

Chapter 29:
Crime Mode Split

Ned Levine
Ned Levine & Associates
Houston, TX

Table of Contents

Theoretical Background	29.1
Utility of Travel and Mode Choice	29.1
Discrete Choice Analysis	29.4
Multinomial Logit Function	29.5
Generalized Relative Utility Function	29.7
Measuring Travel Costs	29.8
Aggregate and Individual Utility Functions	29.9
Tools for Estimating Mode Split in <i>CrimeStat</i>	29.11
Relative Accessibility	29.11
Hierarchical Approaches to Estimating Mode Accessibility	29.12
Spreadsheet for Estimating Mode Split Impedance Values	29.15
Define travel modes	29.15
Define target proportions	29.17
Other studies	29.17
Journey to work census	29.18
Select mode functions	29.20
Select model priorities	29.21
Iteratively estimate parameters	29.22
Examine the graphs in the spreadsheet	29.22
Adapting spreadsheet for travel time or travel cost	29.28
Empirically Estimating the Mode-specific Impedance	29.28
<i>CrimeStat IV</i> Mode Split Tools	29.28
Mode Split Setup	29.30
Constrain Transit Trips to Network	29.30
Default	29.30
Constrain to network	29.30
Accurately defined transit networks	29.31
Utility for creating transit network	29.31
Entering the network parameters	29.32
Minimum absolute impedance	29.32
Applying the Relative Accessibility Function	29.33
Usefulness of Mode Split Modeling of Crime Trips	29.37
Limitations to the Mode Split Methodology for Crime Analysis	29.41
Conclusions	29.43
References	29.44

Chapter 29:

Crime Mode Split

In this chapter, the third modeling step in the crime travel demand model is discussed, mode split. *Mode split* involves separating (splitting) the predicted trips from each origin zone to each destination zone into distinct travel modes (e.g., walking, bicycle, driving, train, bus).

This model has both advantages and disadvantages for crime analysis. At a theoretical level, it is the most developed of the four stages since there has been extensive research on travel mode choice. For crime analysis, on the other hand, it represents the 'weakest link' in the analysis since there is very little available information on travel mode by offenders. Since researchers cannot interview the general public in order to document crimes committed by respondents nor, in most cases, even interview offenders after they have been caught, there is very little information on travel mode by offenders that has been collected.¹ Consequently, we have to depend on the existing theory of travel mode choice and adapt it intuitively to crime data. The approach is solely theoretical and depends on the validity of the existing theory and on the intuitiveness of guesses. Hopefully, in the future, there will be more information collected that would allow the model to be calibrated against some real data. But, for the time being, we are limited in what can be done.

Theoretical Background

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

Utility of Travel and Mode Choice

The key aim of mode choice analysis is to distinguish the travel mode that travelers (or, in the case of crime, offenders) use in traveling between an origin location and a destination location. In the travel demand model, the choice is for travel between a particular origin zone and

¹ There is no reason this data could not be collected. Typically, many police departments collect information on 'Method of departure' from a crime scene. When a police report is taken, the victim is sometimes asked how the offender left the scene. In most cases, the information is not recorded on the police forms, or at least those that have been examined. This information is probably unreliable in any case since many offenders will take the bus or leave their car nearby while they walk/run to the crime scene. Still, if police departments were to put more effort into collecting this information and, perhaps, to validating it with arrested offenders, then it is possible to build up reliable data sets that can be used to model mode split. Until then, unfortunately, we have to rely on theory rather than evidence.

a particular destination zone. Thus, the trips that are distributed from each origin zone to each destination zone in the trip distribution module are further split into distinct travel modes.

With few exceptions, the assumption behind the mode split decision is for a two-way trip. That is, if an offender decides on driving to a particular crime location, we normally assume that this person will also drive back to the origin location. Similarly, if the offender takes a bus to a crime location, then that person will also take the bus back to the origin location. There are, of course, exceptions. A car thief may take a bus to a crime location, then steal a car and drive back. But, in general, without information to the contrary, it is assumed that the travel mode is for a round trip journey.

Underlying the choice of a travel mode is assumed to be a *utility function*. Chapter 21 discusses the economic theory of utility choice, so the discussion here will be brief. Essentially, mode split utility is a function that describes the benefits and costs of travel by that mode (Ortuzar & Willumsen, 2001). This can be written with a conceptual equation:

$$Utility = f(benefits, costs) \tag{29.1}$$

where ‘f’ is some function of the benefits and the costs. The benefits have to do with the advantages in traveling to a particular destination from a particular origin while the costs have to do with the real and perceived costs of using a particular mode. Since the benefits of traveling to a particular destination from a particular origin are probably equal, the differences in utility between travel modes essentially represent differences in costs (Train, 2009). Thus, Equation 29.1 breaks down to:

$$Utility = F(costs) \tag{29.2}$$

where ‘F’ is another function but this time of only the costs. If different travel modes (e.g., driving, biking, walking) are each represented by a separate utility cost function, then they can be compared:

$$Utility_1 = F_1(cost_1 + cost_2 + cost_3 + \dots + cost_k) \tag{29.3a}$$

$$Utility_2 = F_2(cost_1 + cost_2 + cost_3 + \dots + cost_k) \tag{29.3b}$$

$$Utility_3 = F_3(cost_1 + cost_2 + cost_3 + \dots + cost_k) \tag{29.3c}$$

.
.

$$Utility_l = F_l(cost_1 + cost_2 + cost_3 + \dots + cost_k) \tag{29.3l}$$

where $Utility_1$ through $Utility_L$ represents l distinct travel modes, $cost_1$ through $cost_k$ represent k cost components and are variables, and F_1 through F_l represent l different utility functions (one for each mode).

There are several observations that can be made about this representation. First, each of the cost components can be applied to all modes. However, the cost components are variables in that the values may or may not be the same. For example, if $cost_1$ is the operating cost of traveling from an origin to a destination, the cost for a driver is, of course, a lot higher than for a bus passenger since the latter person shares that cost with other passengers. Similarly, if $cost_2$ is the travel time from a particular origin zone to a particular destination zone, then travel by private automobile may be a lot quicker than by public bus. Time differences can be converted into costs by applying some type of hourly wage/price to the time. To take one more example, for driving mode, there could be a cost in parking (e.g., in a central business district); for transit use, on the other hand, this cost component is zero. In other words, each of the travel modes has a different cost structure. The same costs can be enumerated, but some of them will not apply (i.e., they have a value of 0).

Second, the costs can be perceived costs as well as real costs. For example, a number of studies have demonstrated that private automobile use is seen by most people as far more convenient than bus or train (e.g., see Schnell, Smith, Dimsdale, & Thrasher, 1973; Roemer & Sinha, 1974; WASHCOG, 1974; Carnegie-Mellon University, 1975; Johnson, 1978; Levine & Wachs, 1986). 'Convenience' is defined in terms of ease of access and effort involved in travel (e.g., how long it takes to walk to a bus stop from an origin location, the number of transfers that have to be made to reach a final destination, and the time it takes to walk from the last bus stop to the final destination). While it is sometimes difficult to separate the effects of convenience from travel itself, it is clear that most people perceive this as a dimension in travel choice. In turn, convenience can be converted into a monetary value in order to allow it to be calculated in a cost equation, for example how much people are willing to pay in time savings to yield an equivalent amount of convenience (e.g., asking how many more minutes in travel time by bus an individual would be willing to absorb in order to give up having to drive).

Third, these costs can be considered at an aggregate as well as individual level. At an aggregate level, they represent average or median costs (e.g., the average time it takes to travel between zone A and zone B by private automobile, bus, train, walking, or biking; the average dollar value assigned by a sample of survey respondents to the convenience they associate in traveling by car as opposed to bus).

On the other hand, at an individual level, the costs are specific to the individual. For example, travel time differences between car and bus can be converted into an hourly wage using

the individual's income (i.e., someone making \$100,000 a year is going to price that time savings differently than someone making only \$25,000 a year).

Fourth, a more controversial point, the specific mathematical function that ties the costs together into a particular utility function may also differ. Typically, most travel demand models have assumed that a similar mathematical function is used for all travel modes; this is the negative exponential function described below (Ortuzar and Willumsen, 2001; Domencich & McFadden, 1975). However, there is no reason why different functions cannot be used. Thus, the equations above identify different functions for the modes, F_1 through F_L . One can think of this in terms of *weights*. Each of the different mathematical function weigh the cost components differently.

It is an empirical question whether individuals apply different functions to evaluating the different modes. For example, most people would not drive just to travel one block (unless it was pouring rain or unless a heavy object had to be delivered or picked up). Even though it is convenient to get into a vehicle and drive the one block, most people see the effort involved (and, most likely, the fuel and oil costs) as not being worth it.

In other words, it appears that a different utility function is being applied to walking as opposed to driving (i.e., walk up to a certain distance; drive thereafter). A strict utility theorist might disagree with this interpretation saying that the per minute cost of walking the one block and back was less than the monetized per minute cost of operating the vehicle (which may include opening a garage door, getting into the vehicle, starting the vehicle, driving out of the parking spot, closing the garage door, and then driving the one block). In other words, it could be argued that the difference in behavior has to do with the values of the different cost components, rather than the way they are *weighed* together (the mathematical function). In retrospect, one can explain any difference. It is argued in this chapter, however, that crime trips appear to show different likelihoods by travel mode and that treating each of these functions as distinct allows more flexibility in the framework.

Discrete Choice Analysis

No matter how the utility functions are defined, they have to be combined in such a way as to allow a discrete choice. That is, an offender in traveling from zone A to zone B makes a discrete choice on travel mode. There may be a probability for travel by each mode, for example 60% by car and 40% by bus. But, for an individual, the choice is car or bus, not a probability. The probabilities are obtained by a sample of individuals, for example of 10 individuals 6 went by car and 4 went by bus. But, still, at the individual level, there is a distinct choice that was made.

Multinomial Logit Function

A common mathematical framework that used is for mode choice modeling at an aggregate level is the *multinomial logit function* (Ortuzar & Willumsen, 2001; Oppenheim, 1980; Domincich & McFadden, 1975; Stopher & Meyburg, 1975). Chapter 21 on discrete choice modeling discussed this model extensively but for individual decision makers. If there was data available on the individual travel mode choices made by offenders, then an individual level discrete choice model could be constructed. However, till now, such data has not been available. Consequently, the modeling has to occur for zone-to-zone interactions rather than for individuals choosing among zones.

The multinomial logit model used is for aggregate zone-to-zone flows. That is, for all trips from zone i to zone j , a multinomial model can be defined as:

$$P_{ijL} = \frac{e^{V_{ijL}}}{\sum_{j=1}^J e^{V_{ijL}}} \quad (29.4)$$

where P_{ijL} is the probability of using a mode for any particular trip link (particular origin zone i to particular destination j), L is the travel mode, V_{ijL} is the representative utility (that observed by the researcher/analyst as opposed to total utility which includes unobserved factors) for zone i , among j alternative destinations. The representative utility, in turn, is seen as a linear combination of independent variables (predictors) for travel from origin zone i , to destination zone j :

$$V_{ijL} = \beta'_{jL} \mathbf{X}_i = \sum_{j=1}^J \beta_{jL} X_{ijL} \quad (29.5)$$

where V_{ijL} is the utility of traveling from origin zone i , to destination zone j , L is the travel mode, and β_{jL} are coefficients.

Substituting equation 29.5 into equation 29.4, we have:

$$P_{ijL} = \frac{e^{\sum_{j=1}^J \beta_{jL} X_{ijL}}}{\sum_{L=1}^J e^{\sum_{j=1}^J \beta_{jL} X_{ijL}}} \quad (29.6)$$

which relates the linear combination for any one mode to the sum of the linear combinations of all modes.

Several observations can be made about this function. First, the multinomial logit model relates the choice of alternatives to differences in the *characteristics of the zones*, both

origin and destination (the X_{ijl}), rather than to the differences in the destination zones themselves. This is different than the more general conditional logit model which relates the choices to the characteristics of the alternatives (destinations) and to interactions between the origin zones and the destination zones (and is also discussed in Chapter 21).

Second, each travel mode, L , has its own costs and benefits and can be evaluated by itself. That is, each origin zone have a different set of destination alternatives according to the characteristics of both the origin and destination zones. That is, there is a distinct utility function for each mode. This is the numerator of Equation 29.6. However, the choice of any one mode is dependent on its utility value relative to other modes (the denominator of the equation). The more choices that are available, obviously, the lower the probability that a particular origin-destination zone combination will have for that mode. But the value associated with the mode (the utility) does not change. As mentioned above, we generally assume that the benefit of traveling between any two zones is identical for all modes and, hence, any differences are due to costs.

Third, the mathematical form is the negative exponential. The exponential function is a growth function in which growth occurs at a constant *rate* (either positive - growth, or negative - decline). The use of the negative exponential assumes that the costs are related to the likelihood as a function that declines at a constant rate. It is actually a 'disincentive' or 'discount' function rather than a utility function, *per se*. That is, as the costs increase, the probability of using that mode decreases, all other things being equal. Still, for historical reasons, it is still called a utility function.

Fourth, for any one mode, the total cost is a logarithmic function of individual cost components:

$$Utility_{iL} = e^{\sum_{j=1}^J \beta_{jL} X_{ijL}} \quad (29.7)$$

$$Ln(Utility_{iL}) = \sum_{j=1}^J \beta_{jL} X_{ijL} \quad (29.8)$$

where the cumulative cost is made up of independent predictors X_1, X_2 through X_J , and β_1 through β_k are coefficients for the individual cost components. Thus, we see that the utility function is a loglinear model, as was seen in Chapter 13. Thus, the utility function is Poisson distributed, declining at a constant *rate* with increasing cumulative costs. Domincich and McFadden (1975) point out that the error term is not Poisson distributed, but skewed as a Type I extreme value distribution (sometimes called a Gumbel distribution; Train, 2009). As discussed in Chapters 16-17, there are a variety of different Poisson models that incorporate skewed error

terms (Poisson-Gamma, Poisson-lognormal, COM-Poisson) and which could also be used to fine-tune the fit. Nevertheless, the mean utility is a Poisson-type function.

