Network-Based Highway Crash Prediction Using Geographic Information Systems

Dr. John N. Ivan, PI
Dr. Per E. Garder, Co PI
Sumit Bindra, Graduate Assistant
B. Thomas Jonsson, Post-Doctoral Fellow
Hyeon-Shic Shin, Post-Doctoral Fellow
Zuxuan Deng, Graduate Assistant

Prepared for
The New England Transportation Consortium
June 6, 2007
NETCR67 **Project No. 04-5**

ACKNOWLEDGMENTS

The following are the members of the Technical Committee that developed the scope of work for the project and provided technical oversight throughout the course of the research:

Erika B. Smith, Connecticut Department of Transportation, Chairperson
Neil Boudreau, Massachusetts Highway Department
Barbara Breslin, Federal Highway Administration, Connecticut Division
Steve P. Dubois, New Hampshire Department of Transportation
Brian Marquis, Maine Department of Transportation
Paul C. Petsching, Rhode Island Department of Transportation

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NETCR67	N/A	N/A

4. Title and Subtitle	5. Report Date
Network-Based Highway Crash Prediction Using Geographic Information Systems	June 6, 2007
	6. Performing Organization Code: N/A

7. Author(s)	8. Performing Organization Report No.
Dr. John N. Ivan, PI Dr. Per E. Garder, Co PI Sumit Bindra, Graduate Assistant B. Thomas Jonsson, Post-Doctoral Fellow Hyeon-Shic Shin, Post-Doctoral Fellow Zuxuan Deng, Graduate Assistant	NETCR67

9. Performing Organization Name and Address	10 Work Unit No. (TRAIS)
Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT 06269	N/A
	11. Contract or Grant No.: N/A
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
New England Transportation Consortium C/O Advanced Technology & Manufacturing Center University of Massachusetts Dartmouth 151 Martine Street Fall River, MA 02723	FINAL REPORT
	14. Sponsoring Agency Code: NETC 04-5 A study conducted in cooperation with the U.S. DOT

15 Supplementary Notes: N/A

16. Abstract

The objectives of this project were to estimate network-based crash prediction models that will predict the expected crash experience in any given geographic area as a function of the highway link, intersection and land use features observed in the area. The result is a system of GIS programs that permit a polygon to be drawn on a map, or a set of links and intersections to be selected, and then predict the number of crashes expected to occur on the selected traffic facilities. These expected values can then be compared with observed values to identify locations with higher than usual crash incidence and may require attention to improve the safety of the location. Alternatively, this tool could be used to estimate the safety impacts of proposed changes in highway facilities or in different land development scenarios.

A network approach was chosen to solve this problem, in which separate models were estimated for crashes at major intersections, and intersection-related and segment-related crashes on road segments. All three sets of models can then be used to predict the number of crashes for an entire highway facility delineated as the user desires – including all intersections. These models also consider all relevant road features, in particular the intensity of traffic at intersections and driveways resulting from the surrounding land use. Gathering traffic volumes at every intersection and driveway on the road network would preclude the feasibility of such an approach, both for estimation and in practice. Instead, the link between land development and trip generation was exploited to estimate the driveway and minor road volumes. Land development intensity variables were generated from land use inventories organized using Geographic Information Systems (GIS), permitting virtually automatic preparation of the required data sets for model estimation and application and prediction of crash counts on roads. Specifically, population and retail and non-retail employment counts were associated with each analysis segment to represent vehicle exposure to intersection-related crashes.

GIS was used for two purposes in this project: 1) distributing population and employment counts in a traffic analysis zone (TAZ) among all the links in that zone. 2) Visually comparing the predicted and observed accident counts in order to identify higher than usual crash locations.

17. Key Words	18. Distribution Statement
GIS, accidents, safety, land use, demographics, prediction, blackspots	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	82	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimetres	mm
ft	feet	0.305	metres	m
yd	yards	0.914	metres	m
mi	miles	1.61	kilometres	km
		AREA		
in^2	square inches	645.2	millimetres squared	mm^2
ft^2	square feet	0.093	metres squared	m^2
yd^2	square yards	0.836	metres squared	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	kilometres squared	km^2
		VOLUME		
fl oz	fluid ounces	29.57	mililitres	mL
gal	gallons	3.785	Litres	L
ft^3	cubic feet	0.028	metres cubed	m^3
yd^3	cubic yards	0.765	metres cubed	m^3

NOTE: Volumes greater than 1000 L shall be shown in m^3

		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5(F-32)/9	Celcius temperature	°C

* SI is the symbol for the International System of Measurement

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimetres	0.039	inches	in
m	metres	3.28	feet	ft
m	metres	1.09	yards	yd
km	kilometres	0.621	miles	mi
		AREA		
mm^2	millimetres squared	0.0016	square inches	in^2
m^2	metres squared	10.764	square feet	ft^2
ha	hectares	2.47	acres	ac
km^2	kilometres squared	0.386	square miles	mi^2
		VOLUME		
mL	millilitres	0.034	fluid ounces	fl oz
L	litres	0.264	gallons	gal
m^3	metres cubed	35.315	cubic feet	ft^3
m^3	metres cubed	1.308	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams	1.102	short tons (2000 lb)	T
		TEMPERATURE (exact)		
°C	Celcius temperature	1.8C+32	Fahrenheit temperature	°F

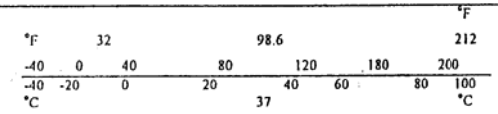


Table of Contents

List of Figures

List of Tables

Executive Summary

The objectives of this project were to estimate network-based crash prediction models that will predict the expected crash experience in any given geographic area as a function of the highway link, intersection and land use features observed in the area. The result is a system of GIS programs that permit a polygon to be drawn on a map, or a set of links and intersections to be selected, and then predict the number of crashes expected to occur on the selected traffic facilities. These expected values can then be compared with observed values to identify locations that are particularly dangerous and require attention for improving safety. Alternatively, this tool could be used to estimate the safety impacts of proposed changes in highway facilities or in different land development scenarios.

A network approach was chosen to solve this problem, in which separate models were estimated for crashes at major intersections, and intersection-related and segment-related crashes on road segments. All three sets of models can then be used to predict the number of crashes for an entire highway facility delineated as the user desires – including all intersections. These models also consider all relevant road features, in particular the intensity of traffic at intersections and driveways resulting from the surrounding land use. Gathering traffic volumes at every intersection and driveway on the road network would preclude the feasibility of such an approach, both for estimation and in practice. Instead, the link between land development and trip generation was exploited to estimate the driveway and minor road volumes. Land development intensity variables were generated from land use inventories organized using Geographic Information Systems (GIS), permitting virtually automatic preparation of the required data sets for model estimation and application and prediction of crash counts on roads. Specifically, population and retail and non-retail employment counts were associated with each analysis segment to represent vehicle exposure to intersection-related crashes.

The most critical data element is the GIS inventories. GIS data were acquired from two different sources. The first is the Capital Region Council of Governments (CRCOG), the metropolitan planning organization for the 29-town region surrounding Hartford, Connecticut. The CRCOG GIS system supports the agency's travel demand forecasting system, and includes population and employment (retail and non-retail) data at the traffic analysis zone level. Data were available for 1122 zones; data were acquired from 902 of these zones and the links adjacent or running through them. The second is the GIS system maintained by Maine Department of Transportation (DOT), which covers the entire State, and includes population and employment in over a dozen categories for each zone. The Maine GIS network is not as detailed as the CRCOG data set, but it was still sufficient for the research.

The population and employment in each network zone had to be allocated to the links representing the roads onto which the traffic generated by it enters and exits the road network. This required identifying the major roads – defined as state-maintained arterials and major collectors – for which traffic volumes were available. The resulting road links were identified as being adjacent to two zones or inside one zone; many links needed to be split in order to do this consistently. Then the population and employment in each zone was allocated among the links adjacent to or inside it in proportion to the length of each link. In cases where this would appear to result in an inappropriate allocation, this allocation was done manually; this was only necessary in 53 out of 901 cases. Finally, the population and employment values allocated to each link were multiplied by trip generation rates published by the Institute of Transportation Engineers to estimate the trips entering and leaving each road segment due to the adjacent land use. The total of this trip generation for each link gave an estimate of intersecting traffic volumes along the link.

Other data items that were required included road characteristics, traffic volumes and crash counts by collision type. The road and traffic data were included in the GIS system for Maine, and the crash data were provided separately. For the CRCOG data, some roadway data were included in the GIS network, including facility type, running speed and number of lanes. Other needed road characteristics were acquired from various sources at the Connecticut Department of Transportation (ConnDOT), including the photo log or the highway inventory. Traffic volumes for the CRCOG links were also acquired from ConnDOT, along with records of accidents occurring on the segments and at the intersections studied.

Crash prediction models were estimated for the following contexts:

1. Major intersection accidents (at intersections of arterials and major collectors)
    a. Three-leg intersections

    b. Four-leg intersections
2. Segment-intersection accidents (along arterials and major collectors at minor collectors, local roads and private access points)
    a. On two-lane rural roads
    b. On two-lane urban or suburban roads
    c. On four-lane undivided roads
3. Segment-related accidents (along arterials and major collectors, not related to minor intersections)
    a. On two-lane rural roads
    b. On two-lane urban or suburban roads
    c. On four-lane undivided roads

Because they were not well-represented in the database, no models were estimated for four-lane divided roads, or roads with 3 lanes, or with two-way left turn lanes. The estimated coefficients and model forms are given in Chapter 5.

Some notable findings in the prediction model results are the following:

- Crash risk is higher with large pavement and shoulder widths on two-lane road segments. This is in contrast with the conventional belief that wider pavements provide more room for preventive maneuvers and thus reduce accidents.
- Population and employment variables were intended to represent exposure to segment-intersection collisions, but they turned out to be significant for predicting segment-related collisions as well.
- Segment length is logically used to represent exposure to segment-related collisions, in that each mile a vehicle drives along a segment increases the accident risk. This normally indicates a simple linear relationship between accidents and segment length. The estimated models, however, indicate a non-linear relationship, so that shorter segments have higher crash risk than longer segments.

These unexpected and difficult to explain results raise issues with the transferability of the segment-related crash models with population and employment variables to other areas. Transferability was also tested by using the models estimated on the CRCOG data to predict crashes in the Maine data. The results of this test show significant but weak correlation, probably because the ranges of key variables such as AADT are different in the two data sets. It is also possible that the trip generation rates in Maine sway more from the national average than Connecticut. Consequently, models for segment-related collisions without the population and employment variables are also provided, and it is suggested that these latter models should be used for more reliable prediction. Similar issues are associated with the models having an exponent on length different from 1.0. These models should be used only when the links are defined as connectors between major intersections as in the CRCOG area.

Further research is needed to estimate models for additional types of roads that may exist in any State road network but for which models could not be estimated here. Our estimation database with all covariates did not include freeways, three-lane roads, roads with two way left turn lanes (TWLTL), four-lane divided roads, or roads with more than two lanes in each direction. Some of these road types were intentionally excluded (e.g., freeways), but the others were not represented in the analysis road network in sufficient numbers to estimate prediction models. Consequently, we were not able to evaluate and comment on the safety of these types of roads. If a network with enough of these missing roads (and the required covariates) is available as a GIS layer, the procedure outlined in Chapter 3 and Appendix A can be used to assign population and employment data from geographic zones to road links and prepare a data set to estimate accident prediction models for these types of roads.

Background

It is hardly necessary to make a case for funding research into highway safety, with 40-45,000 deaths on US highways every year (NHTSA 2004). In response, learning how to predict the expected number of motor vehicle collisions on a road has been the subject of research for almost half a century, with the preferred analysis method evolving over time. At first, linear models were estimated using least squares regression, until it was realized that the assumption required for such models (that the error term follows a normal distribution) is routinely violated on roads with relatively low collision counts (Jovanis and Chang 1986). This realization led to the use of Poisson regression for estimating such models, until it was shown that even a key assumption required for this approach is violated, namely, that the variance is equal to the mean when comparing observed counts at many locations with low expected number of crashes. Now, the preferred estimation method is to use negative binomial regression, which permits estimation of a dispersion parameter that accounts for the violation of this assumption (Miaou and Lum 1993; Hauer 2001).

Another aspect of highway crashes that has been known for some time, but only recently addressed in crash modeling, is the non-linear relationship between crashes and volume. Hauer (1995) was one of the first to draw attention to this issue, and introduced the concept of the *safety performance function* (SPF) for intersections, suggesting the form $N=\alpha V^{\beta}$, where N is the predicted number of crashes, V is the traffic volume, usually the number of entering vehicles per year in millions, and α and β are parameters to be estimated, with α being the normalized crash risk for the location and β being a positive value, usually less than one.

This approach has become the standard for crash prediction modeling (Persaud and Mucsi 1995; Mensah and Hauer 1998; Lord 2002; Qin *et al.* 2003). However, for modeling crashes on highway segments, the length of the segment also needs to be included in the model to properly control for the number of opportunities for a crash to occur on the segment. In other words, it is reasonable to consider intersections to be point locations, but not highway sections. Although it is universally accepted in the crash prediction community that the section length must be included in the crash formula for road segments, whether or not it should take an exponent is the subject of disagreement. The problem is that a non-linear relationship between the number of crashes and the segment length results in the predicted number of crashes for a highway section depending on the criteria used to divide the section into smaller segments. For example, if the exponent on length is less than 1.0, if one divided a segment into sub-segments, one would predict a greater number of crashes. Hence, the exponent estimated for segment length depends on the criteria used to divide the road network from which the estimation data were gathered into segments.

Nevertheless, a recent study found models of the number of crashes on rural highway segments that estimated an exponent parameter for the segment length to perform better than models that did not (Qin *et al.* 2003). This study estimated models for four different crash types: single-vehicle, opposite-direction, same-direction and intersecting-direction. Data sets were collected from the Highway Safety Information System (HSIS), a database maintained under contract for Federal Highway Administration (FHWA) that contains highway characteristic and crash data from eight States spread across the country. Participating states are included on the basis of adherence to clear data variable definition and collection standards, an important one of which is that highway segments are delineated by major intersections. What this means is that segment lengths in the HSIS data are not arbitrary, but in fact are likely inversely proportional to the level of land development in the area – that is, in densely developed areas, there are likely to be more major highway intersections than in areas that are less densely developed. This study found the exponent on segment length to vary across the four types of crashes, but only significantly between two groups: for single-vehicle and opposite-direction crashes the exponent was approximately 1.0, and for same-direction and intersecting-direction crashes the exponent was less than 1.0. What is different between these two groups of crash types is that the latter are more related to interactions between vehicles entering and leaving the roadway via minor road intersections and driveways. This is entirely consistent with the notion that longer segments are related to lower land densities, as the higher the land density, the greater the minor intersection and driveway traffic volumes that would be expected.

This finding would appear to validate the common practice alluded to above, that of predicting crash experience separately for intersections and the segments between them. This is quite logical, since angle collisions can only occur at locations where vehicles approach one another from intersecting paths, and collisions involving turning vehicles can also only occur where vehicles turn on or off the roadway. Both of these situations are only possible at intersections or driveways. In contrast, run-off-road and head-on collisions can occur anywhere, but due to their nature are generally identified as occurring on a segment rather than at an intersection.

The attractive logic behind this approach (of estimating models separately for intersections and segments) breaks down when one attempts to define what an intersection is. The current practice only includes intersections between major roads, largely because these are the facilities for which traffic and other road information is available, but also because minor road intersections are far too numerous, and using them to delineate road segments would result in an enormous information gathering burden. Unfortunately, major road intersections with minor roads and driveways introduce the same kinds of road safety risks as do intersections with other major roads. Defining the crashes resulting from these risks in the same pool as single-vehicle and head-on collisions confounds the modeling process, especially when the segment characteristics provide no information about the minor road intersections and driveways that contribute to the unique crash risk experienced on a particular road.

Consequently, a network approach is proposed for solving this problem, which rather than estimating different models for intersections and segments, instead estimates models by crash type, such as intersection-related and segment-related. Such models would predict the number of crashes for an entire highway facility delineated as the user desires – including all intersections – and consider all relevant road features, in particular, the intensity of traffic at intersections and driveways resulting from the surrounding land development. Estimating such models would require gathering a great deal of qualitative information about every mile of every road to be studied, making the cost to this point prohibitive for such a study.

If traffic volumes at every intersection and driveway on the road network were required for this approach, even estimating such models would be impossible from a financial feasibility standpoint, not to mention applying them in practice. However, it may be possible to exploit the link between land development and trip generation, if not to estimate these volumes, then at least to represent them as a surrogate. Thanks to the spread of electronic mapping and land use inventories organized using Geographic Information Systems (GIS), land development information is now available, not only for preparing data sets for model estimation, but also for application and prediction of crash rates. This project uses GIS land use inventories to generate land development intensity variables associated with specific highway links for estimating crash prediction models using the network approach described above.

Objectives

As stated in the discussion above, one of the objectives of this study was to develop network-based models where accidents are divided into subgroups by occurrence location (relative to major intersections) and then by crash type. Since traffic volumes and other important information for minor roads are not readily available, the accidents which occur further away from a major intersection (>250 feet) are divided into two categories: segment-intersection and segment-related crashes based on the crash type. The crash types which belong to these two categories and the model forms for all three categories are described in Chapter 2.

The main objectives of this study are to demonstrate that:

1. Land development (by type) in the areas surrounding the links can act as a surrogate for exposure to intersection-related collisions in lieu of traffic volume and other information about the minor roads and driveways.
2. In many models, the number of different driveways has been used to represent the land use density in the neighborhood. This approach is expensive and labor intensive. This report demonstrates that, instead, it is possible to use actual land use variables such as population, retail, and non-retail employment from census data and other sources and that these models work as well, if not better, as models using driveways.

