived from principles of *entropy maximization*, the latter part of the equation is a negative exponential function that has a maximum value of 1 (i.e., $e^{-0} = 1$; Wilson, 1970). 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 those by automobile (Bossard, 1993; Foot, 1981). Figure 28.2 shows a typical negative exponential function and one recommended for home-based work trips by the Transportation Research Board as a default value (NCHRP, 1995).

Figure 28.2:

Default Home-Based Work Trip Impedance

(Source: National Cooperative Highway Research Program 365, 1995)



Note that by moving the distance term to the numerator, strictly speaking it no longer is an impedance term since impedance increases with distance. Rather it is a *discount* factor (or *disincentive*); the interaction is discounted with distance. Nevertheless, the term 'impedance' is still used primarily for historical reasons.

There are other distance functions, as well. Chapter 13 explored some of these. For example, we are finding that, for crime trips, the lognormal function may produce better results than the negative exponential primarily because many crimes are committed at short-to-moderate distances. Chapter 17 discusses the MCMC Poisson-lognormal regression model which is useful with a low mean (e.g., very short distance traveled) and small sample sizes. It is possible that the lognormal function is more useful for very localized crime trips than the negative exponential.

Travel Impedance

One of the biggest advances in the negative exponential model of equation 28.10 has been to increase the flexibility of the denominator. In the traditional gravity model, the denominator is distance. This is a proxy for a *discount factor* (or cost); the farther two zones are from each other, the less likely there is to be interaction between them, all other things being equal. Conversely, the closer two zones are, the more likely there is to be interaction, all other things being equal.

Distance v. Travel Time

It has been realized, however, that distance is only an approximation for impedance. In real travel, travel time is a much better indicator of the *cost* of travel in that time varies by the time of day, day of week, direction of travel, type of road used, and other factors. For example, travel across town in any metropolitan area is generally a lot easier at 3 in the morning, say, than at the peak afternoon rush period. The difference in travel time can vary as much as two-to-three times between peak and off-peak hours. Using only distance, however, these variations are never picked up because the distance between locations is invariant.

This realization has led to the concept of *travel impedance* which, in turn, has led to the concept of *travel cost*. 'Impedance' is the resistance (or discounting) in travel between two zones. Using travel time as an impedance variable, the longer it takes to travel between two zones, the less likely there will be interaction between them, all other things being equal. Conversely, a shorter travel time leads to greater interaction between zones, again, all other things being equal. Similarly, a travel route that shortens travel time will generally be selected over one that takes longer even if the first one is longer in distance. For example, it has been

documented that people will change work locations that are farther from their home if traveling to the new work location takes less time (e.g., traveling in the 'opposite' direction to the bulk of traffic; Wachs, Taylor, Levine & Ong, 1993).

If travel time is a critical component of travel, why then don't offenders commit more crimes at, say, 3 in the morning than at the peak afternoon travel times? Since the impedance is less at 3 in the morning than at, say, 5 in the afternoon, would not the model predict more trips occurring in the early morning hours than actually occur in those hours? The answer has to do with the numerator of the gravity equation and not just the denominator. At 3 in the morning, yes, it is easier to travel between two locations, at least by personal automobile (not by bus or train when those services are less frequent). But the attraction side of the equation is also less strong at 3 in the morning. For a street robber, there are fewer potential 'victims' on the street at 3 in the morning than in the late afternoon. For a residential burglar, there is more likely to be someone at home at night than in the afternoon. The travel time component is only one dimension of the likelihood of travel between two locations. The distribution of opportunities and other costs can alter the likelihood considerably.

Nevertheless, shifting to an impedance function allows a travel model to better replicate actual travel conditions. Most travel demand models used by transportation planners use an impedance function, rather than a distance function.[1] Distance would only be meaningful if the standards were invariant with respect to time (e.g., a model calculated over an entire year, 24 hours a day). As will be demonstrated in Chapter 30 on network assignment, a travel time calculation leads to a very different network allocation than a distance calculation. For example, if distance is used as an impedance variable, then the shortest trips will rarely take the freeways because travel to and from a freeway usually makes a trip longer than a direct route between an origin and a destination. But as most people understand, taking a freeway to travel a sizeable distance is usually a lot quicker than traversing an urban arterial system with many traffic lights, stop signs, crossing pedestrians, cross traffic from parking lots and shopping malls, and other urban 'obstacles'. Today, the use of distance in travel demand modeling has virtually been dropped by most transportation planners.

---

[1] Distance can be used as a rough approximation for impedance, but is rarely a good predictor of actual travel behavior. For example, in the mode split mode that will be discussed in Chapter 29, the distance between a location and the nearest bus or rail route can be used to quickly select trip pairs that might travel by transit. However, the actual prediction must be based on a network calculation of travel time or travel cost in traversing the system.

Travel Cost

An even better concept of impedance is that of *travel cost* (sometimes called *generalized cost*) which incorporates real and perceived costs of travel between two locations. Travel time is one component of travel cost in that there is an implicit cost to the trip (e.g., an hourly wage or price assigned). In this case, two different individuals will value the time for a trip differently depending on their hourly 'wage'. For example, for an individual who prices his/her travel at $100 an hour, the per minute cost is $1.67. For another individual who prices his/her travel at $12 an hour, the per minute cost is 20¢. These relative prices assigned to travel will substantially affect individual choices in travel modes and routes. For instance, these two hypothetical individuals will probably use a different travel mode in getting from an airport to a hotel on a trip; the former will probably take a taxi whereas the latter will probably take a bus or train (if available).

But cost involves other dimensions that need to be considered. There are real operating costs in the use of a vehicle - fuel, oil, maintenance, and insurance. Many travel studies have suggested that drivers incorporate these costs as part of their implicit hourly travel price (Ortuzar & Willumsen, 2001; 323-327). But, there are also real, 'out-of-pocket' costs such as parking or toll costs. Parking is particularly a major expense for intra-urban driving behavior. In many built-up business districts, parking costs can be considerable, for example as much as $90 a day in major metropolitan centers. In most busy commercial areas, there are some parking costs, if only at on-street parking meters. Thus, a travel cost model needs to incorporate these real costs as the out-of-pocket costs may overwhelm the implicit value of the travel time. For example, an offender who lives 10 minutes from the downtown area by car would probably not drive into the downtown to commit a robbery since that individual will have to bear the price of parking. There are lots of well known stories that circulate about bank robbers who are caught because they incur parking tickets while committing their crime. How often this has occurred is not known from any study that we are aware, but the story line is cognizant of the actual costs of travel that must be incurred as part of travel.

In addition to real costs are perceived costs. For transit users, particularly, these perceived costs affect the ease and time of travel. One of the standard questions in travel surveys for transit users is the time it takes to walk from their home to the nearest bus stop or intra-urban rail system (if available) and from the last transit stop to their final destination; the longer it takes to access the transit system, the less likely an individual will use it. Similarly, transfers between buses or trains decrease the likelihood of travel by that mode, almost in proportion to the number of transfers. The reason is the difficulty in getting out of one bus or train and into another. But, the time between trains adds an implicit travel cost; the longer the wait between buses, the less likely that mode will be used by travelers. In short, ease of access and convenience are positive incentives in using a mode or a route while difficulty in accessing

it, lack of convenience, and even fear of being vulnerable to crime will decrease the likelihood of using that mode or route.[2]

If the concept is expanded to that of an offender, there are other perceived costs that might affect travel. One obvious one is the likelihood of being caught. It may be easy for one offender to travel to an upscale, high visibility shopping area, but if there are many police and security guards around, the individual is more likely to be caught. Hence, that likelihood (or, more accurately, an assumption that the offender makes about that likelihood since he/she does not really know what is the real likelihood) is liable to affect the choice of a destination and, possibly, even a route.

Another perceived cost is the likelihood of retaliation from other gangs. Bernasco and Block (2009) showed that robbers in Chicago will usually not commit robberies in the territories of rival gangs even if those areas are closer to where the robbers live.

Another perceptual component affecting a likely choice is the reliability of the transportation mode. Many offenders are poor and do not have expensive, well maintained vehicles. If the vehicle is not capable of higher speeds or is even likely to break down while an offence is being committed that vehicle is not liable to be used in making a trip or the choice of destination may be altered. It is well known that many offenders steal vehicles for use in a crime. Fears about not being identified are clearly a major factor in those decisions, but the reliability of their own vehicles may also be a factor.

Thus, in short, a more realistic model of the incentives or disincentives to make a trip between two locations requires a complex function that weights a number of factors affecting the cost of travel - the travel time, implicit operating costs, out-of-pocket costs, and perceived costs. Many travel demand models used by Metropolitan Planning Organizations use such a function, usually under the label of 'generalized cost'. The more complex the pricing structure for parking and travel within a metropolitan area, the more likely a generalized cost function will provide a realistic model of trip distribution.

Travel Utility

The final concept that is introduced in defining impedance is that of *travel utility*. 'Utility' is an individual concept, rather than a zonal one. Also, it is the flip side of cost (i.e., higher cost is associated with less utility). A generalized cost function calculates the objective

2 Most of the research on factors affecting use of transit were conducted in the 1960s and 1970s. These assumptions are more or less assumed by travel demand modelers, rather than documented *per se*. See Schnell, Smith, Dimsdale, & Thrasher, 1973; Roemer & Sinha, 1974; WASHCOG, 1974; Carnegie-Mellon University, 1975; Johnson, 1978; Levine & Wachs, 1986 for some examples.

and average perceived costs of travel between two zones. But the utility of travel for an individual is a function of both those real costs and a number of individual characteristics that affect the value placed on that travel. Thus, two individuals living in the same zone (perhaps even living next door to each other) who travel to the same destination location may 'price' their trip very differently. Aside from income differences which effect the average hourly 'wage', there may be differences due to convenience, attractiveness, or a host of other factors. Other factors are more idiosyncratic. For example, a trip by a gang member into another gang's 'turf' might be expected to increase the perceived costs to the individual of traveling to that location, above and beyond any objective cost factors. Alternatively, a trip to a location where a close friend or relative is located might decrease the perceived cost of travel to that zone. In other words, there are both objective costs as well as subjective costs in travel between two zones.

The concept of utility may be less useful for crime analysis than for general travel behavior. For one thing, since the concept is individual, it can only be identified by individual surveys (Domencich & McFadden, 1975). For crime analysis, this makes it virtually impossible to use since it is very difficult to interview offenders, at least in the United States. But, for completeness sake, we need to understand that the likelihood or disincentive to travel between two locations is a function of individual characteristics as well as objective travel cost components. A mixture of aggregate and individual variables can be used to produce a synthetic utility model for modeling locations where individuals commit crimes (Block & Bernasco, 2009).

The modeling of individual utility can be done with either a multinomial logit model for a limited number of discrete choices or a more general conditional logit model for many choices. Chapter 21 discusses these models while Chapter 22 presents the CrimeStat discrete choice module. At this point, it is impractical to utilize either model for predicting trip distribution links since the number of origin-destination pairs would require an enormous data set. So, we are left for the time being with the gravity function being the only practical approach to trip distribution.

Impedance Function

For a zonal type model, we can think of the gravity function as a generalized impedance function. For travel between any one zone and all other zones, we have:

$$T_i = \alpha P_i^{\lambda} \sum_{j=1}^{N} \frac{A_j^{\tau}}{I_{ij}} \qquad (28.11)$$

where the number of trips from zone i to all other zones is a function of the productions at zone i and the relative attraction of any one zone, j, to the impedance of that zone for i, I_{ij}. The

impedance function, I_{ij} , is some declining function of cost for travel between two zones. It does not have to be any particular form and can be (and usually is) a non-linear function. The costs can be in terms of distance, travel time, speed (which is converted into travel time) or general costs. The greater the separation between two zones (i.e., the higher the impedance), the less likely there will be a trip between them. Generalizing this to all zones, we get:

$$T_{ij} = \alpha P_i^{\lambda} \beta \frac{A_j^{\tau}}{I_{ij}} \tag{28.12}$$

where P_i is the production capacity of zone i, A_j is the attraction of zone j, I_{ij}, is a generalized function that discounts the interaction with increasing separation in distance, time, or cost, α and β are constants that are applied to the productions and attractions respectively, and λ and τ are 'fine tuning' exponents of the productions and attractions respectively. This is the gravity function that we will estimate in the *CrimeStat* trip distribution model.

Alternative Model: Intervening Opportunities

There are alternative allocations procedures to the gravity model. One well known one is that of *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 inter-county migration (Isard, 1979; Isbell, 1944; Bright & Thomas, 1941). Using the gravity type formulation, this can be written as:

$$A_{ji} = \alpha \frac{S_j^{\beta}}{\sum_{k=1}^{O} S_k^{\xi} d_{ij}} \tag{28.13}$$

where A_{ji} is the attraction of location j by residents of location i, S_j is the attractiveness of zone j, S_k is the attractiveness of all other locations, , that are *intermediate* in distance between locations i and j (with there being *O* such locations), d_{ij} is the distance between zones i and j, β is the exponent of S_j, ξ is the exponent of S_k, and λ is the exponent of distance. While the intervening opportunities are implicit in equation 28.7 in the exponents, β and λ, and coefficient, α, equation 28.13 makes the intervening opportunities explicit. The importance of the concept is that travel between two locations becomes a complex function of the spatial environment of nearby areas and not just of the two locations.

In practice, in spite of its more intuitive theoretical model, the intervening opportunities model does not improve prediction much beyond that of the gravity model since it includes the attractions associated with the destination zones. Also, it is a more difficult model to estimate since the attractions of all other zones must be considered for each zone pair (origin-destination combination). Consequently, it is rarely used in actual practice (Ortuzar & Willumsen, 2001).

Another alternative method was conducted by Porojan (2000) in applying the gravity model to international trade flow. He added a spatial autocorrelation component in addition to impedance and obtained a slightly better fit than the pure gravity function. However, whether this approach would improve the fitting of intra-regional crime travel patterns is still unknown. Nevertheless, this and other approaches might improve the predictability of a gravity function for intra-urban crime travel.

Method of Estimation

The *CrimeStat* trip distribution model implements equation 28.12. The specific details are discussed below, but the model is iterative. The steps are as follows:

1. Depending on whether a singly constrained or doubly constrained model is to be estimated, it starts with aN initial guess of the values for α or β (or both for a doubly constrained model). Table 28.1 illustrates the three models.

Table 28.1:
Three Methods of Constraining the Gravity Model

Single constraint

Constrain origins:

$$T_{ij} = \alpha P_i^{\lambda} A_j^{\tau} I_{ij} \quad (28.14)$$

Constrain destinations:

$$T_{ij} = P_i^{\lambda} \beta A_j^{\tau} I_{ij} \quad (28.15)$$

Double constraint

Constrain both origins and destinations:

$$T_{ij} = \alpha P_i^{\lambda} \beta A_j^{\tau} I_{ij} \quad (28.16)$$

2. The routine proceeds to estimate the value for each cell in the origin-destination matrix (see Figure 28.1 above) using the existing estimates for α and β.

3. The routine then sums the rows and columns in the matrix. Then, depending on whether a single- or double-constraint model is to be estimated and, if a single-constraint, whether origins or destinations are to be held constant, it then calculates the ratio of the summed values (row totals or column totals or both) to the initial row or column sum. The inverse of that ratio is the subsequent estimate for α or β (or both for a double-constrained model).

4. The routine repeats steps 2 and 3 until the changes from one iteration to the next are very small.

5. The last estimate of α or β (or both for a double-constrained model) is taken as the final values of these parameters.

6. Once the parameters have been estimated, the model can be applied to the calibration data set or to another data set. Note that the parameters are row or column specific (or both). That is, in the 'constrain origins' model, there is a separate coefficient for each row. In the 'constrain destinations' model, there is a separate coefficient for each column. In the 'constrain both origins and destinations', there is a separate coefficient for each cell (row-column combination).

A comparison can be made between the observed distribution and the predicted (modeled) distribution. Because most origin-destination matrices are very large, the vast majority of cells will have zero in them. Thus, a chi-square test would be inappropriate. Instead, a comparison of the *trip length* distribution is made using two different statistics - a coincidence ratio and the Komologorov-Smirnov Two-sample statistic. Details are provided below.

CrimeStat IV Trip Distribution Module

Next, we examine the actual tools that are available in the *CrimeStat* trip distribution module. The tools are illustrated with examples from Baltimore County. The *CrimeStat* trip distribution module includes one setup screen and five routines that implement the model:

1. **Calculate observed origin-destination distribution**. If there is a file available with the coordinates for individual origins and destinations (e.g., an arrest record), this routine will calculate the empirical trip distribution matrix;

2. **Calibrate impedance function**. If there is a file available with the coordinates for individual origins and destinations, this routine will calibrate an empirical impedance function.

3. **Setup origin-destination model**. This screen allows the user to define the parameters of a trip distribution (origin-destination) model with either a mathematical or empirical impedance function.

4. **Calibrate origin-destination model**. This routine calibrates the parameters of the trip distribution model (equation 28.12) using the parameters defined on the setup page.

5. **Apply predicted origin-destination model**. This routine applies the estimated parameters to a data set. The data set can be either the calibration file or another file.

6. **Compare observed and predicted origin-destination trip lengths.** This routine compares the trip lengths from the observed (empirical) trip distribution with that predicted by the model. Comparisons are made graphically by a coincidence ratio, the Komologorov-Smirnov Two-Sample test, and a Chi square test on the most frequent trip links.

Each of these routines is described in detail below. Figure 28.3 shows a screen shot of the trip distribution module.

Describe Origin-Destination Trips

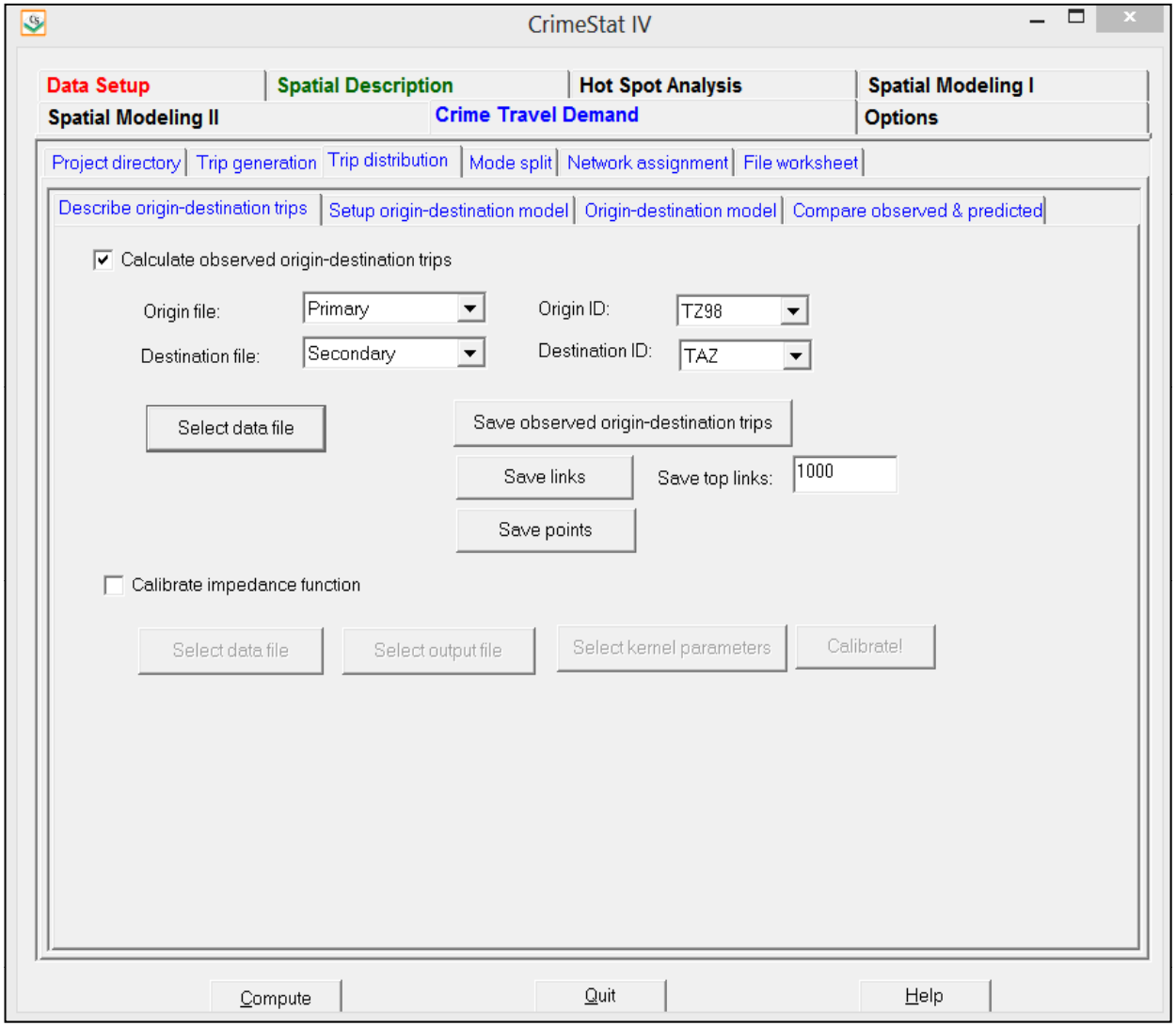
An empirical description of the actual trip distribution matrix can be made if there is a data set that includes individual origin and destination locations. The user defines the origin location and the destination location for each record and a set of zones from which to compare the individual origins and destinations. The routine matches up each origin location with the nearest zone, each destination location with the nearest zone, and calculates the number of trips from each origin zone to each destination zone. This is an *observed* distribution of trips by zone.