Generalized Relative Utility Function

One can generalize this further to allow any type of mathematical function. While the Poisson has a long history and is widely used, allowing other non-linear functions allows greater flexibility. It is possible that individuals apply different *weighting* systems in evaluating different modes (e.g., a negative exponential for walking, but a lognormal function for driving). We certainly see what appear to be different functions when the actual travel behavior of individuals are examined (e.g., homeless individuals don't walk everywhere even though the cost of walking long distances is cheaper in travel time than taking a bus²; people don't drive or take a bus for very short distances, say a block or two). Therefore, if we allow that there are different travel functions for different modes, then more flexibility is possible than by assuming a single mathematical function.

We can, therefore, write a *generalized relative utility function* as:

$$P_{ijL} = \frac{F_L(V_{ijL})}{\sum_{L=1}^L F_L(V_{ijL})} = \frac{I_{ijL}}{\sum_{L=1}^L I_{ijL}} \quad (29.9)$$

where the terms are the same as in 29.4 except the function, F_L , is some function that is specific to the travel mode, L . The numerator is defined as the impedance of mode L in traveling from origin zone i to destination zone j while the denominator is the sum of all impedances.

Notice that the ratio of the cost function for one mode relative to the total costs is also the ratio of the impedance for mode L relative the total impedance. The total impedance was defined in Chapter 28 as the disincentive to travel as a function of separation (distance, travel time, cost). We see that the share of a particular mode, therefore, is the proportion of the total impedance of that mode. This share will vary, of course, with the degree of separation. For any given separation, there will usually be a different share for each mode. For example, at low separation between zones (e.g., zones that are next to each other), walking and biking are much more attractive than taking a bus or a train and, perhaps even driving. At greater separation (e.g., zones that are 5 miles apart), walking and biking are almost irrelevant choices and the likelihood of driving or using public transit is much greater. In other words, the share that any one mode occupies is not constant, but varies with the impedance function.

² In a survey of the travel behavior of homeless persons, it was noted that most homeless walked very short distances over the day even though the value of their time was very low. For longer trips, they still tended to take the bus rather than walk. Survey on the travel behavior of very low income individuals. Urban Planning Program, University of California at Los Angeles, 1987 (with Martin Wachs).

Why then cannot the mode split be estimated directly at the trip distribution stage? If the trip distribution function is:

$$T_{ij} = \alpha P_i^\lambda \beta \frac{A_j^\tau}{I_{ij}} \quad (28.12 \text{ repeat})$$

and if these trips, in turn, are split into distinct modes using equation 29.9, could not 28.12 be re-written as:

$$T_{ijL} = \alpha P_i^\lambda \beta \frac{A_j^\tau}{I_{ijL}} \quad (29.10)$$

where T_{ijL} is the number of trips traveling from origin zone i to destination zone j by mode L , P_i is the production capacity of zone i , A_j is the attraction of zone j , α and β are constants that are applied to the productions and attractions respectively, λ and τ are 'fine tuning' exponents of the productions and attractions respectively, and I_{ijL} is the impedance of using mode L to travel between the two zones? The answer is, yes, it could be calculated directly. If I_{ijL} was a perfectly defined impedance function (with no error), then the mode share could be calculated directly at the distribution stage instead of separating the calculations into two distinct stages. The problem, however, is that the impedance functions are never perfect (far from it, in fact) and that re-scaling is required both to get the origins and destinations balanced in the trip distribution stage and to ensure that the probabilities in equation 29.10 add to 1.0. The effect of these adjustments generally throws off a model such as equation 29.10. Consequently, the trip distribution and mode split stages are usually calculated as separate operations.

Measuring Travel Costs

The next question is what types of travel costs are there that define impedance? As mentioned above, there are real as well as perceived costs that affect a travel mode decision. Some of these can be measured easily, while others are very difficult requiring detailed surveys of individuals. Among these costs are:

1. Distance or travel time. As mentioned throughout this discussion, distance is only a rough indicator of cost since it is invariant with respect to time. Actual travel time is a much better indicator because it varies throughout the day and can be easily converted into a *travel time value*, for example by multiplying by an average unit wage.
2. Other real costs, such as the operating costs of a private vehicle (fuel, oil, maintenance), parking, and insurance. Some of these can be subsumed under travel time value by working out an hourly price for travel.

3. Perceived costs, such as convenience, fear of being caught by an offender, ease of escape from a crime scene, difficulties in moving stolen goods, and fear of retaliation by other offenders or gangs).

Some of these costs can be measured and some cannot. For example, the value of travel time can be inferred from the median household income of a zone for aggregate analysis or from the actual household income for individual-level analysis. Parking can be averaged by zone. Insurance costs can be estimated from zone averages *if* the data can be obtained.

Many perceived costs also can be measured. Convenience, for example, could be measured from a general survey. The fear of being caught can be inferred from the amount of surveillance in a zone (e.g., the number of police personnel, security guards, security cameras). Even though it may be a difficult enumeration process, it is still possible to measure these costs and come up with some average estimate.

Other perceived costs, on the other hand, may not be easily measured. For example, the fear an offender belonging to one gang has about retaliation from another gang is not easily measured. If one could map the 'territory' of a gang, then one could members of one territory would not commit a crime in another territory (Bernasco & Block, 2009). Similarly, the cost of moving stolen goods by a thief is not easily measured; one would need to know the location of the distributors of these goods.

In practice, travel modelers make simple assumptions about costs because of the difficulty in measuring many of them. For example, travel time is taken as a proxy for all the operating costs. Parking costs can be incorporated through simple assumptions about the distribution across zones (e.g., zones within the central business district - CBD, are given an average high parking costs; zones that are central, but not in the CBD, are assigned moderate parking costs; zones that are suburban are assigned low parking costs). It would be just too time consuming to document each and every cost affecting travel behavior, particularly if we are developing a model of offender travel.

Nevertheless, theoretically, these are all potentially measurable costs. They are real and probably have an impact in the travel decisions that offenders make. As researchers and analysts, we have to work towards articulating as many of these costs as possible in order to produce a realistic representation of offender travel.

Aggregate and Individual Utility Functions

One of the big debates in travel modeling is whether to use aggregate or individual utility functions to calculate mode share. The aggregate approach measures common costs for each

zone, assuming an average value. The disaggregate approach (sometimes called ‘second generation’ models) measures unique costs for individuals, then sums upward to yield values for each zone pair. Even though the end result is an allocation of costs to each zone pair, the articulation of unique costs at the individual level can, in theory, allow a more realistic assessment of the utility function that is applied to a region.

The aggregate approach will measure costs by averages. Thus, a typical equation for driving mode might be:

$$Total\ cost_{ij} = \beta_1 T_{ijk} + \beta_2 P_{ijk} \quad (29.11)$$

where T_{ij} is the average travel time between two zones, i and j , and P_{ij} is the average parking cost for parking in zone j . Notice that there are a limited number of variables in an aggregate model (in this case, only two) and that the assigned average is for an entire zone. Notice also that the parking cost is applied only to the destination zone. It is assumed that any traveler will pay that fee in that zone irrespective of which origin zone he/she came from.

A disaggregate approach can allow more cost components, if they are measured. Thus, a typical equation for driving mode might be:

$$Total\ cost_{ijk} = \beta_1 T_{ijk} + \beta_2 P_{ijk} + \beta_3 C_{ijk} + \beta_4 M_{ijk} + \beta_5 S_{ijk} \quad (29.12)$$

where T_{ijk} is the travel time for individual k between two zones, i and j , P_{ij} is the average parking cost for parking in zone j , C_{ijk} is the convenience of traveling to zone j from zone i for individual k , M_{ijk} is the comfort and privacy experienced by individual k in traveling from zone i to zone j , and S_{ijk} is the perceived safety experienced by individual k in traveling from zone i to zone j . Notice that there are more cost variables in the equation and that the model is targeted specifically to the individual, k . Two individuals who live next door to each other and who travel to the same destination may evaluate these components differently. If these individuals have substantially different incomes, then the value of the travel time will differ. If one values privacy enormously while the other does not, then the cost of driving for the first is less than for the second. Similarly, convenience is affected by both travel time and the ease of getting in and out of vehicle. Finally, the perception of safety may differ for these two hypothetical individuals. There are many studies that have documented the significant role played by safety in affecting, particularly, transit trips (Levine & Wachs, 1986).

In other words, the aggregate approach applies a very elementary type of utility function whereas the disaggregate approach allows much more complexity and individual variability.

Of course, one has to be able to measure the individual cost components, a difficult task under most circumstances.

There is also a question about which approach is more accurate for correctly forecasting actual mode splits. Historically, most Metropolitan Planning Organizations have used the aggregate method because it's easier. However, more recent research (McFadden, 2002; Ben-Akiva & Lerman, 1985; Domincich & McFadden, 1975) has suggested that disaggregate modeling may be more accurate. At the very minimum, the disaggregate model is more amenable to policy interpretations because it is more behavioral. *If* one could interview travelers with a survey, then it is possible to explore the variety of cost factors that affect a decision on both destination and mode split, and a more realistic (if not unique) utility function derived.

But, as mentioned above, with crime trips, this is very difficult, if not impossible, to do. Consequently, for the time being, we are stuck with an aggregate approach towards modeling the utility of travel by offenders.

Tools for Estimating Mode Split in *CrimeStat*

CrimeStat has two sets of tools for estimating the mode split model. First, if individual level data on travel modes can be obtained, then the multinomial logit model in the Discrete Choice module can be used (see Chapters 21 and 22). That is, if data on actual mode choices taken by offenders could be obtained with characteristics of both the offenders and the zones in which they lived or committed crimes in being associated with those choices, then the preferred method would be to use the multinomial logit to model the predictors of mode choice.

Second, if individual level data on travel modes is not available (the usual circumstances in most police department), then an approximation to a utility function can be made. The approach, in this case, is to estimate a *relative accessibility* function and then apply that function to the predicted trip distribution. The relative accessibility function is a mathematical approximation to a utility function, rather than a measured utility function by itself. Because the cost components cannot be measured, at least for offenders, we use an inductive approach. Reasonable assumptions are made and mathematical functions are found that fit these assumptions.

Relative Accessibility

The relative accessibility approach produces a plausible model, not an analytical one. The plausibility comes by making reasonable assumptions about actual travel behavior. One can assume that walking trips will occur for short trips, say under two miles. Bicycle trips, on the

other hand, could occur over longer distances, but will still be relatively short (also, there is always the risk of traffic on the safety of bicycle trips). Transit trips (bus and train) will be used for moderately long distances but require an actual transit network. Finally, driving trips are the most flexible because they can occur over any size distance and road network. They are less likely to be used for very short trips, on the other hand, due to reasons discussed above.

Hierarchical Approach to Estimating Mode Accessibility

Using this approach, specific steps can be defined to produce a plausible accessibility model. The Mode Split tab in *CrimeStat* allows the estimation of a relative accessibility model. Figure 29.1 shows the setup page for the mode split module. There are four tabs in the mode split module:

1. The setup tab
2. Calibrate mode split: I
3. Calibrate mode split: II
4. Calibrate mode split: III

The setup tab defines the three files that are used in the module. There is predicted origin file, a predicted destination file, and a predicted origin-destination trip file. Both the predicted origin file and the predicted destination file have to be defined as the Primary file or Secondary file and ID fields have to be defined for each. The predicted origin-destination trip file is input separately on this tab. The variable identifying the predicted number of trips must be defined. Figure 29.1 illustrates this one data set.

The next three tabs define up to five separate travel modes. The first Calibrate mode split tab defines modes 1 and 2. The second tab defines modes 3 and 4 and the third defines mode 5. The user can assign any one mode to each of these available slots. For example, mode 1 could be walking; mode 2 could be driving; mode 3 could be bus; mode 4 could be train; and mode 5 could be bicycle. There is no particular order to the assignment and not all five available modes have to be used. Figure 29.2 illustrates the defining of two modes on the 'Calibrate mode split: I' tab where mode 1 is walking and mode 2 is by bicycle.

For each of the modes that are used, the user must define an impedance function. The impedance function is a mathematical function that approximates the discounting of that mode with distance. The user can use a pre-defined mathematical function or else estimate it using an already-calibrated impedance function (see Chapter 28, page 28.24). In the example in Figure

Figure 29.1:
Mode Split Module

The screenshot shows the 'Mode Split Module' window in CrimeStat IV. The window title is 'CrimeStat IV'. The interface is divided into several sections:

- Navigation Tabs:** Data Setup, Spatial Description, Hot Spot Analysis, Spatial Modeling I, Spatial Modeling II, Crime Travel Demand, and Options.
- Sub-panels:** Project directory, Trip generation, Trip distribution, Mode split (selected), Network assignment, and File worksheet.
- Mode Split Setup:**
 - Setup for mode split model: Calibrate mode split I, Calibrate mode split II, Calibrate mode split III.
 - Predicted origin File: Primary (dropdown), Origin ID: TZ98 (dropdown).
 - Predicted destination file: Secondary (dropdown), Destination ID: TAZ (dropdown).
 - Predicted origin-destination trip file: PredictedTripsDestConstant.dbf (text field), Browse button.
 - Predicted trips: PREDTRIPS (dropdown).
 - Assumed impedance for external zone: 25 (text field), Units: Miles (dropdown).
 - Mode split.
 - Save result, Save links, Save points (buttons).
 - Save top links: 1000 (text field).
 - Assumed coordinates for external zone:
 - Mean center
 - Lower-left corner
 - Upper-right corner
 - Use coordinates
 - X: 0 (text field), Y: 0 (text field).
- Bottom Buttons:** Compute, Quit, Help.

Figure 29.2:
The 'Calibrate Model Split: I' Tab

CrimeStat IV

Data Setup | Spatial Description | Hot Spot Analysis | Spatial Modeling I

Spatial Modeling II | Crime Travel Demand | Options

Project directory | Trip generation | Trip distribution | Mode split | Network assignment | File worksheet

Setup for mode split model | Calibrate mode split: I | Calibrate mode split: II | Calibrate mode split: III

Mode 1 Walk

Default impedance Unit: Miles

Constrain to network: Parameters Minimum absolute impedance: 2

Impedance function:

Use already-calibrated impedance function Browse

Use mathematical formula Distribution: Negative exponential Coefficient: 0.02

Exponent: -6.94

Mode 2 Bicycle

Default impedance Unit: Miles

Constrain to network: Parameters Minimum absolute impedance: 2

Impedance function:

Use already-calibrated impedance function Browse

Use mathematical formula Distribution: Negative exponential Coefficient: 0.002

Exponent: -2.24

Compute | Quit | Help

29.2, a mathematical function is being used for both the walking mode (mode 1) and the bicycle mode (mode 2). For the walking mode, the following negative exponential function is used:

$$Y_{ijL} = 0.02e^{-6.94} \quad (29.13)$$

while bicycle mode uses a negative exponential function of the following form:

$$Y_{ijL} = 0.002e^{-2.24} \quad (29.14)$$

Note that these are not probabilities but frequencies. That is, equation 29.13 estimates the number of walking trips as a function of distance between two zones, i and j . Each equation discounts the likelihood of using that mode as a function of distance between the zones. The probabilities are estimated later when the frequency of all modes have been defined.