In order to achieve this, a procedure linking land use (population and employment) records to the link database was developed and is described in Chapter 3, with supporting data in Appendix A. While these variables (population and employment) are included primarily to represent exposure to segment-intersection collisions, it is possible that they might also help in explaining segment-related accidents. Thus, segment-related crash models with land use as an exposure were also tested. Statistical models using

the Negative Binomial distribution for accident frequency were estimated for all three crash categories: intersection, segment-intersection, and segment-related. The final models proposed are given in the body of the report; the full set of all models estimated and considered are given in Appendix B. An Accident Model User Interface (AMUI) was developed to visually compare predicted and observed accidents on a GIS network. It was programmed using ArcObjects with Visual Basic for Application (VBA). A detailed user manual for this application is presented in Appendix C. The remainder of the report is outlined below:

Chapter 2: ***Model Estimation Methodology*** – Introduces study design, model forms, variables used, and statistical analysis methods.

Chapter 3: ***Allocating Land use to Links*** – Explains how the GIS system and techniques were used to assign population and employment to database links.

Chapter 4: ***Data Sources and Samples*** – Presents the link network and accident database, including how other data were acquired and defined.

Chapter 5: ***Accident Prediction Models*** – Presents the resulting models for all three categories along with recommendations and cautions regarding their use.

Chapter 6: ***Summary and Conclusions*** – Summary of report findings and suggestions for future research.

Appendix A: ***Preparation of Population and Employment Data***

Appendix B: ***Preliminary and Intermediate Model Results***

Appendix C: ***User Guide for the GIS Interface***

Distribution and Regression Methods

Several model forms have been used in the past to predict traffic accidents. In this study Generalized Linear Models (GLM) were used, as past research has shown that certain assumptions in conventional linear regression modeling (normal error structure and constant error variance) are violated by traffic accident data (Jovanis and Chang 1986; Miaou *et al.* 1992; Miaou and Lum 1993). These authors, including others (Joshua and Garber 1990), have shown the advantage of using the Poisson distribution, also known as "the law of rare events" (Wikipedia), over standard linear regression models. This distribution fits well because it is the probability distribution of the number of occurrences of an event that happens rarely but has very many opportunities to happen which is exactly the case with traffic accidents.

However, the Poisson model, although representing a significant advance in accurate and reliable modeling capability, is not without its weaknesses and technical difficulties which must be overcome if it is to be used effectively. Miaou *et al.* (1992) used a Poisson regression model to establish the empirical relationship between truck accidents and highway geometrics on rural interstate highways in North Carolina. Their work suggested that the Poisson constraint that the mean and variance of the accident frequency variable are equal, was violated. This is the case with most accident data sets where the variance of accident frequency exceeds the mean; in such cases, the data sets are said to be overdispersed. The use of the Negative-Binomial distribution relaxes this stipulation and is thus being used widely.

Miaou (1994) studied the relationship between highway geometrics and accidents using negative binomial regression. Miaou tested the performance of the Poisson regression, zero-inflated Poisson regression, and negative binomial regression, and suggested that the Poisson regression model can only be used if no overdispersion is observed; otherwise the negative binomial or zero-inflated Poisson regression models should be used. Several other authors including Shankar, Mannering, and Barfield (1995), Poch and Mannering (1996), Vogt and Bared (1998), Kweon and Kockelman (2004), and Noland and Quddus (2004) have demonstrated the advantage of using Negative-Binomial regression over Poisson regression models. Keeping this in mind, alongside the fact that most accident databases show overdispersion (i.e. variance exceeding the mean, thus violating the underlying assumption in the Poisson model), we decided to use Negative-Binomial regression models for this research project.

Negative-Binomial Regression

As discussed above, the Negative-Binomial distribution is widely used in modeling accident frequency. The Negative-Binomial distribution relaxes the Poisson distribution's constraint of equal variance and mean, and thus can be defined as a more generalized version of the Poisson distribution. In order to understand Negative-Binomial distribution one must begin with the Poisson distribution given in the following equation.

$$P(n_i) = \frac{\lambda_i^{n_i} \exp(\lambda_i)}{n_i!} \tag{1}$$

where $P(n_i)$ is the probability of the occurrence of n accidents on road segment i, and λ_i is the expected accident frequency (i.e. $E(n_i)$) for the road segment i. In applying the Poisson model, the expected accident frequency is assumed to be a function of explanatory variables such that:

$$\lambda_i = \exp(\beta X_i) \tag{2}$$

where X_i is a vector of explanatory variables that can include the geometric and traffic characteristics plus the land use data associated with the road segment i used to determine accident frequency, and β is a vector of estimable coefficients. With this form of λ_i, the coefficient vector β can be estimated by standard maximum likelihood methods with the likelihood function ($L(\beta)$) given by :

$$L(\beta) = \prod_i \frac{\exp[-\exp(\beta X_i)][\exp(\beta X_i)]^{n_i}}{n_i!} \tag{3}$$

In order to relax the constraint of equality of variance and mean from Poisson distribution and thus counter overdispersion, an error term is added to the expected accident frequency (λ_i) such that equation 2 becomes:

$$\lambda_i = \exp(\beta X_i + \varepsilon_i) \tag{4}$$

where $exp(\varepsilon_i)$ is a gamma-distributed error term with mean one and variance α. This gives a conditional probability:

$$P(n_i \mid \varepsilon) = \frac{\exp[-\lambda_i \exp(\varepsilon_i)][\lambda_i \exp(\varepsilon_i)]^{n_i}}{n_i!} \quad (5)$$

Integrating ε out of this expression produces the unconditional distribution of n_i. The formulation of this distribution (the negative binomial) is:

$$P(n_i) = \frac{\Gamma(\theta + n_i)}{[\Gamma(\theta) \cdot n_i!]} \cdot u_i^{\theta} \cdot (1-u_i)^{n_i} \quad (6)$$

where $u_i = \frac{\theta}{\theta + \lambda_i}$ and $\theta = 1/\alpha$. The corresponding likelihood function is:

$$L(\underline{\lambda}) = \prod_{i=1}^{N} \frac{\Gamma(\theta + n_i)}{[\Gamma(\theta) \cdot n_i!]} \left[\frac{\theta}{\theta + \lambda_i}\right]^{\theta} \left[\frac{\lambda_i}{\theta + \lambda_i}\right]^{n_i} \quad (7)$$

where N is the total number of road segments. This function is maximized to obtain coefficient estimates for β and α (dispersion parameter). Note that this model structure allows the variance to differ from the mean such that:

$$\text{var}[n_i] = E[n_i][1 + \alpha E[n_i]] \quad (8)$$

The choice between this negative binomial model and the Poisson model is determined by the statistical significance of the estimated dispersion parameter α. If α is not significantly different from zero, the negative binomial model simply reduces to a Poisson regression with $var[n_i] = E[n_i]$. If α is significantly different from zero, the negative binomial is the correct choice and the Poisson model is inappropriate (Milton and Mannering 1998).

The Model Form

The vector of explanatory variables (X_i) is most important in describing the variation in accident frequency. The more exhaustive the set of explanatory variables, the better will be the accident prediction model. As discussed in Chapter 1, it was suggested that in addition to the geometric variables, land use can also be used as an effective exposure. Some research has been done in this area where number of access points (Ivan *et al.* 2000) or the type of surrounding land use (Kim and Yamashita 2002) have been used to predict accident counts or rates. For this project, separate models are estimated to predict crashes in three categories:

1. *Major Intersection Crashes*, or crashes that occur within the "influence area" (defined below) of the intersection of two "major roads" (also defined below);
2. *Segment-Intersection Crashes*, or crashes that occur at minor intersections contained within segments delineated by major roads; and
3. *Segment-Related Crashes*, or crashes that occur apart from any intersection.

There are two critical definitions here; one is "major road." We define major roads to include US highway routes, state highway routes, and local roads classified as arterials or major collectors, provided that traffic volume – Average Annual Daily Traffic (AADT) – is available. The second critical definition is "influence area." We define this to be the area around an intersection within which one can reasonably assume that all crashes that occur are related to the presence of the intersection. At a minimum, this should include the distance required to stop for the signal or stop sign at the observed running speed, but it might also consider the typical length of the stopped queue at a signalized intersection. For this project an area of 250 feet around a major intersection was considered to be the "influence area"; i.e. all the accidents occurring within 250 feet from the center of the intersection on any leg are assigned to category 1 (Vogt and Bared 1998; Harwood *et al.* 2003; Lyon *et al.* 2003).

Figure 1 illustrates how crashes are assigned to these categories. All major roads on the network to be analyzed are identified; other roads are ignored in the analysis. The intersections between these major roads and their influence areas (250 feet) are then defined, as indicated by the circles labeled 1. The road

sections between them (outside the influence areas) are identified as analysis segments, as indicated by the large oval. All crashes that occur within the influence area of a major intersection (areas labeled 1) are identified with that intersection, and defined as category 1, "major intersection crashes." Crashes that occur on analysis segments but are related to minor intersections or access points (areas labeled 2) are then defined as category 2, "segment-intersection crashes." Other crashes that are not related to intersections or access points, that is sideswipe, backing, parking, run-off-road, roll-over, hit-animal or head-on crashes (areas labeled 3), are defined as category 3, "segment-related crashes." The rear end collisions present a difficulty. Many occur due to vehicles slowing unexpectedly to make a turn into an intersection (left or right), but they can also occur due to drivers following too closely. We chose to put them all in category 2.

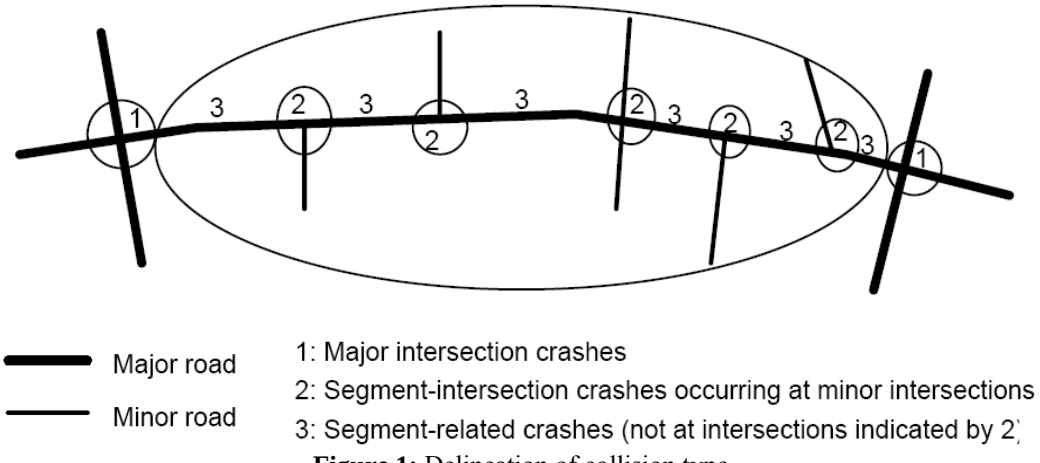


Figure 1: Delineation of collision type

We distinguish between categories 1 and 2 because, typically, actual traffic volumes are often available for major roads, but not for minor roads or access points. This is another important factor in choosing how to distinguish between major and minor roads.

Another important note is that because each of these crash categories is quite different from one another, we must use slightly different model forms for each in terms of the traffic volume and other exposure variables, estimating each using negative binomial regression. Once all the accidents occurring in the "influence area" are identified and assigned to category 1, the remaining accidents are divided between category 2 and 3 by collision type as shown in Table 1. Note that head-on and sideswipe collisions could occur in intersections, or due to intersection-related factors, however for the purposes of this study we chose to identify them only with segments. The following three sub-sections discuss the modeling approach for each crash category.

Category 1 – Major Intersection Crashes: Crashes at intersections are affected less by the geometric design of the intersecting roads than by characteristics of the intersection itself, such as the type of control used (stop sign, signal phasing plan), lane configuration, and adjacent curb cuts. Also, the traffic volumes on the two intersecting roads are both important. The model form proposed for category 1 crashes is:

$$y_1 = V_A^{\beta_A} V_B^{\beta_B} \exp(\beta_1 X_1) \quad (9)$$

Where:

y_1 = The number of category 1 crashes observed in the intersection in the study period (1997-2003);

V_A = The higher of the two intersecting road AADT's in the study period, or the AADT on the road with the higher functional classification;

V_B = The lower of the two intersecting road AADT's in the study period, or the AADT on the road with the lower functional classification;

X_1 = Independent variables for predicting category 1 crashes, potentially including the control type, lane configuration and adjacent land use; and

β's are parameters to be estimated through general linear modeling as discussed above.

Table 1: Distribution among categories by accident type

Category 2: Segment-intersection crashes	*Category 3: Segment-related crashes*
Turning - Same Direction	Head-on
Turning - Opposite Direction	Backing
Turning - Intersecting Paths	Parking
Pedestrian	Sideswipe - Same Direction
Angle	Sideswipe - Opposite Directions
Rear-end	Jackknife
	Fixed Object
	Moving Object
	Miscellaneous Non-Collision
	Overturn
	Unknown

Category 2 – Segment-Intersection Crashes: Crashes at minor intersections are affected by the same factors as major intersection crashes. Unfortunately, we do not know one of the most important ones, namely the traffic volumes of the intersecting minor roads. Of course, the major impetus of this project is to use land use inventories either as a surrogate for these volumes or to estimate them. Therefore, the model form proposed for category 2 crashes is similar to that for category 1, with the volume on the second road replaced by an estimate for this value:

$$y_2 = V^{\beta_V} T^{\beta_T} \exp(\beta_2 X_2) \quad (10)$$

Where:

y_2 = The number of category 2 crashes observed on the segment in the study period (1997-2003);
V = The AADT on the main road in the study period;
T = The estimate of the total number of trips entering and leaving the segment due to the adjacent land use in the study period;
X_2 = Independent variables for predicting category 2 crashes, potentially including the speed limit, shoulder and pavement width; and
β's are parameters to be estimated.

The representation of T was an important item of investigation. The two possibilities considered now are: (1) the number of daily trips predicted as a function of the adjacent land use using the procedures documented in the ITE trip generation manual (ITE 2003), or (2) just using the actual land use inventories directly, with T replaced with the following terms in the X_2 vector:

$$\beta_P P + \beta_R R + \beta_N N \quad (11)$$

where P, R, and N are the population, retail, and non-retail employment, respectively, and the β's are parameters to be estimated. Models were estimated using both options to determine which serves best for predicting the number of category 2 crashes. The equation used for the number of trips generated (T) in the study period (1997-2003) as a function of surrounding land use taken from ITE trip generation manual is presented below.

Population …… $T_P = 7 \times 365 \times \exp[0.91 \times \log(P) + 1.52]$

Retail …… $T_R = 7 \times 365 \times R \times 23.36$

Non-Retail …… $T_N = 7 \times 365 \times \exp[0.84 \times \log(N) + 2.23]$

$$T = T_P + T_R + T_N \quad (12)$$

Some preliminary investigation into segment-intersection models also revealed the importance of using length as an exposure component in these prediction models. It was thought that segments with different lengths but the same land use accessing them should not have different exposure, as the exposure for this type of accident should come from the number of vehicles entering and leaving the link rather than the distance that vehicles drive on the link. However, the segments with shorter length typically are the ones more likely to have intersections and access points, and thus land use, more concentrated than on longer segments. Thus this effect of length as a contribution to accident occurrence might point to a correlation between segment length and land use density. Because of this, the model form in equation 10 was slightly modified and is presented below. Both models with and without length (L) are presented later.

$$y_2 = V^{\beta_V} T^{\beta_T} L^{\beta_L} \exp(\beta_2 X_2) \quad (13)$$

Category 3 – Segment-related Crashes: Segment-related crashes are expected to be more related to the geometric design of the road than intersection-related crashes. They are also clearly related to the length of the segment: the longer the segment, the more opportunities for crashes to occur. Therefore, we proposed the following form for predicting category 3 crashes:

$$y_3 = V^{\beta_V} T^{\beta_T} L \exp(\beta_3 X_3) \quad (14)$$

Since it was observed during initial data investigation that the length may not be linearly related to the accident frequency, a model with an exponent on length was also estimated. This model form is presented in equation 15.

$$y_3 = V^{\beta_V} T^{\beta_T} L^{\beta_L} \exp(\beta_3 X_3) \quad (15)$$

Where:

y_3 = The number of category 3 crashes observed on the segment in the study period in the study period (1997-2003);

V = The AADT on the segment in the study period;

L = The length of the segment in miles;

T = The estimate of the total number of trips entering and leaving the segment due to the adjacent land use in the study period;

X_3 = Independent variables for predicting category 3 crashes, potentially including the pavement or shoulder width; and

β´s are the parameters to be estimated.

Model and Variable Selection

The parameters for the models presented above were estimated using the GENMOD procedure in SAS® (2004). This procedure allows the user to specify the Negative-Binomial distribution for the accident frequency. It also calculates parameters such as Akaike's Information Criterion (AIC) and the Chi-Square significance level for comparing various models and selecting significant parameters respectively. AIC recognizes that the objective of a statistical forecasting model is to convey as much information as possible while limiting the number of parameters estimated. Invariably, the information conveyed by a model increases as one adds more variables to the model. Hence, there is a clear conflict between trying to incorporate more information into a model while at the same time reducing its complexity. AIC helps to trade off between these two competing objectives in evaluating competing models, and is given by:

$$\text{AIC} = 2 \times [\text{LL}(\underline{\beta}) - \text{LL}(\underline{\beta}_{\text{AADT}}) - p - 1] \quad (16)$$

Where:

$LL(\underline{\beta})$ = Partial Log Likelihood of the model under consideration given by SAS;

$LL(\underline{\beta}_{AADT})$ = Partial Log Likelihood of the model with only AADT as exposure; and

p = Degrees of freedom in the model under consideration.

By penalizing the models with a large number of parameters, the AIC permits the selection of models that perform well with fewer parameters. The model with the higher AIC value is preferred. The parameters in the models were selected on the basis of their statistical significance (90%).

Variables Required

As explained in Chapter 1, the main objective of this study is to develop network-based models where the accidents are divided first by location (intersection v. segment) and then by collision type (segment-intersection v. segment-related). But in all the cases traffic volume is the most important factor in accident prediction, since the number of crashes will increase with the number of vehicles on the road, although this relationship is not linear. Thus AADT for all the state highways in the Capitol Region Council of Governments (CRCOG) region is needed for the analysis period of seven years (1997-2003).