The steps in running the model are as follows:

1. **Calculate observed origin-destination trips**. Check if an empirical origin-destination trip distribution is to be calculated.

Figure 28.3:

Trip Distribution Module



2. **Origin file**. The origin file is a list of origin zones with a single point representing the zone (e.g., the centroid). It must be input as either the primary or secondary file. Specify whether the data file is the primary or secondary file.

   A. ***Origin ID***. Specify the origin ID variable in the data file (e.g., CensusTract, Block, TAZ).

3. **Destination file.** The destination file is a list of destination zones with a single point representing the zone (e.g., the centroid). It must be input as either the primary or secondary file. Specify whether the data file is the primary or secondary file.Specify the destination ID variable in the data file (e.g., CensusTract, Block, TAZ).

   A. Note: all destination ID's should be in the origin zone file and must have the same names and both should be character (string) variables.

4. **Select data file**. The data set must have individual origin and destination locations. Each record must have the X/Y coordinates of an origin location and the X/Y coordinates of a destination location. For example, an arrest file might list individual incidents with each incident having a crime location (the destination) and a residence or arrest location (the origin).

   A. Select the file that has the X and Y coordinates for the origin and destination locations. *CrimeStat* can read ASCII, dbase '.dbf', ArcGIS '.shp' and MapInfo 'dat' files.

   B. Select the tab and specify the type of file to be selected. Use the browse button to search for the file. If the file type is ASCII, select the type of data separator (comma, semicolon, space, tab) and the number of columns.

   C. ***Variables***. Define the file which contains the X and Y coordinates for both the origin (residence) and destination (crime) locations.

   D. ***Column***. Select the variables for the X and Y coordinates respectively for *both* the origin and destination locations (e.g., Lon, Lat, HomeX, HomeY, IncidentX, IncidentY.) Both locations must be defined for the routine to work.

   E. ***Missing values***. Identify 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:

a. <blank> fields are automatically excluded. This is the default
b. <none> indicates that no records will be excluded. If there is a blank field, *CrimeStat* will treat it as a 0
c. 0 is excluded
d. −1 is excluded
e. 0 and −1 indicates that both 0 and -1 will be excluded
f. 0, -1 and 9999 indicates that all three values (0, -1, 9999) will be excluded.
g. 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).

F. ***Type of coordinate system and data units.*** The coordinate system and data units are listed for information. If the coordinates are in longitudes and latitudes, then a spherical system is being used and data units will automatically be decimal degrees. If the coordinate system is projected (e.g., State Plane, Universal Transverse Mercator − UTM), then data units could be either in feet (e.g., State Plane) or meters (e.g., UTM.).

5. **Table output**. The full origin-destination matrix is output as a table to the screen including summary file information and:

a. The origin zone (ORIGIN)
b. The destination zone (DEST)'
c. The number of observed trips (FREQ)

6. **Save observed origin-destination trips.** If specified, the full origin-destination matrix output is saved as a 'dbf' file named by the user. The file output includes:

a. The origin zone (ORIGIN)
b. The destination zone (DEST)
c. The X coordinate for the origin zone (ORIGINX)
d. The Y coordinate for the origin zone (ORIGINY)
e. The X coordinate for the destination zone (DESTX)

f. The Y coordinate for the destination zone (DESTY)
g. The number of trips (FREQ)

Note: each record is a unique origin-destination combination. There are M x N records where *M* is the number of origin zones (including the external zone) and *N* is the number of destination zones.

7. **Save links.** The top observed origin-destination trip links can be saved as separate **line** objects for use in a GIS. Specify the output file format (*ArcGIS* '.shp', *MapInfo* '.mif' or *Atlas*GIS* '.bna') and the file name.

8. **Save top links**. Because the output file is very large (number of origin zones x number of destination zones), the user can select a sub-set of zone combinations with the most observed trips. Indicating the top K links will narrow the number down to the most important ones. The default is the top 100 origin-destination combinations. Each output object is a line from the origin zone to the destination zone with an ODT prefix. The prefix is placed before the output file name. The line graphical output for each object includes:

    a. An ID number from 1 to K, where K is the number of links output (ID)
    b. The feature prefix (ODT)
    c. The origin zone (ORIGIN)
    d. The destination zone (DEST)
    e. The X coordinate for the origin zone (ORIGINX)
    f. The Y coordinate for the origin zone (ORIGINY)
    g. The X coordinate for the destination zone (DESTX)
    h. The Y coordinate for the destination zone (DESTY)
    i. The number of observed trips for that combination (FREQ)
    j. The distance between the origin zone and the destination zone.

9. ***Save points***. Intra-zonal trips (trips in which the origin and destination are the same zone) can be output as separate **point** objects as an *ArcGIS* '.shp', *MapInfo* '.mif' or *Atlas*GIS* '.bna' file. Again, the top K points are output (default=100). Each output object is a point representing an intra-zonal trip with an ODTPOINTS prefix. The prefix is placed before the output file name. The point graphical output for each object includes:

a. An ID number from 1 to K, where K is the number of links output (ID)
b. The feature prefix (POINTSODT)
c. The origin zone (ORIGIN)
d. The destination zone (DEST)
e. The X coordinate for the origin zone (ORIGINX)
f. The Y coordinate for the origin zone (ORIGINY)
g. The X coordinate for the destination zone (DESTX)
h. The Y coordinate for the destination zone (DESTY)
i. The number of observed trips for that combination (FREQ)

Example of Observed Trip Distribution from Baltimore County

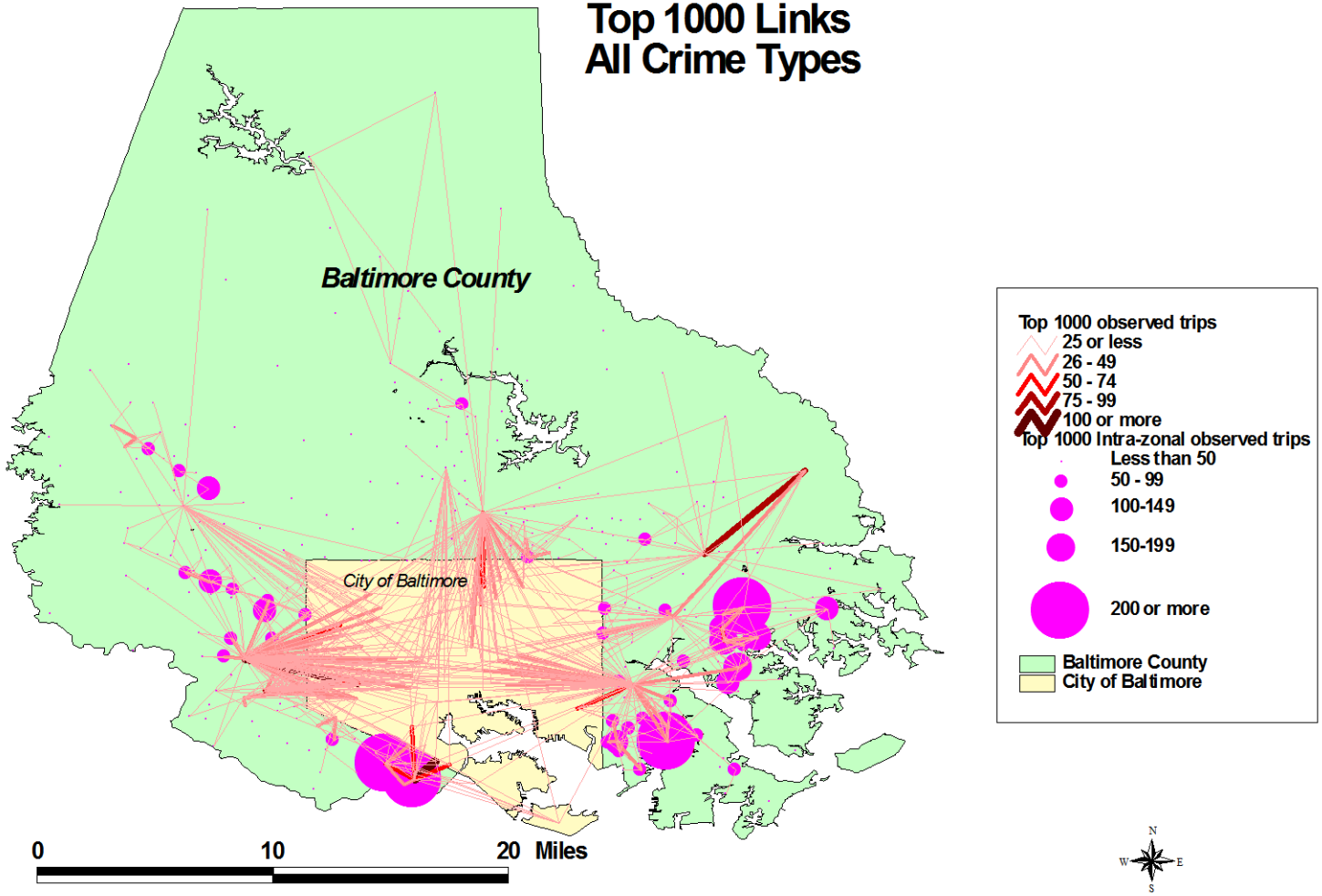
Figure 28.4 shows the output of the top 1000 links for the observed trip distribution from a sample of 41,974 records for incidents committed between 1993 and 1997. The zonal model used was that of traffic analysis zones (TAZ). These were discussed in Chapter 26. Because there are a large number of links (532 origin zones by 325 destination zones), the top 1000 were taken. These accounted for 19,615 crime trips (or 46.7% of all trips). A larger number of links could have been selected, but the map would have become more cluttered. Of the 19,615 trips that are displayed in the map, 7,913 or 40.3% are intra-zonal trips. These were output by the routine as points and have been displayed as circles with the size proportional to the number of trips. The remaining 11,702 trip links were output by the routine as lines and are displayed with the thickness and strength of color of the line being proportional to the number of trips.

There are several characteristics of the trip pattern that should be noted. First, the intra-zonal trips tend to concentrate on the eastern part of Baltimore County. This is an area that is relatively poor with a high number of public housing projects. This suggests that there are a lot of intra-community crimes being committed in these locations. Second, the zone-to-zone pattern, on the other hand, tends to concentrate at five different locations relatively close to border with the City of Baltimore. These five locations are all major shopping malls. Third, the origins for those trips to the shopping mall tend to come from within the City of Baltimore. Fourth, in general, the locations with high intra-zonal trips do not have a large number of zone-to-zone trips. However, there is one exception in the southwest corner of the county.

In other words, the observed distribution of crime trips is complex, but with several patterns being shown. A lot of crime trips occur over very short distances. But there is also a convergence of many crime trips on major shopping malls in the County.

Figure 28.4:

Observed Baltimore County Crime Trips: 1993-1997
Top 1000 Links
All Crime Types



Calibrate Impedance Function

This routine allows the calibration of an approximate travel impedance function based on actual trip distributions. It is used to describe the travel impedance in distance only of an actual sample (the calibration sample). Unlike the remaining routines in this section, the "Calibrate impedance function cannot use travel time, or cost. A file is input which has a set of incidents (records) that include both the X and Y coordinates for the location of the offender's residence (origin) and the X and Y coordinates for the location of the incident that the offender committed (destination.)

The routine estimates a travel distance function using a one-dimensional kernel density method. See the details in Chapter 13. Essentially, for each record, the separation between the origin location and the destination location is calculated and is represented on a distance scale. The maximum impedance is calculated and divided into a number of intervals; the default is 100 equal sized intervals, but the user can modify this. For each impedance point calculated, a one-dimensional kernel is overlaid. For each interval, the values of all kernels are summed to produce a smooth function of travel impedance. The results are saved to a file that can be used for the origin-destination model.

Note, however, that this is an empirical distribution and represents the combination of origins, destinations, and costs. It is not necessarily a good description of the impedance (cost) function by itself. Some of the mathematical functions produce a better fit than the empirical impedance function.

The steps in calculating an empirical impedance function are as follows:

1. **Select data file for calibration**. Select the file that has the X and Y coordinates for the origin and destination locations. *CrimeStat* can read ASCII, dbase '.dbf', ArcGIS '.shp' and MapInfo 'dat' files. Select the tab and select the type of file to be selected. Use the browse button to search for the file. If the file type is ASCII, select the type of data separator (comma, semicolon, space, tab) and the number of columns.

   A. ***Variables***. Define the file which contains the X and Y coordinates for both the origin (residence) and destination (crime) locations

   B. ***Columns***. Select the variables for the X and Y coordinates respectively for *both* the origin and destination locations (e.g., Lon, Lat, HomeX, HomeY, IncidentX, IncidentY.) Both locations must be defined for the routine to work.

C. ***Missing values***. Identify 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:

a. <blank> fields are automatically excluded. This is the default
b. <none> indicates that no records will be excluded. If there is a blank field, *CrimeStat* will treat it as a 0
c. 0 is excluded
d. −1 is excluded
e. 0 and −1 indicates that both 0 and -1 will be excluded
f. 0, -1 and 9999 indicates that all three values (0, -1, 9999) will be excluded.
g. 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).

D. ***Type of coordinate system and data units***. Select the type of coordinate system. If the coordinates are in longitudes and latitudes, then a spherical system is being used and data units will automatically be decimal degrees. If the coordinate system is projected (e.g., State Plane, Universal Transverse Mercator − UTM), then data units could be either in feet (e.g., State Plane) or meters (e.g., UTM.) Directional coordinates are not allowed for this routine.

2. **Select Kernel Parameters**. There are five parameters that must be defined.

A. ***Method of interpolation***. There are fives types of kernel distributions that can be used to estimate point density:

a. The **normal** kernel overlays a three-dimensional normal distribution over each point that then extends over the area defined by the reference file. This is the default kernel function.
b. The **uniform** kernel overlays a uniform function (disk) over each point that only extends for a limited distance.
c. The **quartic** kernel overlays a quartic function (inverse sphere) over each point that only extends for a limited distance.

d. The **triangular** kernel overlays a three-dimensional triangle (cone) over each point that only extends for a limited distance.

e. The **negative exponential** kernel overlays a three dimensional negative exponential function ('salt shaker') over each point that only extends for a limited distance

B. The methods produce similar results though the normal is generally smoother for any given bandwidth.

3. ***Choice of*** **bandwidth**. The kernels are applied to a limited search distance, called 'bandwidth'. For the normal kernel, bandwidth is the standard deviation of the normal distribution. For the uniform, quartic, triangular and negative exponential kernels, bandwidth is the radius of a circle defined by the surface. For all types, larger bandwidth will produce smoother density estimates and both adaptive and fixed bandwidth intervals can be selected.

A. ***Fixed bandwidth***. A fixed bandwidth distance is a fixed interval for each point. The user must define the interval, the interval size, and the distance units by which it is calculated (miles, nautical miles, feet, kilometers, meters.) The default bandwidth setting is fixed with intervals of 0.25 miles each. The interval size can be changed.

B. ***Adaptive bandwidth***. An adaptive bandwidth distance is identified by the minimum number of other points found within a symmetrical band drawn around a single point. A symmetrical band is placed over each distance point, in turn, and the width is increased until the minimum sample size is reached. Thus, each point has a different bandwidth size. The user can modify the minimum sample size. The default for the adaptive bandwidth is 100 points.

4. ***Specify Interpolation Bins***. The interpolation bins are defined in one of two ways:

A. By the number of bins. The maximum distance calculated is divided by the number of specified bins. This is the default with 100 bins. The user can change the number of bins.

B. By the distance between bins. The user can specify a bin width in miles, nautical miles, feet, kilometers, and meters.

5. **Output (Areal) Units**. Specify the density units as points per mile, nautical mile, foot, kilometer, or meter. The default is points per mile.

6. **Calculate Densities or Probabilities**. The density estimate for each cell can be calculated in one of three ways:

   A. **Absolute densities**. This is the number of points per grid cell and is scaled so that the sum of all grid cells equals the sample size.

   B. **Relative densities**. For each grid cell, this is the absolute density divided by the grid cell area and is expressed in the output units (e.g., points per square mile)

   C. **Probabilities**. This is the proportion of all incidents that occur in the grid cell. The sum of all grid cells equals a probability of 1. Unlike the Jtc calibration routine, this is the default. In most cases, a user would want a proportional (probability) distribution as the relative differences in impedance for different costs are what is of interest.

   Select whether absolute densities, relative densities, or probabilities are to be output for each cell. The default is probabilities.

7. **Select Output File**. The output *must* be saved to a file. *CrimeStat* can save the calibration output to either a dbase 'dbf' or ASCII text 'txt' file.

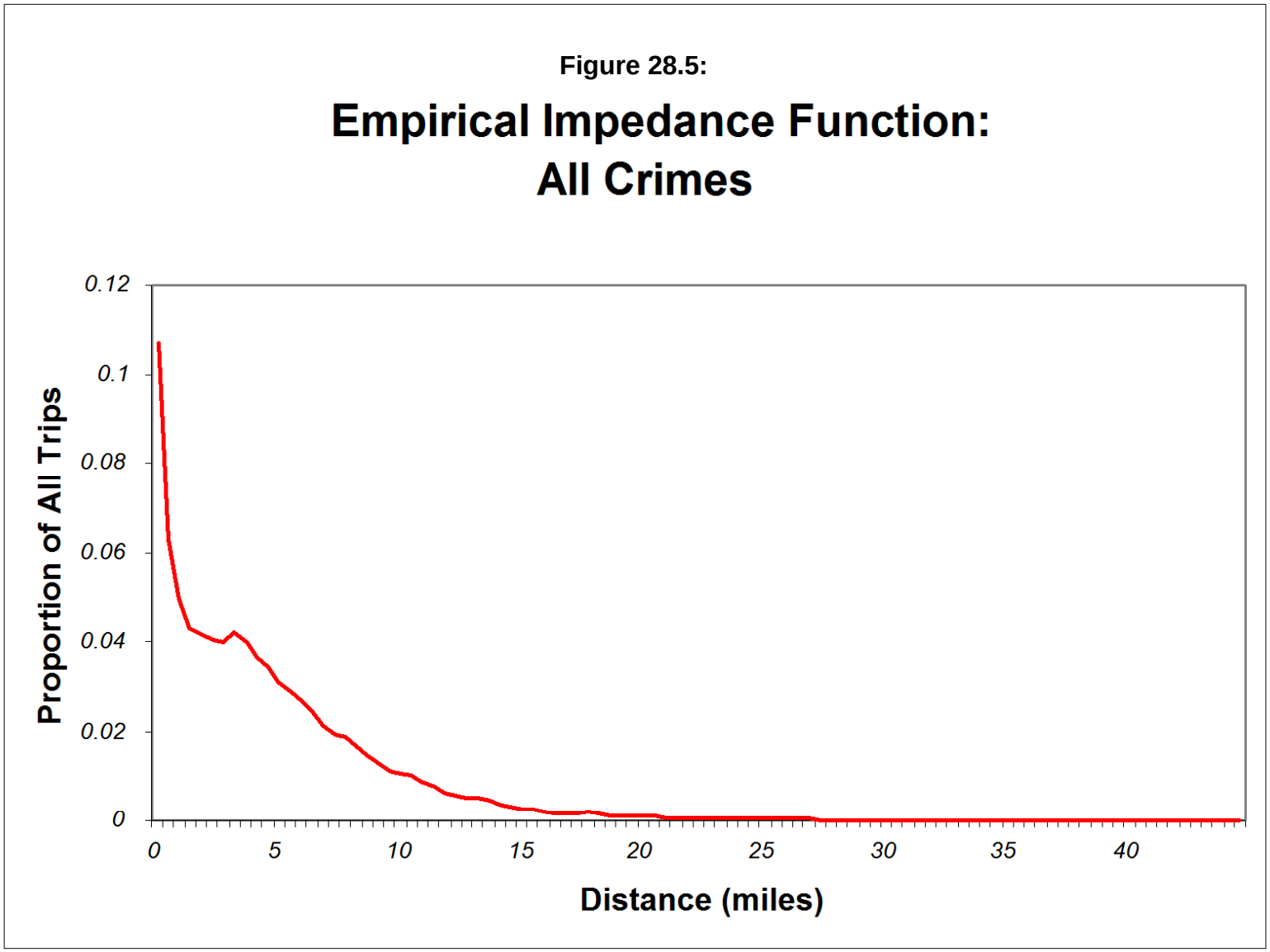
8. **Calibrate!** Click on 'Calibrate!' to run the routine. The output is saved to the specified file upon clicking on 'Close'.

9. **Graphing the travel impedance function**. Click on 'View graph' to see the travel impedance function. The screen view can be printed by clicking on 'Print'. For a better quality graph, however, the output should be imported into a graphics or spreadsheet program.

Example of Empirical Impedance from Baltimore County

An example of an empirical impedance function from Baltimore County is seen in Figure 28.5. This was derived from the 41,974 incidents in which both the crime location and the offender's origin location were known. As seen, the function looks similar to a negative exponential function. But there is a little 'hitch' around 3 miles where the travel likelihood

Figure 28.5:


Empirical Impedance Function:
All Crimes
Proportion of All Trips
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
30
35
40
Distance (miles)

increases, rather than decrease. This could possibly be due to the City of Baltimore border which abuts much of the southern part of the County.