Spreadsheet for Estimating Mode Split Impedance Values

To identify the parameters that produce a plausible model of mode frequency, an Excel spreadsheet has been developed for making these calculations (*Mode Split Impedance Defaults.xls*). It is part of the “Crime Travel Demand Sample Data.zip” file and can be downloaded from the *CrimeStat* download page. Figure 29.3 shows part of the spreadsheet.

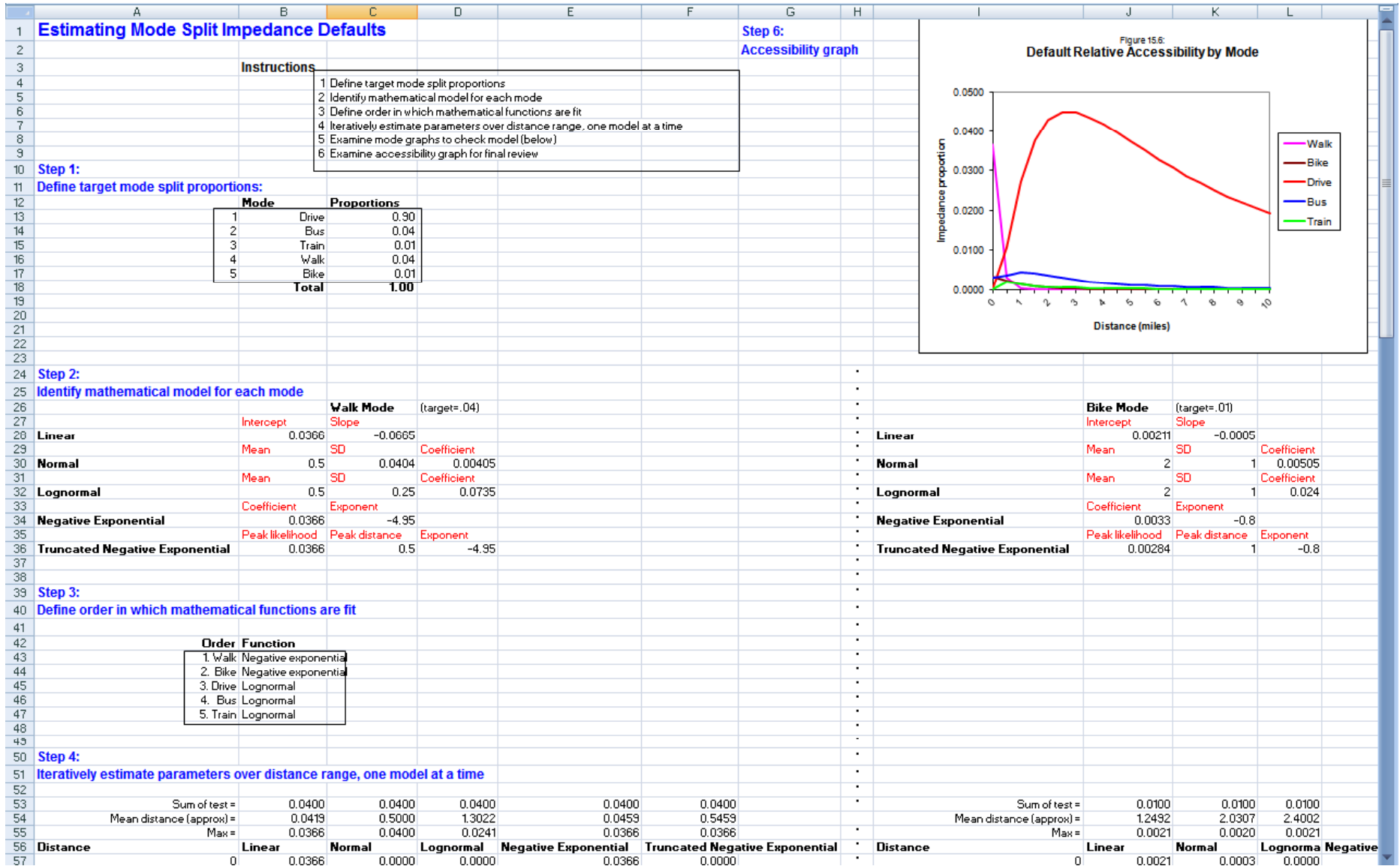
The spreadsheet has been defined with distance, but it can be adapted for travel time or travel cost as well. A spreadsheet has been used because it is more flexible than incorporating it as a routine in *CrimeStat* to estimate the parameters. There is not a single solution to the parameters estimates and the different choices can be seen more easily.

Define travel modes

The following provides instructions on estimating the impedance values with the spreadsheet. In the *CrimeStat* mode split routine, up to five different modes are allowed. First, the user should define the travel modes to be modeled. In the spreadsheet, these have default names of “Walk”, “Bike”, “Drive”, “Bus”, and “Train”. The user is not required to use these names nor all five modes. Clearly, if there is not a train system in the study area, then the “Train” mode does not apply. Travel modelers use variations on these, such as “drive alone”, “carpool”, “automobile”, “motorcycle”, and so forth.

Figure 29.3:

Estimating Mode Split Impedance Values



Define target proportions

Second, define the *target proportions*. These are the expected proportions of travel for each mode. Where would such proportions come from? There have been many studies of driving and transit behavior, but relatively few studies of bicycle and pedestrian use (Schwartz et al, 1999; Porter, Suhrbier & Schwartz, 1999; Turner, Shunk, & Hottenstein, 1998). There are not simple tables that one can look up default values.

Other studies

To solve this problem, examples were sought from different size metropolitan areas. Estimates of travel mode share for all trip purposes (work and non-work) were obtained from Ottawa (McCormick Rankin, 2011; Ottawa, 2008), Portland (Portland, 1998); and Houston³. Table 29.1 shows the estimated shares. The Houston data does not include walking and biking shares, and transit trips are not distinguished by mode in the Portland and Ottawa data.

Table 29.1:
Estimated Mode Share for Three Metropolitan Areas
All Trip Purposes

	Ottawa	Portland	Houston
Population:	725 thousand (1995)	2.0 million (2001)	4.6 million (2000)
Percent of trips by:	(1995)	(1994)	(2025 forecast)
Driving	73.5%	88.6%	98.3%
Transit	15.2%	3.0	1.7% (bus 1.1%; rail 0.6%)
Walking	9.6%	4.6%	-
Bicycle	1.7%	1.0%	-
Other	-	2.8%	-

While it is difficult to generalize, walking is very much dependent on both the compactness of the city and the existence of an extensive transit system. In Houston, the transit system is primarily a commuter system whereas in Portland and Ottawa, it serves multiple purposes. Clearly, the more compact is the urban area, the more likely that trips will occur by transit, walking or biking. But, even in the case of Ottawa where almost 10% of trips are by

³ Houston-Galveston Area Council. Personal communication. 2004.

walking, the majority of trips are by private vehicle. In the United States and Canada, for metropolitan areas with extensive transit facilities (New York, Chicago, Boston, Montreal), a majority of regional trips are still by automobile.

Based on this, some default values were selected and put into the spreadsheet. The spreadsheet requires that they are entered as proportions (not percentages). The defaults values were (Table 29.2):

The user can modify these in the spreadsheet. It is important that a user contact the local Metropolitan Planning Organization to find out what would be reasonable values for the urban area. The default values are guesses based on a limited amount of data.

Table 29.2:
Default Mode Share Values
Proportions

Mode	Share
Driving	.90
Bus	.04
Train	.01
Walk	.04
Bicycle	.01

Journey to work census

An alternative approach is to use the Journey to Work data of the U.S. Census Bureau (2004). During every census, the Census Bureau documents home-to-work ‘commute’ trips and breaks down these data by mode share. They release these data under the title “Journey to Work”. The 2010 Journey to Work data set has not been released. However, in 2000 in the United States, 87.9% of all home-to-work trips were by private vehicle (automobile, van, truck), 4.7% were by public transit (bus 2.5%; rail 2.1%; other 0.1%), 2.9% were by walking, 0.4% were by bicycle, 0.1% were by motorcycle, 0.7% were by other means, and 3.3% worked at home.

National journey to work statistics for 1990 and 2000 and for metropolitan areas in 1990 can be found at U.S. Census, 2009). Data on metropolitan areas for 2000 can be found in McGuckin and Srinivasan (2003). In 2000, the home-to-work mode share for a sample of large metropolitan (including the 15 largest) areas is shown in Table 29.3. They are rank-ordered by the 2000 population of the metropolitan area.

As can be seen, the larger metropolitan areas generally have a higher share of transit use and walking than smaller metropolitan areas, but the differences are not that dramatic. Further,

Table 29.3:**Mode Share of Journey to Work Trips: 2000**

(From McGuckin & Srinivasan, 2003)

Greater Metropolitan Area	2000 Pop (M)	<u>Mode Share</u>					
		<u>Walk</u>	<u>Bicycle</u>	<u>Drive</u>	<u>Bus</u>	<u>Rail</u>	<u>Other</u>*
New York	21.1	5.6%	0.3%	65.7%	6.8%	17.1%	4.5%
Los Angeles	16.4	2.6%	0.6%	87.6%	4.3%	0.3%	4.6%
Chicago	9.2	3.1%	0.3%	81.5%	4.6%	6.6%	3.9%
Washington DC	7.6	3.0%	0.3%	83.2%	4.1%	5.0%	4.4%
San Francisco	7.0	3.3%	1.1%	81.0%	5.7%	3.5%	5.4%
Philadelphia	6.2	3.9%	0.3%	83.6%	5.3%	3.3%	3.6%
Detroit	5.5	1.8%	0.2%	93.4%	1.7%	0.0%	2.9%
Boston	5.8	4.1%	0.4%	82.7%	3.2%	5.5%	4.1%
Dallas	5.2	1.5%	0.1%	92.7%	1.6%	0.1%	4.0%
Houston	4.7	1.6%	0.3%	91.3%	3.1%	0.0%	3.7%
Atlanta	4.1	1.3%	0.1%	90.6%	2.4%	1.1%	4.5%
Miami	3.9	1.8%	0.5%	90.1%	3.2%	0.5%	3.9%
Seattle	3.6	3.2%	0.6%	84.4%	6.2%	0.0%	5.6%
Phoenix	3.3	2.1%	0.9%	90.0%	1.9%	0.0%	5.1%
Minneapolis/ St Paul	3.0	2.4%	0.4%	88.4%	4.4%	0.0%	4.4%
Cleveland	2.9	2.1%	0.2%	91.1%	3.1%	0.3%	3.2%
San Diego	2.8	3.4%	0.6%	86.9%	3.1%	0.2%	5.8%
St Louis	2.6	1.6%	0.1%	92.5%	2.1%	0.2%	3.5%
Denver	2.6	2.4%	0.7%	87.1%	4.2%	0.1%	5.5%
Pittsburgh	2.4	3.6%	0.1%	87.1%	6.0%	0.1%	3.1%
Portland	2.3	3.0%	0.8%	85.2%	5.1%	0.5%	5.4%
Cincinnati	2.0	2.3%	0.1%	91.4%	2.8%	0.0%	3.4%
Sacramento	1.8	2.2%	1.4%	88.9%	2.4%	0.3%	4.8%
Kansas City	1.8	1.4%	0.1%	93.2%	1.2%	0.0%	4.1%
Milwaukee	1.7	2.8%	0.2%	90.0%	3.9%	0.0%	3.1%
Indianapolis	1.6	1.7%	0.2%	93.3%	1.2%	0.0%	3.6%
Orlando	1.6	1.3%	0.4%	92.7%	1.6%	0.0%	4.0%
San Antonio	1.6	2.4%	0.1%	90.9%	2.8%	0.0%	3.8%
Norfolk	1.6	2.7%	0.3%	91.0%	1.7%	0.0%	4.3%
Las Vegas	1.6	2.4%	0.5%	89.5%	3.9%	0.0%	3.7%
Charlotte	1.5	1.2%	0.1%	93.8%	1.3%	0.0%	3.6%

Table 29.3: (continued)

Greater Metropolitan Area	2000 Pop (M)	<u>Mode Share</u>					
		<u>Walk</u>	<u>Bicycle</u>	<u>Drive</u>	<u>Bus</u>	<u>Rail</u>	<u>Other</u>*
New Orleans	1.3	2.7%	0.6%	87.7%	5.2%	0.0%	3.8%
Salt Lake City	1.3	1.8%	0.4%	90.3%	2.7%	0.3%	4.5%
Memphis	1.1	1.3%	0.1%	93.9%	1.6%	0.0%	3.1%
Rochester	1.1	3.5%	0.2%	90.9%	1.9%	0.0%	3.5%
Oklahoma City	1.1	1.7%	0.2%	93.8%	0.5%	0.0%	3.8%
Louisville	1.0	1.7%	0.2%	92.9%	2.2%	0.0%	3.0%

* Includes taxi, ferry, and working at home

for even the largest metropolitan areas, the majority of the home-to-work trips are by private vehicle.

The problem with these data, however, is that they only examine work trips. Nationally, home-to-work trips represent only about 15% of all daily trips (BTS, 2002). On the other hand, 45% of daily trips are for shopping and errands and 27% are social and recreational. Further, non-work trips are even more likely to occur by automobile, and are generally shorter. For example, in Houston, for home-based non-work trips, only 1% of trips were by transit compared to 3.1% for home-to-work trips in 2004. These home-based non-work trips may be a better analogy to crime trips than work trips since they tend to be of similar trips lengths as crime trips.

Thus, unless the user is willing to assume that a crime trip is like a work trip (which is questionable), then the Journey to Work tables are probably not the best guide for the target proportions. Nevertheless, an examination of them is valuable to see how work trips are split among the various travel modes.