Geometric variables like pavement width, shoulder width, number of lanes, median width and posted speed limit also play an important role in accident prediction models. Understanding the effect of geometric variables is also necessary for implementing the improvement measures since the geometrics of the road is easier to change out of all the factors.

Zeeger *et al.* (1981) studied the effects of lane and shoulder width on crashes and concluded that for two-lane roadways the crash rate decreased with an increase in lane width until a width of about 12 feet, after which the rate started increasing again. They excluded rear-end and intersection or driveway-related accidents and studied only the run-off-road and opposite direction crashes, stating that the former are unrelated to lane width. Harwood (1990) examined whether narrower lanes on urban arterials affect safety adversely. He concluded that whenever there is an increase in the number of lanes, the number of intersection accidents also increase, but this has little to do with lane width. Hadi *et al.* (1995) used the NB model with the functional form $exp(\beta \times (Lane\ Width))$ to evaluate crash data in Florida. They observed negative values on β for all road types, where it was significant, concluding that the accident frequency will decrease with increase in lane width no matter how wide the lane is. Similar results were observed by Vogt and Bared (1998) and several other authors.

Most of these studies have modeled pavement or shoulder width as continuous variables. A coefficient obtained from continuous data will not be able to explain a "U" curve phenomenon for pavement width, i.e. if the number of accidents decreases with an increase in pavement width only to a certain value but then starts increasing again. Thus, it is safer to model pavement width as a categorical variable instead of a continuous variable or include the possibility for a non-linear effect so as to be able to capture such an effect if it exists. Categorical values work well for pavement width because these values tend to be identified as even numbers of feet (10, 12, 16) rather than continuously (10.6, 11.9, 12.7).

Also, since most studies model all accidents together it is possible that a positive effect of lane width on certain type of accidents is negated by the negative effect on some other accident types. For example, on two lane roads without a left turn lane, a wider pavement will allow other drivers to pass around the vehicle waiting to make the left turn. This could reduce the number of rear-end crashes but might increase the number of same direction side-swipes. Intersection geometry suggests that this effect is more likely to be observed in segment models (Category 2 and 3).

It is also likely that the number of crashes at an intersection will vary by the control type. King and Goldblatt (1975) studied the relationship of accident patterns by the type of intersection control but found no evidence that signalization, by itself, would decrease the net accident-related disutility. However, it is possible that in combination with other factors, the control type may help in explaining the variation in accidents at an intersection. Thus, for the purpose of this study, AADTs for both the roads at an intersection along with certain geometric variables (number of lanes, intersection configuration, and skewness) and type of control are required to model intersection crashes (Category 1).

As stated above, although some of the variables affecting segment crashes are similar to those affecting intersection crashes (AADT, speed limit), there are some notable differences. For example, unlike intersection crashes, pavement width or shoulder width definitely have a significant effect on a certain type of accidents, and factors such as control type will not be applicable in case of segment crashes. Bared and Vogt (1996) studied mid-block crashes apart from intersection crashes using different exposure variables for both. Ivan and O'Mara (1997) studied two-lane roadways in the state of Connecticut. They found that the frequency of intersections on the segment was one of the most important predictor variables in crash prediction. This is probably because as the frequency of the intersections increases, so do the conflict opportunities. Several other studies have also shown that as the number of access points along a street increases, so does the number of crashes. On a street with many access points, traffic may have to slow and stop often to accommodate vehicles entering and exiting. This kind of traffic movement can result in more crashes. Table 2 presents some of the studies presented in Levinson and Gluck (2000) regarding accidents and access spacing (or driveway density).

Table 2: Studies of Accidents vs. Access Spacing

No.	*Year*	*Author and Area*	*Conclusions*
1	1992-93	Sokolow et al., Long et al., Florida	Accident rates doubled when driveways exceeded 20 per mile (Sokolow). Accident rates increased 70% as driveways per mile increased from less than 13 to more than 20. It was estimated that one driveway adds 0.02 crashes per year.
2	1993	Millard, Florida	Doubling access points from 20 to 40 per mile doubled the accident rate. Doubling signals from 2 to 4 per mile more than doubled the accident rate
3	1994	Michigan	Midblock accident rates generally increased as the number of intersections per mile (including driveways) and the number of lanes increased
4	1995	Fitzpatrick and Balke, Texas	Total and midblock accidents generally increased as driveways became more numerous
5	1996	Norwalk-Wilton, Connecticut (Route 7)	Accident rate per mile increased along roadways carrying 20,000 to 25,000 vehicles per day as access density increased
6	1996	Garber and White, Virginia (10 mi, 30 locations)	Multiple regression analysis assessed effects of ADT/lane, average speed, number of access points, left-turn lane availability, average access spacing and average difference in access spacing

Using number of access points as a predictor variable may appear quite attractive, but it does have some limitations. For example, counting the number of residential, retail and non-retail driveways for each road segment is an excessively labor intensive, and thus expensive, task. As well, the accuracy of the driveway models is suspicious since the access point count does not give an estimate of the number of vehicles accessing a road segment through each driveway. For example, a driveway serving a convenience store is generally counted as a "retail" driveway the same as is a large supermarket. The number of vehicles accessing the major artery will vary significantly in both cases and so will the exposure to traffic accidents. Thus, another important aspect of this study is to show that models using land use variables such as population, retail and non-retail employment can act as a surrogate for, and potentially improve, the expensive driveway models.

Thus in order to compare the land use models with the driveway models, both the number of access points and land development variables will be needed. Furthermore, since the land development data has to be assigned to the links it is important that it has the same spatial distribution as the links. For example, if there is a large number of links in an area, the land development variables should be available for finer areas. For this purpose it is ideal to have the land development variables by traffic analysis zones (TAZ's) since the size of a TAZ becomes smaller if there are more roads (Figure 2).

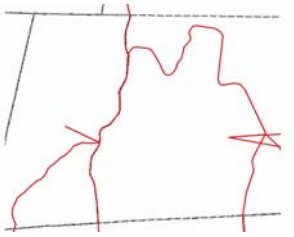
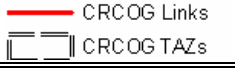
More Roads, Smaller Zones	*Fewer Roads, Bigger Zones*
	
	

Figure 2: Variation in TAZ size with link density (at same scale)

Chapter Summary

From the discussion in this chapter we can conclude that the Negative Binomial (NB) distribution works better for accident databases than the Poisson distribution and, thus, NB regression models were developed for this project. Also, it is clear from the discussion above that the accidents need to be modeled in separate categories, namely: Intersection, segment-intersection, and segment-related accidents on the basis of their location and collision type. The potential variables identified from the literature review for these models are presented in Table 3.

Table 3: Potential Variables for the Models

Type	*Data*	*Definition*
General	AADT	AADT over the analysis period (1997-2003)
	Speed Limit	The posted speed limit on the roads
	Links & Intersections Database	GIS layers
Geometric	Median	Reduces mid-block left turns and changes distribution of fixed object and head-on collisions
	Skewness	Configuration of the roads at intersections
	Control Type	Signal/Stop/None
	Configuration	Three or four leg intersections
	Number of Lanes	Helps in predicting intersection accident
	Pavement and Shoulder Width	Important to predict segment accidents
Land Use	Type of Land Use	Categorical (for intersections)
	Population	Population accessing a link
	Retail Employment	Retail employment accessing a link
	Non-Retail Employment	Non-Retail employment accessing a link
	Residential Driveways	Number of residential driveways on a link
	Retail Driveways	Number of retail driveways on a link
	Non-Retail Driveways	Number of non-retail driveways on a link

Road segments differ significantly from each other due to surrounding land use type, number of lanes, and the presence of median. For example, the rural highways are generally undivided two lane highways without significant shoulder width. Unlike them, urban/suburban highways tend to have more variety in number of lanes and the shoulder width. Thus it was decided that these roads should be modeled in different groups. Out of the total 1378 state highways in the CRCOG region (Table 4), only three categories had enough links needed to reliably develop regression models. They are:

1. Rural two-lane undivided (326 links)
2. Urban/Suburban two-lane undivided (587 links)
3. Urban/Suburban four-lane undivided (233 links)

The total number of intersections between the CRCOG state highways was 133. They were modeled in two different categories:

1. Three-leg (61 intersections)
2. Four-leg (72 intersections)

The next chapter describes the procedure for allocating the population and employment to road segments.

Table 4: Distribution of Links

Type	***Lanes***	***Div/Undivided***	***Frequency***
Rural	1	Div	0
		Undivided	1
	2	Div	5
		Undivided	**326**
	3	Div	0
		Undivided	6
	4	Div	23
		Undivided	18
Urban/Suburban	2	Div	3
		Undivided	**587**
	3	Div	1
		Undivided	38
	4	Div	94
		Undivided	**233**
	5	Div	1
		Undivided	20
	6	Div	13
		Undivided	9
Total			**1378**

Background and Context

This chapter describes a procedure by which GIS land use inventories are used to identify the land development intensity associated with specific highway links to be used as an estimate of exposure in models for predicting crashes at minor intersections on road segments. Specifically, this procedure takes land use data (population and employment by category) identified with geographic zones and allocates it to links representing the road segments passing through or abutting the corresponding zones. The resulting population and employment identified with each road segment can then be used to estimate the volume of traffic entering and exiting the main road as a measure of exposure to collisions related to intersections along the segment for crash prediction models. Several sub-procedures have been developed: splitting of links into smaller segments when needed due to the link and zone topology, identifying the links passing through and abutting each zone, and calculating weights to specify how much of the zone is associated with each link. The procedure was developed for use in predictive accident models, but parts or all of the procedure can also be applied to other spatially-related transportation planning problems.

The data used for this project was obtained from the Capitol Region Council of Governments (CRCOG), the largest of Connecticut's regional planning agencies, serving the City of Hartford (the state capital and third largest city) and 28 towns surrounding it. CRCOG is the only planning agency in Connecticut to maintain its own travel demand forecasting model. Conveniently, CRCOG uses ArcGIS by ESRI® to manage the land use data and link network for the model, and made both the land use and road network layers available to the project. Additional data about the spatial distribution of land development in the study area was obtained from the Center for Land Use Education and Research (CLEAR) at the University of Connecticut.

The data set obtained from CRCOG provides numbers of residents, retail employees and non-retail employees in each of 1122 traffic analysis zones (TAZ's) covering the CRCOG region as well as some towns bordering the region (including some in Massachusetts, e.g., the City of Springfield). The zones have a wide range of degree of development, ranging from rural sparsely populated zones to very densely developed zones in the center of Hartford; 902 out of the total 1122 TAZs actually lying in the CRCOG area were selected for this project. Figure 3 depicts the area covered by the model, indicating the size and distribution of the TAZ's. The sizes of the selected zones vary between 0.03 and 8.50 square miles, with a mean value of 0.84 square miles.

The CRCOG network includes centroids representing the TAZs as nodes connected to the road network, but the latitude and longitude coordinates of the centroids do not necessarily correspond to the weighted center of the land development in the zone. To more accurately distribute the land use activity of each TAZ to the correct links, additional land cover data was obtained from CLEAR. These data are extracted from aerial photographs that were digitally processed to convert them into different land cover types (Wilson *et. al.* 2003); see Figure 4.

The procedure for allocating zone data to links is carried out using three sub-procedures:

1. Splitting links in the GIS database so that each link segment is coincident with a single zone on both sides for its entire length;
2. Recognizing which links are associated with each zone, that is those that are coincident with the edge of the zone, or are fully contained in it; and
3. Allocation of the zone attributes to the links associated with each zone.

ArcGIS by ESRI® was used as the development platform for this project. Figure 5 shows a map generated over a part of the CRCOG network. Even when a street acts as a boundary of a TAZ, the match of the zone boundary and the link is far from perfect, since the zone boundaries and the street links were digitized separately. The procedure must take this into account and in some cases identify links as belonging to a zone when in the GIS network they are located slightly outside the zone, as in reality they coincide with the border of the zone.

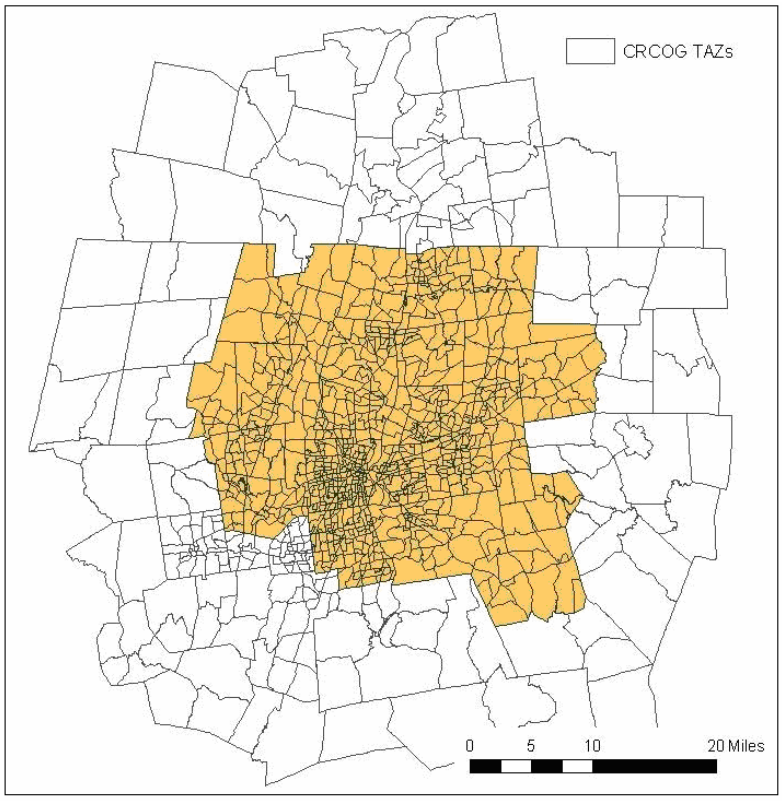


Figure 3: CRCOG Traffic Analysis Zone Map, with Member Towns Shaded.

Figure 4: Roadway network and land use map from CLEAR.

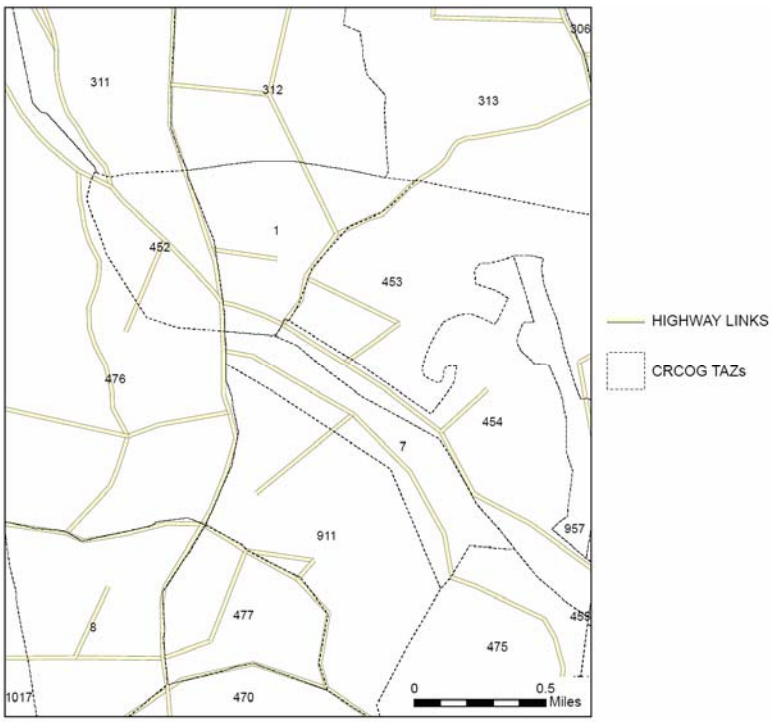


Figure 5: Road network with TAZ boundaries, generated from CRCOG data

Splitting Links

In some cases a link does not border, or fall within, the same zone for its entire length. Figure 6 a, b and c show some examples of this. To be able to easily associate links and zones we implemented a first step in which any links that borders more than one zone on one side or the other is identified and split into smaller segments so that each segment borders no more than one zone on each side for its entire length. Each segment of the link takes on the same physical attributes as the original link except for the link length which is recalculated by the ArcGIS system.

There are two primary ways in which this situation is identified, one for the case when a link acts as a border between zones (Situation 1, Figure 6a), and one for when the link changes from being fully within one zone to being fully within another (Situation 2, Figure 6b). Situations can also arise when the link changes from being fully within one zone and continues on to become a border link (Situation 3, Figure 6c); this case is handled in the same way as Situation 1.

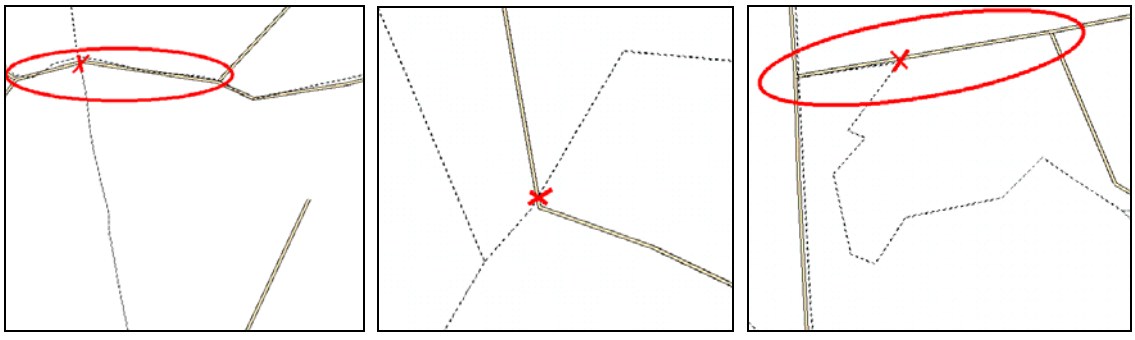


Figure 6 a-c: Visualization of link splitting, Situation 1: Passing several zones (a), Situation 2: Intersecting zone boundary (b), Situation 3: Changing from boundary link to internal link (c). Each link is split at the X.