Whatever the reason, the empirical impedance function can be used as a proxy for travel 'cost' by offenders. As we shall see, however, it may not produce as good a fit in the gravity model as some of the mathematical functions, particularly the lognormal. The reason is that it is a behavioral description. Consequently, the pattern reflects both the existence of crime opportunities (attractions) as well as costs. While an empirical description is useful for guessing where a serial offender might live, for a trip distribution model it apparently does not cleanly estimate the real costs to an offender. Nevertheless, it is a tool that can be used.

Setup of Origin-Destination Model

The page is for the setup of the origin-destination model. All the relevant files, models and exponents are input on the page and it allows the trip distribution model to be calibrated and allocated. Figure 28.6 shows the setup screen. There are a number of parameters that have to be defined:

1. **Predicted origin file.** The predicted origin file is a file that lists the origin zones with a single point representing the zone (e.g., the centroid) and also includes the predicted number of crimes by origin zone. The file must be input as either the primary or secondary file. Specify whether the data file is the primary or secondary file.

    A. ***Origin variable.*** Specify the name of the variable for the predicted origins (e.g., PREDICTED, ADJORIGINS).

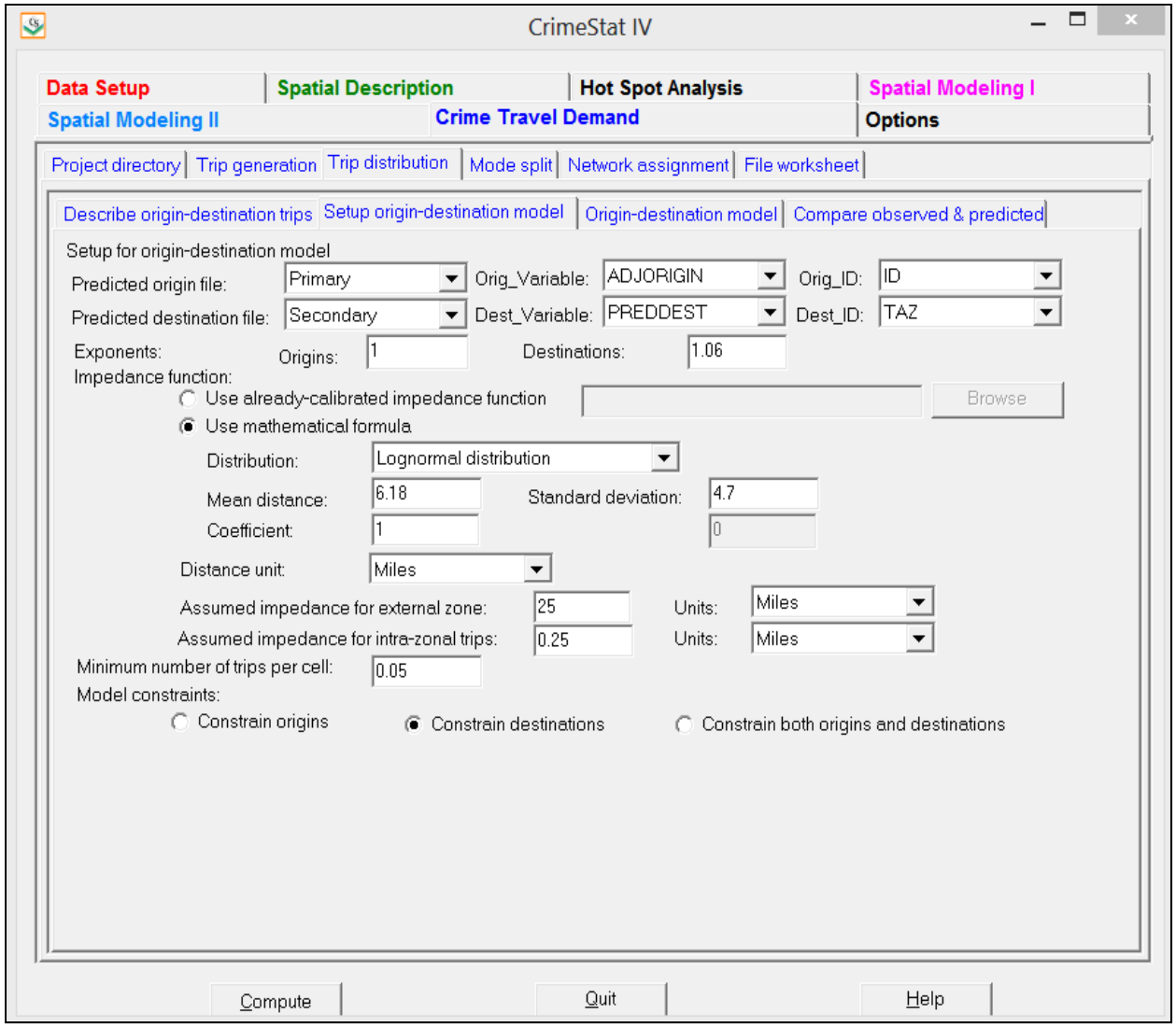
    B. ***Origin ID***. Specify the origin ID variable in the data file (e.g., CensusTract, Block, TAZ).

2. **Predicted destination file**. The predicted destination file is a list of destination zones with a single point representing the zone (e.g., the centroid) and also includes the predicted number of crimes by destination zone. It must be input as either the primary or secondary file. Specify whether the data file is the primary or secondary file.

    A. ***Destination variable***. Specify the name of the variable for the predicted destination (e.g., PREDICTED, ADJDEST).

Figure 28.6:

Trip Distribution Model Setup



B. **Destination ID**. Specify the destination ID variable in the data file (e.g., CensusTract, Block, TAZ).

Note: with a 32 bit operating system (e.g., Windows XP, 32 bit Windows 7), there is maximum allowable of 4 Gb. If M is the number of rows and N is the number of columns, then the total number of grid cells (M x N) cannot be greater than $\sqrt{\frac{(RAM-64)}{56}}$ where RAM is the available RAM. With a 64 bit operating system, on the other hand, 32 Gb are addressable.

3. **Exponents**. The exponents are power terms for the predicted origins and destinations. They indicate the relative strength of those variables. For example, compared to an exponent of 1.0 (the default), an exponent greater than 1.0 will strengthen that variable (origins or destinations) while an exponent less than 1.0 will weaken that variable. They can be considered 'fine tuning' adjustments.

   A. ***Origins.*** Specify the exponent for the predicted origins. The default is 1.0.

   B. ***Destinations.*** Specify the exponent for the predicted origins. The default is 1.0.

4. **Impedance function**. The trip distribution routine can use two different travel distance functions:

   A. ***Use an already-calibrated distance function.*** If a travel distance function has already been calibrated (see 'Calibrate impedance function' above), the file can be directly input into the routine. The user selects the name of the already-calibrated travel distance function. *CrimeStat* reads dbase 'dbf', ASCII text 'txt', and ASCII data 'dat' files.

   B. ***Use a mathematical formula.*** A mathematical formula can be used instead of a calibrated distance function. Similar to the Journey-to-crime module (see chapter 13), there are five mathematical functions. They measure a *separation* between two zones and estimate a likelihood value. 'Separation' can be in terms of distance, travel time, speed (which is converted into travel time), or travel costs.

5. **Mathematical functions.** Briefly, the five functions are:

A. **Linear**. The simplest type of distance model is a linear function. This model postulates that the likelihood of traveling to a zone from another by an offender 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;

$$f(d_{ij}) = \alpha + \beta S_{ij} \quad (28.17)$$

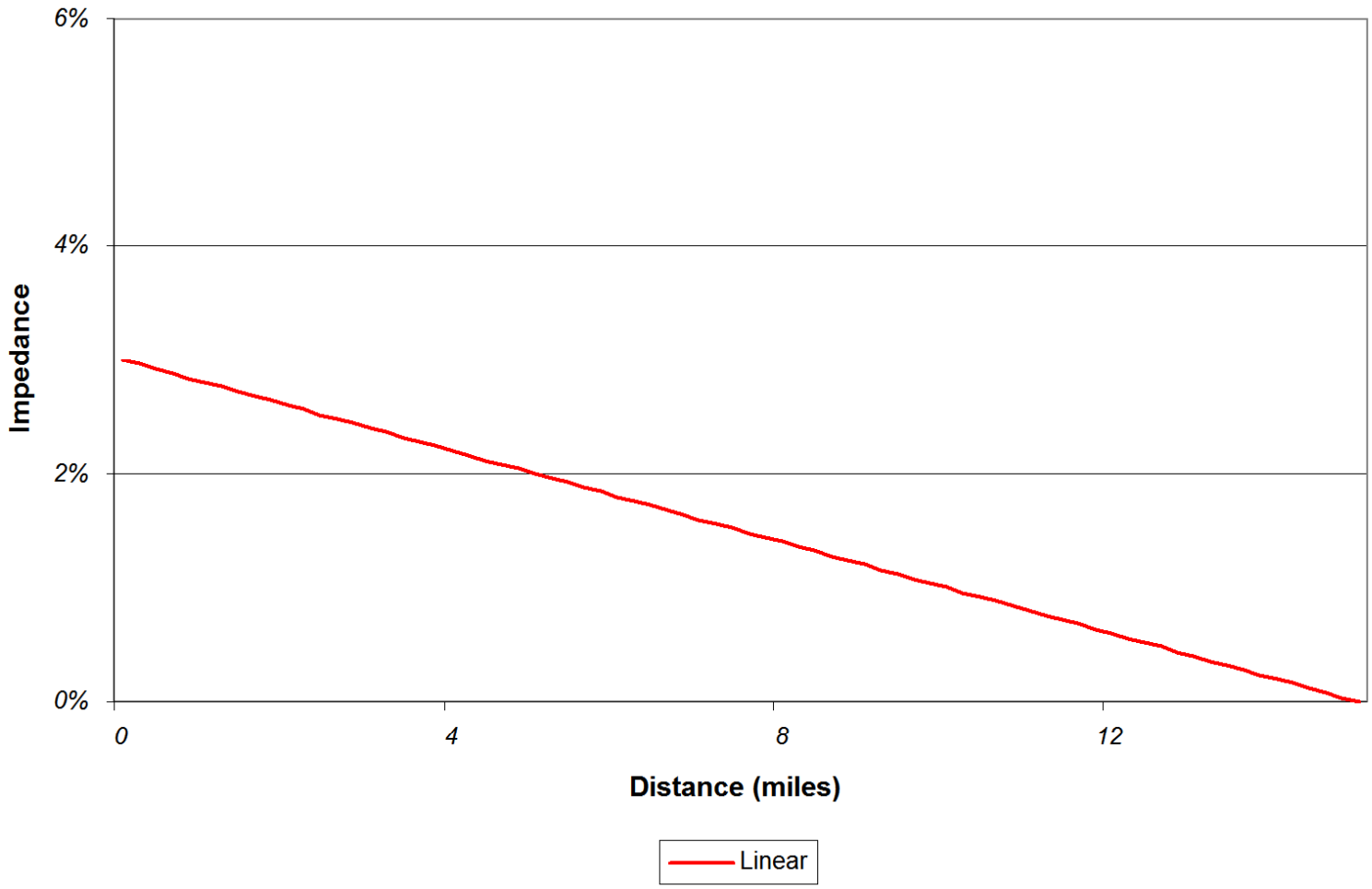
where $f(d_{ij})$ is the likelihood that the offender will travel from an origin zone, i, to a destination zone, j, S_{ij} is the *separation* in distance, time or cost between the offender's residence, i, and location j, α is a slope coefficient which defines the fall off in distance, and β is a constant. It would be expected that the coefficient β would have a negative sign since the likelihood should decline with separation. The user must provide values for *A* and β. The default for A is 10 and for β is -1. When the function reaches 0 (the X axis), the routine automatically substitutes a 0 for the function. Figure 28.7 illustrates this function.

B. ***Negative Exponential***. A slightly more complex function is the negative exponential. In this type of model, the likelihood of travel also 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:

$$f(d_{ij}) = \alpha e^{-\beta S_{ij}} \quad (28.18)$$

where $f(d_{ij})$ is the likelihood that the offender will travel from an origin zone, i, to a destination zone, j, S_{ij} is the *separation* in distance, time or cost between the offender's residence, i, and location j, e is the base of the natural logarithm, α is the coefficient and β is an exponent of e. The user inputs values for α - the coefficient, and β - the exponent. The default for α is 10 and for β is 1.

Figure 28.7:
Linear Impedance Function



a. This function is the one most used by travel demand modelers. It has been recommended for use by the Federal Highway Administration (NCHRP, 1995). Figure 28.8 illustrates a typical negative exponential impedance function.

C. ***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:

$$Z_{ij} = \frac{(S_{ij} - \bar{d})}{\sigma_d} \tag{28.19}$$

$$f(d_{ij}) = \alpha \frac{1}{\sigma_d \sqrt{2\pi}} e^{-0.5 Z_{ij}^2} \tag{28.20}$$

where $f(d_{ij})$ is the likelihood that the offender will travel from an origin zone, i, to a destination zone, j, S_{ij} is the *separation* in distance, time or cost between the offender's residence, i, and location j, $\bar{d}$ is the mean distance input by the user, σ_d is the standard deviation of distances, e is the base of the natural logarithm, and α is a coefficient. The user inputs values for $\bar{d}$, σ_d, and α. The default values are 1 for each of these parameters.

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., $\bar{d} = 1, S_d = 2$), the distribution will be skewed to the left because the left tail of the normal distribution is not evaluated. Figure 28.9 illustrates a possible normal impedance function.

Figure 28.8:
Negative Exponential Impedance Function

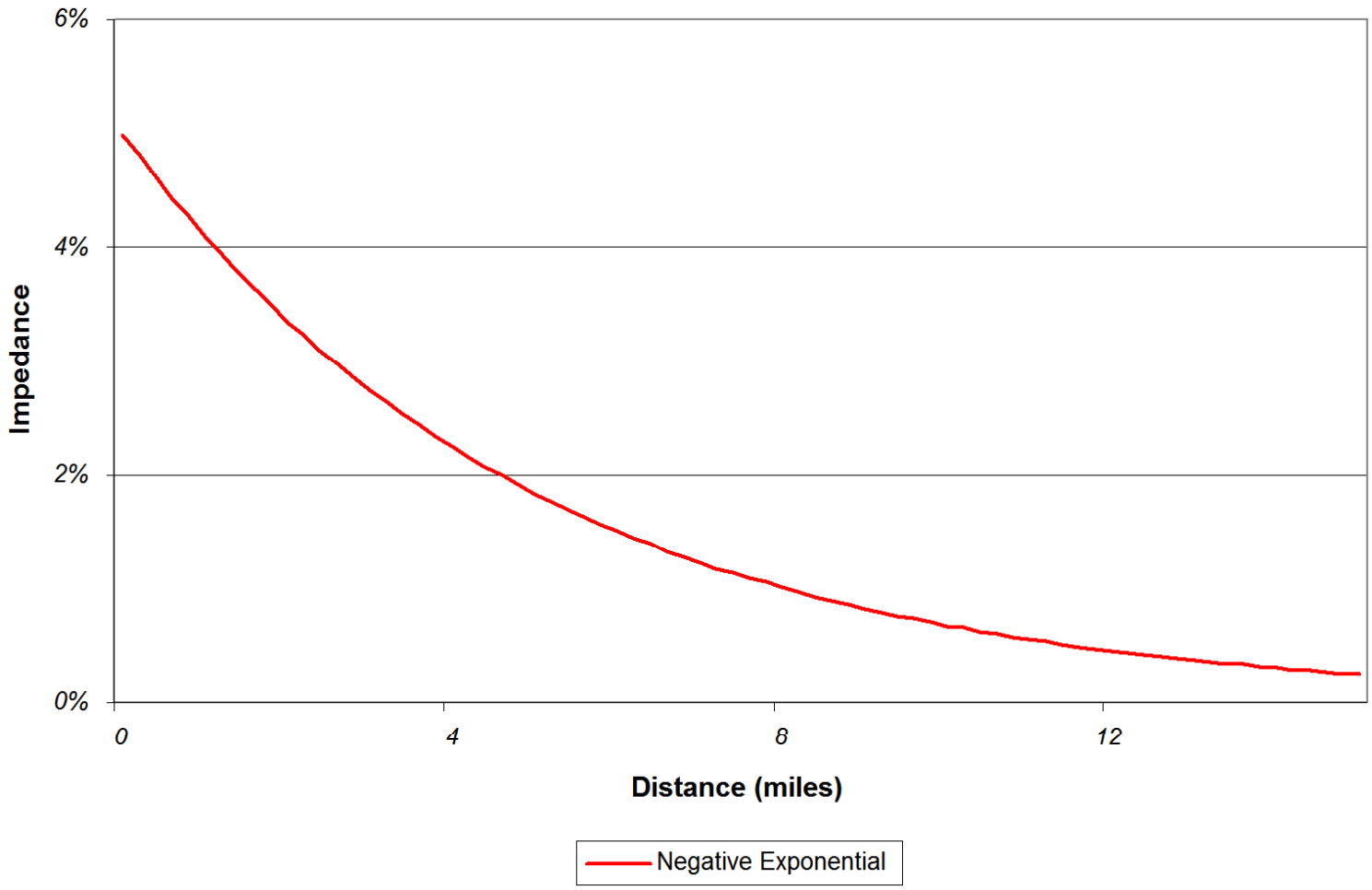
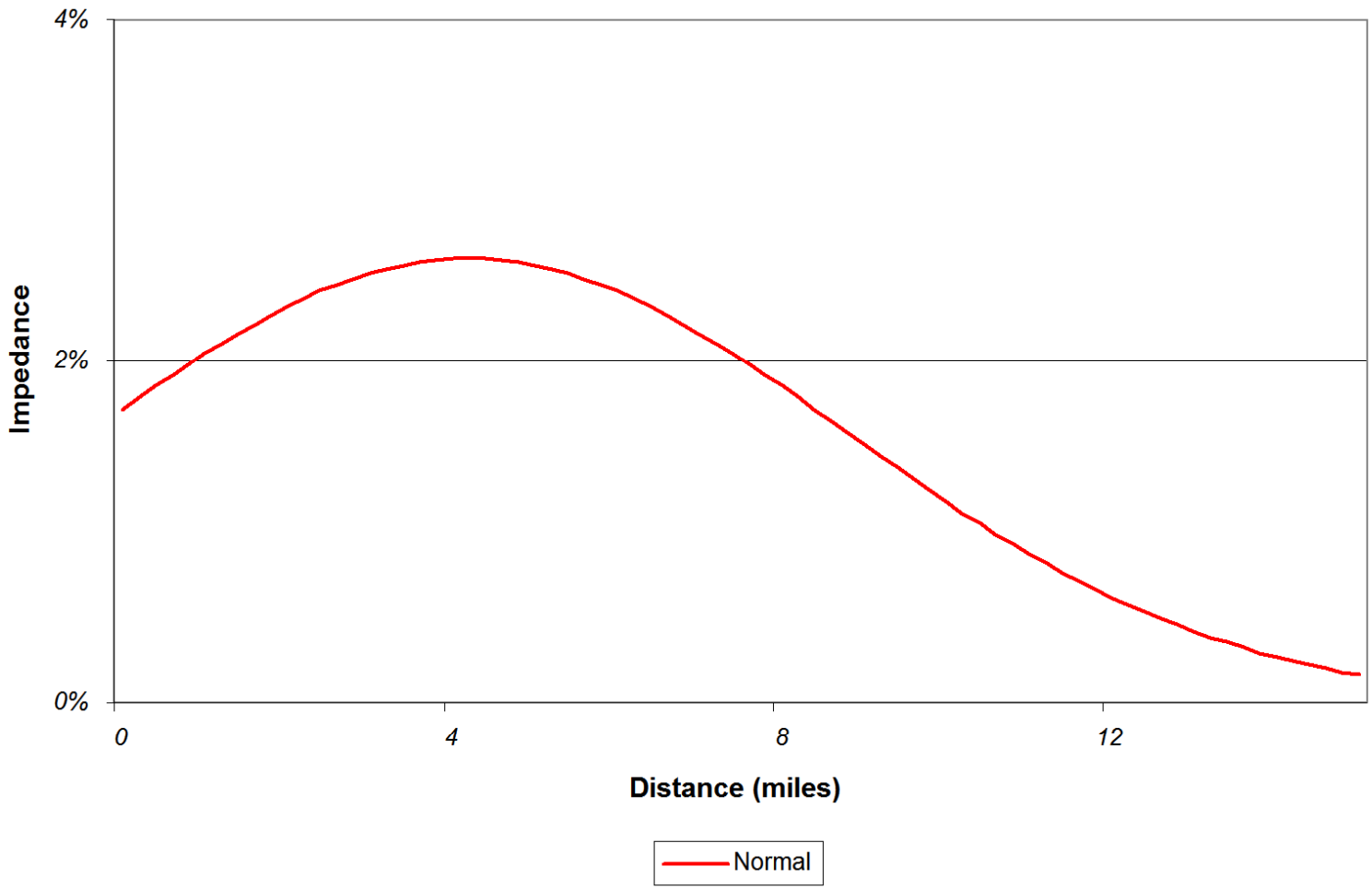


Figure 28.9:
Normal Impedance Function



D. ***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 origin with a more gradual decline from a location of peak likelihood. The mathematical form of the function is:

$$\left(d_{ij}\right) = \alpha \frac{1}{S_{ij}^2 \sigma_d \sqrt{2\pi}} e^{-\frac{-(\ln\left(S_{ij}^2\right)-\bar{d})^2}{2\sigma_d^2}} \qquad (28.21)$$

where $f(d_{ij})$ is the likelihood that the offender will travel from an origin zone, i, to a destination zone, j, S_{ij} is the *separation* in distance, time or cost between the offender's residence, i, and location j, $\bar{d}$ is the mean distance input by the user, σ_d is the standard deviation of distances, e is the base of the natural logarithm, and α is a coefficient. The user inputs values for $\bar{d}$, σ_d, and α. The default values are 1 for each of these parameters. Figure 28.10 illustrates a log-normal impedance function that had wide utility in several studies that are discussed below.