Select mode functions

Third, select mathematical functions that approximate accessibility utility. Again, some plausible assumptions need to be made. In *CrimeStat*, the user can select among five different mathematical functions (linear, negative exponential, normal, lognormal, truncated negative exponential). The default functions are shown in Table 29.4 below.

**Table 29.4:
Default Mode Share Functions**

Mode	Function
Walk	Negative exponential
Bicycle	Negative exponential
Driving	Lognormal
Bus	Lognormal
Train	Lognormal

The reasoning behind this is that walking and biking are relatively short trips whereas transit modes involve intermediate length trips. Finally, driving can be used for any length trip other than very short trips (e.g., less than one or two blocks). Thus, it is unlikely that an automobile will be used for very short trips (less than a quarter mile) and it is very unlikely that transit will be used for short trips (less than a half mile or more). Nevertheless, the user can modify these choices and examine the appropriate column in the spreadsheet.

Select model priorities

Fourth, select the priorities for modeling the target. Unfortunately, there may not be a single solution that will yield the target proportions. Therefore, a decision needs to be made on which **order** the spreadsheet will be calculated. The default order is shown Table 29.5.

**Table 29.5:
Default Mode Share Functions**

Mode	Order of Iteration
Walk	1
Bicycle	2
Driving	3
Bus	4
Train	5

The reasoning is that the offender first makes a decision on the length of the trip (short, medium, long, or the equivalent in travel time). Then, within each category, the offender makes a decision on which mode to choose. For very short trips, the default mode is walking. For intermediate to long trips, the default choice is driving. However, the user can change this order.

Iteratively estimate parameters

Fifth, in the spreadsheet, iteratively adjust the parameters until the target proportion is reached. Do this in the order selected in the above step. Again, there is not a single solution that will produce the target proportion. For example, each of the mathematical functions has two or three parameters that can be adjusted:

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 target proportion can be achieved by adjusting any or all of the parameters. For example, to achieve a target proportion of 0.05 (i.e., 5%) using the negative exponential, an infinite number of models can yield this, for example coefficient=0.0366, exponent=-2.63; coefficient=0.0459 or exponent=-5; coefficient=0.01966, exponent=-1; and so forth. Therefore, there must be additional criteria to constrain the choices.

One criterion is to set an approximate mean distance. For example, with walking trips, the mean distance can be set to a half mile or for driving, the mean distance can be set to 6 miles. Then, check the approximate mean distance of the selected function. Though rarely will the exact mean distance be replicated, the calculated mean distance should be close to the ideal. The one exception is for very short trips. Since the intervals in the spreadsheet are a half mile each, there is considerable error for very short distances.

Examine the graphs in the spreadsheet

Another diagnostic tool is to examine the graph of the function in the spreadsheet (below the calculations). Does the typical trip approximate the expected mean distance? Does the selected function produce something that looks intuitive? Admittedly, these are subjective decisions. But, if the function looks strange, it can be caught and re-calculated.

In short, the aim should be to produce a function that not only captures the target proportion, but looks plausible. Several examples are shown below. Figure 29.4 shows the default walking model using a negative exponential. Figure 29.5 shows the default biking model, also using a negative exponential. Figure 29.6 shows the default driving mode using a lognormal function. Figure 29.7 shows the default bus mode, also using a lognormal function, and Figure 29.8 shows the default train mode using a lognormal function.

Figure 29.4:

Negative Exponential Function: Walk Mode

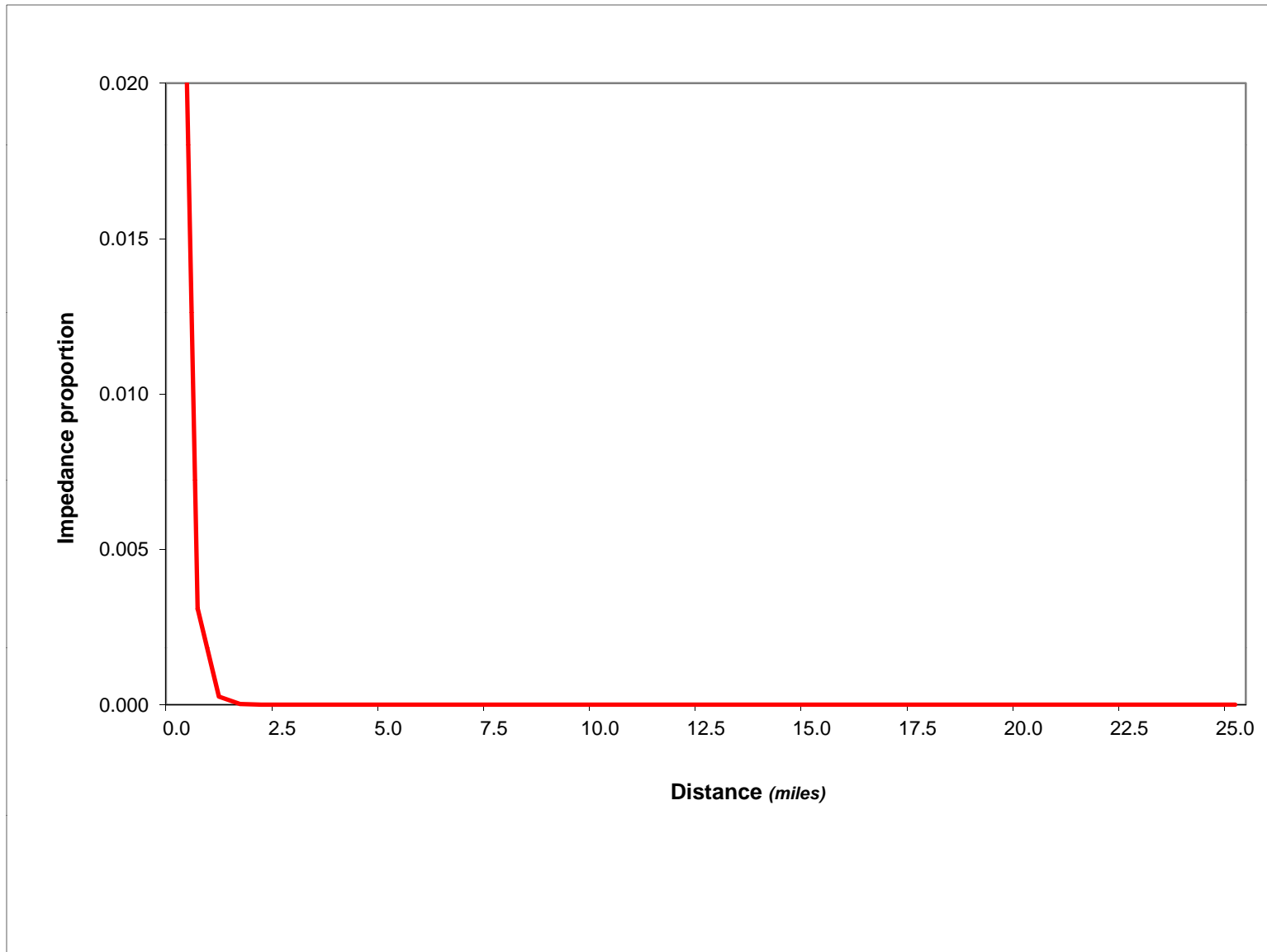


Figure 29.5:

Negative Exponential Function: Bike Mode

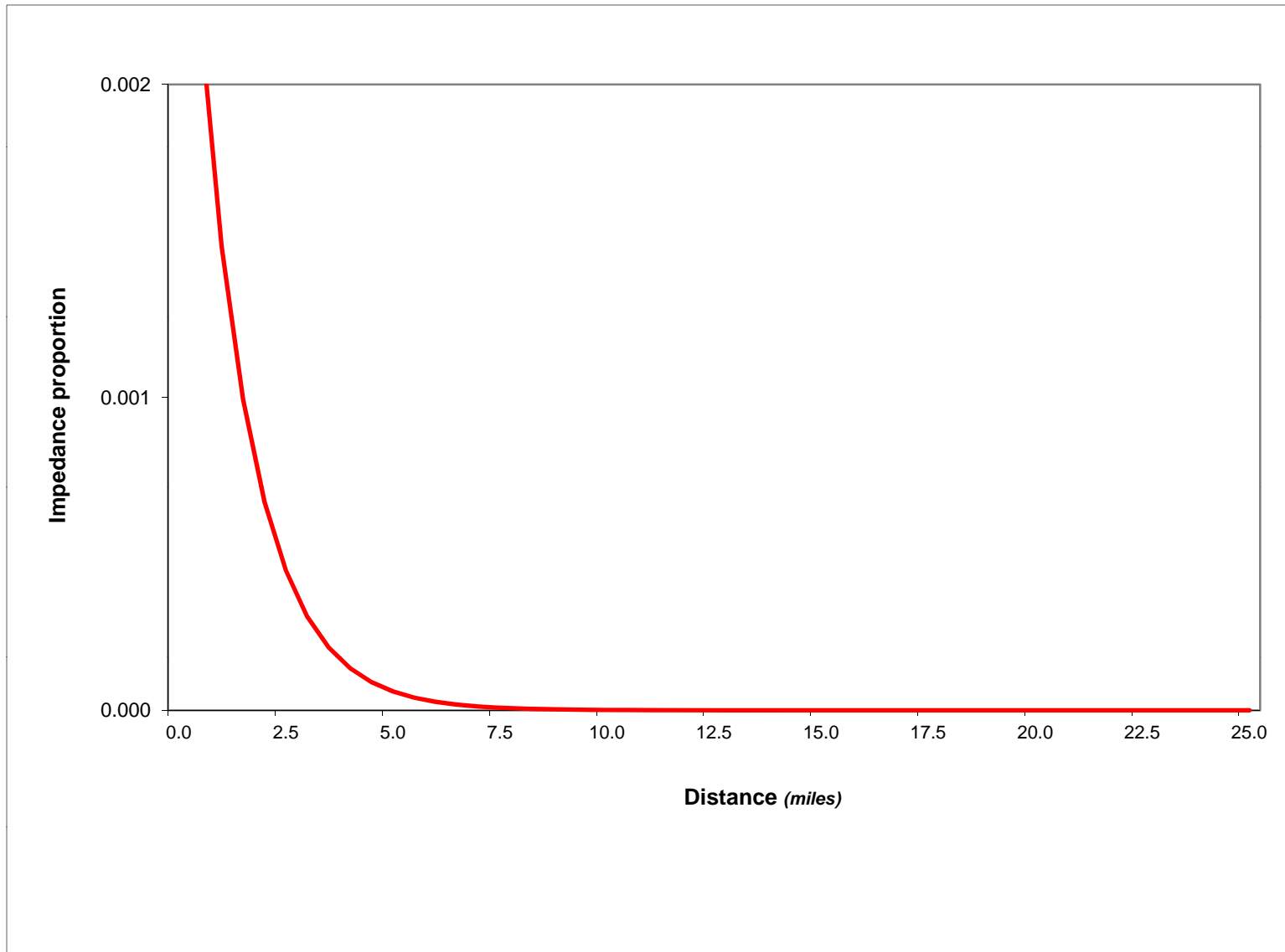


Figure 29.6:

Lognormal Function: Drive Mode

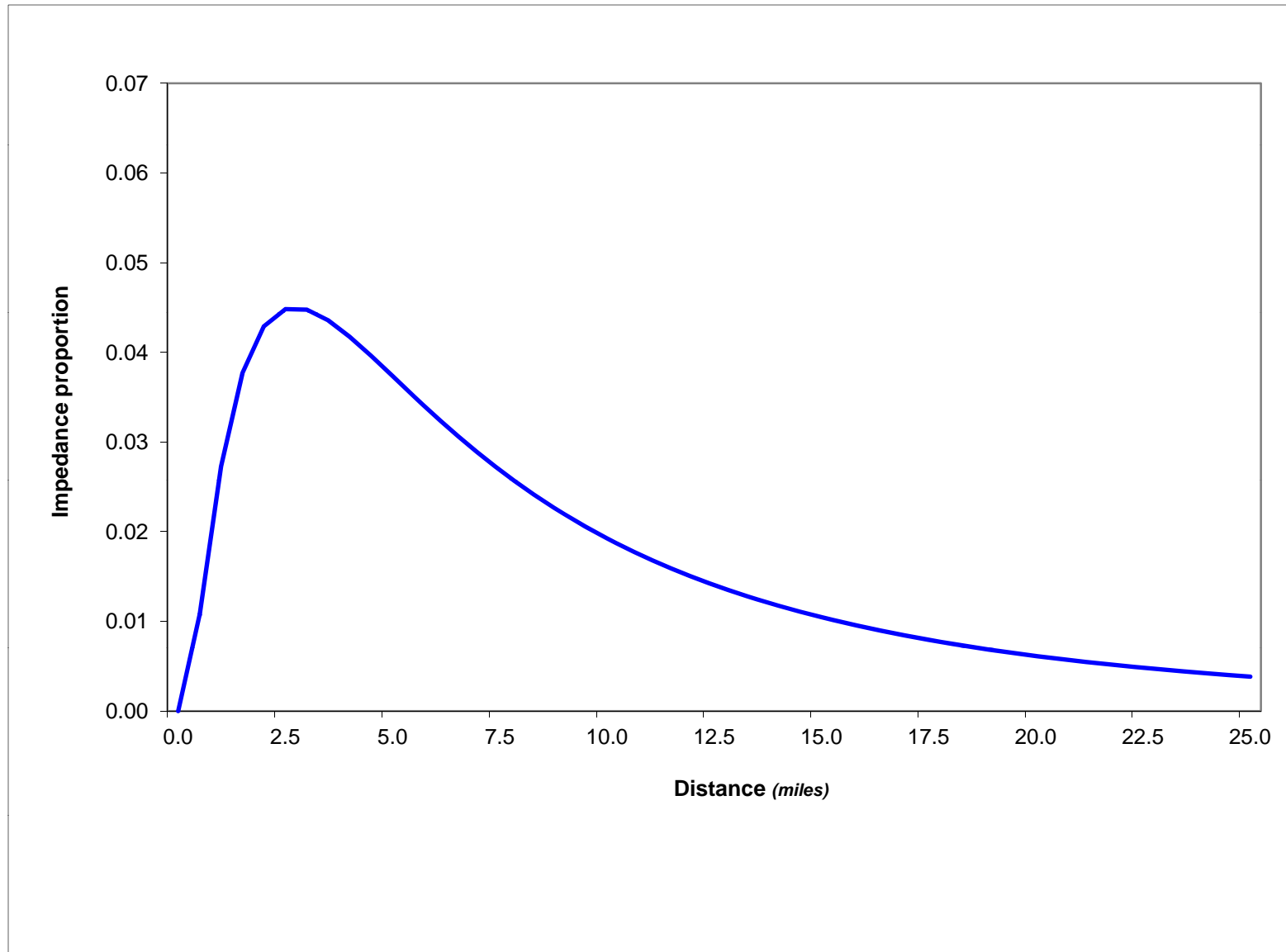


Figure 29.7:

Lognormal Function: Bus Mode

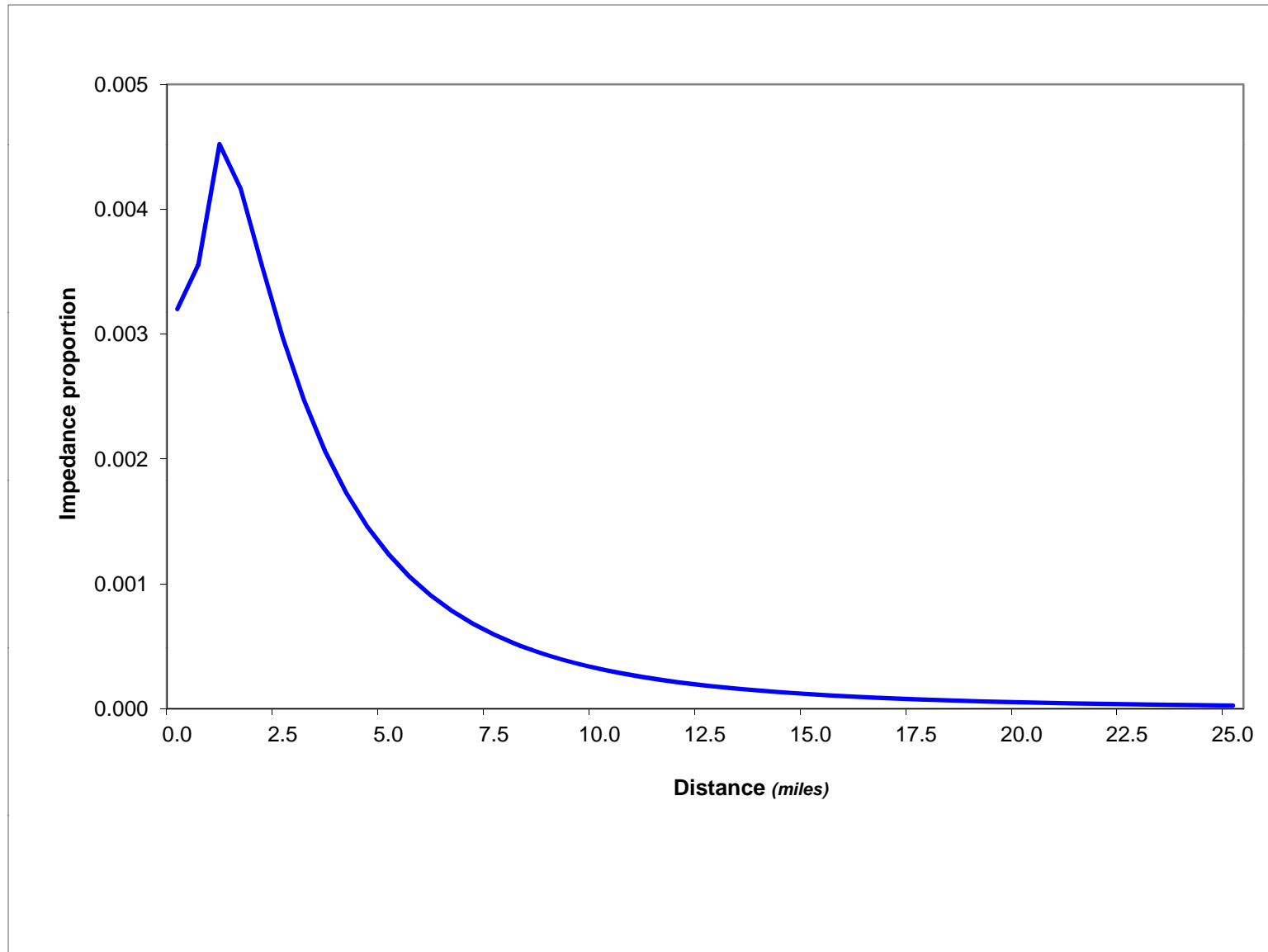


Figure 29.8:

Truncated Negative Exponential Function: Train Mode

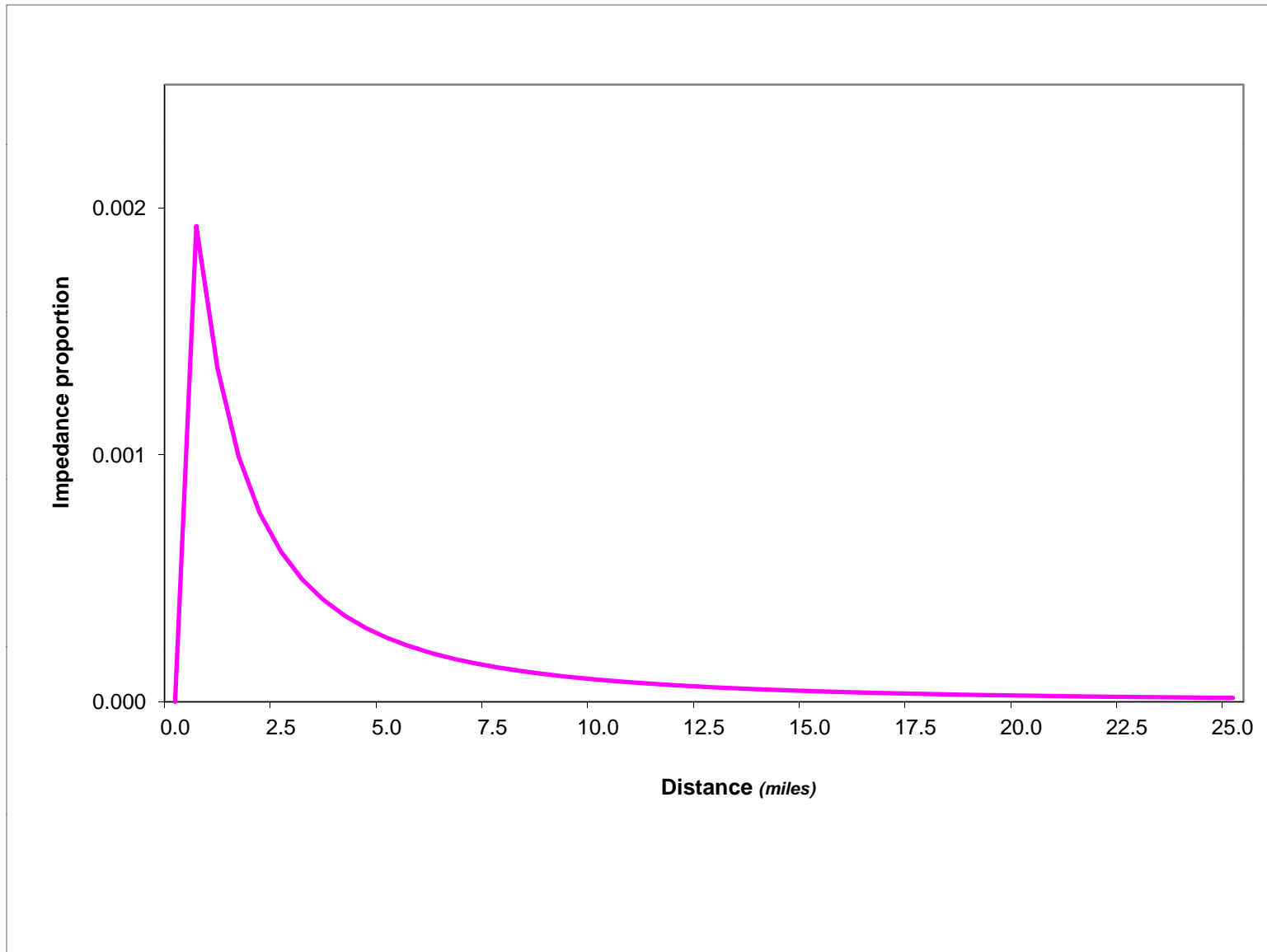


Figure 29.9 shows the cumulative results of the default values. This is also graphed in the spreadsheet, starting in cell I1. Notice how the relative accessibility function works. As distance increases, the mode proportions change. At very short distances, walking trips predominate with biking trips also getting a moderate share. As the distance increases, the proportions increasingly shift toward driving. Even though the likelihood of driving declines with distance, the other modes decline even faster. In other words, the relative accessibility function is estimating the relative shares of each mode as a function of the impedance (in this case, distance). Note also the relative differences in the frequency of trips. Driving trips are far more frequent than any other mode. Thus, compared to the individual graphs (figure 29.4-29.8), the other modes are more muted than driving.

Adapting spreadsheet for travel time or travel cost

The illustrations to this point have used distance as an impedance unit. However, other impedance units, such as travel time and generalized travel cost, can also be used. These generally require a network (see below) in that weights have to be assigned to segments. Nevertheless, the same logic applies. For each travel mode, a specific impedance function is estimated and then applied to the trip distribution matrix.

Empirically Estimating the Mode-specific Impedance

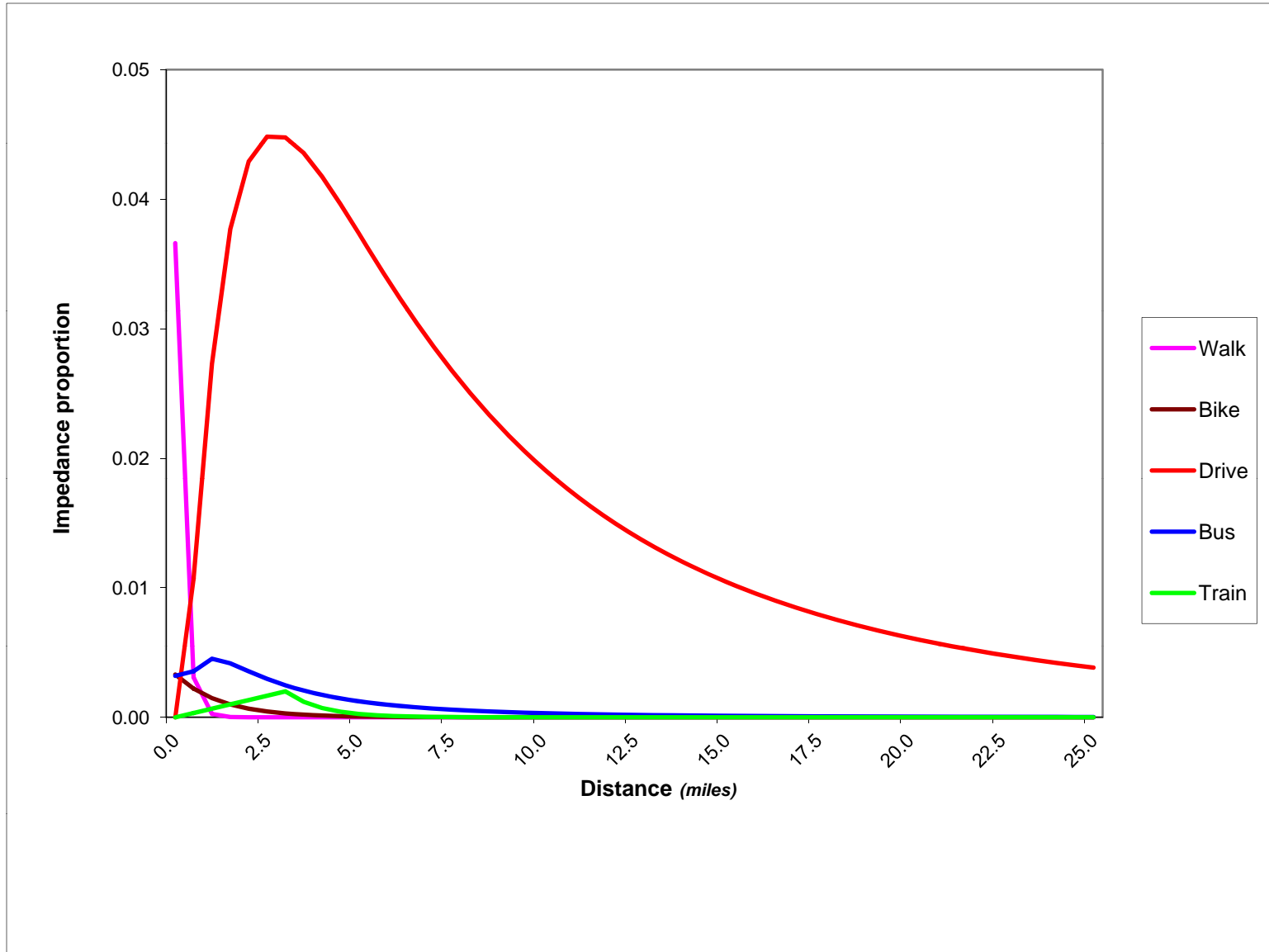
As mentioned at the beginning of this chapter, the lack of information about offender travel modes has necessitated the use of mathematical 'guesses' about travel behavior. However, if it were possible to obtain actual information on travel modes by offenders, then this information could be utilized directly to estimate a much more accurate impedance function. The multinomial logit model in the Discrete Choice module could be used for this purpose (see Chapters 21 and 22). The advantage would be enormous. Instead of guesses about likely impedance functions of specific travel modes, the user would have a function that was based on real data. There should be a substantial improvement in modeling accuracy. However, these data have to be first collected.

CrimeStat IV Mode Split Tools

The *CrimeStat* mode split module allows the relative accessibility function to be calculated. The following provides detailed instructions on running the model. Figure 29.2 above showed the setup page for the mode split routine and Figure 29.3 showed the setup for modes 1 and 2, in the example "Walk" and "Bicycle". The setup for modes 3, 4, and 5 are similar.

Figure 29.9:

Relative Accessibility by Mode



Mode Split Setup

On the mode split setup page, the predicted origin and predicted destination files must be input as the primary and secondary files. If the origin and destination files are identical (i.e., all the origin zones are included in the destination zones), then the file must be input as the primary file.