To identify links matching situation 1, a buffer is created around all edges of the polygon which denotes a zone. If a link passes through this buffer without having a start or end point very close to the buffer, it is split at the point closest to the polygon corner (Figure 6a). To handle links in situation 2, every link that cleanly intersects a zone boundary is split at the point of intersection (Figure 6b). Segments in situation 3 are handled in the same way as situation 1. Each link is handled according to only one of the three situations.

Connecting Zones and Links

In order to associate each link to a specific zone, not only the links within the TAZ polygon need to be considered, but also those just outside it. This is because the link network and the TAZ polygons don't match up perfectly even when the link in reality is the zone border. A buffer of 200 feet is defined around each TAZ in which links are also identified as belonging to the TAZ. Several different buffer widths (50, 75, 100 and 200 feet) were tested to identify the ideal buffer width; this ideal buffer width would need to be re-estimated for different networks since the optimal buffer zone width will depend on how well the zone boundaries and links agree in the GIS network used.

Table 5: Test of Buffer Widths for Associating Links with Each TAZ

#	*Zone No.*	*50ft*	*75ft*	*100ft*	*200ft*
1	524	OK	OK	OK	OK
2	60	OK	OK	OK	OK
3	713	OK	OK	OK	+400ft
4	274	OK	OK	OK	OK
5	354	-600ft	OK	OK	OK
6	604	-2500ft	-1300ft	-1200ft	-500ft
7	895	OK	OK	OK	OK
8	281	OK	OK	OK	+200ft
9	60	-3800ft	-3500ft	-2400ft	OK
10	327	-800ft	-600ft	-300ft	+400ft
11	934	-2400ft	-1200ft	-900ft	+200ft
12	82	-850ft	-700ft	OK	OK
13	581	-600ft	-600ft	-400ft	OK
14	275	OK	OK	OK	OK
15	791	-2600ft	-800ft	-150ft	OK
16	135	OK	OK	OK	OK
17	310	-3500ft	-3000ft	-2400ft	OK
18	701	OK	OK	OK	OK
19	344	OK	OK	OK	OK
20	265	OK	OK	OK	OK

Table 5 shows the results of a test for 20 randomly selected zones where each buffer width was applied and the length of links not correctly assigned to zones was totaled (both missed links and false matches). A small number of link segments have been identified as belonging to the TAZ when they shouldn't; these are marked with a positive value. Conversely, the total length of links that should have been identified but were not is listed with a negative sign. "OK" means all links associated with that TAZ, and no excess links, were identified. For no TAZ were there both missed and excess links.

The excessive links identified when using the 200-ft buffer mostly intersect a bordering link (the correctly identified one), joining it at a sharp angle and thus lying in the buffer zone for a length greater than the width of the buffer zone (Figure 7).

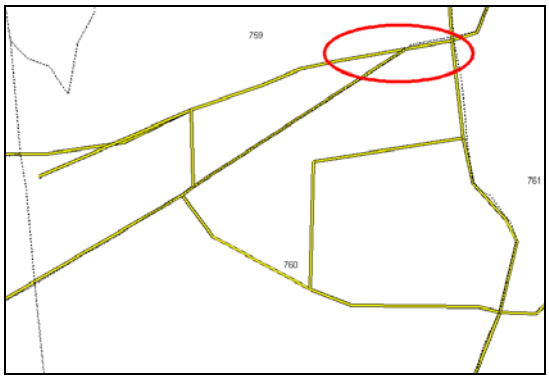


Figure 7: Example of excessive link identified with zone 760, joining the actual border of the zone at a sharp angle.

Assigning the Land Use Data to the Links

In the third step the land use data, in the form of number of residents, retail employees and non-retail employees, is allocated to the links. The population and employment data should ideally be allocated according to where the traffic it generates actually enters the major road network. However, the population and employment data is only given for an entire TAZ, and thus it is necessary to derive a procedure in which link weights are assigned more carefully according to the portion of the land use activity that accesses the network through each link.

Because, as noted above, the TAZ centroids generally do not represent the weighted centers of each TAZ, using distance from the centroid to each link assigned to the TAZ to calculate the allocation weight for each link was not appropriate. Instead, a rather straightforward procedure was used where land cover maps were checked against road maps for each zone, and if the developed areas of a zone are homogeneously spread along the neighboring or internal links, then only the link length was used as the weight. In the case of non-homogeneously spread land use, the weights were determined manually by approximating the portion of the land use area that should be associated with each of the links. The manual assignment of the land use in a zone to each link is very time consuming, and is therefore infeasible to repeat for a large number of zones. In the CRCOG network, it turned out that the developed areas are fairly evenly spread in nearly all zones; of the 1122 TAZ's only 53 warranted a manual calculation of link weights.

Figure 8 illustrates the process of assigning weights for non-homogeneous zones. The land cover and the minor streets are used to estimate how much of the land use exits to each assigned link (major roads). The eastern boundary of the zone consists of four different links, delineated by the intersections. The southern border consists of one long link. For this zone no links have been split. If the land cover were homogeneously spread, the land use would have been assigned by weights only according to link length. In that case the southern link would have been allotted the majority of the land use. In this case the land cover was manually examined and the weights were set equal among the four links on the eastern boundary as well as the one on the southern boundary, as approximately the same area of developed land is connected to each link.

$$p_{ijk} = \frac{w_{ijk}}{\sum_{j' \in J_i} w_{ij'k}} \tag{17}$$

Where:

J_i = Set of links adjacent or inside TAZ_i, $i \in I$

p_{ijk} = Proportion of zone i allocated to link j for land use type k

w_{ijk} = Weight of land use k on link j for zone i

The weights are set equal to the link length for zones with homogeneous land development spread. For zones with non-homogeneous spread, the weights are assigned manually according to the portion of development that would exit the zone to that link.

In Figure 8 it can also be seen that there is one street in the northern part of the developed area not assigned to the zone (marked white). This is a centroid connector which is automatically excluded. The

location of the centroid illustrates the previously mentioned problem with the somewhat arbitrary location of the centroids. In the common case of a link that acts as a border between two zones, the link is assigned land use activity from both zones, which are added together.

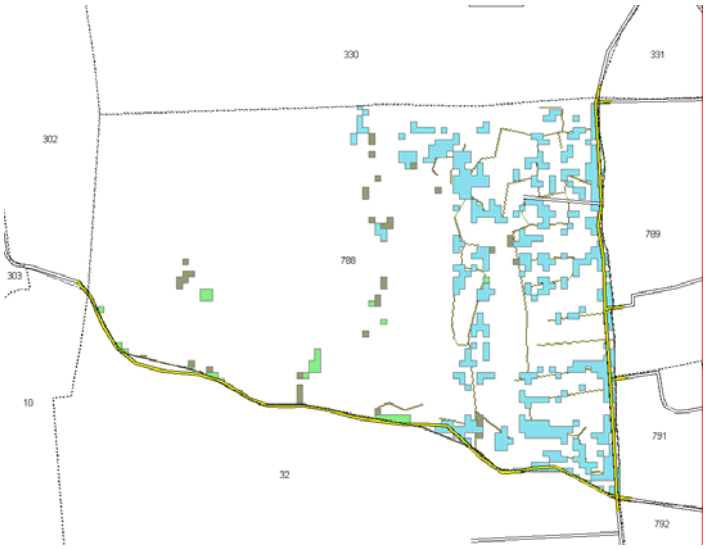


Figure 8: TAZ with non-homogeneous land development, shaded links are the links associated with this zone; developed areas and minor streets are included in plot.

The procedure works well in so far as it is able to allocate the zone data to the links. It would be preferable to validate the procedure against actual data regarding how many vehicles are entering and exiting the zones through each link. This, however, requires very detailed data for which adequate resources are not available in the funded project. An estimate of the functionality of the procedure was, however, available after the use of the assigned data in accident modeling in so far as how much the assigned zone data can improve the accident models. The buffer zone used for the identification of links associated with each zone was found to perform best when set to 200 feet, this size is dependent on how well the zone boundaries and the links are in agreement and needs to be recalibrated for different datasets.

The procedure still requires some manual work for establishing the weights. It would be desirable to limit the need for time consuming manual work further. One possibility for automating the process would be to assign the land use by automatic image processing of land cover maps. The developed areas are however not always connected to the nearest major streets, thus calling for the need of a more complicated analysis where the minor street network is used to aid in the determination of which major street the land use activities generated will be exiting and entering through.

To conclude, the developed procedure provides the possibility to allocate zonal data to a link network in a semi-automated manner. Further development is still needed to enhance the procedure; for instance the calibration of the procedure against actual trip data on a detailed level. Although the CRCOG model offers a wide variety in zone and link density it is still desirable to test the procedure on other networks to test the transferability of the results.

As discussed in Chapter 2, there are several variables required to estimate the parameters for the models presented in Equations 9, 10, 13, and 15. The population and employment data and link network for the model provided by CRCOG and the additional data about the spatial distribution of land development obtained from CLEAR at the University of Connecticut were described in Chapter 3. Other databases used to compute these parameters are listed below and are discussed subsequently.

1. Connecticut (CT) Collision Analysis System;
2. CT Traffic Log;
3. CT Photo Log;
4. CT Highway Log; and
5. ConnDOT Road Geometrics Database.

Connecticut (CT) Collision Analysis System

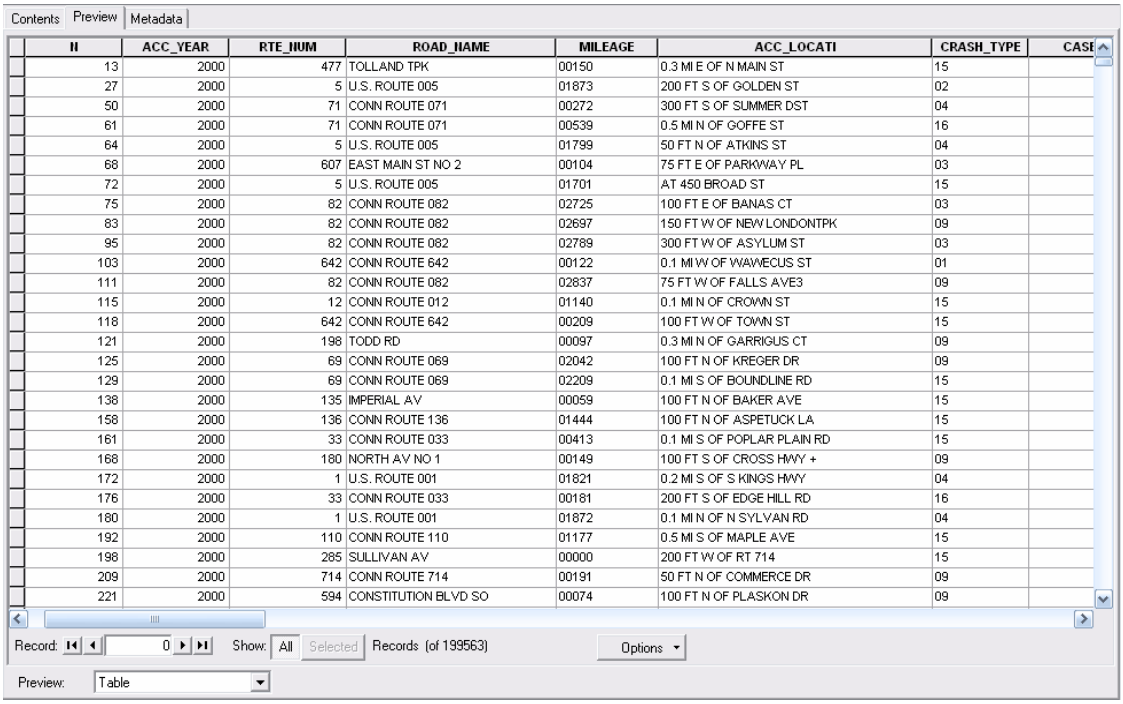
ConnDOT's Accident Record Section maintains databases containing details (time, location on road network, number of vehicles involved, collision type, etc.) of every accident occurring each year on state-maintained roads in Connecticut, and all accidents resulting in an injury or fatality anywhere in the State (Figure 9). For this project accident records from seven years (1997 to 2003) were used. These databases contain a lot of details about each crash but for our purpose the variables required were:

a. ***Route number*** of the road segment where the collision occurred;
b. ***Mileage*** at which the collision occurred; and
c. ***Collision type,*** coded numerically from the list defined in Table 6.

The route number and the mileage help to assign each accident in the accident database to the links in the CRCOG layer. The collision type is used to divide the accidents into the three crash categories, namely: intersection, segment-intersection, and segment-related accidents defined in Chapter 2.

Table 6: Crash Type Codes in Connecticut Accident Records

Coll. Type	*Description*	*Coll. Type*	*Description*
01	Turning-Same Direction	10	Head-on
02	Turning-Opposite Direction	11	Backing
03	Turning-Intersecting Paths	12	Parking
04	Sideswipe-Same Direction	13	Pedestrian
05	Sideswipe-Opposite Direction	14	Jackknife
06	Miscellaneous Non-Collision	15	Fixed Object
07	Overturn	16	Moving Object
08	Angle	17	Unknown
09	Rear-end		



Contents | Preview | Metadata

II	ACC_YEAR	RTE_NUM	ROAD_NAME	MILEAGE	ACC_LOCATI	CRASH_TYPE	CASE
13	2000	477	TOLLAND TPK	00150	0.3 MI E OF N MAIN ST	15	
27	2000	5	U.S. ROUTE 005	01873	200 FT S OF GOLDEN ST	02	
50	2000	71	CONN ROUTE 071	00272	300 FT S OF SUMMER DST	04	
61	2000	71	CONN ROUTE 071	00539	0.5 MI N OF GOFFE ST	16	
64	2000	5	U.S. ROUTE 005	01799	50 FT N OF ATKINS ST	04	
68	2000	607	EAST MAIN ST NO 2	00104	75 FT E OF PARKWAY PL	03	
72	2000	5	U.S. ROUTE 005	01701	AT 450 BROAD ST	15	
75	2000	82	CONN ROUTE 082	02725	100 FT E OF BANAS CT	03	
83	2000	82	CONN ROUTE 082	02697	150 FT W OF NEW LONDONTPK	09	
95	2000	82	CONN ROUTE 082	02789	300 FT W OF ASYLUM ST	03	
103	2000	642	CONN ROUTE 642	00122	0.1 MI W OF WAWECUS ST	01	
111	2000	82	CONN ROUTE 082	02837	75 FT W OF FALLS AVE3	09	
115	2000	12	CONN ROUTE 012	01140	0.1 MI N OF CROWN ST	15	
118	2000	642	CONN ROUTE 642	00209	100 FT W OF TOWN ST	15	
121	2000	198	TODD RD	00097	0.3 MI N OF GARRIGUS CT	09	
125	2000	69	CONN ROUTE 069	02042	100 FT N OF KREGER DR	09	
129	2000	69	CONN ROUTE 069	02209	0.1 MI S OF BOUNDLINE RD	15	
138	2000	135	IMPERIAL AV	00059	100 FT N OF BAKER AVE	15	
158	2000	136	CONN ROUTE 136	01444	100 FT N OF ASPETUCK LA	15	
161	2000	33	CONN ROUTE 033	00413	0.1 MI S OF POPLAR PLAIN RD	15	
168	2000	180	NORTH AV NO 1	00149	100 FT S OF CROSS HWY +	09	
172	2000	1	U.S. ROUTE 001	01821	0.2 MI S OF S KINGS HWY	04	
176	2000	33	CONN ROUTE 033	00181	200 FT S OF EDGE HILL RD	16	
180	2000	1	U.S. ROUTE 001	01872	0.1 MI N OF N SYLVAN RD	04	
192	2000	110	CONN ROUTE 110	01177	0.5 MI S OF MAPLE AVE	15	
198	2000	285	SULLIVAN AV	00000	200 FT W OF RT 714	15	
209	2000	714	CONN ROUTE 714	00191	50 FT N OF COMMERCE DR	09	
221	2000	594	CONSTITUTION BLVD SO	00074	100 FT N OF PLASKON DR	09	

Record: 0 | Show: All | Selected | Records (of 199563) | Options

Preview: Table

Figure 9: Sample of ConnDOT Accident Record Database (2000)

Traffic Log

The Traffic Log lists the estimated average daily traffic volumes (ADT) for the entire state-maintained highway network in Connecticut. Volumes are counted once every three years on highway segments delineated by intersections at which significant traffic volume changes occur. These ADT values are estimates of the number of vehicles passing through the defined section of highway on an average day in each year for both directions of travel combined, except for one-way ramps or other one-way roadways (Connecticut State Traffic Log 2005).

Figure 10 shows a sample of the log covering a portion of Route 5. Note that the estimated volumes are sorted by the route number and milepost, which are used to associate the segments in the Traffic Log to the road segments in the GIS layers. The estimated ADTs for each study segment were extracted for the same years as the accident data, 1997 to 2003. The extracted ADTs for each segment were averaged over the time period.

Photo Log

Some of the descriptor variables such as the posted speed limit, intersection control type, number of lanes, and other geometric features were obtained from the CT Photo Log. Annually, the entire state-maintained roadway network containing approximately 6,155 route kilometers is videorecorded using two Automatic Road Analyzer (ARAN) systems. The two cameras in this system, recording the forward and right side views, record high quality videos of the road segment and its surrounding. The Photo Log is classified by route number and milepost for each route, permitting precise location of each intersection and road segment to match them to the segments in the GIS layers. This system was also used to calculate the number of residential, retail, non-retail access points. The type of land use around an intersection (urban, suburban, or rural) was judged by visual inspection of its surroundings. A number of residential, retail and/or non-retail driveways in the area around the intersection will make it an urban environment and if the driveways are far apart and there is small to no activity, the intersection is assumed to be in rural environment.