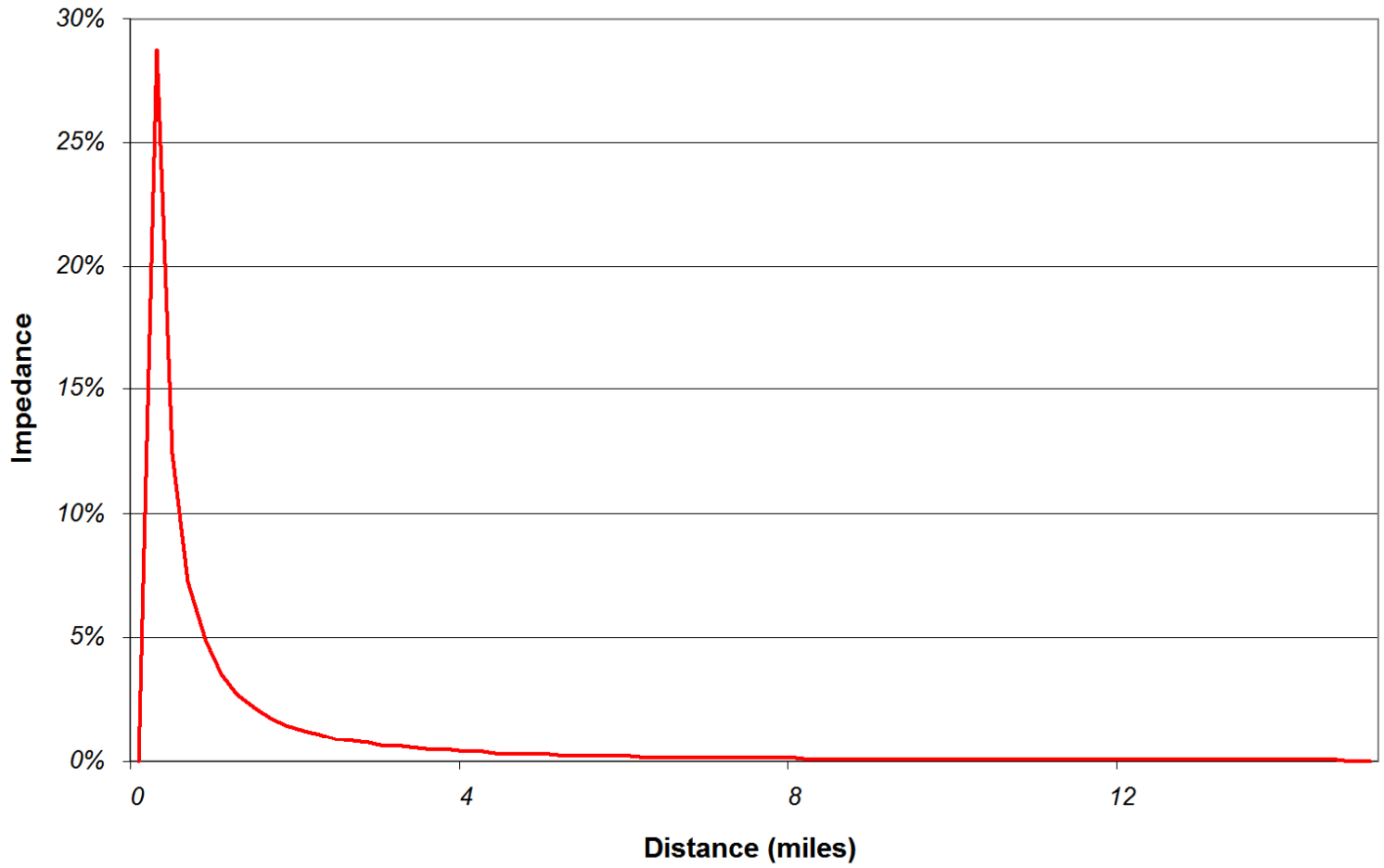
E. ***Truncated Negative Exponential***. Finally, the truncated negative exponential is a joined function made up of two distinct mathematical functions - the linear and the negative exponential. For the near distance, a positive linear function is defined, starting at zero likelihood for distance 0 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: $f\left(d_{ij}\right) = 0 + \beta S_{ij} = \beta d_{ij}$ for $S_{ij} \geq 0$, $S_{ij} \leq S_p$ (28.22)

Negative Exponential: $f\left(d_{ij}\right) = \alpha e^{-\xi S_{ij}}$ for $S_{ij} \geq S_p$ (28.23)

where $f(d_{ij})$ is the likelihood that the offender will travel from an origin zone, i, to a destination zone, j, S_{ij} is the *separation* in distance, time or cost between the offender's residence, i, and location j, β is the slope of the linear function (default=+1) and α is a coefficient and ξ is an exponent

Figure 28.10:
Lognormal Impedance Function



for the negative exponential function. Since the negative exponential only starts at a particular distance, S_p, α, is assumed to be the intercept if the Y-axis were transposed to that distance. Figure 28.11 illustrates a truncated negative exponential impedance function.

F. ***Model parameters***. For each mathematical model, two or three different parameters must be defined:

1. For the negative exponential, the coefficient and exponent
2. For the normal distribution, the mean distance, standard deviation and coefficient
3. For lognormal distribution, the mean distance, standard deviation and coefficient
4. For the linear distribution, an intercept and slope
5. For the truncated negative exponential, a peak distance, peak likelihood, intercept, and exponent.

The parameters will be obtained either from a previous analysis or from an iterative process of experimentation. See the example below under "Compare observed and predicted trips".

G. **'Fine Tuning' Exponents**. In addition, for each function, exponents for the attraction and production terms can adjusted. This allows a 'fine tuning' of the impedance function to better fit the empirical distribution.

5. **Distance Units.** The routine can calculate impedance in four ways, by:

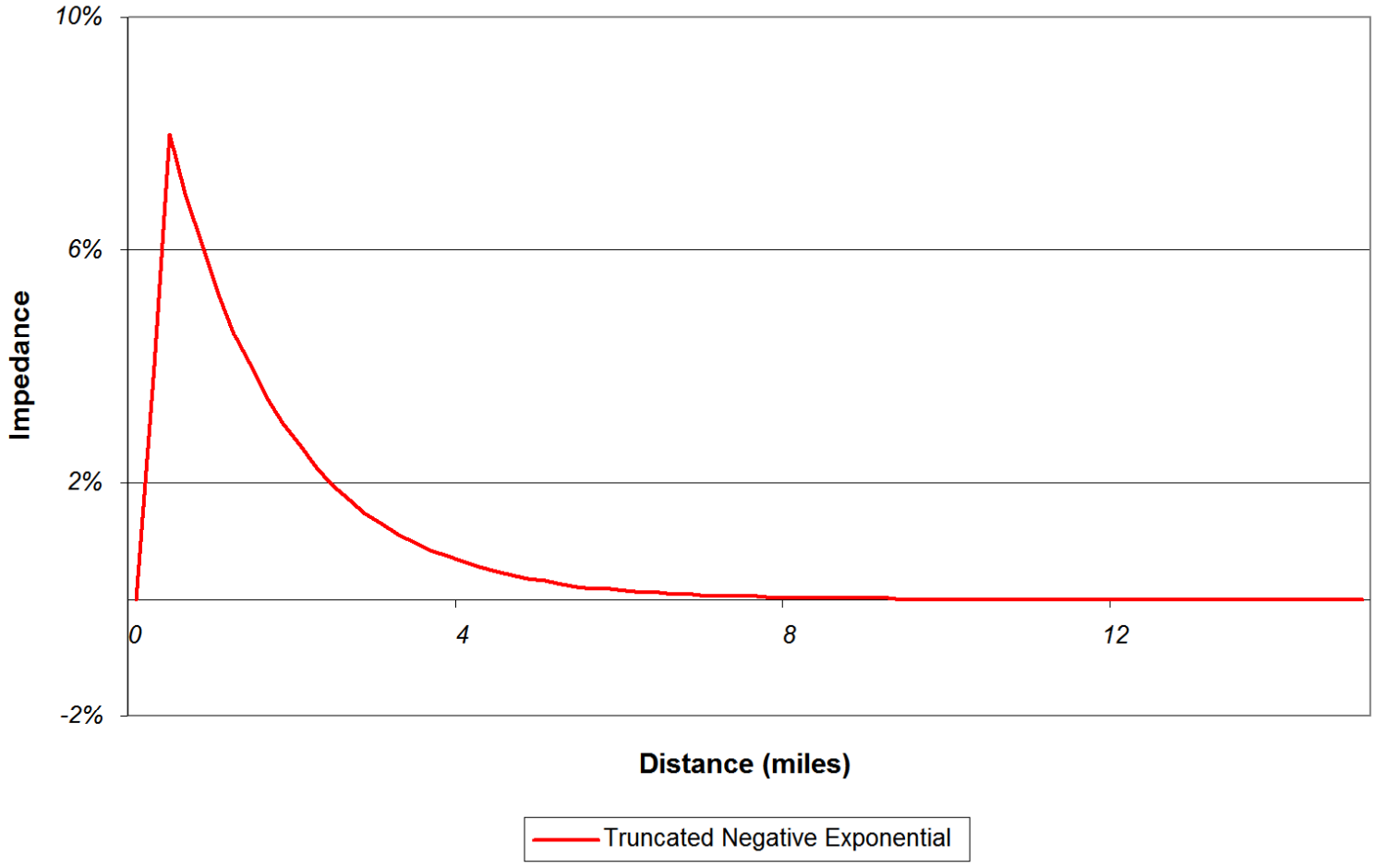
A. Distance (miles, nautical miles, feet, kilometers, and meters)
B. Travel time (minutes, hours)
C. Speed (miles per hour, kilometers per hour). Speed is then converted into travel time, in minutes.
D. General travel costs (unspecified units).

These must be setup under 'Network Distance' on the Measurement Parameters page. In the Network Parameters dialogue, specify the measurement units. The default is distance in miles.

6. **Assumed Impedance for External Zones.** For trips originating outside the study area (external trips), specify the amount and the units that will be assumed for these trips. The default is 25 miles.

Figure 28.11:

Truncated Negative Exponential Impedance Function



7. **Assumed Impedance for Intra-zonal Trips**. For trips originating and ending in the same zone (intra-zonal trips), specify the amount and the units that will be assumed for these trips. The default is 0.25 miles.

8. **Model Constraints.** In calibrating a model, the routine must constrain either the origins or the destinations (single constraint) or constrain both the origins and the destinations (double constraint). In the latter case, it is an iterative solution. The default is to constrain destinations as it is assumed that the destination totals (the number of crimes occurring in each zone) are probably more accurate than the number of crimes originating in each zone. Specify the type of constraint for the model.

   A. ***Constrain origins.*** If 'constrain origins' is selected, the total number of trips from each origin zone will be held constant.

   B. ***Constrain destinations.*** If 'constrain destinations' is selected, the total number of trips from each destination zone will be held constant.

   C. ***Constrain both origins and destinations.*** If 'constrain both origins and destinations' is selected, the routine works out a balance between the number of origins and the number of destinations.

Fitting the Impedance Function

The impedance function is fit in an iterative manner. First, either an empirical impedance or a mathematical impedance is chosen. Second, the particular mathematical function is selected. For example, with the lognormal function, which has been found to produce the best fit for three different data sets, there are three parameters: 1) the mean distance; 2) the standard deviation of distance; and 3) the coefficient.

Third, initial values of the parameters are chosen; one suggestion is to use the defaults available in the *CrimeStat* routines. The "Compare observed and predicted trips" routine can be used to evaluate the fit of the model. Fourth, the parameters are adjusted in small increments, one at a time, on both side of the initial guess in order to improve the fit. For example, with the lognormal function, the mean distance is fit first because it has the greatest impact on the overall fit. Then, after a "best" mean distance has been found, the standard deviation of distance is adjusted until it produces a "best" fit. Then, the coefficient is adjusted until it produces a "best" fit. Fifth, and finally, the 'fine tuning' exponents of the production and attraction functions are adjusted. Typically, these change the final fit only slightly. Hence, they represent a final adjustment.

This process is illustrated below in the discussion on the comparison of the observed and predicted trips. Essentially, the empirical (observed) distribution is being used as a calibration sample in order to find that impedance model and parameters that best approximate it.

Running the Origin-Destination Model

The trip distribution (origin-destination) model is implemented in two steps. First, the coefficients are calculated according to the exponents and impedance functions specified on the setup page. Second, the coefficients and exponents are applied to the predicted origins and destinations resulting in a predicted trip distribution. Because these two steps are sequential, they cannot be run simultaneously.

Calibrate Origin-Destination Model.

In this routine, the row or column parameters (or both if double constraint is used) are estimated using a calibration file. The steps are as follows:

1. **Check** the 'Calibrate origin-destination model' box to run the calibration model.

2. **Save Modeled Coefficients (parameters).** The modeled coefficients are saved as a 'dbf' file. Specify a file name.

Apply Predicted Origin-Destination Model

In this routine, the coefficients that were calibrated in the above routine can be applied to a data set. The data set can be the same as the calibration file or a different one. The reason for separating the calibration from application steps is that the coefficients can be used for many different data sets. The steps are as follows:

1. **Check** the 'Apply predicted origin-destination model' box to run the trip distribution prediction.

2. **Modeled Coefficients File**. Load the modeled coefficients file saved in the 'Calibrate origin-destination model' stage.

3. **Assumed Coordinates for External Zone**. In order to model trips from the 'external zone' (trips from outside the study area), specify coordinates for this zone. These coordinates will be used in drawing lines from the predicted origins to the predicted destinations. There are four choices:

A. Mean center (the mean X and mean Y of all origin file points are taken). This is the default.
B. Lower-left corner (the minimum X and minimum Y values of all origin file points are taken).
C. Upper-right corner (the maximum X and maximum Y values of all origin file points are taken).
D. User coordinates (user-defined coordinates). Indicate the X and Y coordinates that are to be used.

Because an arbitrary location is taken to represent the 'external zone', any lines that are shown from that zone will not necessarily represent any real travel behavior. However, if a very high proportion of all crime trips fall within the modeled origin zones (i.e., 95% or more), then it is very unlikely that any of the top trip links will come from the 'external zone'.

4. **Table Output**. The table output includes summary file information and:

   A. The origin zone (ORIGIN)
   B. The destination zone (DEST)
   C. The number of predicted trips (PREDTRIPS)

5. **Save Predicted Origin-destination Trips**. Define the output file. The output is saved as a 'dbf' file specified by the user.

6. **File Output**. The file output includes:

   A. The origin zone (ORIGIN)
   B. The destination zone (DEST)
   C. The X coordinate for the origin zone (ORIGINX)
   D. The Y coordinate for the origin zone (ORIGINY)
   E. The X coordinate for the destination zone (DESTX)
   F. The Y coordinate for the destination zone (DESTY)
   G. The number of predicted trips (PREDTRIPS)

   Note: each record is a unique origin-destination combination and there are M x N records where M is the number of origin zones (including the external zone) and N is the number of destination zones.

7. **Save Links.** The top predicted origin-destination trip links can be saved as separate **line** objects for use in a GIS. Specify the output file format (*ArcGIS* '.shp', *MapInfo* '.mif' or *Atlas*GIS* '.bna') and the file name.

Save Top Links

Because the output file is very large (number of origin zones x number of destination zones), the user can select a sub-set of zone combinations with the most predicted trips. Indicating the top K links will narrow the number down to the most important ones. The default is the top 100 origin-destination combinations. Each output object is a line from the origin zone to the destination zone with an ODT prefix. The prefix is placed before the output file name.

The graphical output includes:

A. An ID number from 1 to K, where K is the number of links output (ID)
B. The feature prefix (ODT)
C. The origin zone (ORIGIN)
D. The destination zone (DEST)
E. The X coordinate for the origin zone (ORIGINX)
F. The Y coordinate for the origin zone (ORIGINY)
G. The X coordinate for the destination zone (DESTX)
H. The Y coordinate for the destination zone (DESTY)
I. The number of predicted trips for that combination (PREDTRIPS)
J. The distance between the origin zone and the destination zone.

8. **Save Points**

Intra-zonal trips (trips in which the origin and destination are the same zone) can be output as separate **point** objects as an *ArcGIS* '.shp', *MapInfo* '.mif' or *Atlas***GIS* '.bna' file. Again, the top K points are output (default=100). Each output object is a point representing an intra-zonal trip with an ODTPOINTS prefix. The prefix is placed before the output file name.

The graphical output for each includes:

A. An ID number from 1 to K, where K is the number of links output (ID)
B. The feature prefix (POINTSODT)
C. The origin zone (ORIGIN)
D. The destination zone (DEST)
E. The X coordinate for the origin zone (ORIGINX)
F. The Y coordinate for the origin zone (ORIGINY)
G. The X coordinate for the destination zone (DESTX)
H. The Y coordinate for the destination zone (DESTY)
I. The number of predicted trips for that combination (PREDTRIPS)

Example of the Predicted Trip Distribution from Baltimore County

The predicted origins and predicted destinations from Baltimore County were input into a trip distribution model and a predicted trip distribution was output. The impedance function was a lognormal distribution, which produced a good fit to the observed (empirical) distribution (see discussion below).

Figure 28.12 outputs the top 1000 links from the model. The top 1000 links account for 14,271.9 trips, or 34.0% of the total number of trips. Compared to the observed distribution, the top 1000 links account for a smaller proportion of the total trips (14,272 v. 19,615). This suggests that the actual distribution is slightly more concentrated than the model suggests. Like the observed distribution, however, a sizeable number of the top links are intra-zonal trips (5,428 or 12.9%). The intra-zonal trips have been displayed as circles in the figure.

Comparing the predicted trip distribution to the observed trip distribution, some similarities and differences are seen. Figure 28.13 compares the top 1000 zone-to-zone links for the predicted and observed distributions. The model has captured many of the major links. For the five shopping malls that received a sizeable number of actual crime trips, the model has captured the majority of trips for three of them and some trips for a fourth. For the mall in the southeast corner of the county, on the other hand, the model has not allocated a large number of trips. Similarly, for a zone near the western edge of the county, the model has allocated more trips than actually occurred.

There are, of course, only 325 intra-zonal trip links (one for each destination zone). Looking at a comparison of the intra-zonal trips (Figure 28.14), some similarities and differences are seen. Generally, the model captured the location of many intra-zonal trips, but it did not capture the quantity very accurately. Zones that had many intra-zonal trips are shown as having only some by the model and, conversely, the model predicts many intra-zonal trips for two zones which had only some.

In other words, the fit between the actual distribution and the model is not perfect. Considering that only 1000 of the 172,900 trip links (532 origin zones x 325 destination zones) are shown, the model has still done a reasonable job of capturing the major links.

It is not surprising that the model is not perfect. The model is a simple analogue using only three variables (productions, attractions, impedance) whereas the actual distribution represents a very complex set of individual decisions made by offenders. What is perhaps remarkable is that the model has done a decent job of capturing some of these relationships at all.