In addition, the user must input a predicted origin-destination trip file from the trip distribution module. Finally, an assumed impedance value for trips from the “External zone” must be specified. The default is 25 miles. Choose a value that would represent a ‘typical’ trip from outside the study region.

For each mode, the user must provide a label for the name and define the mathematical function which is to be applied and specify the parameters. The first time the routine is opened, the default values are listed. However, the user can change these.

Hint: Once the parameters are entered, they can be saved on the Options page. Then, they can be re-entered by loading the saved parameters file.

Constrain Transit Trips to Network

The impedance will be calculated either directly or is constrained to a network. The default impedance is defined with the type of distance measurement specified on the Measurement Parameters page (under Data setup). On the other hand, if the impedance is to be constrained to a network, then the network has to be defined.

Default

The default impedance is that specified on the Measurement parameters page. If direct distance is the default distance (on the measurement parameters page), then all impedances are calculated as a direct distance. If indirect distance is the default, then all impedances are calculated as indirect (Manhattan) distance. If network distance is the default, then all impedances are calculated using the specified network and its parameters; travel impedance will automatically be constrained to the network under this condition.

Constrain to network

An impedance calculation should be constrained to a network when there are limited choices. For example, a bus trip requires a bus route; if a particular zone is not near an existing bus route, then a direct distance calculation will be misleading since it will probably

underestimate true distance. Similarly, for a train trip, there needs to be an existing train route. Otherwise, the routine will assign transit trips where those are not possible (i.e., it will assign train trips where there are no train stations and it will assign bus trips where there are no bus routes). The routine does not 'know' whether there are transit routes and must be told where they are. Even for walking, bicycling and driving trips, an existing network might produce a more realistic travel impedance than simply assuming a direct travel path.

If the impedance calculation is to be constrained to a network, then the network must be defined. A more extensive discussion of a network is provided in Chapter 3 (under Type of distance measurement on the Measurement Parameters page) and in Chapter 30 in the discussion of the Trip Assignment module. Essentially, a network is a series of connected segments that specify possible routes. Each segment has two end nodes (in *CrimeStat*, they are called 'FromNode' and "ToNode"). Depending on the type of network, the segments can be bi-directional (i.e., travel is allowed in either direction) or single directional (i.e., travel is allowed only from the "FromNode" to the "ToNode").

A critical component of a network for the mode split routine is that travel can only pass through nodes. This means that two segments that are connected can allow a trip to pass over those two segments whereas two segments that are not connected cannot allow a trip to pass directly from one to the other. From outside the network, a trip connects to it at a node. For a transit network, this can be critical. For a bus route, it may or may not be important. A precise bus network defines nodes by bus stops so that a trip can 'enter' or 'leave' the bus system at a real stop. A less precise bus network defines nodes by the ends of segments (e.g., the end nodes of a TIGER segment). The routine will not know whether the node it enters or leaves from is a real bus stop or not. In the case of bus routes, it probably does not matter since they generally make very regular stops (every two or three blocks).

Accurately defined transit networks

For train networks, however, it is absolutely critical that the network be defined accurately. The nodes must be legitimate stations; a trip can only enter or leave the train system through a station (i.e., it cannot enter or leave a train network at the end of an arbitrary segment node). Most travel demand models use very precise bus and train networks that have been carefully checked; where errors occur, the networks are edited and updated.

Utility for creating transit network

If the user does not have an edited transit network, one can be made in the trip assignment module. There is a "Create a transit network from primary file" routine that will draw segments between input primary file points; the user inputs the station or bus stop locations

as the primary file and the routine creates a network from one point to the next in the *same* order as in the primary file (i.e., the primary file needs to be properly sorted in order to travel). See Chapter 30 for more information about creating a transit network.

Entering the network parameters

The network is input by selecting “Constrain to network” and click on the ‘Parameters’ button. A dialogue is brought up that allows the user to specify the network to be used. The network file can be either a shape line or polyline file (the default) or another file, either dBase IV ‘dbf’, Microsoft Excel ‘xls/xlsx’, Microsoft Access ‘mdb’, Ascii ‘dat’, or an ODBC-compliant file. If the file is a shape file, the routine will know the locations of the nodes. All the user needs to do is identify a weighting variable, if used, and possible one way routes (‘flags’). For a dBase IV or other file, the X and Y coordinate variables of the end nodes must be defined. These are called the “From” node and the “End” node, though there is no particular order.

An optional weight variable is allowed for both a shape or dbf file. The routine identifies nodes and segments and finds the shortest path. By default, the shortest path is in terms of distance though each segment can be weighted by travel time, travel speed, or generalized cost; in the latter case, the units are minutes, hours, or unspecified cost units.

Finally, the number of graph segments to be calculated is defined as the network limit. The default is 50,000 segments. This can be changed, but be sure that this number is greater than the number of segments in your network.

Minimum absolute impedance

If a mode is constrained to a network, an additional constraint is needed to ensure realistic allocations of trips. This is the minimum absolute impedance between zones. The default is 2 miles. For any zone pair that is closer together than the minimum specified (in distance, time interval, or cost), no trips will be allocated to that mode. This constraint is to prevent unrealistic transit trips being assigned to intra-zonal trips or trips between nearby zones.

CrimeStat uses three impedance components for a constrained network:

1. The impedance from the origin zone to the nearest node on the network (e.g., nearest rail station);
2. The impedance along the network to the node nearest to the destination; and
3. The impedance from that node to the destination zone.

Since most impedance functions for a mode constrained to a network will have the highest likelihood some distance from the origin, it's possible that the mode would be assigned to, essentially, very short trips (e.g., the distance from an origin zone to a rail network and then back again might be modeled as a high likelihood of a train trip even though such a trip is very unlikely).

For each mode that is constrained to a network, specify the minimum absolute impedance. The units will be the same as that specified by the measurement units. The default is 2 miles. If the units are distance, then trips will only be allocated to those zone pairs that are equal to or greater in distance than the minimum specified. If the units are travel time or speed, then trips will only be allocated to those zone pairs that are farther apart than the distance that would be traveled in that time at 30 miles per hour. If the units are cost, then the routine calculates the average cost per mile along the network and only allocates trips to those zone pairs that are farther apart than the distance that would be traveled at that average cost.

Applying the Relative Accessibility Function

To apply the relative accessibility function, the parameter choices for each mode are entered into the mode split routine. All transit modes are then constrained. Once the mode split setup has been defined and all transit modes have been constrained to a proper network, the mode split routine can be run.

Figure 29.10 shows the top 300 walking crime trips in Baltimore County estimated with the default accessibility functions. As seen, the vast majority of walking trips are intra-zonal (local). There are only a couple of inter-zonal walking trip links shown. The default impedance function assigned approximately 4% of the trips to this mode and the result is many intra-zonal trips.

Figure 29.11 shows the top 300 bicycle crime trips in Baltimore County. There are fewer trips by bicycle and they also tend to be quite local. The impedance function used for bicycle trips allocated approximately 1% of all trips to this mode. Thus, it's less frequent than walking mode. There are proportionately more inter-zonal trips among the top 300 than for walking trips, but these tend to be quite short (travel between adjacent zones).

On the other hand, driving is the predominant travel mode for the crime trips (Figure 29.12). The impedance function used allocated approximately 90% of the trips to driving. Among the top 300 links, there were no intra-zonal driving trips. The use of a lognormal function minimized intra-zonal travel.

Figure 29.10:
Mode Split: Walking Crime Trips
Intra-zonal (local) and Zone-to-zone Trips

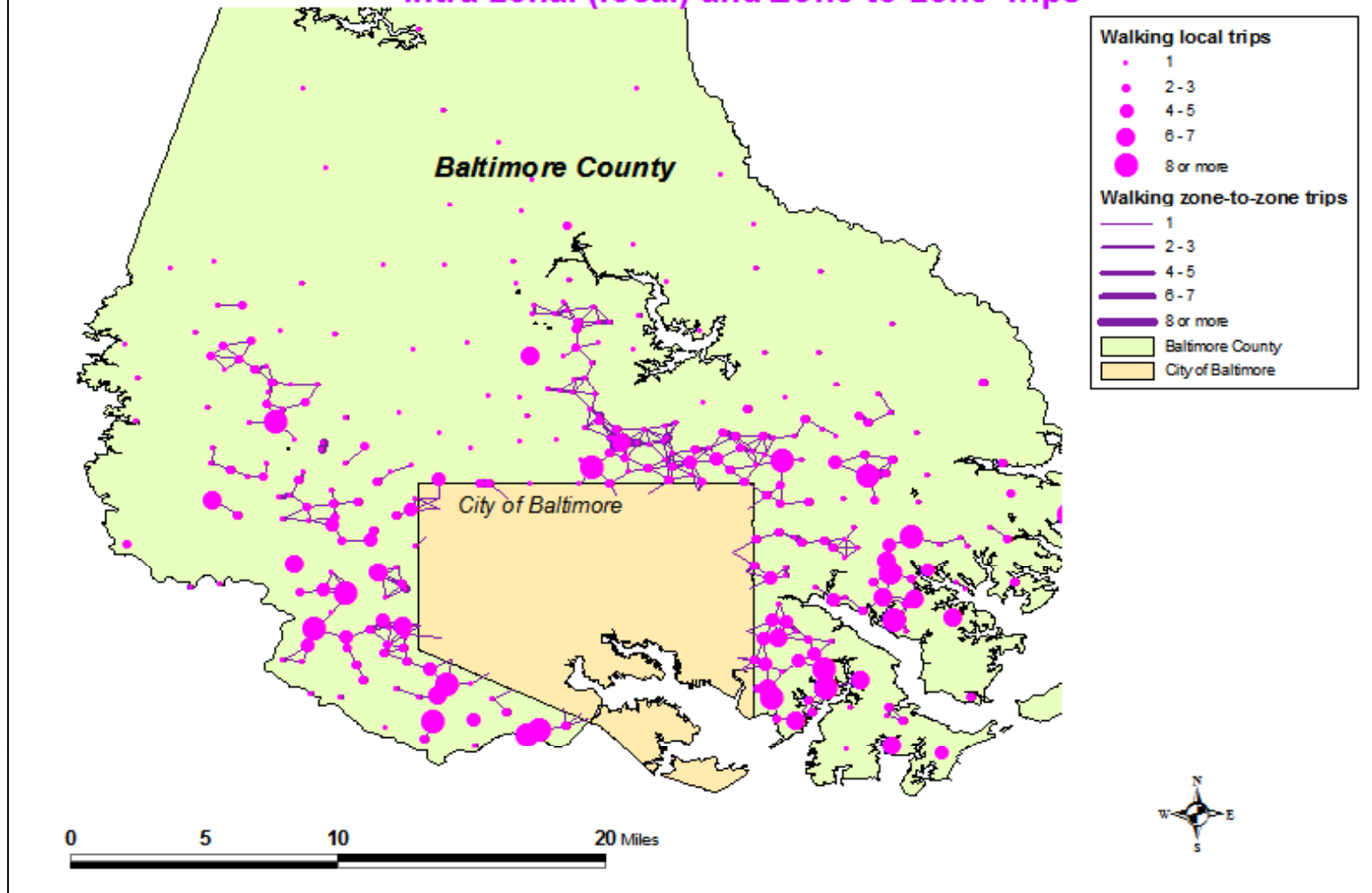


Figure 29.11:
Mode Split: Bicycle Crime Trips
Intra-zonal (local) and Zone-to-zone Trips