ROUTE - 005 NB I-91 (NEW HAVEN) TO MASSACHUSETTS STATE LINE (ENFIELD) PAGE 14

LOGGED DIRECTION/NORTH

FROM	CUM MILES	TO	CUM MILES	SECT LENGTH	2005 ADT
EXIT FR NB I-91(138)	.00	WILLOW ST	.17	.17	13900
WILLOW ST	.17	ACC TO SB I-91(142)	.29	.12	16100
ACC TO SB I-91(142)	.29	FERRY ST	.37	.08	21000
FERRY ST	.37	GRACE ST(ONE-WAY SB)	.42	.05	22600
GRACE ST(ONE-WAY SB)	.42	NEW HAVEN - HAMDEN TL	.98	.56	17700
NEW HAVEN - HAMDEN TL	.98	PARK RD	1.13	.15	17700
PARK RD	1.13	RIDGE RD(SB)	1.56	.43	13200
RIDGE RD(SB)	1.56	HAMDEN - NORTH HAVEN TL	3.51	1.95	9000
HAMDEN - NORTH HAVEN TL	3.51	SKIFF ST	3.71	.20	9000
SKIFF ST	3.71	SR 717(DIXWELL AVE)	4.65	.94	15100
SR 717(DIXWELL AVE)	4.65	SR 720(DEVINE ST)	4.81	.16	12000
SR 720(DEVINE ST)	4.81	SR 729(BROADWAY)	5.69	.88	8300
SR 729(BROADWAY)	5.69	RTE 22(BISHOP ST)	5.82	.13	8700
RTE 22(BISHOP ST)	5.82	VALLEY SERVICE RD(DE)	6.08	.26	18300
VALLEY SERVICE RD(DE)	6.08	EXIT FR NB I-91(024)	6.36	.28	20100
EXIT FR NB I-91(024)	6.36	RTE 103(WASHINGTON AVE)	6.48	.12	23800
RTE 103(WASHINGTON AVE)	6.48	EXIT FR NORTH HAVEN PLAZA	6.55	.07	26700
EXIT FR NORTH HAVEN PLAZA	6.55	FRANKLIN ST	6.68	.13	22700
FRANKLIN ST	6.68	ACC TO NB I-91(029)	7.13	.45	19800
ACC TO NB I-91(029)	7.13	EXIT FR SB I-91(028)	7.28	.15	20200
EXIT FR SB I-91(028)	7.28	WADSWORTH AVE	7.70	.42	19900
WADSWORTH AVE	7.70	BRADLEY ST	8.36	.66	16700
BRADLEY ST	8.36	NORTH HAVEN - WALLINGFORD TL	9.26	.90	14500
NORTH HAVEN - WALLINGFORD TL	9.26	SR 702(WB)	9.44	.18	14500
SR 702(WB)	9.44	SOUTH ELM ST	10.05	.61	17700
SOUTH ELM ST	10.05	SOUTH ORCHARD ST	10.58	.53	15700
SOUTH ORCHARD ST	10.58	WARD ST	11.09	.51	12500
WARD ST	11.09	N JCT RTE 150(HALL AVE)(ONE-WAY NB)	11.43	.34	11400
N JCT RTE 150(HALL AVE)(ONE-WAY NB)	11.43	CHRISTIAN ST	11.82	.39	14200
CHRISTIAN ST	11.82	NORTH PLAINS HWY(DE)	12.39	.57	16700
NORTH PLAINS HWY(DE)	12.39	CON TO RTE 68	13.61	1.22	18900
CON TO RTE 68	13.61	ACC TO NB RTE 15(219)	13.92	.31	26200
ACC TO NB RTE 15(219)	13.92	EXIT FR SB RTE 15(144)	14.04	.12	28300
EXIT FR SB RTE 15(144)	14.04	RTE 71(OLD COLONY RD)	14.42	.38	26700
RTE 71(OLD COLONY RD)	14.42	RTE 150(SOUTH BROAD ST)	14.88	.46	15900
RTE 150(SOUTH BROAD ST)	14.88	WALLINGFORD - MERIDEN TL	15.07	.19	19200
WALLINGFORD - MERIDEN TL	15.07	GREEN RD	15.60	.53	19200
GREEN RD	15.60	HALL AVE	16.06	.46	15400
HALL AVE	16.06	ANN ST	16.47	.41	12800
ANN ST	16.47	CURTIS ST	16.83	.36	13900
CURTIS ST	16.83	EAST MAIN ST #1	17.10	.27	18700
EAST MAIN ST #1	17.10	WALL ST	17.30	.20	14500
WALL ST	17.30	ACC TO EB I-691(002A)	17.73	.43	17000

ROUTE NO 005

Figure 10: Connecticut Traffic Log 2005

Highway Log

The Connecticut Highway Log (Figure 11) was primarily used to cross check the milepost information obtained from the CRCOG GIS layers and the Photo Log. It also provided information for differentiating between the major and the minor road to define major and minor AADTs for Category 1 collisions. For example, from mileage 7.64 to 7.89 in the table below US 7 overlaps Route 4 thus it is of higher priority than Route 4. Thus, for example, in the case of an intersection between US 7 and Route 4, US 7 will be treated as the major road and Route 4 will be the minor road. At an intersection, a road segment with higher AADT is prioritized as a major road when there is no information available about the order of the links in the Highway Log.

Road Geometrics Database

This database was used to obtain the pavement and shoulder width of the road segments to be used as exposures in the regression models. This step was necessary because lane and shoulder widths could not be determined accurately using the photo log. These databases also contained the route and the milepost information for each road segment along with various road widths such as lane width, pavement width, and the shoulder width in both directions. In some cases the value of the pavement or the shoulder width was not constant over the entire link in the CRCOG database. In those cases an average value weighted by the partial link length was calculated.

Table 7 presents all the variables collected from the sources listed above to compute the parameters in each of the three categories, and the modifications needed to convert the data to a more usable form. The variables are arranged by the category they were used in, for example, a categorical land use type (urban, suburban, or rural) was only used in intersection accidents (Category 1) and thus is listed accordingly. Variables such as pavement width and number of driveways were not included in intersection crashes thus are only listed under category 2 and 3.

```
TWN              ROUTE &                I    M     M   HPMS M SYS R FUNCT CLASS A LNS BRDGE D EXIT
                                        N  C I   K E        A    -                          I
 #          LOCATION DESCRIPTION        T  U L   I T   A  # I S N U             C R L  #  S S  # S
                                        C  M E   L E   R    N T H               T E O     U T    U
                                        H    S   O R   E    T A S               L V G     F      F
                                                   S   A
_____________________________________________________________________________________________________

RTE   4
--------
125 RTE 41(MAIN ST)                         .00    .00 X285 S P   R MINOR ART   1   2        4
    RTE 343(AMENIA RD)                      .00    .00
                                            .79   1.27 X286
    MITCHELLTOWN RD                        1.78   2.86
                                           2.00   3.22 ----
                                           2.71   4.36
    BGN SCENIC RD                          3.42   5.50
    OP CATTLE CROSSING(CONC BOX)(01929)    4.77   7.68                           6511
    OP MILL RIVER                          5.21   8.38                           1930
    OP GUINEA BK                           6.27  10.09                           0421
    DR TO GRISTMILL REST AREA              6.42  10.33
    DR TO GRISTMILL REST AREA              6.48  10.43
    OP GUINEA BK                           6.58  10.59                           1931
    OP GUINEA BK                           7.26  11.68                           0422
    BGN OVLP US 7                          7.64  12.30      ----------------------------------
    US 7 & RTE 4(CORNWALL BRIDGE RD)       7.64  12.30
    W JCT US 7                             7.64  12.30
    RTE 4(CORNWALL BRIDGE RD)              7.64  12.30
    END SCENIC ROAD                        7.64  12.30
031 SHARON - CORNWALL TL                   7.71  12.41
     (RTE 4 - FURNACE BROOK RD)            7.89  12.70
    E JCT US 7(KENT RD)                    7.89  12.70 X057 S P   R MINOR ART   1   2        4
    END OVLP US 7                          7.89  12.70
    US 7 & RTE 4(KENT RD)                  7.89  12.70
```

Figure 11: Connecticut Highway Log (2004)

Since the project crash database spanned a seven year period (1997-2003), all the factors in the database with a *per day* value, such as AADT or number of trips generated per day from the trip generation manual, were multiplied by a factor of 2555 (7 x 365) to bring them to the same time scale as the number of crashes. AADT was also multiplied by 10^{-6} to convert AADT to million vehicles.

Table 7: Variables in the Models

Category	***Data***	***Source***	***Modification***
1, 2, & 3	Links & Intersections	CRCOG	None
	AADT	Traffic Log	2556 x 10^{-6} x 7-year Avg.
	Posted Speed Limit (mi/h)	Photo Log	None
	Median	Photo Log	None
1	Control Type	Photo Log	None
	Configuration	Photo Log	None
	Skewness	Photo Log	None
	Type of Land Use	Photo Log	None
	Number of Lanes	Photo Log	None
2 & 3	Pavement Width (feet)	Road Geometrics Database	Average(Segment)
	Shoulder Width (feet)	Road Geometrics Database	Average(Segment)
	Population	CRCOG	Assigned to Links[1]
	Retail Employment	CRCOG	Assigned to Links[1]
	Non-Retail Employment	CRCOG	Assigned to Links[1]
	Residential Driveways	Photo Log	None
	Retail Driveways	Photo Log	None
	Unsignalized Intersections	Photo Log	None

[1] See Chapter 3 or Jonsson, Deng, and Ivan (2006)

This chapter presents the recommended crash prediction models to be used in the GIS network-based crash prediction system. As discussed in Chapter 2, the system predicts accidents in three categories: intersection accidents, segment-intersection accidents, and segment-related accidents. There are separate models for intersection accidents distinguished by the number of legs in the intersection (3 or 4), and for the other two categories by the type of road: rural two-lane, urban/suburban two-lane, and four-lane undivided. As noted in Chapter 2, there were not enough observations for other types of roads to estimate reliable prediction models. If needed, prediction models for these road types will have to be found from other sources, such as the Highway Safety Manual (HSM 2007) and research being conducted for it, with the given models calibrated to local conditions according to the instructions.

All estimated coefficients in each model are significantly different from zero with 90% or greater confidence. Table 8 presents a description of the variables used in the regression models. Table 9 lists the range of values observed in the estimation data set for all continuous variables used in each model. It is not recommended to use the models to predict crashes when any of the variables are outside the indicated ranges, as there is no guarantee that the relationships extrapolate consistently outside these ranges. Similarly, frequency distributions are presented for all categorical variables in Table 10; if the distributions of variables in the prediction data set differ substantially from the given distributions, then the prediction results also may be invalid.

The rest of this chapter presents the models in three sections:

- Segment-Intersection Crashes
- Segment Related Crashes
- Major Intersection Crashes

Table 8: Definition of Predictor Variables Used

Variable	Description
Ln(AADT)	Natural log of AADT
Ln(Length)	Natural logarithm of segment length (miles)
Posted Speed	Posted speed limit (mph)
LU_Pop	Population associated with the link
LU_Retl	Retail employment associated with the link
LU_Nretl	Non-retail employment associated with the link
Res_Driway	Number of residential driveways on the link
Retl_Driway	Number of retail driveways on the link
Unsignal_Inter	Number of unsignalized intersections on the link
Ln(Trips)	Natural logarithm of trips generated per day
Pav	Total pavement width (feet)
Shld	Shoulder width (feet)

Segment-Intersection Crashes

The segment-intersection models were developed using the variables described in Table 7 for Category 2 for all three road types and are presented in Table 11. For the rural two-lane undivided roads, 319 (out of total 326) observations were used. Because the population and employment variables entered the equation for number of trips generated as logarithms, values of zero could not be used for estimation, so seven observations had to be removed in order to have an equal number of observations to compare the models using population and employment directly with those using estimated trips.

Models for this road type were estimated using number of trips as a variable as well as with and without population, retail and non-retail employment variables. The models using number of trips performed best. For segment-intersection crashes, we expected the crash exposure to be completely represented by the volume on the segment and the number of intersecting trips or the population and employment, and the length would be unimportant. This is because segment-intersection collisions must occur at minor intersections and driveways along the segment, and the length theoretically should not contribute. However, for the rural two-lane roads, when length was included, it took an exponent of 0.276, and the model had an AIC value much higher than the comparable model without length, so it is hard to dismiss this unexpected result. What this means is that segment length accounts for a significant portion of the variation in crash risk that is not explained by the exposure or the covariates. Consequently, in order to apply this model to another location (i.e., other than the CRCOG region), it is necessary to assume that the contribution of segment length will be the same there as in this data set. For this reason, both models with and without segment length are included to avoid the transferability issue; the model without segment length does not perform as well, but its coefficients are more reliably transferred to another application context.

Pavement width rather than shoulder width was significant for these models, suggesting the full pavement width is more important than how the pavement is allocated to lane and shoulder on rural roads. The model also shows that the number of accidents increases when pavement width is greater than 40 feet. This is in contrast with the finding of Hadi *et al.* (1995) and the general conception that the number of accidents decreases with an increase in pavement width, since there is more room for preventive maneuvers. Because we are dealing with segment-intersection collisions, it is possible that the wider road width encourages speeds that are unsafe for the presence of intersections and driveways. In any case, this finding suggests that the idea that widening a road is always safer should be re-examined.

Posted speed limit was also very significant and showed a decrease in the number of accidents for higher speed limits. However, this should not be interpreted as "increasing the speed limit will decrease the number of accidents." Speed limits are usually set higher on roads that are designed with higher design speeds, and roads known to have safety problems are often posted at lower speed limits. Consequently, this effect more likely captures the effect of geometric design and other general safety issues, which are not easy to represent otherwise.

Similar results were observed for the urban/suburban two-lane undivided roads. For this road type, 14 out of 587 observations had to be removed due to values of zero for the population and employment. Unlike rural two-lane roads, the shoulder width was a better predictor than overall roadway width and showed an increase in the number of accidents if the shoulder width was above 6 feet. On these roads the speeds are lower and the shoulder may be used to pass turning vehicles when it is not safe to do so. Accident frequency was also lower for roads with speed limits higher than 35 mph. Number of trips was a better predictor than population and employment variables, with the number of accidents increasing with increasing number of trips. Again, the AIC value is higher for the model with segment length, so both models are provided.

The estimated urban/suburban four-lane undivided road models are substantially different from those for the two-lane road types. Models with retail employment performed the best for these roads; population and non-retail employment were not significant. Also, there is a decrease in the number of accidents with increasing pavement width. This is probably because on four-lane roads it is not necessary to use the shoulders to pass slower or turning vehicles and instead of encouraging unsafe speeds they actually provide a margin of safety as they are designed to. The accident frequency decreases for roads with speed limits higher than 35 mph. The AADT has an exponent closer to one (unlike rural roads), suggesting that for these more heavily traveled (than rural two-lane) roads, accidents increase linearly with the traffic volume. In this context, retail employment predicts better than the total estimated trips; on such roads, retail driveways are usually much more common than residential land access, and thus appear to contribute more to the exposure to intersection accidents on these segments.

Other models for these three road types without population and employment variables are presented in Appendix B. Side by side comparison of the full models with population and employment, number of trips and number of residential, retail and non-retail driveway models are also presented in Appendix B. As discussed in Chapter 2, using number of access points as a predictor variable may appear quite attractive, but it does have some limitations. For example, counting the number of residential, retail and non-retail driveways for each road segment is an excessively labor intensive, and thus expensive, task. As well, the consistency of driveway variables is suspect since the access point count does not give an estimate of the number of vehicles accessing a road segment through each driveway. For example, a driveway serving a convenience store is generally counted as a "retail" driveway the same as is a large supermarket. The number of vehicles accessing the major artery will vary significantly in both cases and so will the exposure to traffic accidents. These model results show that population, retail, and non-retail employment counts at a TAZ level can effectively be used as surrogates for the volume entering and exiting the segment at intermediate access points. In fact, in all cases they work better than the access point models demonstrating that it is indeed logical to use the population and employment variables since they better describe the intensity of the traffic accessing the major artery.