Figure 28.12:

Predicted Baltimore County Crime Trips: 1993-1997
Top 1000 Links
All Crime Types

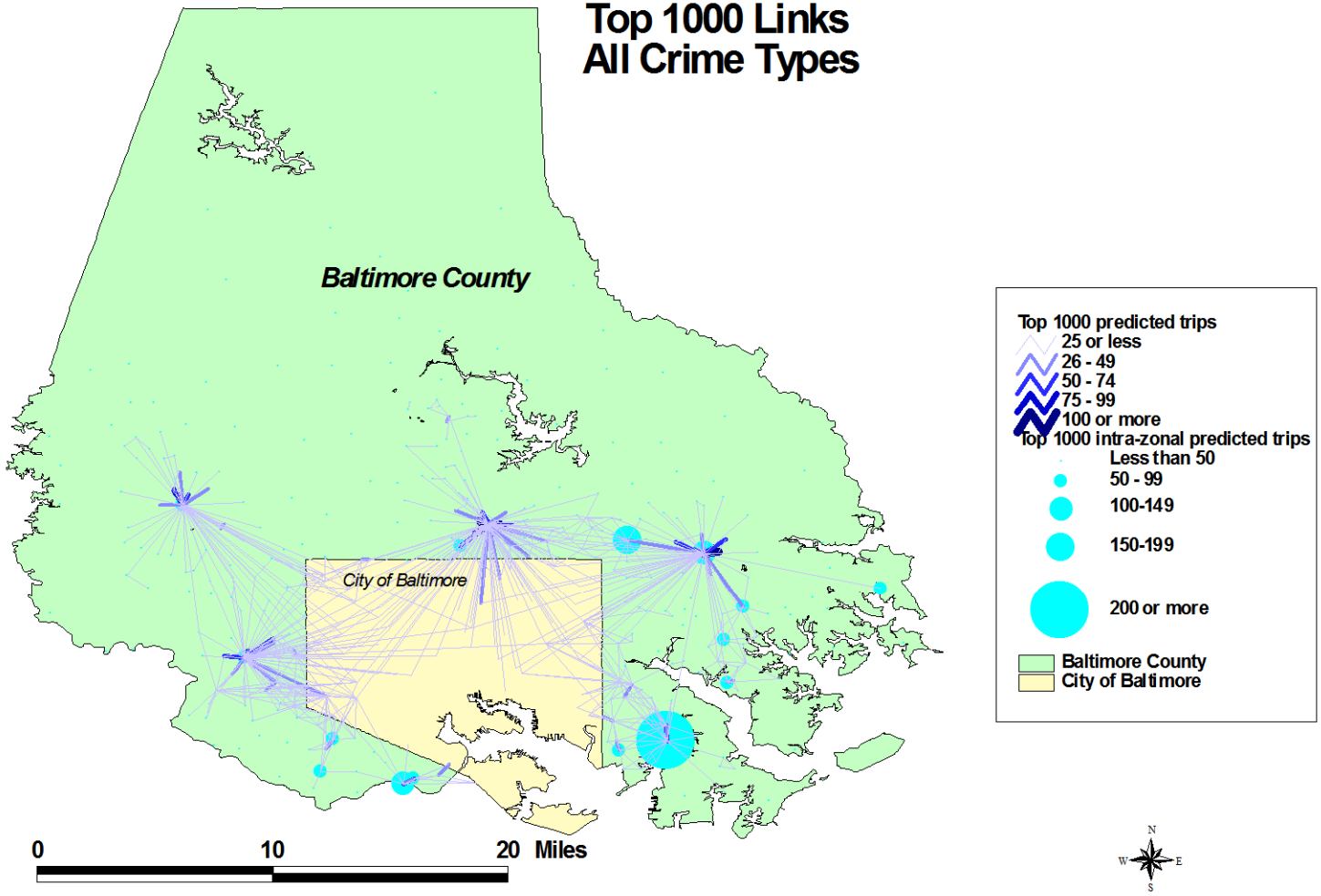


Figure 28.13:

Comparison of Predicted and Observed Crime Trips
1000 Top Zone-to-Zone Trips
All Crime Types

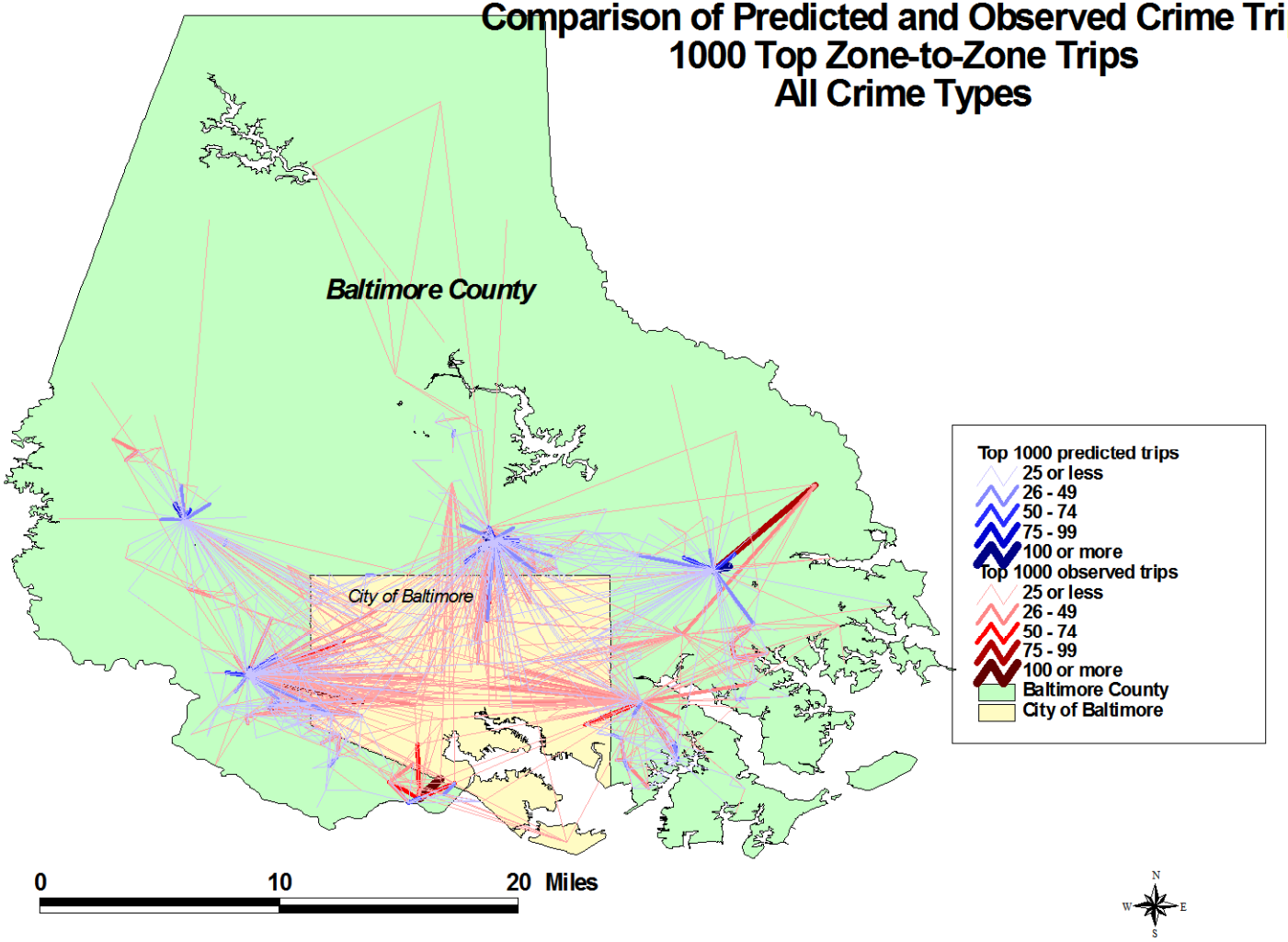
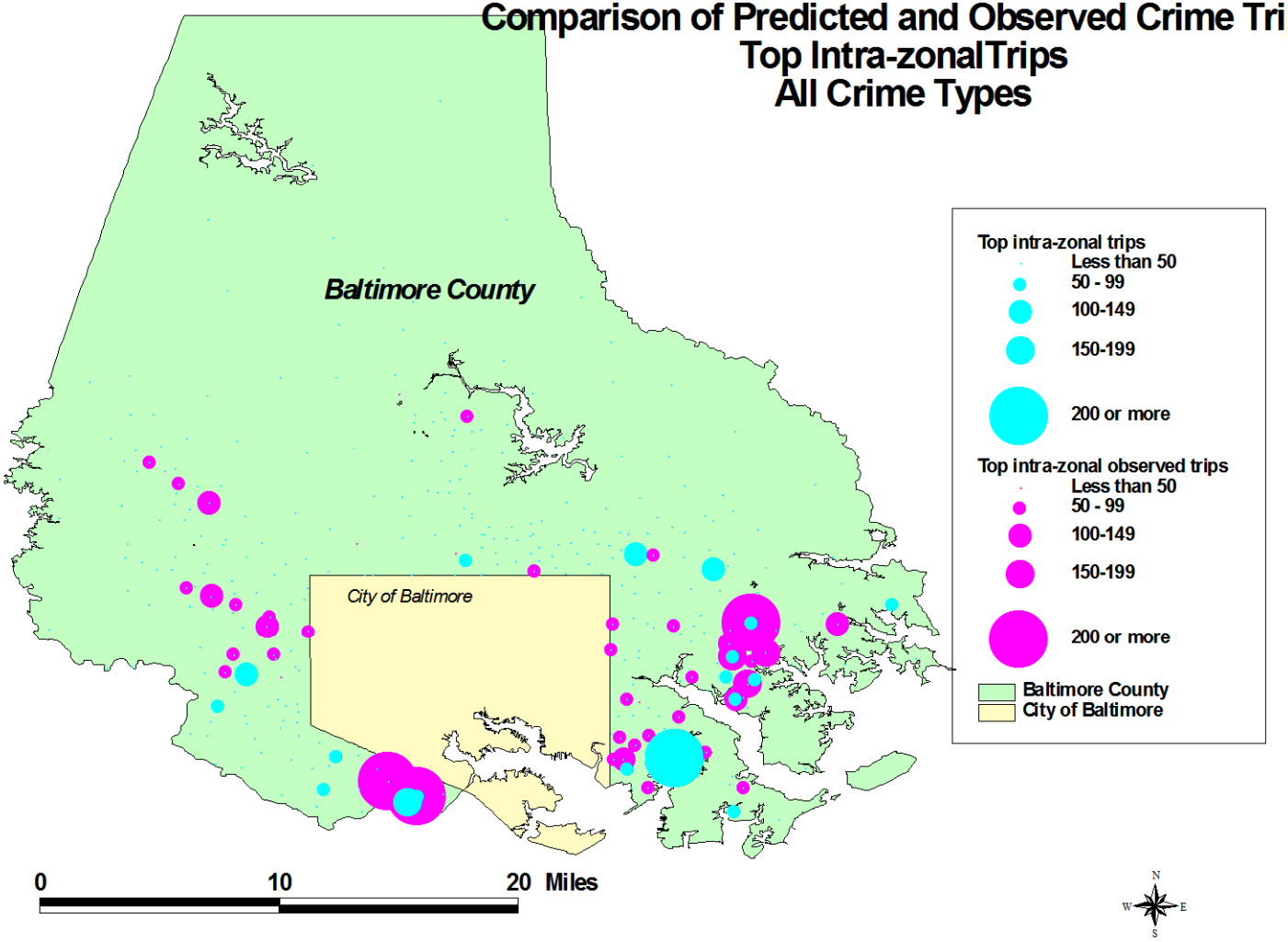


Figure 28.14:
Comparison of Predicted and Observed Crime Trips
Top Intra-zonalTrips
All Crime Types



This brings up an important point, namely that a model is not reality. It is only a simplified set of relationships that approximates reality (in this case, the observed distribution). It is important in developing any model to evaluate it relative to an observed set of facts, and this applies no less to the trip distribution model. One has to understand, however, that a good model will not capture all the relationships. Hopefully, it captures enough of them to make the model useful for prediction and evaluating policy options.

Comparing Observed & Predicted Trips

It is important to conduct a number of tests on the predicted model to ensure that it is capturing the most important elements of the observed distribution. These are conducted by comparing the predicted distribution with the observed (empirical) distribution. Figure 28.15 shows the setup page for comparing the observed with the prediction distribution

There are a number of tests that can be used to evaluate a model by comparing the predicted distribution with the observed one. *CrimeStat* includes three of these and the steps are as follows:

1. Estimate the parameters of the model and apply them to the calibration data set

2. Examine the intra-zonal trips to be sure that the predicted number corresponds to the observed number

3. Compare the trip lengths of the observed and predicted distributions using two tests:

   A. The Coincidence Ratio

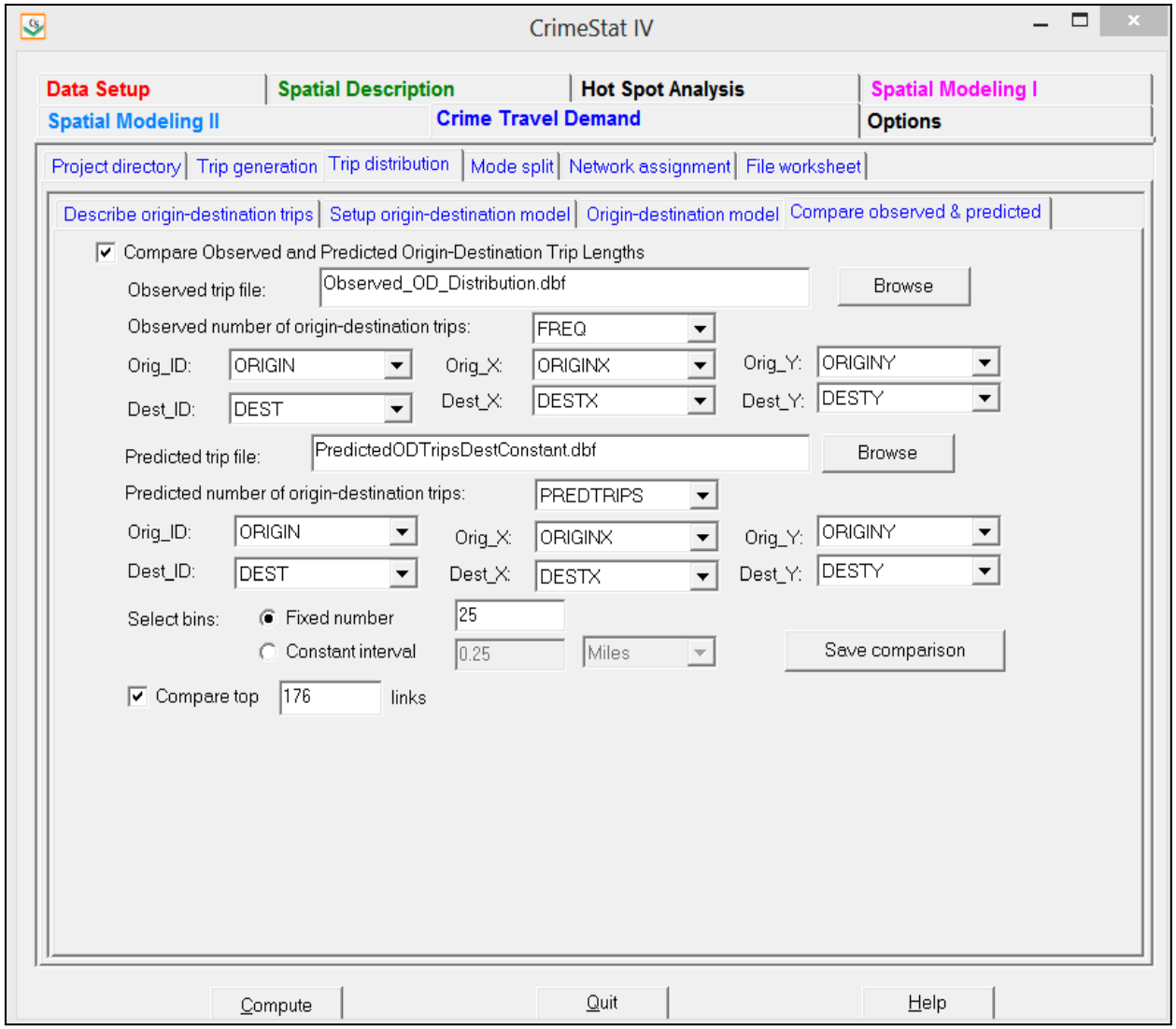
   B. The Komolgorov-Smirnov Two-sample Test

4. Compare the number of trips for the top links using a pseudo-Chi square test. That is, the number of trips for the most frequent links in the observed distribution is compared to the number predicted by the model for the same links.

Unfortunately, not one of these tests is sufficient to validate a model. Further, minimizing the discrepancy for only one of them may distort the others. It is very unlikely that there will be a model that minimizes the errors for all three tests. Consequently, the user will have to choose a model that balances these factors in a desirable way (an *optimum* model).

Figure 28.15:

Comparing Observed and Predicted Trip Lengths



Estimating Impedance Parameters and Exponents of the Gravity Model

While this is not strictly an evaluation test, this step is essential in estimating the particular impedance parameters that are used in the first place. Typically, an analyst will approximate an impedance function. Using a comparison between the observed and predicted models, the parameters can be adjusted to produce a better fit. The steps are as follows:

1. The model is estimated with a calibration data set. There is a file of predicted origins and another file of predicted destinations; typically, these are defined as the primary and secondary files respectively, though the order could be reversed or the same file used for both origins and destinations (if the number of origins zones was identical to the number of destination zones).

2. On the trip distribution setup page, select the type of impedance function that is to be used, already-calibrated (empirical) or mathematical. For the journey-to-crime routine, generally the empirical function led to better results than the mathematical. However, with a trip distribution function, a mathematical function may be as good, if not better. This was tested with three data sets for Baltimore County, Las Vegas, and Chicago and, in all cases, a mathematical function (the lognormal) gave a much better fit than an empirically-derived function (see Chapters 31 and 32).

3. *If* a mathematical function is to be used, select the type of distribution. The default value is a lognormal, but the user can choose a negative exponential, a normal, a linear, or a truncated negative exponential function.

4. For the particular mathematical function, select initial guesses for the parameters. For each mathematical model, two or three different parameters must be defined:

    A. For the negative exponential, the coefficient and exponent
    B. For the normal distribution, the mean distance, standard deviation and coefficient
    C. For lognormal distribution, the mean distance, standard deviation and coefficient
    D. For the linear distribution, an intercept and slope
    E. For the truncated negative exponential, a peak distance, peak likelihood, intercept, and exponent.

5. In addition, there are exponents of the production and attraction side that can be made to 'fine tune' the model. In general, these exponents will only affect the

results slightly, compared to the basic choices of the type of model and the selection of values for the main parameters.

6. Calibrate and apply the model to the calibration data set. Examine the three criteria discussed below to minimize the error between the actual distribution and that predicted by the model.

7. Modify the parameter values slightly.

8. Repeat steps 4 through 7 until a good fit is found between the actual and predicted distribution and in which the errors are minimized and optimized. The process by which this is done is discussed below.

Comparing Intra-zonal Trips

The first evaluation test is to compare the percentage of trips that occur within the same zone - intra-zonal trips. The Travel Model Improvement Program manual indicates that intra-zonal trips should represent typically no more than 5% of all trips for home-to-work trips; that is, commuting trips (FHWA, 1997, chapter 4). However, given that most crime trips are quite short, the proportion of trips that are intra-zonal is liable to be much higher. In Baltimore County, for example, 19.7% of all crime trips were intra-zonal. Ideally, the predicted model should also have 19.7% of all crime trips being intra-zonal.

The "Compare observed and predicted trip lengths" routine is discussed below. The routine outputs the number of trips that are intra-zonal in both the observed and predicted distributions. A good model should produce approximately the same number of intra-zonal trips in the predicted distribution as what actually occurred.

Illustration

For example, in the Baltimore County model displayed in Figure 28.12 above, there were 8,272 intra-zonal trips in the actual distribution (out of 41,979). On the other hand, there were only 5,428 predicted intra-zonal trips in the model. In other words, the predicted model assigned fewer intra-zonal trips than actually occurred.

It may be necessary to modify the model to produce a closer fit for the intra-zonal trips. A simple way to do this to increase or decrease the relative impedance parameter in the model. So, to use the example, if the predicted model is assigning too few intra-zonal trips, then the cost function can be strengthened (i.e., making travel more expensive). In this case, in the original model the lognormal function was used with a mean distance of 6.18 miles. If the mean

distance of the impedance function is reduced to 3.5, then the number of predicted intra-zonal trips increases to 8,275, almost the same number as occurred in the observed distribution.

In other words, by decreasing the mean distance for the lognormal function, the impedance function was strengthened (i.e., made more expensive) and a better fit was created between the observed and predicted distributions.

In and of itself, a mismatch for intra-zonal trips between the predicted model and what actually occurred does not necessarily require a modification of the gravity function. Other criteria must be considered, namely how well the predicted model fits the trip length distribution and how well the predicted models captures the most frequent inter-zonal (zone-to-zone) trip links. Later in the discussion, the issue of optimizing a model by balancing these different criteria will be described.

Comparing Trip Length Distributions

The second evaluation test in comparing the observed with the predicted distribution is a calculation of the trip length distribution (see steps below). Because the trip distribution matrix will typically be very large, most cell values will be zero. Rarely will there be enough data to cover all the cells and, even if there was, the skewness in a crime distribution will leave most cells with no data. For example, for the Baltimore County model, with 532 origin zones and 325 destination zones, there will be 172,900 cells (325 x 532). The calibration data set had only 41,974 cases. Thus, the number of cells is more than four times the sample size and it is not possible to fill all cells with a number.

Consequently, because of the large number of cells with zero counts, one cannot use the Chi square test to compare the observed and predicted distributions. The Chi square test assumes that, first, the distribution is relatively normal (which it is not since the data are highly skewed) and, second, that there are at least 5 cases per cell. The latter condition is impossible given the large number of cells.

Therefore, what is usually done is to compare the *trip length* distribution of the observed and predicted models. 'Trip length' is the length in distance, travel time, or cost of each trip. It is measured by the actual length (or separation) between two zones times the number of cases for that zone pair. For example, in Figure 28.1, there were 15 trips from zone 1 to zone 2 and 7 trips in the opposite direction (from zone 2 to zone 1). Let's assume that the distance between zone 1 and zone 2 is 1.5 miles. Thus, there are 22 trips that fall into a trip length of 1.5 miles (15 in the direction of zone 1 to zone 2 and 7 in the direction of zone 2 to zone 1).

If travel time is used, the calculation uses time rather than distance. For example, if a vehicle was traveling 30 miles per hour, then it would take 3 minutes to cover 1.5 miles (1.5 miles ÷ 30 miles per hour = 0.05 hours x 60 minutes per hour = 3 minutes). Thus, there are 22 trips that fall into a trip 'length' of 3 minutes. A similar logic would apply to travel cost categories.

This process is repeated for all cells and the distribution of trips is allocated to the distribution of trip lengths (in distance, travel time, or travel cost). In general, one uses many intervals (or bins) for trip length (25 or more). In *CrimeStat*, the default number of trip lengths is 25, but it is not unknown to use up to 100. The problem in using too many is that the distributions become unreliable and differences that appear may not be real.