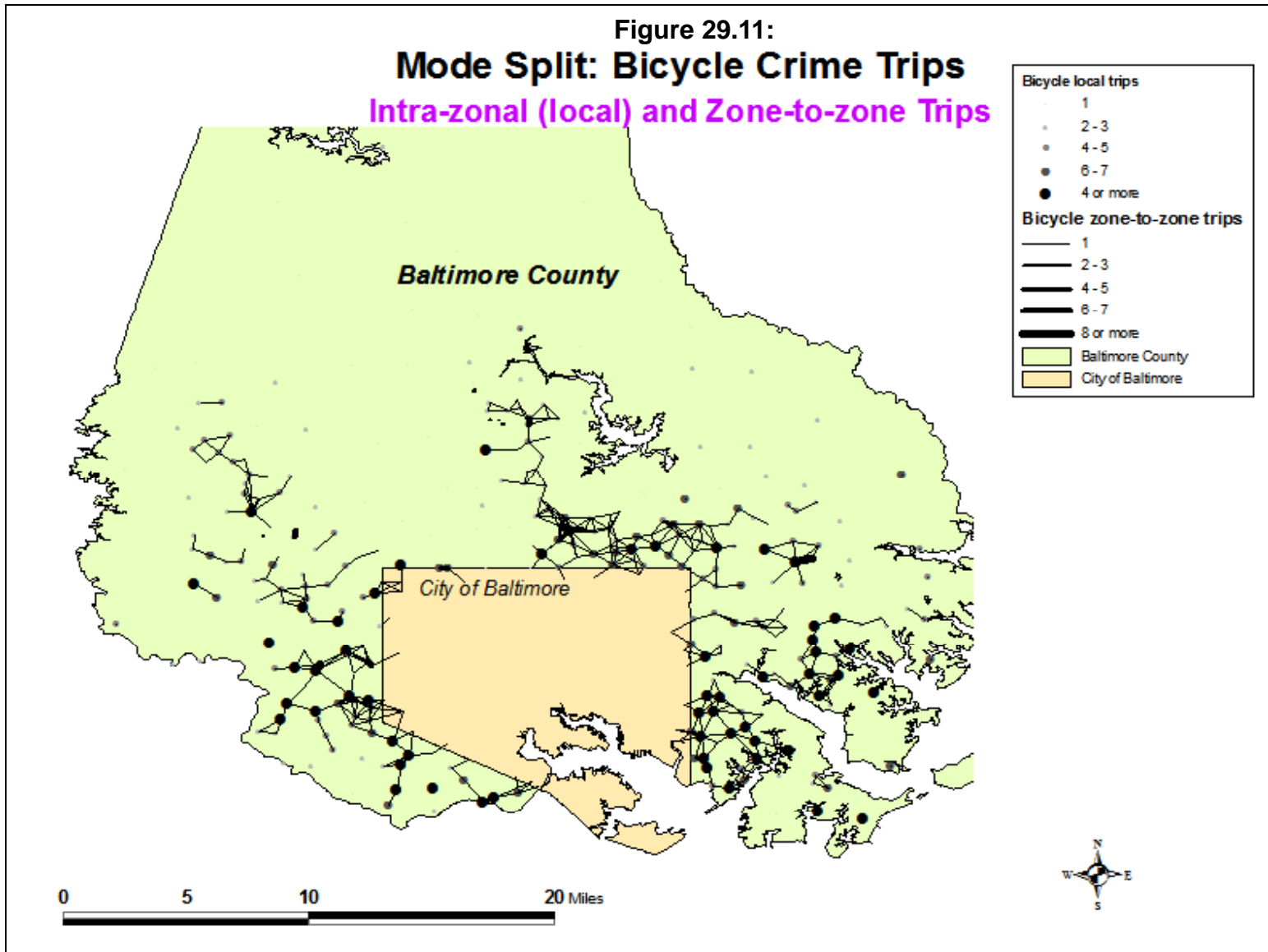
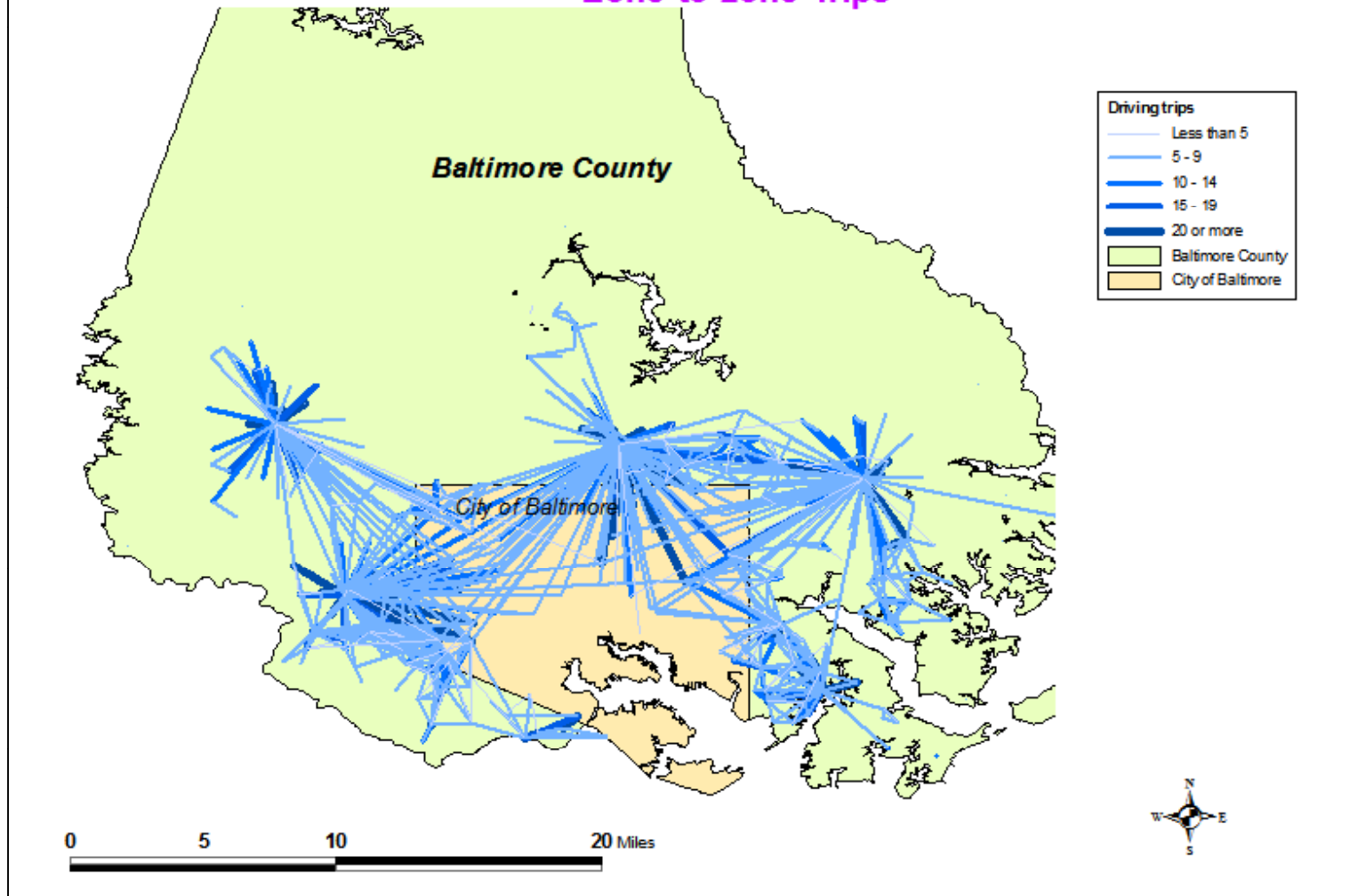


Figure 29.12:
Mode Split: Driving Crime Trips
Zone-to-zone Trips



To allocate bus and train trips, however, it was necessary to constrain them to a network. Separate bus and train networks were obtained from the Baltimore Metropolitan Council. Figure 29.13 shows the Baltimore bus network and Figure 29.14 shows the predicted bus trips superimposed over the bus network. Overall, about 4% of the total trips were allocated to the bus mode by the accessibility function. As seen, the trips tend to be moderate distances and tend to be close to the bus network. Constraining these trips by the network decreased the likelihood that the routine would assign a particular trip link that was far from the bus work to a bus trip.

Finally, train crime trips were constrained to the train network. Figure 29.15 superimposes the assigned train trips over the intra-urban rail network. Overall, only 1% of the total trips were allocated to train mode. Therefore, the number of trips for any zone pair is quite small. The trips are generally longer than the bus trips, as might be expected, and they also tend to fall along the major rail lines. Some of the trips start quite far from the rail lines, so it's possible that these are not realistic representations. Keep in mind that this is a mathematical model and is far from perfect.

Overall, the mode split routine has produced a reasonable approximation to travel modes for crime trips. Since there was no data upon which to calibrate the functions, reasonable guesses were made about the accessibility function. The mathematical model produced a plausible representation of these assumptions, generally fitting into what we know about crime travel patterns.

Usefulness of Mode Split Modeling of Crime Trips

The mode split model is a logical extension of the travel demand framework. For transportation planning, it is an important step in the process. But, it also is important for crime analysis. First, it addresses the complexity of travel by separating the trips from specific origins to specific destinations into distinct modes. In this sense, it adds more realism to our understanding of criminal travel behavior. The Journey-to-crime literature, which has been used by crime analysts and criminal justice researchers to "understand" criminal travel behavior, is simplistic in this respect. It assumes a single mode, though that is rarely articulated by the researchers. By pointing out typical travel distances by offenders circumvents the critical question of how they made the trip. This was, perhaps, not as critical 50-60 years ago when most crimes were committed within a smaller community and it could be assumed that most offenders walked to the crime location. But in the post- World War era, automobile travel has become increasingly dominate. This model assumes that the vast majority of crime trips are taken by automobile. While there is currently no data to prove that assertion, it follows from the transportation patterns that have become widespread in the U.S. and elsewhere.