Table 9: Observed Ranges in Estimation Data: Continuous Variables

Variable	**Rural – Two-lane – Undivided**			**Urban/Suburban – Two-lane – Undivided**			**Urban/Suburban – Four-lane – Undivided**		
	Minimum	***Maximum***	***Mean***	***Minimum***	***Maximum***	***Mean***	***Minimum***	***Maximum***	***Mean***
AADT	1014	21184	7941.23	1470	33371	10694.49	7685	34300	18729.82
Length	0.04	5.56	0.67	0.01	1.82	0.36	0.01	0.97	0.21
LU_Pop	1.43	1560.07	154.83	0.38	1261.88	153.68	0	1087.19	104.94
LU_Retl	0	57.36	5.23	0	274.19	12.44	0	198.40	19.15
LU_Nretl	0.06	825.10	37.35	0.17	1352.31	62.99	0.27	1412.05	96.95
Res_Driway	0	72	9.78	0	67	9.19	0	25	2.09
Retl_Driway	0	15	1.18	0	25	1.85	0	33	4.07
Unsignal_Inter	0	9	1.65	0	10	1.78	0	6	0.85

Table 10: Observed Frequencies in Estimation Data: Categorical Variables

Rural – Two-lane – Undivided		**Urban and Suburban–Two-lane– Undivided**		**Urban and Suburban – Four-lane – Undivided**	
Variable	***Frequency (%)***	***Variable***	***Frequency (%)***	***Variable***	***Frequency (%)***
Posted Speed < 40	21.8	Posted Speed < 35	18.9	Posted Speed < 35	20.6
Posted Speed = 40	27.3	Posted Speed = 35	38.7	Posted Speed = 35	42.9
Posted Speed > 40	50.9	Posted Speed > 35	42.4	Posted Speed > 35	36.5
Total	100	Total	100	Total	100
Pav < 30 ft	32.8	Pav < 30 ft	25.6	Pav < 50 ft	18.0
Pav >= 30 to < 40 ft	51.5	Pav >= 30 to < 40 ft	53.8	Pav >= 50 to < 60 ft	52.8
Pav >= 40 ft	15.6	Pav >= 40 ft	20.6	Pav >= 60 ft	29.2
Total	100	Total	100	Total	100
Shld < 3 ft	45.1	Shld < 3 ft	34.8	Shld < 2 ft	21.5
Shld >= 3 to < 6 ft	35.9	Shld >= 3 to < 6 ft	46.6	Shld = 2 ft	42.0
Shld >= 6 ft	19.0	Shld >= 6 ft	18.6	Shld > 2 ft	36.5
Total	100	Total	100	Total	100

Table 11: Estimated Coefficients and Fit Diagnostics: Segment-Intersection Models

Parameter	Rural–Two-lane–Undivided		Urban and Suburban – Two-lane – Undivided		Urban and Suburban – Four-lane – Undivided	
	Land Use / Driveway / Trips		***Land Use / Driveway / Trips***		***Land Use / Driveway / Trips***	
	w/ Length	w/o Length	w/ Length	w/o Length	w/ Length	w/o Length
Intercept	-4.289 (1.309)*	-6.815 (0.987)	-3.803 (0.712)	-6.190 (0.595)	***0.871 (0.858)***	***-0.380 (0.857)***
Ln(AADT)	0.647 (0.115)	0.590 (0.114)	1.103 (0.092)	0.904 (0.088)	0.985 (0.221)	1.118 (0.232)
Ln(Length)	0.276 (0.097)		0.373 (0.061)		0.431 (0.093)	
Ln(Trips)	0.387 (0.089)	0.563 (0.066)	0.258 (0.049)	0.434 (0.040)		
LU_Retl / 1000					8.026 (2.475)	11.95 (2.534)
Posted Speed < 40	***-0.031 (0.174)****	***-0.092 (0.175)***				
Posted Speed = 40	0.000 (0.000)	0.000 (0.000)				
Posted Speed > 40	-0.687 (0.161)	-0.651 (0.160)				
Posted Speed < 35			***0.119 (0.099)***	***0.042 (0.101)***	***0.165 (0.168)***	***-0.004 (0.171)***
Posted Speed = 35			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Posted Speed > 35			-0.294 (0.077)	-0.189 (0.079)	-0.449 (0.140)	-0.464 (0.146)
Pav < 30 ft	***0.205 (0.186)***	***0.265 (0.188)***				
Pav >= 30 to < 40 ft	0.000 (0.000)	0.000 (0.000)				
Pav >= 40 ft	0.496 (0.173)	0.523 (0.176)				
Pav < 50 ft					***-0.285 (0.188)***	***-0.177 (0.195)***
Pav >= 50 to < 60 ft					0.000 (0.000)	0.000 (0.000)
Pav >= 60 ft					-0.482 (0.151)	-0.653 (0.152)
Shld < 3 ft			***0.061 (0.087)***	***0.100 (0.090)***		
Shld >= 3 to < 6 ft			0.000 (0.000)	0.000 (0.000)		
Shld >= 6 ft			0.184 (0.095)	***0.157 (0.097)***		
Dispersion	1.058 (0.088)	1.087 (0.090)	0.606 (0.039)	0.646 (0.041)	0.777 (0.076)	0.850 (0.081)
AIC	80.86	74.89	139.57	105.43	55.98	38.11

* Estimate (Standard Error)

** Bold Italic formatting represents statistically insignificant variables (95% confidence)

Segment-Related Crashes

The models for segment-related crashes on the rural two-lane undivided roads are presented in Table 12. Neither the population and employment variables nor the number of trips generated by the population and employment were significant for these types of accidents, which is not surprising, as these variables were intended to represent exposure to segment-intersection collisions, not to segment-related collisions. The possible effect of these variables as covariates for the segment-related crashes is most likely as density per mile of the segment. Thus these variables were divided by the length of the segment to represent intensity of land development in all segment-related crash models.

Similar to the segment-intersection accidents for this road type, the frequency of segment-related accidents increased when pavement width was greater than 40 feet. The traffic volume had a low exponent as compared to the other road types. Due to better design and more space for preventive maneuvers, the roads with posted speed limit greater than 40 mph have lower accident risk. Models were estimated both with and without the exponent on length fixed at 1.0 (entered as an offset). Issues related to an exponent on length other than 1.0 are explained later. The estimated value is 0.889, just barely significantly different from 1.0, indicating that the crash risk is slightly higher on shorter segments than on longer segments with all other factors identical.

Table 12: Estimated Coefficients and Fit Diagnostics: Segment-related Crashes on Rural Two-lane Undivided Roads

Parameter	Natural Log of Length as Covariate	Natural Log of Length as Offset
Intercept	1.992 (0.254)*	2.037 (0.258)
Ln(AADT)	0.393 (0.079)	0.416 (0.080)
Ln(Length)	0.889 (0.053)	1.000
Posted Speed < 40	***-0.027 (0.126)*** **	***-0.034 (0.128)***
Posted Speed = 40	0.000 (0.000)	0.000 (0.000)
Posted Speed > 40	-0.466 (0.111)	-0.525 (0.110)
Pav < 30 ft	***0.118 (0.117)***	***0.085 (0.118)***
Pav >= 30 to < 40 ft	0.000 (0.000)	0.000 (0.000)
Pav >= 40 ft	0.296 (0.128)	0.339 (0.129)
Dispersion	0.468 (0.047)	0.475 (0.047)
AIC	210.51	208.15

* Estimate (Standard Error)

** Bold Italic formatting represents statistically insignificant variables

The category 3 crash models for urban/suburban two-lane undivided road segments are presented in Table 13. The model results showed that the number of accidents is higher for roads with posted speed limits below 35 and lower for posted speeds above that. The models with population and employment variables work only slightly better than the models without them, showing that the population and employment does not have a great effect on segment-related crashes. Shoulder width is a better predictor for this road type than overall pavement width (similar to segment-intersection accidents for this road type), probably because there is more variation in shoulder width for urban/suburban two-lane roads than there is for pavement width in the dataset. The models with length as offset perform poorly as compared to the models where the exponent of length was allowed to vary, again suggesting a correlation between segment length and some unobserved factor that influences crash incidence.

Although population and employment variables improve the regression models slightly, the clear effect of these variables on the segment-related crashes is not understood. Thus both the models with and without land use are presented for better transferability outside the CRCOG region. The models with length as an exponent have an inherent problem. Whenever length is raised to a power other than one, the prediction is affected by the procedure used to split the segments i.e. more short segments increase the prediction.

Table 13: Estimated Coefficients and Fit Diagnostics: Segment-related Crashes on Urban/Suburban Two-Lane Undivided Roads

Parameter	Natural Log of Length		Natural Log of Length as offset	
	No Land Use	***Land Use***	***No Land Use***	***Land Use***
Intercept	2.141 (0.261)*	2.174 (0.260)	2.221 (0.269)	2.245 (0.266)
Ln(AADT)	0.305 (0.078)	0.243 (0.079)	0.361 (0.079)	0.284 (0.080)
Ln(Length)	0.799 (0.046)	0.825 (0.046)	1.000	1.000
LU_Pop/1000*Len		0.263 (0.084)		0.294 (0.086)
LU_Retl/1000*Len		0.894 (0.473)		1.059 (0.481)
Posted Speed < 35	0.218 (0.090)	0.246 (0.089)	0.263 (0.091)	0.288 (0.090)
Posted Speed = 35	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Posted Speed > 35	-0.216 (0.072)	-0.166 (0.072)	-0.285 (0.072)	-0.220 (0.072)
Shld < 3 ft	0.174 (0.079)	0.184 (0.078)	0.166 (0.081)	0.180 (0.079)
Shld >= 3 to < 6 ft	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Shld >= 6 ft	0.247 (0.087)	0.235 (0.086)	0.265 (0.089)	0.250 (0.087)
Dispersion	0.431 (0.034)	0.415 (0.033)	0.445 (0.034)	0.424 (0.033)
AIC	245.73	258.13	229.29	246.11

* Estimate (Standard Error)

Table 14 presents the segment-related crash models for urban/suburban four-lane undivided roads. The results were very similar to other urban/suburban roads with slight improvement in the model performance if the exponent on length was allowed to vary. The expected relation with posted speed limit was observed (i.e. the number of accidents decrease with increasing speed limits due to the fact the higher speed limits are posted on roads designed for higher speeds).

Table 14: Estimated Coefficients and Fit Diagnostics: Segment-related Crashes on Urban/Suburban Four-Lane Undivided Roads

Parameter	Natural Log of Length		Natural Log of Length as offset	
	No Land Use	***Land Use***	***No Land Use***	***Land Use***
Intercept	2.304 (0.715)*	2.143 (0.708)	3.008 (0.693)	2.835 (0.690)
Ln(AADT)	0.416 (0.182)	0.412 (0.179)	0.327 (0.185)	0.323 (0.182)
Ln(Length)	0.776 (0.079)	0.780 (0.076)	1.000	1.000
LU_Pop/1000*Len		0.193 (0.100)		0.192 (0.104)
LU_Nretl/1000*Len		0.126 (0.068)		0.134 (0.071)
Posted Speed < 35	0.360 (0.151)	***0.220 (0.152)***	0.468 (0.151)	0.326 (0.152)
Posted Speed = 35	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Posted Speed > 35	-0.377 (0.123)	-0.294 (0.123)	-0.370 (0.127)	-0.284 (0.127)
Pav < 50 ft	-0.461 (0.165)	-0.472 (0.163)	-0.516 (0.169)	-0.527 (0.166)
Pav >= 50 to < 60 ft	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Pav >= 60 ft	***-0.189 (0.128)*****	***-0.176 (0.125)***	***-0.139 (0.131)***	***-0.128 (0.128)***
Dispersion	0.519 (0.062)	0.487 (0.060)	0.545 (0.064)	0.512 (0.062)
AIC	86.70	91.90	80.85	85.90

* Estimate (Standard Error)

** Bold Italic formatting represents statistically insignificant variables (95% confidence)

As stated earlier, the effect of population and non-retail employment on segment-related accidents is not clearly understood. Thus it is suggested that outside the CRCOG region the model with these variables should not be used. Similarly, if the delineation of the links is not similar to the CRCOG region

(i.e. as connectors between major intersections), then the models with the natural log of length as an offset (exponent = 1.0) should be used for reliable predictions.

Intersection Crashes

Models for three and four leg intersections were estimated separately and several variables were tested for these models including skewness of the intersection, control type on each leg, presence/absence of median, etc. Most of these variables were collected using the Photo Log. Very few of these variables were actually significant.

Table 15 presents the model containing the most significant variables (which also had the best AIC) for three-leg intersections. The models were estimated by backward elimination of the insignificant variables from the full model containing all the variables defined in Table 7 for category 1 accidents. Exclusion of any more variables from the final model reduced the AIC significantly. As expected, the traffic volume on both major and minor links was significant and the exponent on the minor road volume was smaller than that on the major road volume. The categorical area type variables (not the population and employment) showed that the crash risk in urban areas is not significantly different from the suburban area and the crash risk is lower in rural areas. Also, the number of accidents is lower if the major link has two lanes than if it has three or more lanes. The intersections with speed limits lower than 30 mph on the major road have higher crash risk, probably because they are more likely to have lower design speeds, often due to narrower lanes or other restrictive geometry or roadside objects.

Table 15 also presents the model for four-leg intersection crashes, which is similar to the three-leg model. Both major and minor traffic volumes were significant and the urban land use was not significantly different from the suburban land use. Only the major control type was significant for this intersection type and, as expected, intersections with some control (signal or stop) have lower crash risk than intersections with no control on the major road approaches, but stop control on the minor road approach.

Table 15: Estimated Coefficients and Fit Diagnostics: Intersection Crashes at Three-leg and Four-leg Intersections

Parameter	Three-leg Estimate (SE)	Four-leg Estimate (SE)
Intercept	1.139 (0.529)*	***0.044 (0.428)***
Ln(Major AADT)	0.585 (0.145)	0.686 (0.122)
Ln(Minor AADT)	0.201 (0.114)	0.572 (0.110)
LU_Urban	***0.274 (0.244)*****	***0.136 (0.120)***
LU_SubUrban	0.000 (0.000)	0.000 (0.000)
LU_Rural	-0.290 (0.147)	-0.342 (0.165)
Major Lanes = 2	-0.280 (0.168)	
Major Lanes = 3	0.000 (0.000)	
Major Lanes >= 4	***0.068 (0.166)***	
Major Speed <= 30	0.382 (0.180)	
Major Speed > 30	0.000 (0.000)	
Maj. Control = None		0.783 (0.352)
Maj. Control = Stop		0.000 (0.000)
Maj. Control = Signal		***0.081 (0.308)***
Dispersion	0.183 (0.040)	0.160 (0.029)

* Estimate (Standard Error)

** Bold Italic formatting represents statistically insignificant variables (95% confidence)

Validation on Maine Data

The crash models developed for the Connecticut dataset were used to predict accidents in the Maine data set; the results of comparing the predicted with the observed values are presented here. The Maine data set consists of 10,000 rural two-lane road segments, 4600 urban two-lane road segments, and only 60 urban four-lane road segments. No intersection data was available, so those models are not tested here. Tables 16 and 17 present the Pearson coefficient of correlation between the predicted and observed

crashes for two categories: segment-intersection and segment-related crashes. This shows that when conservative models (without length for segment-intersection crashes and with length as an offset and no land use for segment-related crashes) are used, there is significant correlation between the observed and predicted crashes, although the relatively low values indicate the correlation is not very strong. The correlation is best for the rural two-lane road models, and worst for the urban/suburban four-lane road models; and generally better for the segment-related crashes than the segment-intersection crashes. Surprisingly, the models with the highest AIC for Connecticut data have slightly higher correlation than the conservative models, contrary to expectation due to concerns about transferability of the relationships between crashes and segment length or population and employment.

Figure 12 through 17 present the cumulative residual (observed value less the predicted value) plotted against the AADT for the Maine data. In order to generate a cumulative residual plot for a variable (AADT in our case), the residual is computed for each road segment. The segments are then sorted in the ascending order of the variable. Once so sorted, the residuals are cumulated and plotted against the variable (i.e. AADT) as shown in figures below (Hauer 2004). For example in Figure 12, the sum of all residuals up to AADT=5000 is about -550 accidents. This means that we have predicted an additional 550 accidents in total on all road segments with AADT less than 5000.

These plots show that the Connecticut models either severely over- or under-predict the accidents, and there is a systematic bias in all cases. This is likely because the ranges of exposure values (i.e., AADT) in the Maine data on each type of road are significantly different from those in Connecticut. For example, the AADT's in Maine range from near 0 to 10,000 while for similar road types in Connecticut the AADT ranges from 1000 to 20,000. These results show that it is not advisable to use the Connecticut models in a state which primarily consists of rural roads and thus has significant different driver behavior and traffic characteristics. In this case, mismatch in the range of AADT is more problematic than the unexpected covariate relationships mentioned above. Also, the relationship between crashes and the population and employment values appears to be quite different for Maine than for Connecticut. This is not surprising, and may be due to possible differences in per person and employee trip generation in the two States. These results indicate that it is necessary to estimate separate models for Maine to properly capture the effects of each variable on crash incidence.

Table 16: Validation results – Conservative Models Suggested for General Use

Category	***Road Type***	***Pearson Correlation Coeff.***
Segment-intersection	Rural Two-lane	0.404
	Urban/suburban Two-lane	0.360
	Urban/suburban Four-lane	0.083
Segment-related	Rural Two-lane	0.700
	Urban/suburban two-lane	0.622
	Urban/suburban Four-lane	0.467

Table 17: Validation results – Models with Highest AIC for CT Data

Category	***Road Type***	***Pearson Correlation Coeff.***
Segment-intersection	Rural Two-lane	0.469
	Urban/suburban Two-lane	0.402
	Urban/suburban Four-lane	0.324
Segment-related	Rural Two-lane	0.704
	Urban/suburban two-lane	0.563
	Urban/suburban Four-lane	0.536

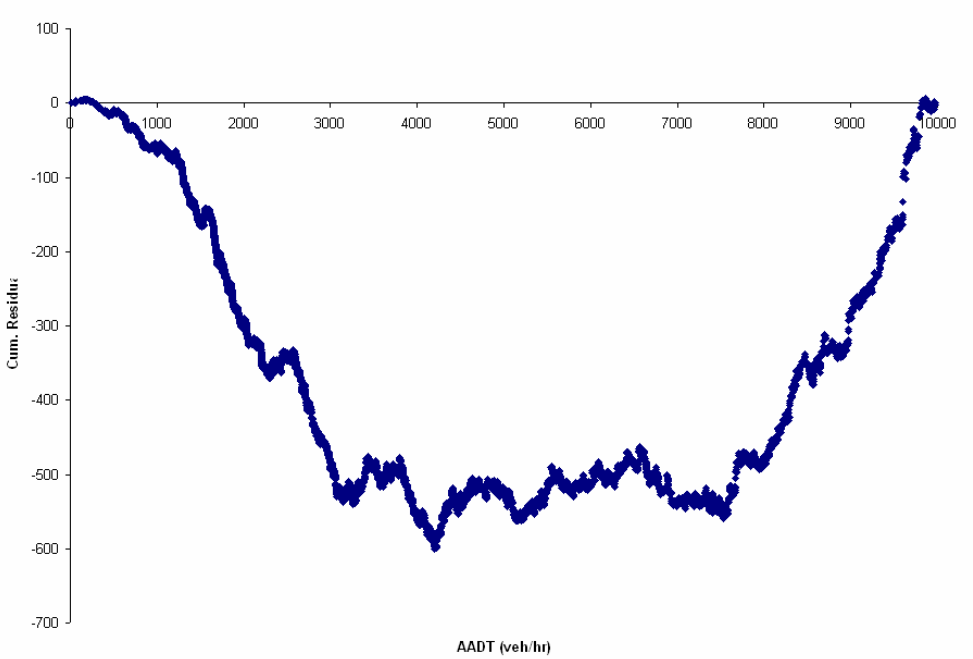


Figure 12: Cumulative residual plot Segment-intersection crashes – Rural two-lane roads – Conservative Models