Graphical fit

Once the trip length distribution is calculated for both the observed and predicted distributions, it is possible to compare them. *CrimeStat* outputs a graph showing the fit of the two distributions. In general, they should be very close. An examination of differences between the distributions can indicate at what trip lengths the model is failing. This might allow the parameters to be adjusted in order to improve the fit on the next iteration. Examples will be given below of the graphing of the two distributions. But, it is important to come up with a model in which the two distributions 'look' similar.

Coincidence ratio

The *coincidence ratio* compares the two trip length distributions by examining the ratio of the total area of those distributions that coincide, that are in common (FHWA, 1997, chapter 4). It is defined as:

$$Coincidence = \sum_{k=1}^{K} \min \left(\frac{f^O}{F^O}, \frac{f^P}{F^P} \right) \quad (28.24)$$

$$Total = \sum_{k=1}^{K} \max \left(\frac{f^O}{F^O}, \frac{f^P}{F^P} \right) \quad (28.25)$$

$$Coincidence\ ratio = \frac{Coincidence}{Total} \quad (28.26)$$

The steps are as follows:

1. Essentially, the two distributions are broken into K bins (or intervals). That is, the number of trips in each bin is enumerated (see example above).

2. Each of the two distributions is converted into a proportional distribution by dividing the bin count by the total number of trips in the distribution. This step is not absolutely essential as the test can be conducted of the raw counts. However, by converting into proportions, the two distributions are standardized.

3. A cumulative count is conducted of the *minimum* proportion in each interval. That is, starting at the lowest interval, the smaller of the two proportions is taken. At the next interval, the smaller of the two proportions is added to the count. This is repeated for all K bins. This is called the *coincidence* and measure the overlapping proportions over all intervals.

4. A similar cumulative count is conducted of the *maximum* proportion in each interval. That is, starting at the lowest interval, the larger of the two proportions is taken. At the next interval, the larger of the two proportions is added to the count. This is repeated for all K bins. This is called the *total* and measures the unique proportion over all intervals.

5. Finally, the coincidence ratio is defined as the ratio of the minimum count to the total count.

The coincidence ratio is a proportion from 0 to 1. It is analogous to the R^2 statistic in regression analysis in that it measures the 'explained' (or overlapping) variance. According to the Travel Model Improvement Program manual (FHWA, 1997, chapter 4), the higher the coincidence ratio, the better. A value of 0.9 would generally be considered good.

Komolgorov-Smirnov two-sample test

The Komolgorov-Smirnov Two-Sample Test is similar to the coincidence ratio, but it examines the maximum difference across all bins (Kanji, 1993). For each distribution, a cumulative sum is created. At each interval, the difference between the two cumulative sums is calculated. The ***maximum*** difference between the two distributions is taken as the test statistic:

$$D = |O_i - P_i| \tag{28.27}$$

where D is the maximum difference found, O_i is the cumulative sum of the actual (observed) trip lengths, and P_i is the cumulative sum of the predicted trip lengths. There are tables of critical values for the Komolgorov-Smirnov Two-Sample Test which are a function of the number of intervals, K (Smirnov, 1948; Massey, 1951; Siegel, 1956; Kanji, 1993).

Illustration

To illustrate the trip length comparison, figures 28.16 through 28.19 show the results for four different impedance models - an empirical impedance function, a negative exponential impedance function, a truncated negative exponential impedance function, and a lognormal impedance function. As seen, the fit of the empirical impedance function is not particularly good, but gets progressively better with the three different mathematical functions.

The best fit is clearly was with the lognormal function. With these parameters (mean center = 6.0 miles, standard deviation = 4.7 miles, coefficient = 1, origin exponent = 1, and destination exponent = 1.06), the Coincidence Ratio was 0.93.

But, again, this is just one criterion, though it fits most of the distribution matrix. As with the number of intra-zonal trips, minimizing the error for a trip length distribution will not necessarily minimize the error for the other two criteria (intra-zonal trips and the top links). But, it is important that the trip length comparison be reasonably close.

Comparing the Trips of the Top Links

The third evaluation test focuses on the top links. That is, it evaluates how well the predicted model captures the major trip links, both intra-zonal and inter-zonal. Since crime trip distributions are skewed (i.e., a handful of zones contribute to most crime origins and a handful of zones attract many crimes), capturing the most important links is essential for a good crime distribution model. This is particularly true since a model that produces the best fit for the overall trip length distribution may not capture the top links very well.

Therefore, simply comparing the trip length distribution may not adequately capture the top links. That is, on average a particular model may produce a good fit between the predicted and observed distributions, but may do this by minimizing error across the entire matrix of trip pairs without necessarily minimizing the error for the top links.

Consequently, it is important to also compare the fit of the model for the top links. One of the lines in the dialogue for the “Compare observed and predicted trip lengths” is “Compare top links”. The user should specify the number of top links to be compared; the default is 100. The top links are the trip pairs that have the most number of actual trips, starting from the pair with the most trips and sorting in descending order. The routine calculates a pseudo-Chi square test on just those links. Since the top links will all have a sufficient number of trips, it is possible to calculate a Chi square statistic. However, since not all links are being considered in this test, a significance test of this statistic cannot be calculated since the sampling error is not known.

Figure 28.16:

Comparing Observed and Predicted Crime Trip Lengths

Empirical Impedance Function

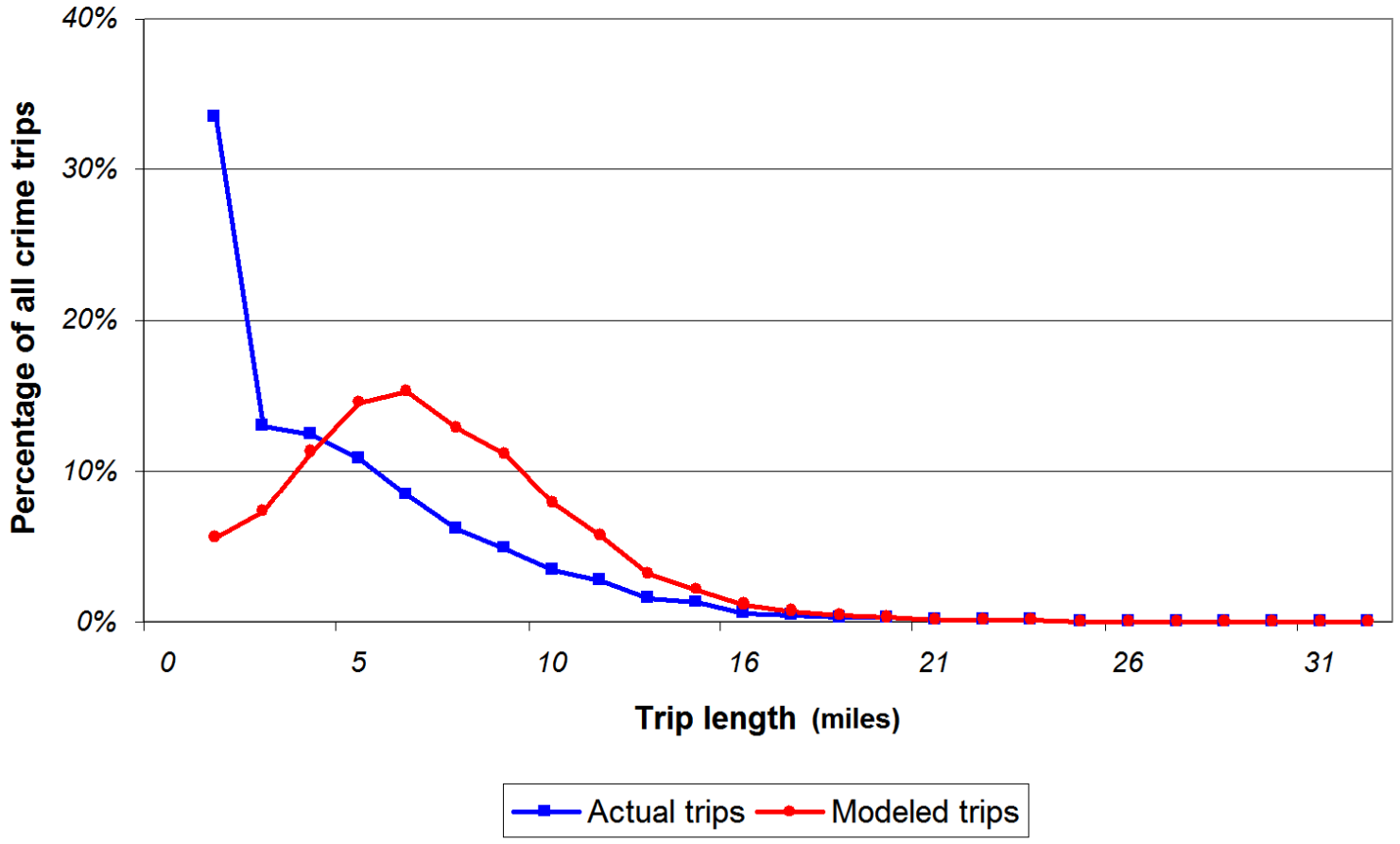


Figure 28.17:

Comparing Observed and Predicted Crime Trip Lengths

Negative Exponential Impedance Function

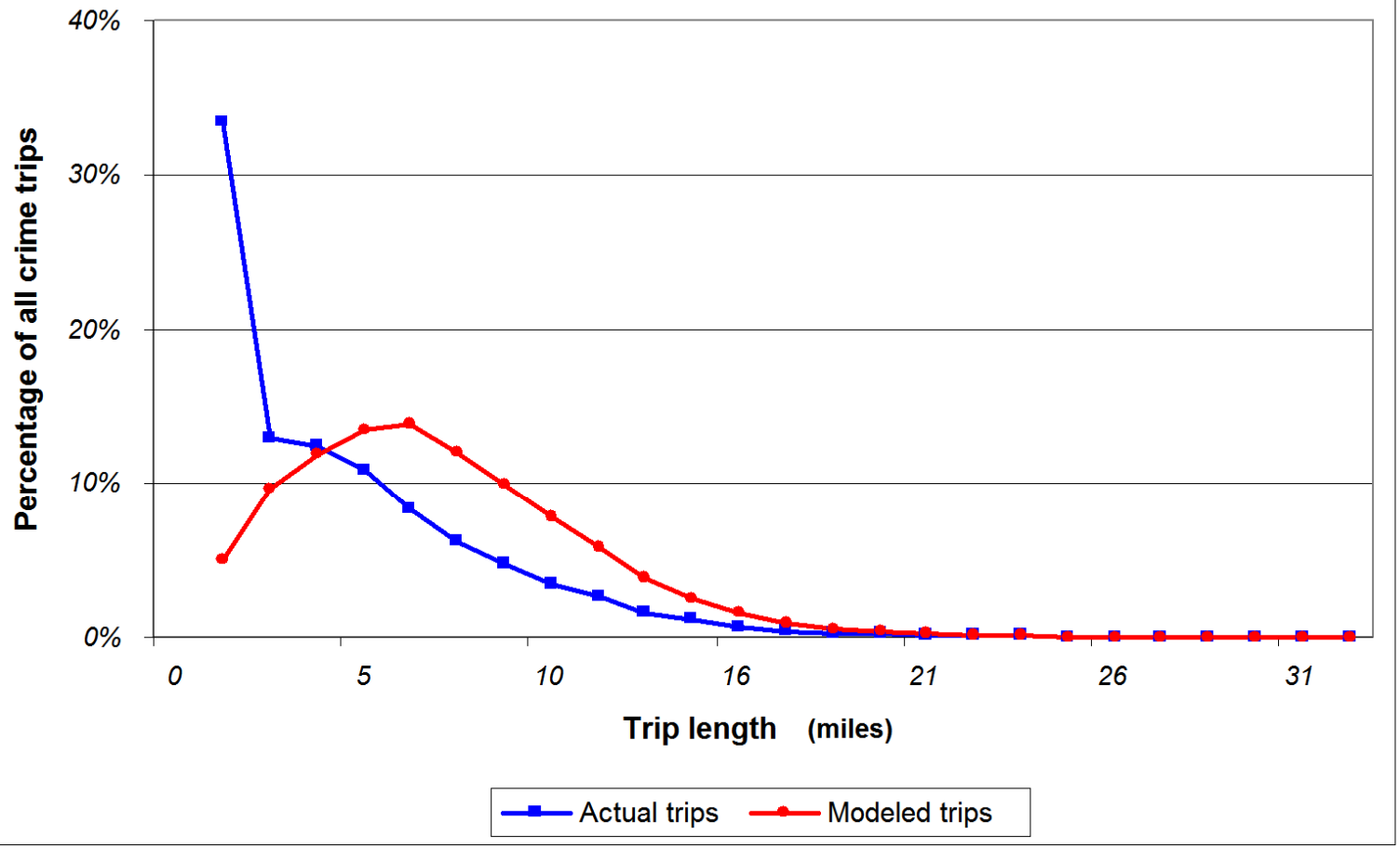


Figure 28.18:

Comparing Observed and Predicted Crime Trip Lengths

Truncated Negative Exponential Impedance Function

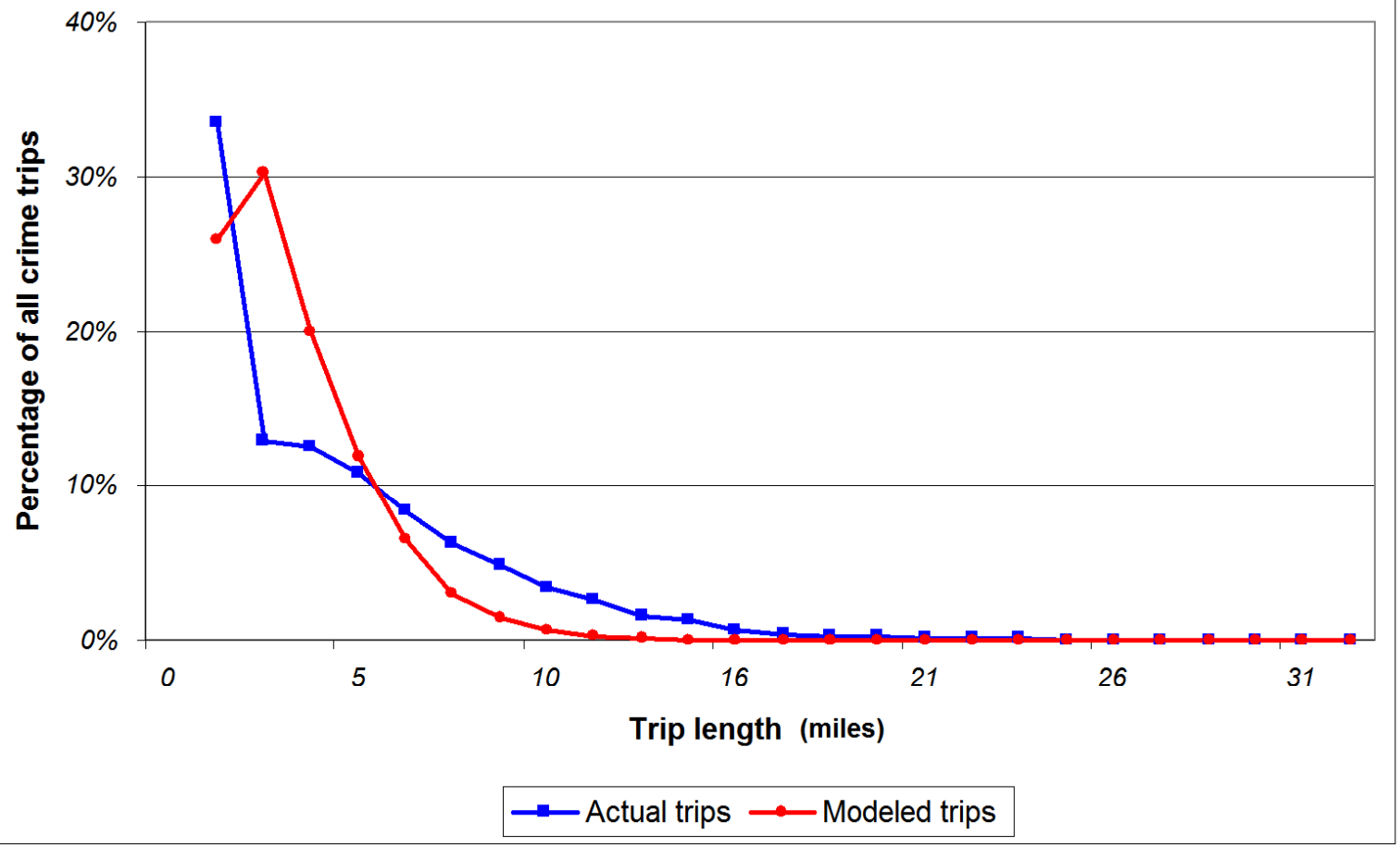
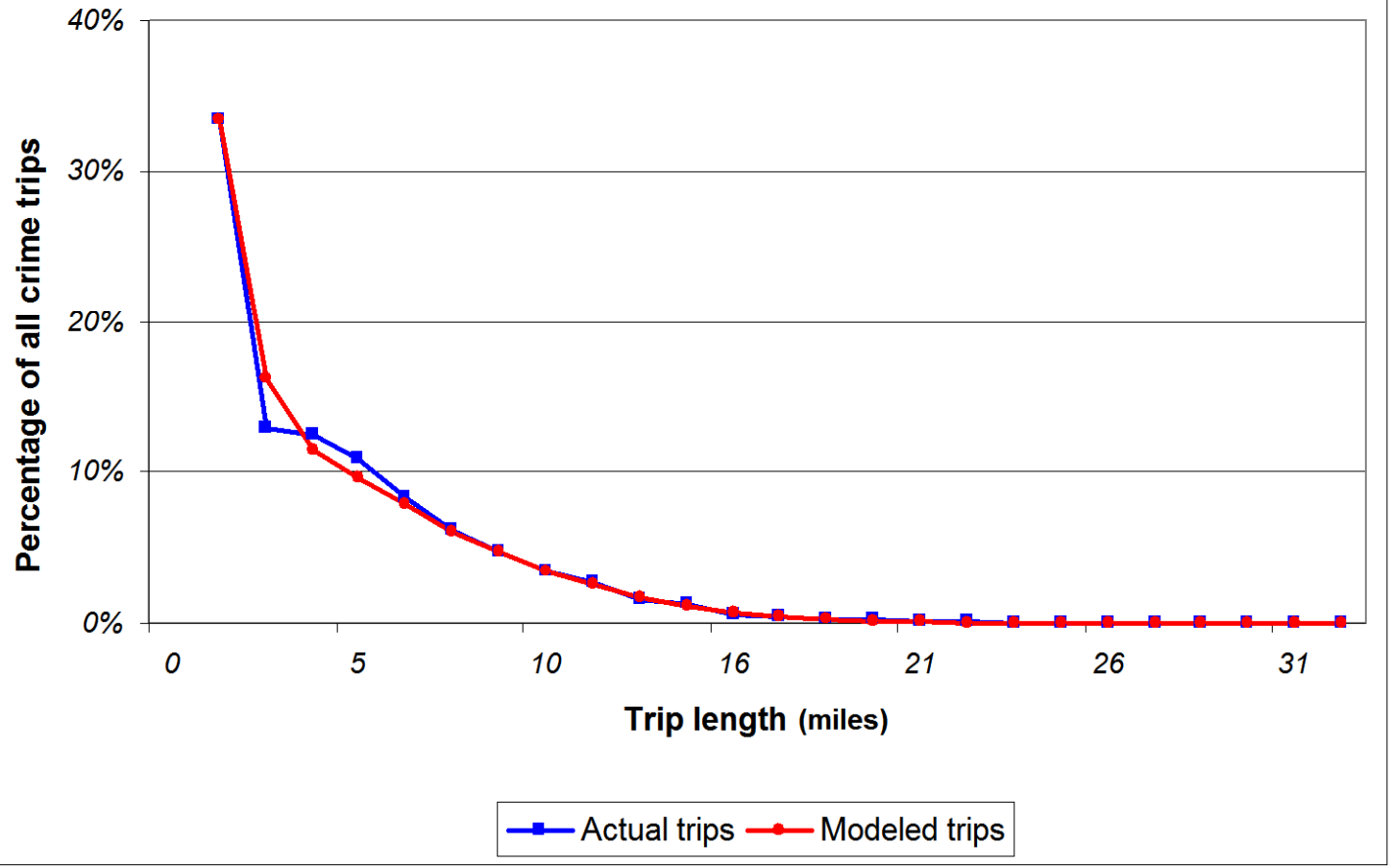


Figure 28.19:

Comparing Observed and Predicted Crime Trip Lengths

Lognormal Impedance Function



Using the observed (actual) links as the reference, the test calculates:

$$Pseudo - \chi^2 = \sum_{k=1}^{K} \frac{(O_i - P_i)^2}{O_i} \quad (28.28)$$

where O_i is the observed (actual) number of trips for trip pair, i, P_i is the predicted number of trips for trip pair, i, and i is the number of trip pairs that are compared up to K comparisons, where K is selected by the user.