**Figure 29.13:
Baltimore Bus Network**

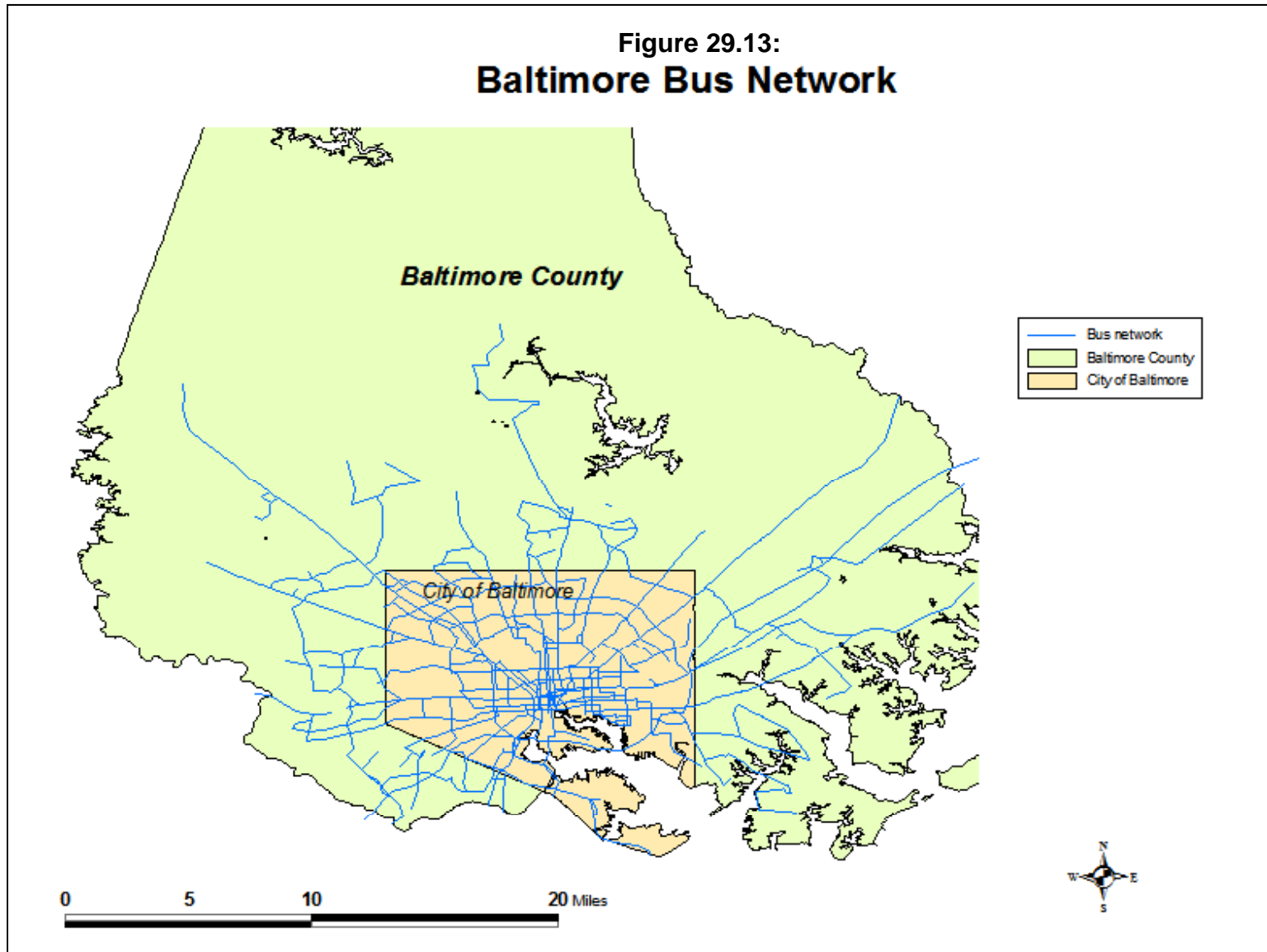


Figure 29.14:
Mode Split: Bus Crime Trips
Zone-to-zone Trips

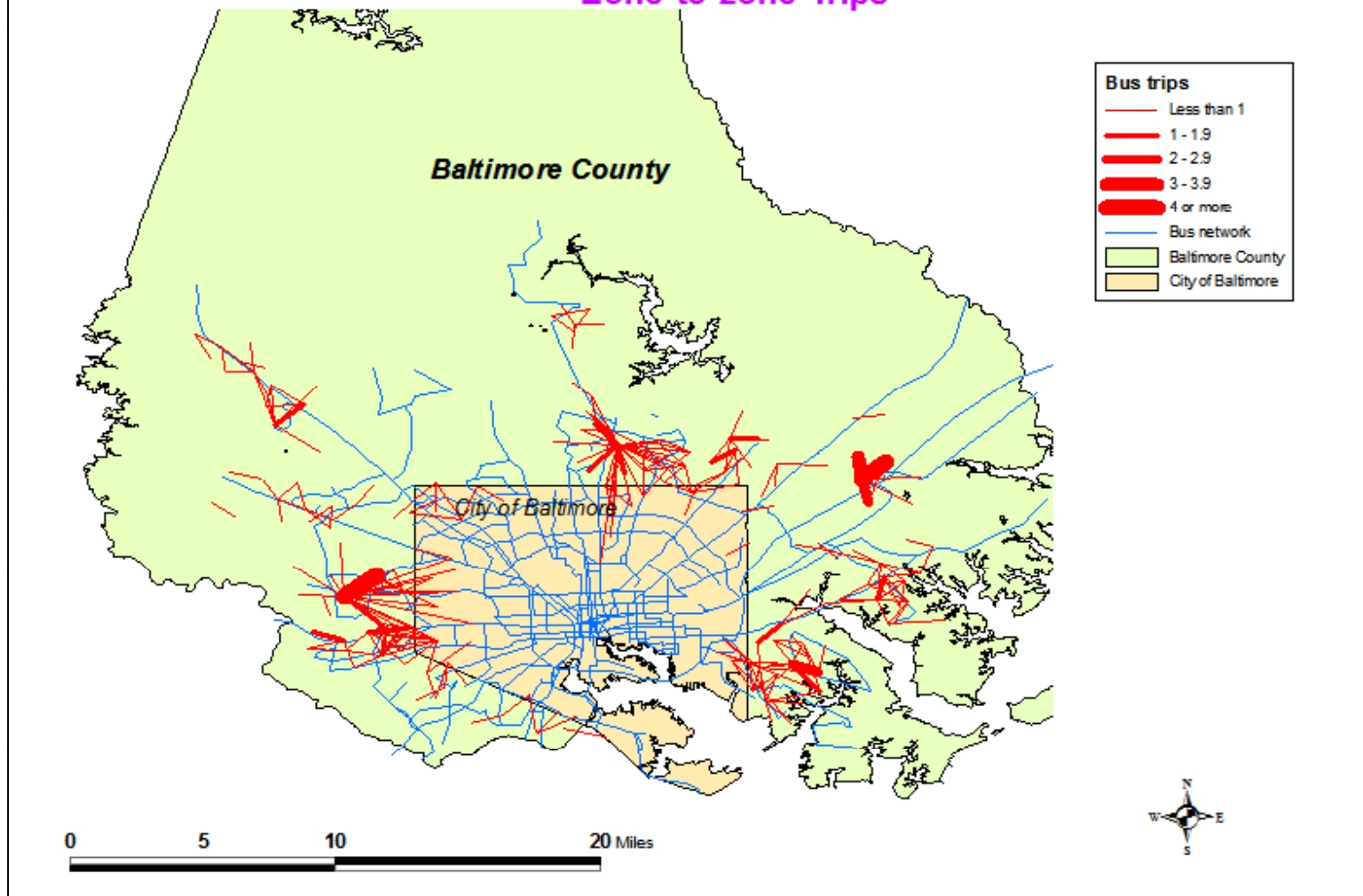
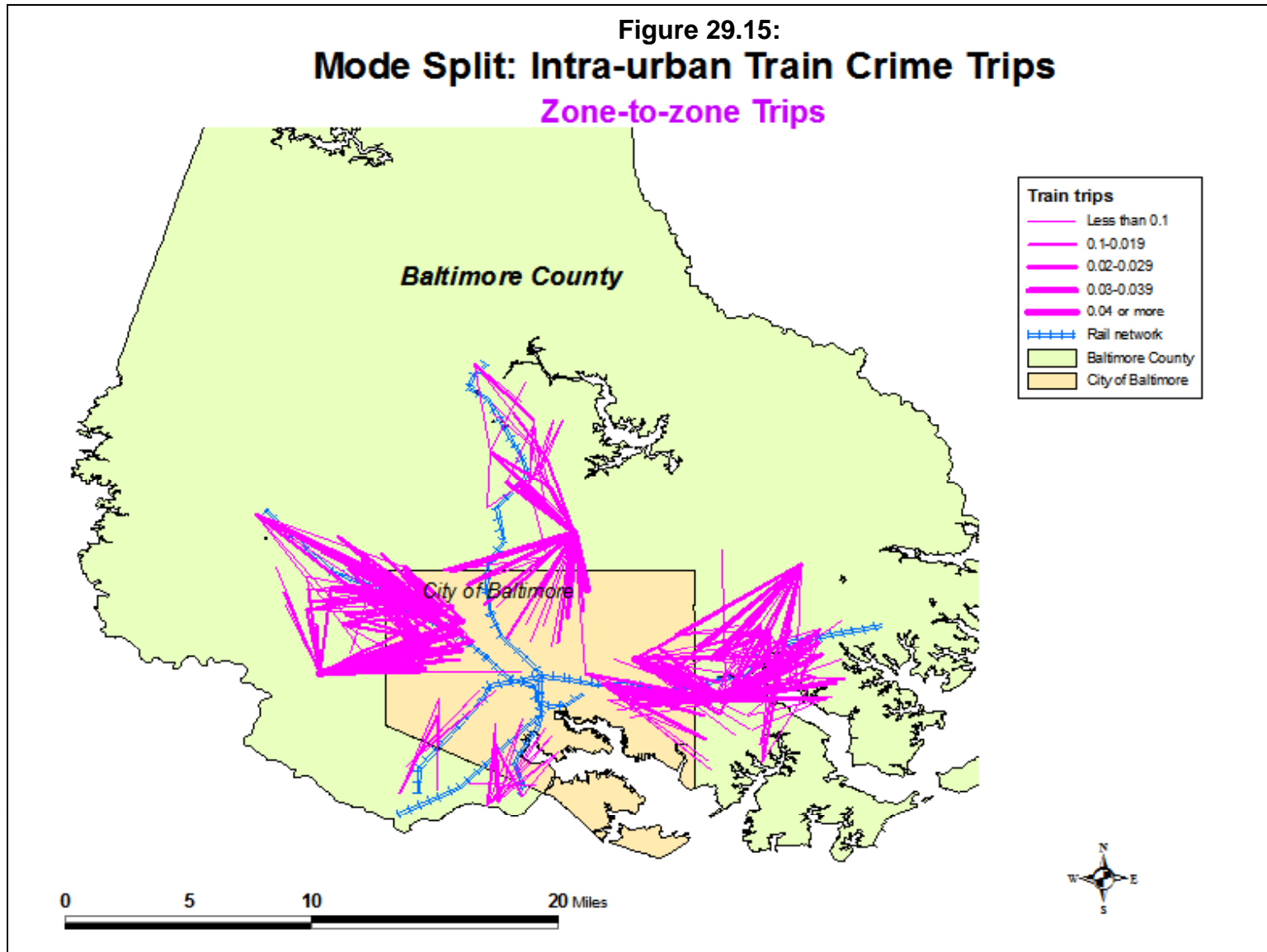


Figure 29.15:
Mode Split: Intra-urban Train Crime Trips
Zone-to-zone Trips



There is a second reason why an analysis of crime travel mode can be important. *If* the limitations of travel mode information could be improved through better and more careful data collection by police and other law enforcement agencies, this type of analysis could be very useful for policing. For one thing, it could allow more focused police deployment. For neighborhoods with a predominance of walking crime trips, then a police foot patrol could be most effective. Conversely, for neighborhoods with a predominance of driving crime trips, then patrol cars are probably the most effective. Police intuitively understand these characteristics, but the crime mode split model makes this more explicit.

For another thing, a mode split analysis of crime can better help crime prevention efforts. As the Baltimore data suggest, many of the local (intra-zonal) crime trips are committed around housing projects and in very low income communities. Most likely, this is a by-product of poverty, lack of local employment opportunities, deteriorated housing, and even poor surveillance. Since teenagers are more likely to *not* own vehicles, it might be expected that the majority of these local crime trips are committed by very young offenders (see Levine & Lee, 2012). This can be useful in crime prevention. Again, “Weed and Seed” and after-school programs are generally targeted to youth from very low income neighborhoods. What is shown by the mode split analysis is probably the crime patterns associated with these neighborhoods. Even though it is intuitively understood, the mode split analysis quantifies these relationships in an explicit manner.

In short, a mode split analysis of crime trips is an important tool for crime analysts and criminal justice researchers. If correctly calibrated, it can help focus police enforcement and crime prevention efforts more specifically and can improve the theory of criminal travel behavior.

Hopefully, police departments will start to improve the quality of data in capturing likely travel modes while taking incident reports. Even though most police departments have an item similar to “Method of departure”, there has not been a lot of emphasis on this information and most crime data sets are deficient on this information. However, with improved data will come more accurate accessibility functions and, hopefully, even real utility functions where actual costs are measured. The expectation is that this will happen and we should work towards accelerating the process.

Limitations to the Mode Split Methodology for Crime Analysis

There are also limitations to the method, particularly the aggregate approach. First, the aggregate approach does not consider individuals, only properties associated with zones (e.g., average travel time between two zones). As mentioned earlier, the accessibility function used

(or the underlying utility theory) is much simpler for zones than for individuals. Consequently, the analysis is cruder at an aggregate level than at an individual level. Policy scenarios are much more limited with aggregate mode split than with individual-level models. For example, if an analyst wanted to explore what was the likely effect of increased public surveillance on walking behavior by pickpockets, it is more difficult to do with aggregate data than with individual data. For example, it could be hypothesized that actual pickpockets are more sensitive to increased public surveillance than, say, car thieves, but this cannot be tested at the aggregate level. Instead, some general characteristics are assigned to all individuals (e.g., the number of security personnel in a zone).

Second, the zonal model for mode split (as with trip distribution) cannot explain intra-zonal travel. For intra-zonal trips, it is inaccurate and generally defaults to simple choices (e.g., walking, biking or driving). For example, bus or train mode can rarely be applied at an intra-zonal level because there are usually too few network segments that traverse a zone and the segments rarely stop within the zone. While this deficiency also applies to the trip distribution model, the dependence on a network for transit modes, particularly, leads to underestimation of transit use for very short trips.

Third, the zonal mode split model cannot explain individual differences. This goes back to the first point that a single utility function is being applied at the zonal level. Thus, the value of time to different individuals living in the same zone cannot be examined; instead, everyone is given the same value.

Fourth, the aggregate mode split model does not analyze time of day very well. The probabilities are assigned to all trips, rather than to trips taken at particular times of the day. To conduct that analysis, an analyst has to break down crimes by time of day and model the different periods separately. Aside from being awkward, the summed trips need to be balanced to ensure that they sum to the total number of trips.

Fifth, and finally, the mode split model, both aggregate and disaggregate, cannot explain *linked trips* (sometimes called *trip chaining*). An offender might leave home one day, go out to eat, visit a friend, commit a street robbery, go to a 'fence' to distribute the goods, buy drugs from a drug dealer, and then finally go home. The mode split model treats each of these as separate trips; in the case of crime mode split, there are three distinct crime trips - committing the robbery, selling the stolen goods to the 'fence', and buying the drugs from the drug dealer. The model does not understand that these are related events, but instead assigns separate mode probabilities to each trip. Thus, it is possible to produce absurd choices, such as driving to the crime scene, taking the bus to the drug dealer, and then biking home. In this respect, the disaggregate approach is equally flawed as the aggregate since both treat each trip as independent events. The solution to this lies in a 'third generation' of travel modeling in which individual trip makers are simulated over a day; *activity-based modeling*, as it is known, is still in

a research stage (FHWA, 2009; Culp & Lee, 2005; Miller & Salvini, 1999). But, it will eventually emerge as the dominant travel demand modeling algorithm.

Conclusions

Nevertheless, mode split modeling can be a very useful analysis step for crime analysis. It represents a new approach for crime analysis and one with many useful possibilities. It will require building more systematic databases in order to document travel modes. But, the possibilities that it offers up can be important for crime analysts and criminal justice researchers alike.

References

- Ben-Akiva, M. & Lerman, S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press: Cambridge.
- Bernasco, W. & Block, R. (2009). Where offenders choose to attack: A discrete choice model of robberies in Chicago. *Criminology* 47(1): 93-130.
- BTS (2002). *National Household Travel Survey: Daily Travel Quick Facts*. Bureau of Transportation Statistics, U.S. Department of Transportation: Washington, DC.
<http://nhts.ornl.gov/download.shtml#2009>. Accessed June 1, 2012.
- Carnegie-Mellon University (1975). *Security of Patrons on Urban Public Transportation Systems*. Transportation Research Institute, Carnegie-Mellon University: Pittsburgh, PA.
- Johnson, M.A. (1978). Attribute importance in multiattribute transportation decisions, *Transportation Research Record*, 673, 15-21.
- Culp, M. & Lee, E. J. (2005). Improving travel models through peer review. *Public Roads*, 68 (6), FHWA-HRT-05-005. Federal Highway Administration, U.S. Department of Transportation: Washington, DC.
<http://www.fhwa.dot.gov/publications/publicroads/05may/07.cfm>. Accessed April 28, 2012.
- 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 (2009). Integrated Urban Systems Modeling, *The Exploratory Advanced Research Program Fact Sheet*, FHWA-HRT-09-042. Federal Highway Administration, U.S. Department of Transportation: Washington, DC.
<http://www.fhwa.dot.gov/advancedresearch/pubs/interurbsys.pdf>. Accessed April 28, 2012.
- Levine, N. & Lee, P. (2009). Bayesian journey to crime modeling of juvenile and adult offenders by gender in Manchester. *Journal of Investigative Psychology & Offender Profiling*. 6(3), 237-251.
- Levine, N. & Wachs, M. (1986). Bus Crime in Los Angeles: II - Victims and Public Impact. *Transportation Research*. 20 (4), 285-293.
- McCormick Rankin (2011). *Transportation Demand Management Plan: Final Report*. Ottawa.
<http://ottawa.ca/cs/groups/content/@webottawa/documents/pdf/mdaw/mdc3/~edisp/cap078202.pdf>. Accessed June 1, 2012.
- McFadden, D. L. (2002). The path to discrete-choice models. *Access*, No. 20, Spring. 20-25.
<http://www.uctc.net/access/access20.shtml>. Accessed April 28, 2012.

References (continued)

- McGuckin, N. A. & Srinivasan, N. (2003). *Journey to Work in the United States and its Major Metropolitan Areas*. FHWA-EP-03-058, Office of Planning, Federal Highway Administration: Washington, DC.
- Miller, E. J. & Salvini, P. A. (1999). Activity-based travel behavior modeling in a microsimulation framework. Paper presented at IATBR Conference, Austin, TX. December. http://www.civ.utoronto.ca/sect/traeng/ilute/downloads/conference_papers/miller-salvini_iatbr-97.pdf. Accessed May 4, 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.
- Ottawa (2008). *Transportation Master Plan*. Regional Municipality of Ottawa-Carleton. http://ottawa.ca/en/city_hall/planningprojectsreports/master_plans/tmp/. Accessed June 1, 2012.
- Porter, C., Suhrbier, J. & Schwartz, W. L. (1999). Forecasting bicycle and pedestrian travel: State of the practice and research needs. *Transportation Research Record*, 1674, 94-101.
- Portland (1998). *Bicycle Master Plan*. Resolution 35515, Office of Transportation, City of Portland: Portland, OR. <http://www.portlandonline.com/transportation/index.cfm?a=369990&c=49304>. Accessed June 1, 2012.
- Roemer, F. & Sinha, K. (1974). Personal security in buses and its effects on ridership in Milwaukee, *Transportation Research Record*, 487, 13-25.
- 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.
- Schwartz, W.L., C. D. Porter, G.C. Payne, J.H. Suhrbier, P.C. Moe, & W.L. Wilkinson III (1999). *Guidebook on Methods to Estimate Non-Motorized Travel: Overview of Methods*. Turner-Fairbanks Highway Research Center, Federal Highway Administration: McLean, VA. July. <http://www.fhwa.dot.gov/publications/research/safety/pedbike/98165/index.cfm>. Accessed June 1, 2012.
- Stopher, P. R. & Meyburg, A. H. (1975). *Urban Transportation Modeling and Planning*. Lexington, MA: Lexington Books.

References (continued)

Turner, S., Shunk, G. & Hottenstein, A. M. (1998). *Development of a Methodology to Estimate Bicycle and Pedestrian Travel Demand*. Report 1723-S, Texas Transportation Institute: College Station. <http://tti.tamu.edu/publications/catalog/record/?id=146>. Accessed April 28, 2012.

Train, K. (2009). *Discrete Choice Methods with Simulation* (2nd edition). Cambridge University Press: Cambridge.

U.S. Census Bureau (2009). *Commuting (Journey to Work) Main*. U.S. Census Bureau, U.S. Department of Commerce: Washington, DC. <http://www.census.gov/hhes/commuting/>. Accessed June 1, 2012.

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