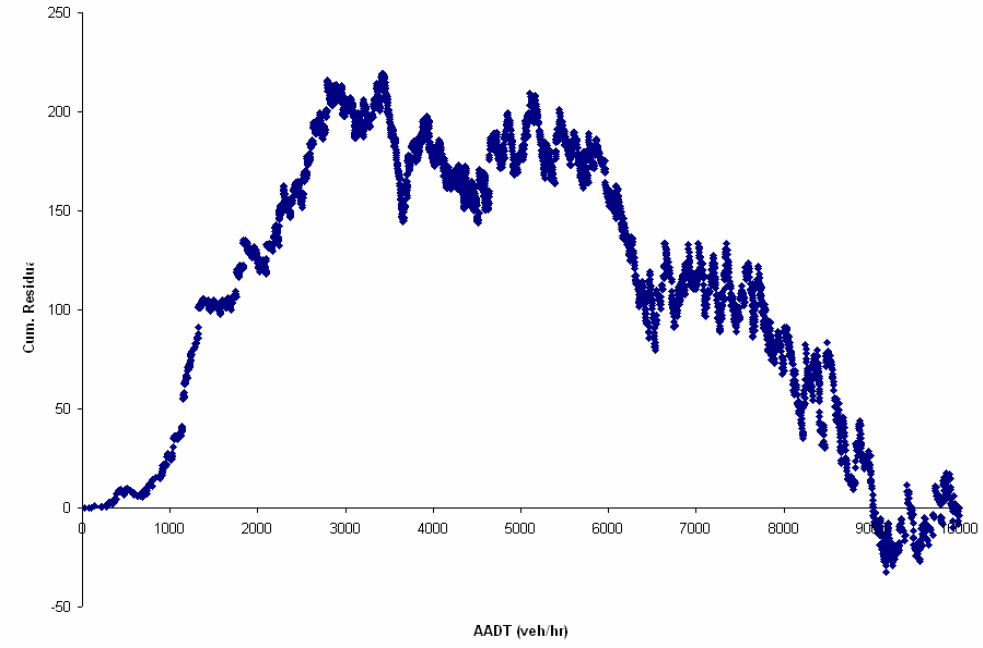


Figure 13: Cumulative residual plot Segment-intersection crashes – Urban two-lane roads – Conservative Models

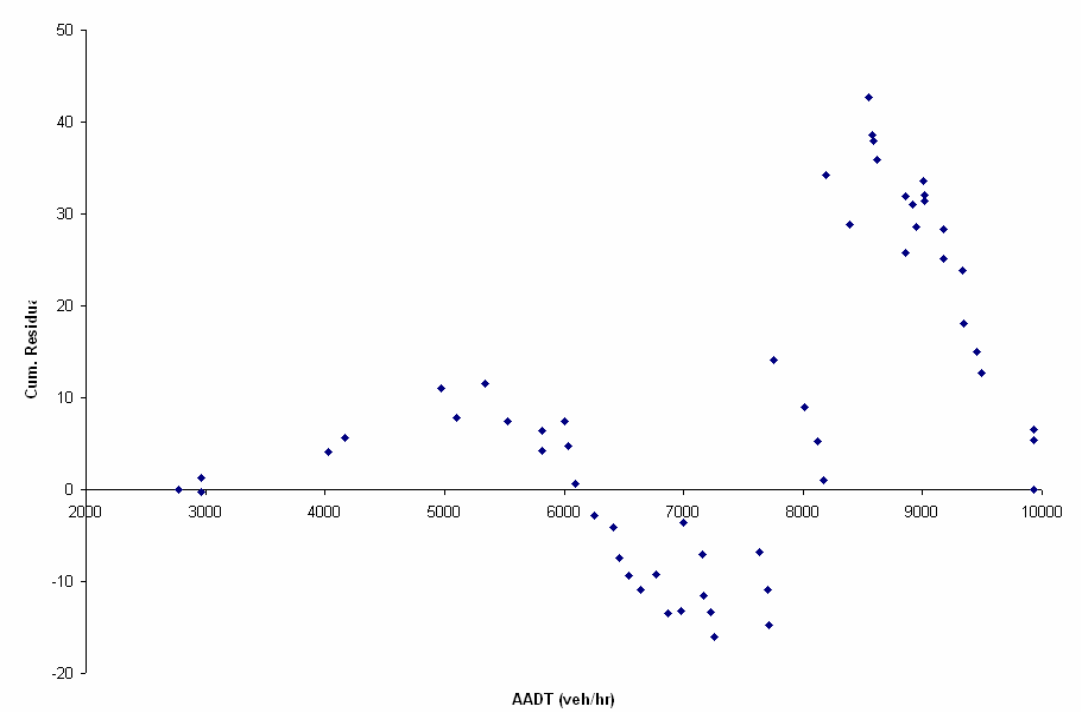


Figure 14: Cumulative residual plot Segment-intersection crashes – Urban four-lane roads – Conservative Models

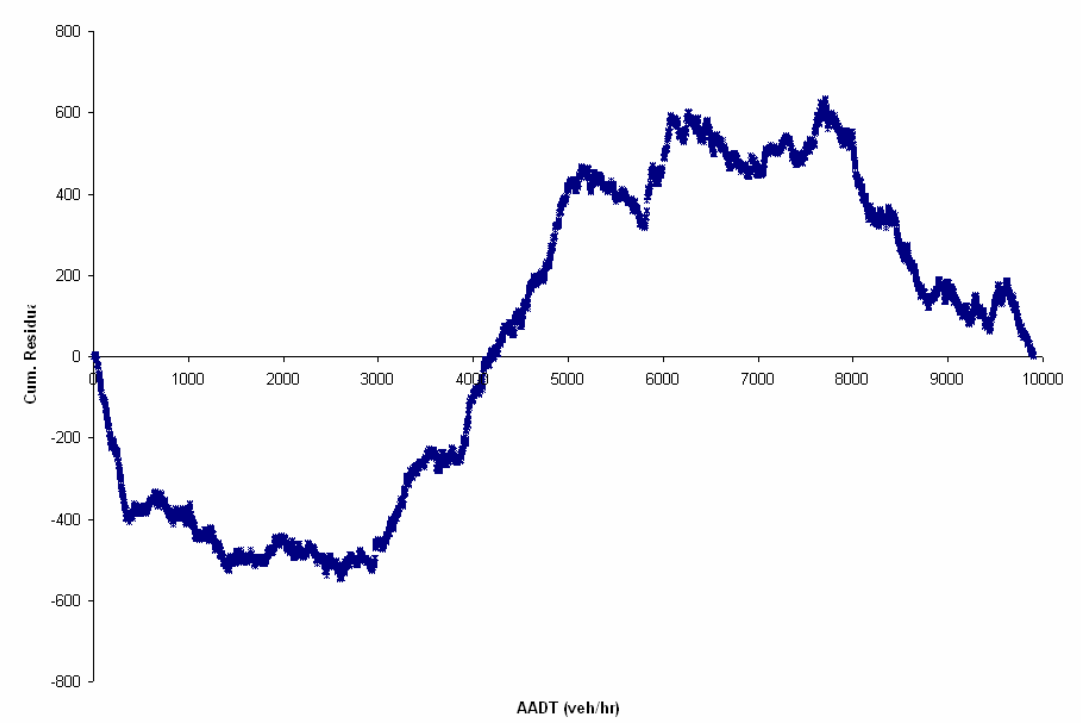


Figure 15: Cumulative residual plot Segment-related crashes – Rural two-lane roads – Conservative Models

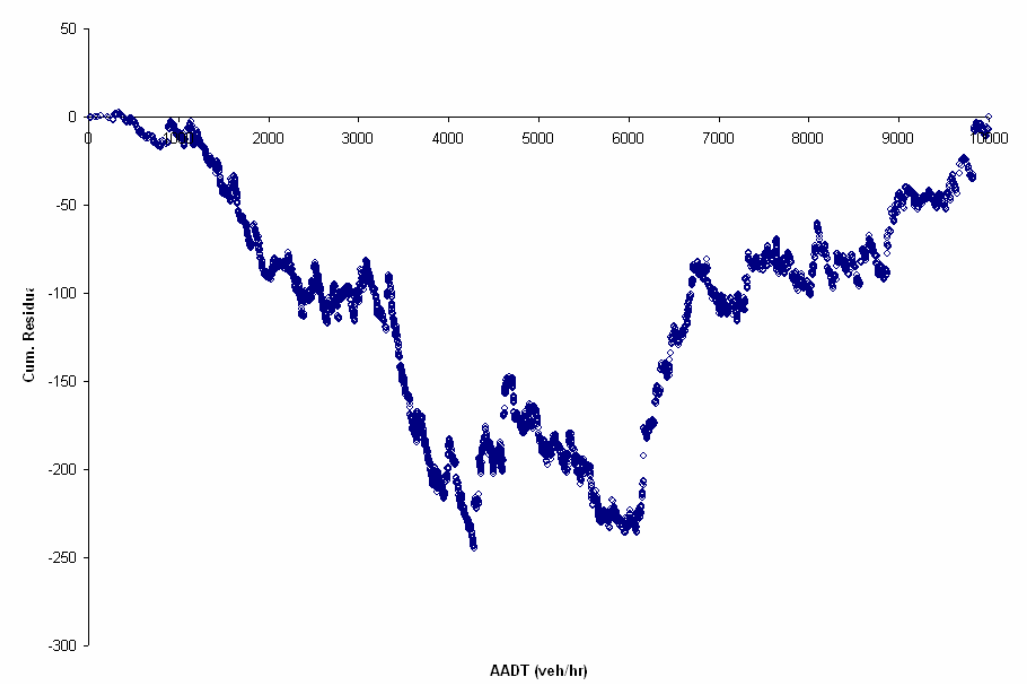


Figure 16: Cumulative residual plot Segment-related crashes – Urban two-lane roads – Conservative Models

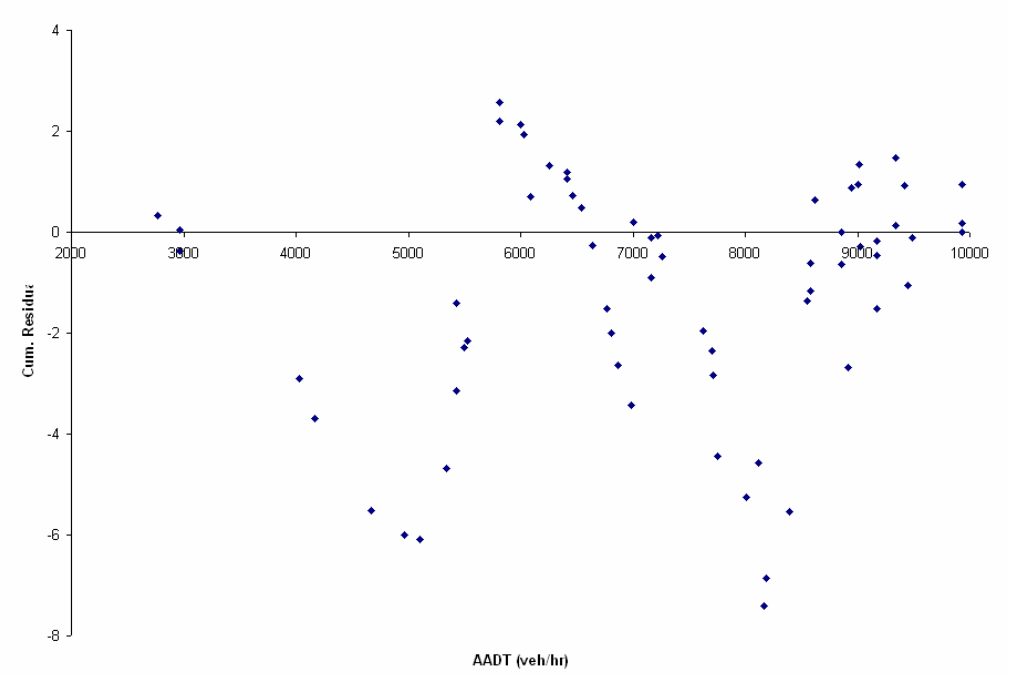


Figure 17: Cumulative residual plot Segment-related crashes – Urban four-lane roads – Conservative Models

Because the validation exercise showed that the Connecticut models did not replicate the observations in the Maine data well, we decided to estimate models using the Maine data to see what differences might arise in the parameters. Only four models were estimated: for segment-intersection and segment-related collisions on two-lane rural and urban/suburban roads. Insufficient observations were available for estimating any four-lane road models. We decided to only estimate the conservative models (i.e., more reliable, without questionable variables) using the same covariates as for the Connecticut models. Tables 18 and 19 present the results of the model estimation along with the comparable Connecticut models.

There are numerous differences between the Maine and Connecticut models, especially in the segment-intersection models. The intercepts in the Maine models are only about half the value of the Connecticut models. For the segment-intersection models they are all negative, indicating that the unexplained factors reduce the crash risk less in Maine than in Connecticut. Conversely, the intercepts are positive on the segment-related models, so that the unexplained factors increase the crash risk more in the Connecticut data than in the Maine data.

Table 18: Comparing Maine and Connecticut Models: Segment-Intersection Crashes

Parameter	**Rural–Two-lane–Undivided**		**Urban and Suburban – Two-lane – Undivided**	
	Maine	***Connecticut***	***Maine***	***Connecticut***
Intercept	-3.349 (0.149)*	-6.815 (0.987)	-3.842 (0.249)	-6.190 (0.595)
Ln(AADT)	0.467 (0.020)	0.590 (0.114)	0.377 (0.031)	0.904 (0.088)
Ln(trips)	0.234 (0.010)	0.563 (0.066)	0.280 (0.016)	0.434 (0.040)
Posted Speed < 40	***0.080 (0.052)****	***-0.092 (0.175)***		
Posted Speed = 40	0.000 (0.000)	0.000 (0.000)		
Posted Speed > 40	***-0.039 (0.050)***	-0.651 (0.160)		
Posted Speed < 35			***0.075 (0.063)***	***0.042 (0.101)***
Posted Speed = 35			0.000 (0.000)	0.000 (0.000)
Posted Speed > 35			***-0.031 (0.080)***	-0.189 (0.079)
Pav < 30 ft	***0.001 (0.027)***	***0.265 (0.188)***		
Pav >= 30 to < 40 ft	0.000 (0.000)	0.000 (0.000)		
Pav >= 40 ft	***0.022 (0.028)***	0.523 (0.176)		
Shld < 3 ft			-0.119 (0.046)	***0.100 (0.090)***
Shld >= 3 to < 6 ft			0.000 (0.000)	0.000 (0.000)
Shld >= 6 ft			***-0.080 (0.061)***	***0.157 (0.097)***
Dispersion	0.157 (0.009)	1.087 (0.090)	0.305 (0.016)	0.646 (0.041)

* Estimate (Standard Error)

** Bold Italic formatting represents statistically insignificant variables (95% confidence)

The exponents on AADT are comparable between the two State models in all four cases except for the segment-intersection collision model for urban/suburban roads, where the exponent on the Maine model is less than half that of the one in the Connecticut model. The same is true for the exponent on trips (entering and exiting minor intersection and driveways) in the rural road segment-intersection collision models. In more practical terms, when the exponent on volume for one case is smaller than that for another, the crash risk in the first case is higher than the second at lower traffic volumes, and lower than the second at higher traffic volumes. This effect is illustrated in Figure 18 for the exponent on AADT for the urban and suburban road segment-intersection collision models. The slope of a line drawn from the origin to any point on one of these lines is the crash risk at that level of AADT for that model, so that a steeper slope indicates a greater crash risk. The risk of segment-intersection collisions on urban-suburban roads in Maine is much higher at lower values of AADT than at higher, possibly because speeds on these roads are higher, or because drivers entering and exiting driveways and minor roads do not expect to meet vehicles on the main road and thus take greater risks in making gap acceptance decisions. Similarly, the lower exponent on minor intersection and driveway entering and exiting trips in the Maine models indicates a higher risk at

lower values of this factor, suggesting drivers on the main road do not expect traffic to enter or leave the road at these locations and are thus less prepared to react when a vehicle turns in or out of the road.

Table 19: Comparing Maine and Connecticut Models: Segment-Related Crashes

Parameter	Rural–Two-lane–Undivided		Urban and Suburban – Two-lane – Undivided	
	Maine	***Connecticut***	***Maine***	***Connecticut***
Intercept	1.225 (0.048)*	2.037 (0.258)	1.924 (0.101)	2.221 (0.269)
Ln(AADT)	0.595 (0.012)	0.416 (0.080)	0.330 (0.032)	0.361 (0.079)
Ln(Length)	1.000	1.000	1.000	1.000
Posted Speed < 40	***0.024 (0.042)*** **	***-0.034 (0.128)***		
Posted Speed = 40	0.000 (0.000)	0.000 (0.000)		
Posted Speed > 40	***-0.036 (0.037)***	-0.525 (0.110)		
Posted Speed < 35			0.099 (0.058)	0.263 (0.091)
Posted Speed = 35			0.000 (0.000)	0.000 (0.000)
Posted Speed > 35			0.119 (0.068)	-0.285 (0.072)
Pav < 30 ft	0.169 (0.018)	***0.085 (0.118)***		
Pav >= 30 to < 40 ft	0.000 (0.000)	0.000 (0.000)		
Pav >= 40 ft	-0.171 (0.023)	0.339 (0.129)		
Shld < 3 ft			***0.053 (0.044)***	0.166 (0.081)
Shld >= 3 to < 6 ft			0.000 (0.000)	0.000 (0.000)
Shld >= 6 ft			-0.116 (0.058)	0.265 (0.089)
Dispersion	0.165 (0.006)	0.475 (0.047)	0.189 (0.015)	0.445 (0.034)

* Estimate (Standard Error)

** Bold Italic formatting represents statistically insignificant variables (95% confidence)

The coefficients on posted speed and pavement or shoulder width are substantially different from one State to the other, even taking opposite signs in three cases. Posted speed was not significant in the Maine models except for segment-related collisions on urban and suburban roads. On these roads, the accident risk is higher for any speed limits greater or less than 35 mph, indicating a U-shape curve in the plot. The higher risk on roads with lower speed limits may be observed because roads that are known to be unsafe are assigned lower speed limits; for roads with higher speed limits, the higher risk may be due to the higher speed actually not being safe for the urban conditions. This effect was not observed in the Connecticut model, which indicates a decreasing relationship with speed limit for this road type and this type of collision. On the other hand, for the same type of collision, the accident risk has a decreasing trend for shoulder width in the Maine model, while for the Connecticut model the U-shape was found. Thus, in Connecticut, shoulders wider than 6 ft on urban and suburban roads actually increase the risk of segment-related collisions, while in Maine they reduce the risk. The conflicting combinations of these two effects suggests that the Connecticut and Maine driving environments are different from one another with respect to speed limit and shoulder width, which could include driver behavior and/or local road design and traffic practice. Without a detailed crash causality analysis, no more can be inferred about these effects.