Number of links to test

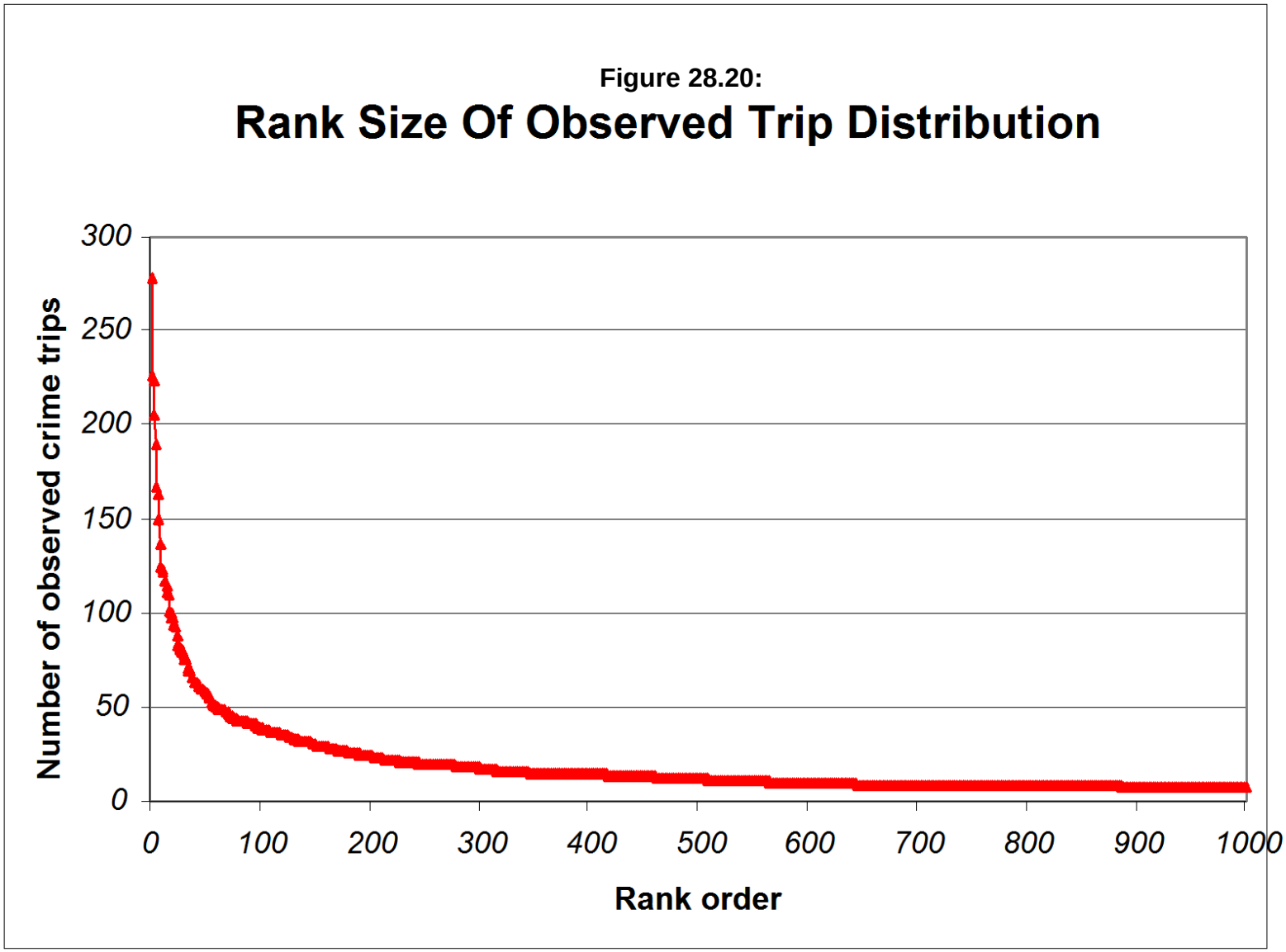
The number of top links that are to be compared depends on how skewed is the distribution. One good way to look at this is to plot the *rank size* distribution of the observed trips. Using the output 'dbf' file for the observed trip distribution (see "Calculate observed origin-destination trips" above), import the file into a spreadsheet. Sort the file in descending order of the trip frequency and create a new variable called "Rank order", which is simply the descending order of the trip frequencies. Then, plot the frequency of trips (FREQ) on the Y axis against the rank order of the trip pairs on the X axis.

Figure 28.20 below shows the rank size distribution of the Baltimore County crime trips. Notice how the distribution is very skewed for the top crime trip pairs, but declines substantially after that. That is, the top trip link (which was an intra-zonal trip pair - zone 654 to itself) had 278 trips. The second top link (also an intra-zonal pair - zone 714 to itself) had 226 trips. The third had 223; the fourth had 205; and so forth. As mentioned above, the top 1000 trip links account for about 47% of all the trips in the matrix, but the first 176 account pairs account for half of that. In other words, if the top 150 to 200 trip pairs are examined, the highest volume links will be included and most of the skewness in the distribution will be accounted for. The remaining distribution, which is not fitted, will be less skewed.

Illustration

An illustration of how comparing the top links can modify a trip distribution model can be given. The same model as shown in Figure 28.12 was run. The pseudo-Chi square test for the first 176 pairs was 5,832 (rounding-off to the nearest integer). However, by modifying the mean distance of the lognormal function a lower Chi square value was obtained. After several iterations, the lowest Chi square value was obtained for a mean distance of 5.2 miles (χ^2 = 5,448). Again, the top links represents only one criterion out of the three mentioned. A good model should balance all three of these.

Figure 28.20:


Rank Size Of Observed Trip Distribution
Number of observed crime trips
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700
800
900
1000
Rank order

Optimizing the Three Evaluation Criteria

The ideal solution would be to have all three evaluation criteria minimized. That is, with an ideal model, there should be very little error between the predicted model and the observed distribution for the number of intra-zonal trips, the trip length distribution, and the top links.

In practice, it is unlikely that any one model will minimize all three types of errors. Thus, a balance (a compromise) must be obtained in order to produce an optimal solution. Since a balance can be obtained in different ways, there are multiple solutions possible.

> **Hint**: In CrimeStat, it is very easy to run through different models. The parameters are input on the "Setup origin-destination model page". The coefficients are calibrated in the "Calibrate origin-destination model" routine on the "Origin-Destination Model" page. The coefficient file which is output is then input into the "Apply predicted origin-destination model" routine on the same page. The comparison between the observed and predicted values is found in the "Compare observed and predicted origin-destination trip lengths" routine. Once set up, iterations of the models can be run very easily. A change is made on the setup page. The model is calibrated. It is then applied to the calibration data set. Finally, a comparison is made. Since the file names remain constant, an entire iteration will take less than a minute on a fast computer.

To illustrate the multiple criteria, Table 28.2 shows the best models for each of the three tests with variations on the mean distance in the model shown in Figure 28.12. All other parameters were held constant. Many models were run to produce this table including testing other functions. These are the three best.

As seen, different models produce the lowest error for each of the criteria. For obtaining the closest fit to the number of intra-zonal trips, the mean distance of the lognormal function was 3.5 miles. For producing the best fit to the top 176 links, the mean distance for the best model was 5.2 models. For producing the best fit for the entire trip length distribution, the mean distance of the best model was 6.0 miles. The question is which one to use?

Table 28.2:

Multiple Criteria in Selecting a Distribution Function

Lognormal function
Standard deviation = 4.7 miles
Coefficient = 1
Origin exponent = 1.0
Destination exponent = 1.06

Mean distance	**Number of Intra-zonal Trips**	**Chi square for top 176 Links**	**Coincidence Ratio**
Observed	8272	-	-
6.0	5463	5814	**0.93**
5.2	6296	**5777**	0.87
3.5	**8275**	5986	0.74

One solution for optimizing decisions

One possible solution is to optimize in the following way:

1. *If* the trip distribution matrix is highly skewed (which will occur with most crime data sets), then it is essential that the top links be replicated closely. This would take priority over the second criterion which is minimizing the error for the trip length distribution, and the third criterion which is minimizing the error in predicting intra-zonal trips.

2. Next fit the model to minimize the Chi square value for the top links. In the example above, this would be the top 176 pairs. Typically, the mean distance has the biggest impact for a lognormal or normal function and this would be adjusted first. For a negative exponential function, the exponent has the strongest impact. For a linear function, the slope has the strongest impact and for a truncated negative exponential, both the peak distance, for the near distance, and the exponent, for the far distance, has the biggest impacts (see Chapter 13). Again, the aim is to produce the Chi square for the top links with the lowest value.

4. Then, while trying to maintain a Chi square value as close to this minimal value as possible, adjust the model to minimize the error in the trip length comparison. In this case, the model with the highest Coincidence Ratio is that which minimizes the error. For lognormal and normal functions, the standard deviation

is the next parameter to adjust. For a negative exponential function, the coefficient should be adjusted next. For a linear function, the intercept would be adjusted next and for a truncated negative exponential the slope would be adjusted next. Again, the aim should be to obtain the highest Coincidence Ratio without losing the fit for the top links.

5. Finally, if it is possible, adjust the exponents of the origins and destinations and the other parameters (e.g., the coefficient in the lognormal and normal distributions) to reduce the error in the total number of intra-zonal trips. Typically, however, these do not alter the results very much. They can be thought of as "fine tuning" adjustments.

Notice that this hierarchy fits the highest volume trip links first, then fits the overall trip length distribution, and finally fits the number of intra-zonal trips.

Illustration

To illustrate, we first start with the model that produced the lowest Chi square. That model used a lognormal function with a mean distance of 5.2 miles, a standard deviation of 4.7 miles, a coefficient of 1, an origin exponent of 1.0 and a destination exponent of 1.06. Varying the standard deviation of the lognormal function produced the following results (Table 28.3).

Table 28.3:
Minimizing the Second Criteria in Selecting a Distribution Function

Lognormal function
Mean distance = 5.2 miles
Standard Deviation = 4.6 miles
Coefficient = 1
Origin exponent = 1.0
Destination exponent = 1.06

Standard deviation	**Number of Intra-zonal Trips**	**Chi square for top 176 Links**	**Coincidence Ratio**
4.5	5809	5789	0.90
4.6	**6057**	**5779**	**0.88**
4.7 (baseline)	6296	5777	0.87
4.8	6526	5780	0.86
4.9	6746	5788	0.84

As the standard deviation was increased, the Coincidence Ratio decreased while the number of intra-zonal trips increased. Of these five different standard deviations, 4.5 produced the highest Coincidence Ratio, but also increased the Chi square statistic for the 176 top links. Since that criterion was set first, we do not want to loosen it substantially during the second adjustment. Consequently, a standard deviation of 4.6 was selected because this increased the Coincidence Ratio slightly while not substantially worsening the Chi square test.

Subsequent tests varying the coefficient of the lognormal function and the exponents of the origin and destination terms did not alter these values. Consequently, the final model that was selected is listed in Table 28.4.

Table 28.14:
Baltimore County Crime Trips: 1993-1997
Optimal Model Selected

Lognormal function
Mean distance = 5.2 miles
Standard deviation = 4.6
Coefficient = 1
Origin exponent = 1.0
Destination exponent = 1.06

The model was re-run with the new parameters used. The top 176 predicted trip links were output and were compared to the top 179 observed trip links (which exceeded 176 because of tied values). The top predicted 176 links accounted for 7,241 trips, or 17.3% of the total number of trips. The top observed 179 links accounted for 9,900 trip, or 23.6% of the total. Compared to the observed distribution, the top 176 predicted links accounted for a smaller proportion of the total trips.

However, the fit was generally better. Figure 28.21 shows the top predicted inter-zonal trip links and compares them to the top observed links while Figure 28.22 shows the top predicted intra-zonal (local) trip links and compares them to the top observed intra-zonal links. Comparing these maps to Figure 28.12 and 28.13 (which mapped the top 1000 links, not the top 176), the fit is a bit better for the major links, which is what we optimized. The fit is not perfect; it probably will never be. But, it is reasonably close.

Of course, this is not the only way to optimize and different users might approach it differently (e.g., minimizing the intra-zonal trips first, then the overall trip length distribution, and finally the top links). It has to be realized that optimizing in a different order will probably produce varying results; there is not, unfortunately, a single optimum solution to these three

Figure 28.21:

Comparison of Predicted and Observed Crime Trips
Top Zone-to-Zone Trips from Optimized Model
All Crime Types

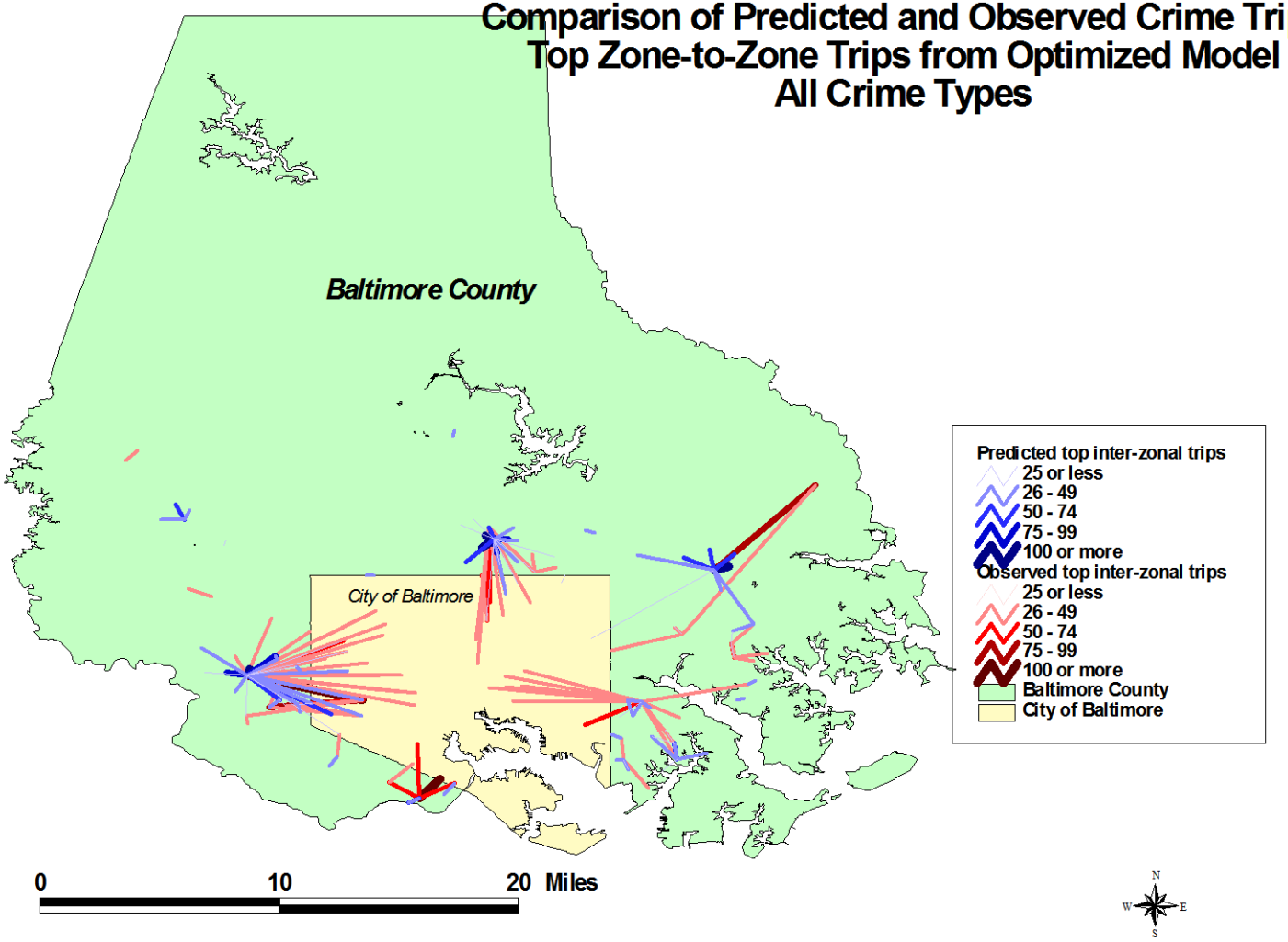
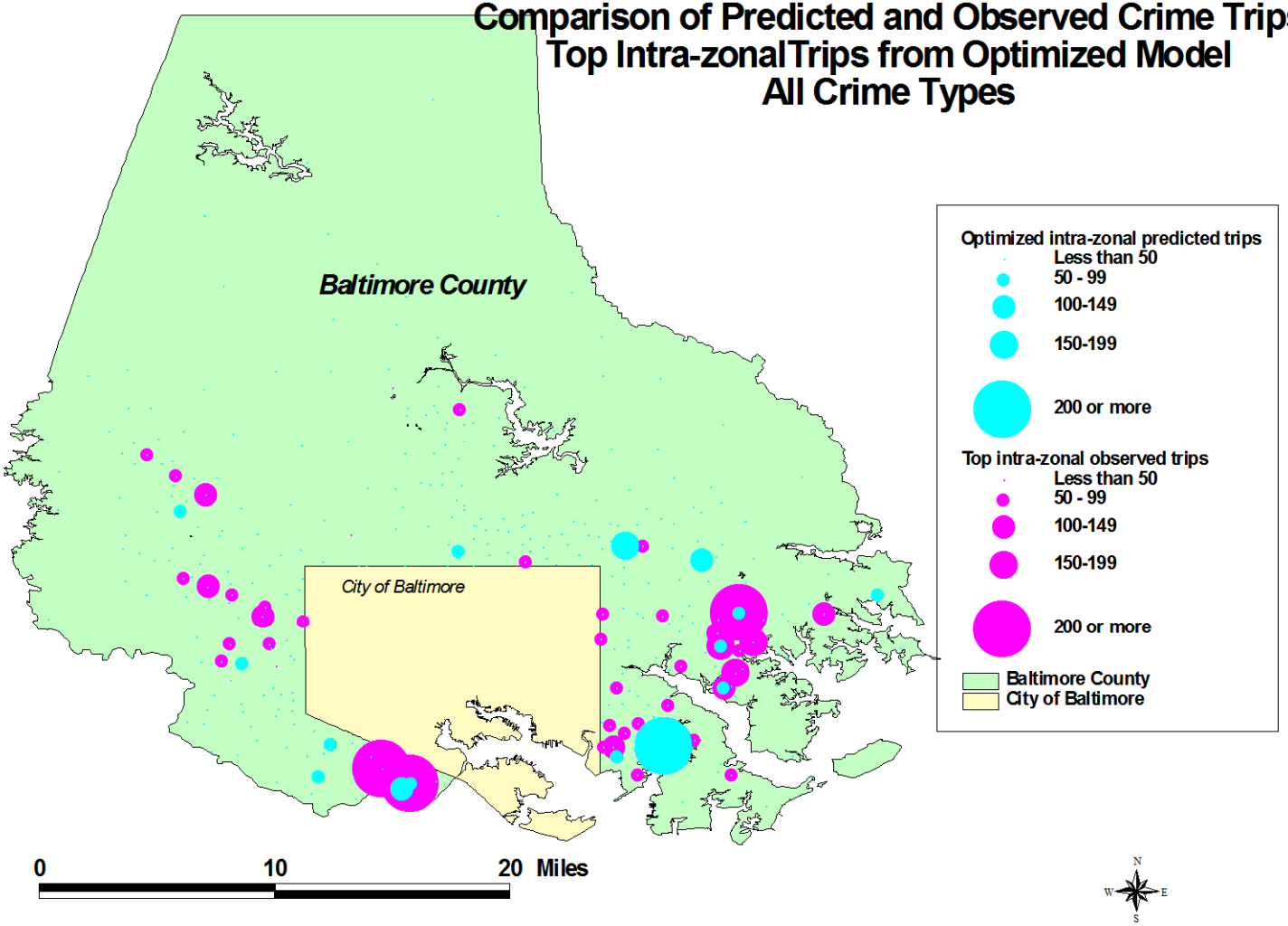


Figure 28.22:

Comparison of Predicted and Observed Crime Trips
Top Intra-zonalTrips from Optimized Model
All Crime Types



criteria. That is why it is important to explicitly define how an optimal solution will be obtained. In that way, users of the model can be cognizant of where the model is most accurate and where it is probably less accurate.

Implementing the Comparisons in *CrimeStat*

The mechanics of conducting the tests is fairly straightforward. The three tests are implemented in the "Compare Observed and Predicted Trip Lengths" routine on the last page of the Trip distribution module.

Observed trip file

Select the observed trip distribution file by clicking on the Browse button and choosing the appropriate file.

Observed number of origin-destination trips

Specify the variable for the observed number of trips. The default name is FREQ.

Orig_ID

Specify the ID name for the origin zone. The default name is ORIGIN. Note that the ID's used for the origin zones must be the same as in the destination file and the same as in the predicted trip file if the top links are to be compared.

Orig_X

Specify the name for the X coordinate of the origin zone. The default name is ORIGINX.

Orig_Y

Specify the name for the Y coordinate of the origin zone. The default name is ORIGINY.

Dest_ID

Specify the ID name for the destination zone. The default name is DEST. Note that all destination ID's should be in the origin zone file and must have the same names and the same as in the predicted trip file if the top links are to be compared.

Dest_X

Specify the name for the X coordinate of the destination zone. The default name is DESTX.

Dest_Y

Specify the name for the Y coordinate of the destination zone. The default name is DESTY.

Predicted trip file

Select the predicted trip distribution file by clicking on the Browse button and choosing the appropriate file.

Predicted number of origin-destination trips

Specify the variable for the observed number of trips. The default name is PREDTRIPS.

Orig_ID

Specify the ID name for the origin zone. The default name is ORIGIN. Note that the ID's used for the origin zones must be the same as in the destination file and the same as in the observed trip file if the top links are to be compared.

Orig_X

Specify the name for the X coordinate of the origin zone. The default name is ORIGINX.

Orig_Y

Specify the name for the Y coordinate of the origin zone. The default name is ORIGINY.

Dest_ID

Specify the ID name for the destination zone. The default name is DEST. Note that all destination ID's should be in the origin zone file and must have the same names and the same as in the observed trip file if the top links are to be compared.

Dest_X

Specify the name for the X coordinate of the destination zone. The default name is DESTX.

Dest_Y

Specify the name for the Y coordinate of the destination zone. The default name is DESTY.

Select bins

Specify how the bins (intervals) will be defined. There are two choices. One is to select a fixed number of bins. The other is to select a constant interval.

Fixed number

This sets a fixed number of bins. An interval is defined by the maximum distance between zone divided by the number of bins. The default number of bins is 25. Specify the number of bins.

Constant interval

This defines an interval of a specific size. If selected, the units must also be chosen. The default is 0.25 miles. Other distance units are nautical miles, feet, kilometers, and meters. Specify the interval size.

Compare top links

The “Compare top <value> links” dialogue implements a comparison of the top links. The user specifies the number of links to be compared. The default is 100. The routine

calculates a Chi square statistic for these links. Note that in order to make the comparison, the origin and destination ID's must be the same for both the observed and predicted trip files.

Save comparison

The output is saved as a 'dbf' file specified by the user.

Table output

The table output includes summary information and:

1. The number of trips in the observed origin-destination file
2. The number of trips in the predicted origin-destination file
3. The number of intra-zonal trips in the observed origin-destination file
4. The number of intra-zonal trips in the predicted origin-destination file
5. The number of inter-zonal trips in the observed origin-destination file
6. The number of inter-zonal trips in the predicted origin-destination file
7. The average observed trip length
8. The average predicted trip length
9. The median observed trip length
10. The median predicted trip length
11. The Coincidence Ratio (an indicator of congruence varying from 0 to 1)
12. The D value for the Komolgorov-Smirnov two-sample test
13. The critical D value for the Komolgorov-Smirnov two-sample test
14. The p-value associated with the D value of Komolgorov-Smirnov two-sample test relative to the critical D value.
15. The pseudo-Chi square test for the top links

and for each bin:

16. The bin number
17. The bin distance
18. The observed proportion
19. The predicted proportion

File output

The saved file includes:

1. The bin number (BIN)

2. The bin distance (BINDIST)
3. The observed proportion (OBSERVPROP)
4. The predicted proportion (PREDPROP)

Graph

While the output page is open, clicking on the graph button will display a graph of the observed and predicted trip length proportions on the Y-axis by the trip length distance on the X-axis. This would produce a similar graph to that seen in Figures 28.16 through 28.19 above.

Uses of Trip Distribution Analysis

There are a number of uses for the trip distribution analysis. First, for policing, an analysis of the actual (observed) trip distribution can be valuable. Second, the predicted model has value, above-and-beyond the analysis of the actual distribution.

Utility of Observed Trip Distribution Analysis

This information by itself can be very useful for police. Two applications will be discussed.

Crime prevention efforts

A major application is using the data shown in a trip distribution map to guide enforcement efforts. For example, in Baltimore County, with the crimes occurring at the five shopping malls, the origin locations can be more easily seen. This has utility for police. First, the police can intervene more effectively on the routes leading from likely origin locations. They can patrol those routes more heavily and, perhaps, intervene more frequently. By using the information from the trip distribution analysis, they make their enforcement efforts smarter. Second, they can conduct crime prevention efforts more effectively. By knowing the likely origin of offenders, intervention efforts in the origin zones may head off some of these incidents. Programs such as *weed-and-seed* and after-school programs depend on providing alternative facilities for youth, hoping to redirect them to more constructive activities. These facilities can be placed in locations where many crimes originate.

Improved Journey-to-crime analysis

A second application is in guessing the likely origin of a serial offender. In Chapter 13, theories of travel behavior by a serial offender was discussed. The resulting analysis (geographic profiling, Journey-to-crime analysis) utilized information on the distribution of

incidents committed by the offender. On the other hand, the trip distribution pattern seen in Figure 28.4 provides a probability map of offender locations and gives more information than was evident in the Journey-to-crime model. That model assigned a likelihood of the offender living at a location (the origin) on the basis of the distribution of the incidents. There was no additional information used about likely origin locations. This trip distribution map, on the other hand, points to certain zones as being the likely origin for offenses committed at the major destination locations. There is more 'structure' in this analysis than in the Journey-to-crime logic. This is the basis for the Bayesian Journey-to-crime approach discussed in Chapter 14.

One can think of this in terms of a quasi-Bayesian approach to guessing the likely origin of an offender. The geographic profiling/Journey-to-crime logic assumes no *prior probabilities*. The only information that is used is the distribution of crimes committed by a serial offender and a model of crime travel distance (essentially, an impedance function). The trip distribution map, on the other hand, points to certain locations as being the likely origin for incidents. Admittedly, this is based on a large sample of cases rather than one particular serial offender. But, the map points to certain prior probabilities for an origin location. The Bayesian Journey-to-crime routine combines those two pieces of information. As mentioned in Chapter 14, tests on more than 1000 serial offenders in four cities (in three countries) showed that the method was 10-15% more accurate than the traditional journey-to-crime approach and as precise.

In other words, the empirical description of crime travel patterns is useful for policing, above-and-beyond any modeling that is developed.

Utility of Predicted Trip Distribution Analysis

The model also has a lot of utility for both policing and crime analysis. A number of examples will be given. First, it can be used for **forecasting**. By calibrating the model on one data set, it be applied to a future data set. As mentioned in Chapter 26, much of the population and employment data that form the basis of a trip generation model comes from a Metropolitan Planning Organization (MPO). Most MPOs in the United States also make forecasts of future population and employment. Those forecasts can be, in turn, converted into forecasts of future crime origins and crime destinations. Thus, on the assumption that the distribution trends will remain the same over time, the trip distribution model can be applied to the forecast set of origins and destinations. This could allow an examination of possible changes in the crime distribution (assuming that the future forecasts are correct and that the trip distribution coefficients remain constant).

Second, a model of crime trip distribution can be useful for modeling **changes in land uses**. For example, if a new shopping mall is being planned, one can take the existing trip generation model and adjust it to fit the planned situation (e.g., adding 500 retail jobs to the zone

in which the mall is being developed). Then, the trip generation model is re-run with the new expected data, and the trip distribution model is applied to the predicted crime origins and crime destinations. The result would be a model of likely crime trips to the new shopping mall. This can be useful to the mall developers, to future businesses, and to the police. If it turns out that the model forecasts there will be a sizeable number of crime trips to that mall, then preventive actions can be developed before the mall is built (e.g., improving security design in the mall; improving the parking lot arrangement).

Third, a model of crime trip distribution can help in analyzing **future interventions**. For example, increasing police patrols in a high crime attraction area can be examined as to possible effectiveness before taking the trouble to reorganize deployment. Or, adding a new drug treatment center or a new youth center can be modeled as to its possible effectiveness in changing the nature of crime trips. Again, the input is at the data level, which affects the trip generation model. But the trip distribution model is applied to the new outputs from the trip generation model. The advantage of a model is that it explores a set of interventions without having to actually having to implement them; it is a 'thinking' tool for planning change.

Fourth, and finally, a crime trip distribution model is helpful in developing **crime theory**. As indicated in Chapter 25, the theory of crime travel has been very elementary up to now. The primary focus of analysis has been only on the destinations and on the trip lengths as measured by distance traveled. A trip distribution model, on the other hand, analyzes both trip destinations and trip origins, and can include a more sophisticated measure of impedance than simple distance. Because analysis is conducted over a larger area (a jurisdiction or a metropolitan area), the hierarchy of crime trips can be analyzed as an interaction between origins and destinations. In short, a crime trip distribution model is a 'quantum leap' in sophistication and complexity compared to the usual Journey-to-crime types of models. Hopefully, it will generate even more sophisticated types of models. The attachment illustrates how the crime travel demand model was used to examine possible interventions to reduce DWI trips ending in crashes in Baltimore County.

The next chapter continues the travel demand model by examining how crime trip links are split into different travel modes. That is, the trip distribution model estimates the number of trips flowing from each origin zone to each destination zone. The mode split model then breaks these trips into distinct travel modes.

References

Andersson, T. (1897). *Den Inre Omflyttningen*. Norrland: Mälmo.

Bernasco, W. & Block, R. (2009). Where offenders choose to attack: A discrete choice model of robberies in Chicago. *Criminology* 47(1): 93-130.

Bossard, E. G. (1993). RETAIL: Retail trade spatial interaction. In Richard E. Klosterman, Richard K. Brail & Earl G. Bossard, *Spreadsheet Models for Urban and Regional Analysis*. Center for Urban Policy Research, Rutgers University: New Brunswick, NJ, 419-448.

Bright, M. L. & Thomas, D. S. (1941). Interstate migration and intervening opportunities, *American Sociological Review*, 6, 773-783.

Carnegie-Mellon University (1975). *Security of Patrons on Urban Public Transportation Systems*. Transportation Research Institute, Carnegie-Mellon University: Pittsburgh, PA.
Cliff, A. D. & Haggett, P. (1988). *Atlas of Disease Distributions*. Blackwell Reference: Oxford.

Domencich, T. & McFadden, D. (1975). *Urban Travel Demand: A Behavioral Analysis*. North Holland Publishing Company: Amsterdam & Oxford (republished in 1996). Also found at http://emlab.berkeley.edu/users/mcfadden/travel.html. Accessed April 28, 2012.

FHWA (1997). *Model Validation and Reasonableness Checking Manual*. Prepared by Barton-Aschman Associates, Inc and Cambridge Systematics, Inc for the Travel Model Improvement Program, Federal Highway Administration, U.S. Department of Transportation: Washington, DC. http://ops.fhwa.dot.gov/freight/publications/qrfm2/sect08.htm. Accessed May 31, 2012.

Field, B. & MacGregor, B. (1987). *Forecasting Techniques for Urban and Regional Planning*. UCL Press, Ltd: London.

Foot, D. (1981). *Operational Urban Models*. Methuen: London.

Hägerstrand, T. (1957). Migration and area: survey of a sample of Swedish migration fields and hypothetical considerations on their genesis. *Lund Studies in Geography, Series B, Human Geography*, 4, 3-19.

Huff, D. L. (1963). A probabilistic analysis of shopping center trade areas. *Land Economics*, 39, 81-90.

Isbel, E. C. (1944). Internal migration in Sweden and intervening opportunities, *American Sociological Review*, 9, 627-639.

References (continued)

Isard, W. (1979). *Location and Space-Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure* (originally published 1956). Program in Urban and Regional Studies, Cornell University: Ithaca, NY.

Johnson, M.A. (1978). Attribute importance in multiattribute transportation decisions, *Transportation Research Record*, 673, 15-21.

Kanji, G. K. (1993). *100 Statistical Tests*. Sage Publications: Thousand Oaks, CA.

Levine, N. & Canter, P. (2011). "Linking origins with destinations for DWI Motor Vehicle Crashes: An application of crime travel demand modeling". *Crime Mapping*, 3, 7-41.

Levine, N. & Wachs, M. (1986). Bus Crime in Los Angeles: II - Victims and Public Impact. *Transportation Research*. 20 (4), 285-293.

Massey, F. J., Jr (1951). The distribution of the maximum deviation between two sample cumulative step functions. *Annuals of Mathematical Statistics*, 22, 125-128.

NCHRP (1995). *Travel Estimation Techniques for Urban Planning*. Project 8-29(2). National Cooperative Highway Research Program, Transportation Research Board: Washington, DC. http://www.trb.org/main/blurbs/160284.aspx. Accessed May 29, 2012.

Oppenheim, N. (1980). *Applied Models in Urban and Regional Analysis*. Prentice-Hall, Inc.: Englewood Cliffs, NJ.

Ortuzar, J. D. & Willumsen, L. G. (2001). *Modeling Transport* (3rd edition). J. Wiley & Sons: New York.

Porojan, A. (2000). Trade flows and spatial effects: the Gravity Model revisited. Conference on Managing Economic Transition in Eastern Europe: Emerging Research Issues. The Manchester Metropolitan University: Manchester, England, January.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Royal Statistical Society*. 48.

Reilly, W. J. (1929). Methods for the study of retail relationships. *University of Texas Bulletin*, 2944.

Roemer, F. & Sinha, K. (1974). Personal security in buses and its effects on ridership in Milwaukee, *Transportation Research Record*, 487, 13-25.

References (continued)

Schnell, J. B., A. J. Smith, K. R. Dimsdale, & L. J. Thrasher (1973). *Vandalism and Passenger Security: A Study of Crime and Vandalism on Urban Mass Transit Systems in the United States and Canada*. Prepared by the American Transit Association for the Urban Mass Transportation Administration (now Federal Transit Administration), U. S. Department of Transportation. National Technical Information Service: Springfield, VA. PB 236-854.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill: New York.

Smirnov, N. V. (1948). Table for estimating the goodness of fit of empirical distributions. *Annuals of Mathematical Statistics*, 19, 279-281.

Stewart, J. Q. (1950). The development of social physics. *American Journal of Physics*, 18, 239-53.

Stouffer, S. A. (1940). Intervening opportunities: a theory relating mobility and distance. *American Sociological Review*, 5, 845-67.

Wachs, M., Taylor, B., Levine, N. & Ong, P. (1993). The Changing Commute: A Case Study of the Jobs/Housing Relationship Over Time. *Urban Studies*. 30 10, 1711-1729.

WASHCOG (1974). *Citizen Safety and Bus Transit*. Metropolitan Washington Council of Governments. National Technical Information Service, Springfield, VA. PB 237-740/AS.

Wilson, A. G. (1970). *Entropy in Urban and Regional Planning*. Leonard Hill Books: Buckinghamshire.

Zhao, F., Chow, L-F, Li, M-T, Gan, A., & Shen, D. L. (2001). *Refinement of FSUTMS Trip Distribution Methodology*. Lehman Center for Transportation Research, Florida International University: Miami, FL..

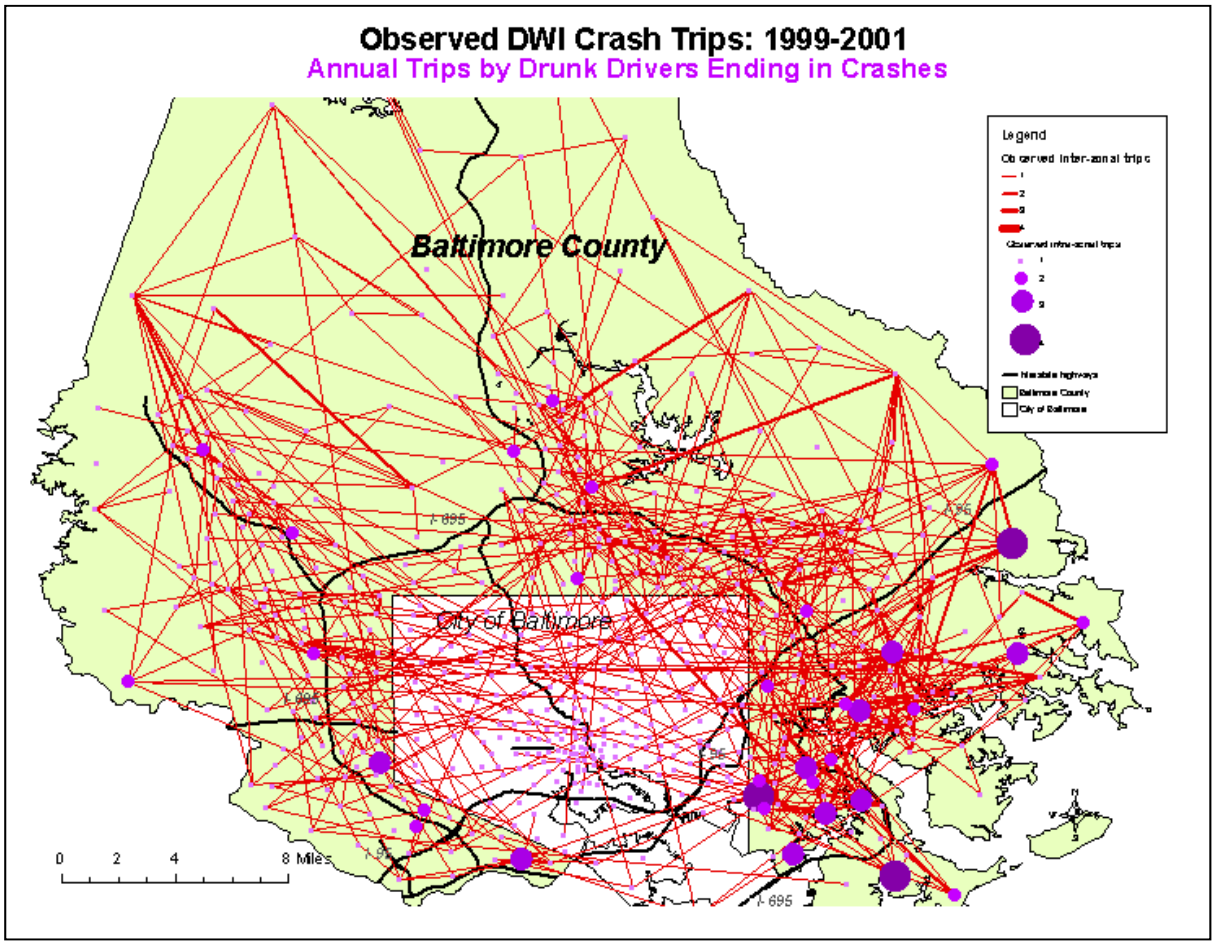
Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort*. Cambridge.

Modeling DWI Trips That End in Crashes in Baltimore County, MD

Ned Levine
Ned Levine & Associates
Houston, TX

Phil Canter
Towson University
Towson, MD

A crime travel demand study was conducted on 862 Driving While Intoxicated (DWI) motor vehicle crash trips that occurred in Baltimore County, Maryland between 1999 and 2001. Factors associated with both the residence location of the drivers and the crash location were identified. The crime travel demand model was used to simulate the likely outcome of concentrating on a few zones with targeted interventions. It was estimated that a 7.5% reduction in DWI crashes could be obtained by targeting 3% of the origin zones and 6% of the destination zones with anti-DWI efforts. The full study can be found in Levine, N. & Canter, P. (2011), Linking origins with destinations for DWI Motor Vehicle Crashes: An application of crime travel demand modeling". *Crime Mapping*, 3, 7-41.

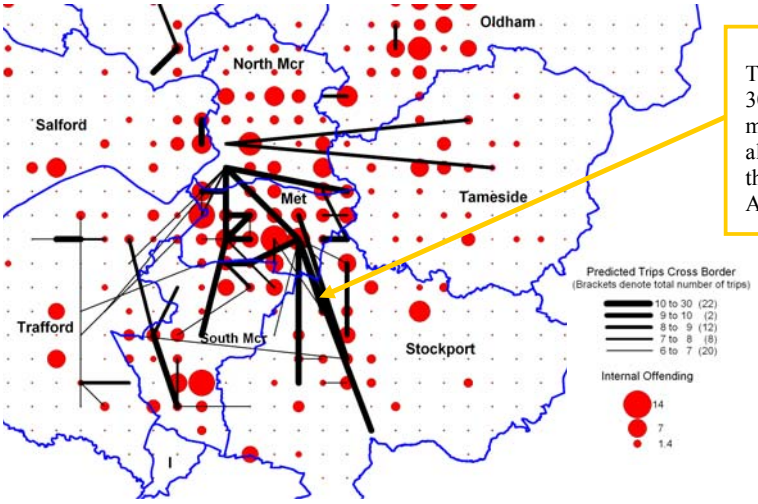


Targeting Crime on Public Transport:
An Example from Greater Manchester, England

Daisy Smith & Steph Winstanley
Strategic Analytical Partnership Co-ordinators
Greater Manchester Against Crime Central Team

The aim of the Greater Manchester Against Crime Central Team was to provide GMPTE (Greater Manchester Passenger Transport Executive) with an evidence base for their resources to address incidents of crime and anti-social behaviour on public transport during a Greater Manchester Partnership Day of Action. The analysis made use of the *Crimestat* Crime Travel Demand module to map the 'journey to crime' (home address to offence location) taken by personal robbery offenders within Greater Manchester. As a result GMPTE were able to identify their role in the partnership operation as they could easily visualise the bus and Metrolink tram system routes that ran coterminous with the most frequent journeys taken by offenders.

Personal Robbery: Internal Offending and Predicted Cross Border Trips



The map shows a high number of predicted trips (10 to 30) between Stockport and central Manchester. This mirrors the route of the 192 bus service, which travels along the A6 between Stockport and Manchester that was then identified as a main target for the Partnership Day of Action.

During the Day of Action, Gateway checks were conducted on the key public transport routes (bus routes and the Metrolink tram system) that were identified through *Crimestat* analysis. The Gateway checks consisted of staff from a range of agencies deployed on static and mobile patrols in order to identify fare evasion/ fraud and conduct intelligence checks. The agencies involved included Greater Manchester Police, GMPTE, Carlisle Security (independent enforcement agency) and the UK Border Agency. The public transport routes identified through *Crimestat* were targeted with much success and resulted in 7058 passengers being checked, 496 buses boarded, 76 people identified without valid tickets, 28 intelligence checks, 22 Bus Operator penalties issued and 22 arrests (including possession of illegal substances, robbery, fraud, outstanding warrant). The total fraud prevented through the Gateway Checks was estimated to be £3784.50 and extremely positive feedback was received from all agencies involved.