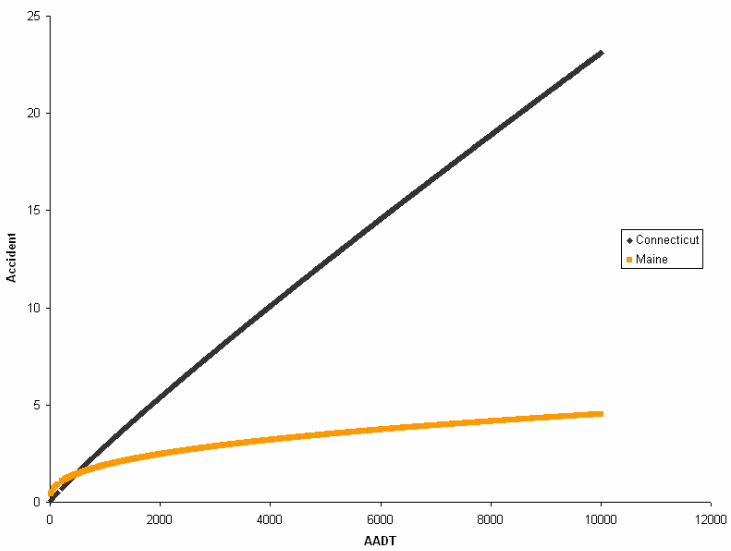


Figure 18: Urban/suburban Segment-Intersection Crashes – Connecticut v. Maine

Consequently, we have decided to report two sets of models for two-lane roads, one estimated from Connecticut data and one from Maine data. The user must decide which set of models best matches the conditions for which he/she wishes to predict accidents. One logical conclusion would be to apply the Connecticut models in the southern tier states (Massachusetts, Connecticut and Rhode Island) and the Maine models in the northern tier states (Vermont, New Hampshire and Maine). The Maine models might also be more appropriate for western Massachusetts and northwestern Connecticut. It is up to the user to decide which models to use for his/her context.

Chapter Summary

This chapter presented the models for the three crash categories subdivided into the three road types, and the two intersection types. Some interesting results were observed from these models such higher crash risk with large pavement and shoulder widths for two-lane road segments, which is in contrast with the conventional belief that wider pavements provide more room for preventive maneuvers and thus reduce accidents. Another important issue is the transferability of the segment-related crash models with population and employment variables to other areas. It is suggested that since this effect is not clearly understood, the models without these variables should be used for more reliable prediction. Similar issues are associated with the models having an exponent on length different from 1.0. These models should be used only when the links are defined as connectors between major intersections as in the CRCOG area. Presented below is the list of simplest, most reliable models (with length as an offset and without population and employment for the segment models) for each road type in equation form for easy application.

In order to use these equations for different time periods a proportionality constant and a variable for the number of years used in prediction is also added in these models. The proportionality constant allows the user to simply input the AADT in vehicles per hour per day without any modification for the number of years or the number of days in a year. The parameter for number of years allows the prediction for different time periods. Kindly note that the values for traffic volume and the number of trips in the models presented in the tables are scaled for seven year crash data but the values in the equations presented below are per day (AADT and number of trips per day).

Three-leg intersection:

$$y_1 = (1.87 \times 10^{-4})(Z^2)(MAADT)^{0.585}(mAADT)^{0.201} \exp[1.139 + 0.274(URB) - 0.290(RUR) - 0.280(ML2) + 0.068(ML4) + 0.382(MS30)]$$

Four-leg intersection:

$$y_1 = (1.12 \times 10^{-5})(Z^2)(MAADT)^{0.686}(mAADT)^{0.572} \exp[0.044 + 0.136(URB) - 0.342(RUR) + 0.783(MCN) + 0.081(MCS)]$$

Where:

- y_1 = The number of category 1 crashes;
- Z = Number of years used for prediction;
- *MAADT* = The higher of the two intersecting road AADT's, or the AADT on the road with the higher functional classification;
- *mAADT* = The lower of the two intersecting road AADT's, or the AADT on the road with the lower functional classification;
- *URB* = 1 if the surrounding land use is urban;
- *RUR* = 1 if the surrounding land use is rural;
- *ML2* = 1 if the total number of lanes on the major road is equal to 2;
- *ML4* = 1 if the total number of lanes on major road is greater than or equal to 4;
- *MS30* = 1 if the posted speed limit on major road is less than or equal to 30 mph;
- *MCN* = 1 if there is no control on the major road;
- *MCS* = 1 if there is a traffic signal control on the major road;

<u>*Rural undivided two-lane segment-intersection and segment-related crashes*</u>

Connecticut:

$$y_2 = (5\times10^{-2})(Z^2)(AADT)^{0.590}(T)^{0.563}\exp[-6.815 - 0.092(PSLT40) - 0.651(PSGT40) + 0.265(PWLT30) + 0.523(PWGT40)]$$

$$y_3 = (1.19\times10^{-2})(Z)(AADT)^{0.416}(L)^{1.0}\exp[2.037 - 0.034(PSLT40) - 0.525(PSGT40) + 0.085(PWLT30) + 0.339(PWGT40)]$$

Maine:

$$y_2 = (7.88\times10^{-3})(Z^2)(AADT)^{0.467}(T)^{0.234}\exp[-3.349 + 0.080(PSLT40) - 0.039(PSGT40) + 0.001(PWLT30) + 0.022(PWGT40)]$$

$$y_3 = (4.1\times10^{-3})(Z)(AADT)^{0.595}(L)^{1.0}\exp[1.225 + 0.024(PSLT40) - 0.036(PSGT40) + 0.169(PWLT30) - 0.171(PWGT40)]$$

Where:

- y_2 = The number of category 2 crashes;
- y_3 = The number of category 3 crashes;
- Z = Number of years used for prediction;
- *AADT* = The AADT on the main road;
- T = The estimate of the total number of trips entering and leaving the segment per day due to the adjacent land development;
- L = The length of the segment in miles;
- *PSLT40* = 1 if the posted speed limit on the major road is less than 40 mph;
- *PSGT40* = 1 if the posted speed limit on the major road is greater than 40 mph;
- *PWLT30* = 1 if the total pavement width is less than 30 ft;
- *PWGT40* = 1 if the total pavement width is greater than or equal to 40 ft;

<u>*Urban/suburban undivided two-lane segment-intersection and segment-related crashes*</u>

Connecticut:

$$y_2 = (2.79\times10^{-3})(Z^2)(AADT)^{0.904}(T)^{0.434}\exp[-6.190 + 0.042(PSLT35) - 0.189(PSGT35) + 0.100(SWLT3) + 0.157(SWGT6)]$$

$$y_3 = (1.66\times10^{-2})(Z)(AADT)^{0.361}(L)^{1.0}\exp[2.221 + 0.263(PSLT35) - 0.285(PSGT35) + 0.166(SWLT3) + 0.265(SWGT6)]$$

Maine:

$$y_2 = (1.93\times10^{-2})(Z^2)(AADT)^{0.377}(T)^{0.280}\exp[-3.842 + 0.075(PSLT35) - 0.031(PSGT35) - 0.119(SWLT3) - 0.08(SWGT6)]$$

$$y_3 = (1.99\times10^{-2})(Z)(AADT)^{0.330}(L)^{1.0}\exp[1.924 + 0.099(PSLT35) + 0.119(PSGT35) + 0.053(SWLT3) - 0.116(SWGT6)]$$

Where:

- y_2 = The number of category 2 crashes;

- y_3 = The number of category 3 crashes;
- Z = Number of years used for prediction;
- $AADT$ = The AADT on the main road;
- T = The estimate of the total number of trips entering and leaving the segment per day due to the adjacent land development;
- L = The length of the segment in miles;
- $PSLT35$ = 1 if the posted speed limit on the major road is less than 35 mph;
- $PSGT35$ = 1 if the posted speed limit on the major road is greater than 35 mph;
- $SWLT3$ = 1 if the outer shoulder width is less than 3 ft;
- $SWGT6$ = 1 if the outer shoulder width is greater than or equal to 6 ft;

Urban/suburban undivided four-lane segment-intersection and segment-related crashes

$$y_2 = (1.80 \times 10^{-4})(Z)(AADT)^{1.118} \exp[-0.380 - 0.004(PSLT35) - 0.464(PSGT35) + 11.95(REmp) - 0.177(PWLT50) - 0.653(PWGT60)]$$

$$y_3 = (2.03 \times 10^{-2})(Z)(AADT)^{0.327}(L)^{1.0} \exp[3.008 + 0.468(PSLT35) - 0.370(PSGT35) - 0.516(PWLT50) - 0.139(PWGT60)]$$

Where:

- y_2 = The number of category 2 crashes;
- y_3 = The number of category 3 crashes;
- Z = Number of years used for prediction;
- $AADT$ = The AADT on the main road;
- $REmp$ = The number of retail employment associated with the segment multiplied by 10^{-3};
- L = The length of the segment in miles;
- $PSLT35$ = 1 if the posted speed limit on the major road is less than 35 mph;
- $PSGT35$ = 1 if the posted speed limit on the major road is greater than 35 mph;
- $PWLT50$ = 1 if the total pavement width is less than 50 ft;
- $PWGT60$ = 1 if the total pavement width is greater than or equal to 60 ft;

In order to visually compare the results obtained from these models to the observed accidents, an Accident Model User Interface (AMUI) was developed (Appendix C). AMUI is a Geographic Information System (GIS)-integrated interface. The purpose of this interface is to provide users with a tool that presents accident model results on a map. AMUI is a map document that contains two customized modules: Link Accident Module and Intersection Accidents Module. The Link Accident Module adds the segment-intersection and segment-related crashes and compares them against the observed accidents on the road segment atleast 250 feet away from any major intersection. The detailed example of this application can be found in Appendix C.

The objective of this study was to develop network-based models where accidents are divided into subgroups by occurrence location (relative to major intersections) and then by crash type. Since traffic volumes and other important information for minor roads were not readily available, the accidents which occur farther away from a major intersection (>250 feet) were divided into two categories: segment-intersection and segment-related crashes based on the crash type. The Negative Binomial distribution was assumed for accident frequency since it relaxes the Poisson distribution's constraint of equal variance and mean. Models with and without segment length as a covariate were estimated for the segment-intersection crashes, and with and without population and employment data for the segment crashes, for more flexibility in transferring models to other contexts, since the respective effect of these variables on each type of accident was not clearly understood. The road types in the CRCOG region with a sufficient number of observations to estimate regression models were rural two-lane, urban/suburban two-lane, and urban/suburban four-lane undivided roads. Separate models were also estimated for crashes at three and four leg intersections.

It was shown that the land development (population, retail, non-retail employment and the number of trips generated by them) in the areas surrounding the links can act as a surrogate for exposure to segment-intersection collisions in lieu of traffic volume and other information about the minor roads and driveways. It was observed that the number of trips (for rural two-lane and urban/suburban two-lane undivided roads) and retail employment (for urban/suburban four-lane undivided roads) were better predictors for the crash models than the number of driveways (by type). For segment-intersection crashes, we expected the crash exposure to be completely represented by the volume on the segment and the number of intersecting trips or the population and employment, and the length would be unimportant. This is because segment-intersection collisions must occur at minor intersections and driveways along the segment, and the length theoretically should not contribute. However, for all road types, when length was included, it took an exponent different from 1.0, and the model had an AIC value much higher than the comparable model without length, so it is hard to dismiss this unexpected result. What this means is that segment length accounts for a significant portion of the variation in crash risk that is not explained by the exposure or the covariates.

The two-lane road models also showed that the number of accidents increases when pavement (for rural) or shoulder (for urban/suburban) width is greater than 40 ft and 6 ft, respectively. This is in contrast with the general conception that the number of accidents decreases with an increase in pavement width, since there is more room for preventive maneuvers. Because we are dealing with segment-intersection collisions, it is possible that the wider road width encourages speeds that are unsafe for the presence of intersections and driveways. In any case, this finding suggests that the idea that widening a road is always safer should be re-examined. For urban/suburban four lane roads there was a decrease in the number of accidents with pavement width greater than 60 ft. This is probably because on four-lane roads it is not necessary to use the shoulders to pass slower or turning vehicles and instead of encouraging unsafe speeds they actually provide a margin of safety as they are designed to. Similar effects for pavement and shoulder width were observed for segment-related crashes.

Segment-related crashes for urban/suburban roads performed slightly better with the population and employment data, even though this effect is not intuitive. Thus it is suggested that outside the CRCOG region the model with these variables should not be used. Similarly, if the delineation of the links is not similar to the CRCOG region (i.e. as connectors between major intersections), then the models with the natural log of length as an offset (exponent = 1.0) should be used for more reliable prediction.

For the intersection crashes, the traffic volume on both major and minor links was significant and the exponent on the minor road volume was smaller than that on the major road volume. The categorical area type variables showed that the crash risk in urban areas is not significantly different from the suburban area and the crash risk is lower in rural areas. For three-leg intersections the number of lanes and the speed limit on the major road were important in explaining variation in the accident count. The control type on the major road was important for the four-leg models, showing that the intersections with no control on the major road and stop control on the minor road (two-way stop control) were more dangerous.

These models were tested on crash data from Maine and the results show significant but weak correlation. This is probably because the ranges of exposure values (i.e., AADT) in the Maine data on each type of road are significantly different from those in Connecticut. Also the trip generation rates in Maine

sway more from the national average (obtained from the ITE Trip Generation Manual) than Connecticut. Thus, it is suggested that these models should be used with caution. These models can only be used for the data ranges given in Table 9. In order to apply these models to another location (i.e., other than the CRCOG region), the delineation of the links has to be similar to the CRCOG region (i.e. as connectors between major intersections). If not, then the conservative models presented as equations in Chapter 5 should be used for better transferability (without length for segment-intersection crashes and with length as an offset and without population and employment for the segment-related crashes).

Further research is needed to cover all the various types of roads that may exist in any State road network. Our estimation database with all covariates did not include freeways, three-lane roads, roads with two way left turn lanes (TWLTL), four-lane divided roads, or roads with more than two lanes in each direction. Some of these road types were intentionally excluded (e.g., freeways), but the others were not represented in the analysis road network in sufficient numbers to estimate prediction models. Consequently, we were not able to evaluate and comment on the safety of these types of roads. If a network with enough of these missing roads (and the required covariates) is available as a GIS layer, the procedure outlined in Chapter 3 and Appendix A can be used to assign population and employment data from geographic zones to road links and prepare a data set to estimate accident prediction models for these types of roads.

Bared, J. G. and A. Vogt (1996). "Highway Safety Evaluation System for Planning and Preliminary Design of Two-Lane Rural Highways." Accident Analysis Workshop, Federal Highway Administration.

Connecticut State Highway Log (2004).
www.ct.gov/dot/LIB/dot/Documents/dpolicy/hwylog/hwylog.pdf

Connecticut State Photo Log (2003).
http://www.ct.gov/dot/cwp/view.asp?a=1387&q=259618

Connecticut State Traffic Log (2005).
www.ct.gov/dot/LIB/dot/Documents/dpolicy/traflog/traflog.pdf

Hadi, M. A., J. Aruldhas, L. F. Chow, and J. A. Wattleworth (1995). "Estimating safety effects of cross-section design for various highway types using negative binomial regression." *Transportation Research Record*, Vol. 1500, pp 169-177.

Harwood, D. W. (1990). "Effective Utilization of Street Width on Urban Arterials." *330 National Cooperative Highway Research Program*, Washington, D.C.

Harwood, D., K. Bauer, I. Potts, D. Torbic, K. Richard, E. Kohlman Rabbani, E. Hauer, L. Elefteriadou, M. Griffith (2003). "Safety Effectiveness of Intersection Left- and Right-Turn Lanes." *Transportation Research Record,* Vol. 1840, pp 131-139.

Hauer, E. (1995). "On Exposure and Accident Rate." *Traffic Engineering and Control*, Vol. 36, No. 3, pp 134-138.

Hauer, E. (2001). "Overdispersion in Modeling Accidents on Road Sections and in Empirical Bayes Estimation." *Accident Analysis and Prevention,* Vol. 33, pp 799-808.

Hauer, E. (2004). "Statistical Road Safety Modeling." *Transportation Research Record*, Vol. 1897, pp 81-87.

Highway Safety Manual Website (2007). www.highwaysafetymanual.org.

Ivan, J. N., C. Wang, and N. R. Bernardo (2000). "Explaining two-lane highway crash rates using land use and hourly exposure." *Accident Analysis and Prevention,* Vol. 32, pp 787-795.

Ivan, J. N. and P. J. O'Mara (1997). "Prediction of Traffic Accident Rates Using Poisson Regression." 76th Annual Meeting Transportation Research Board, Washington, D. C.

Jonsson, T., Z. Deng, and J. N. Ivan (2005). "A Procedure for Allocating Zonal Attributes to a Link Network in a GIS Environment." 85th Annual Meeting Transportation Research Board, Washington, D. C.

Joshua, S. C., and N. J. Garber (1990). "Estimating Truck Accident Rate and Involvements using Linear and Poisson Regression Models." *Transportation Planning and Technology*, Vol. 15, pp 41-58.

Jovanis, P. P., and H. L. Chang (1986). "Modeling the Relationship of Accidents to Miles Traveled." *Transportation Research Record*, Vol. 1068, pp 42-51.

Kim, K. and E. Yamashita (2000). " Motor Vehicle Crashes and Land Use: Empirical Analysis from Hawaii." *Transportation Research Record*, Issue No. 1784, pp 73-79.

King, G. F., and R. B. GoldBlatt (1975). "Relationship of Accident Patterns to Type of Intersection Control." *Transportation Research Record*, Issue No. 540, pp 1-12.

Kweon, Y. J., and K. M. Kockelman. (2004). "Spatially disaggregate panel models of crash and injury counts: the effect of speed limits and design." 83rd Annual Meeting Transportation Research Board, Washington, D. C.

Levinson, H. S. and J. S. Gluck (2000). "Access Spacing and Safety: Recent Research Results." *Fourth National Access Management Conference*, Transportation Research Board, Portland.

Lord, D. (2002). "Issues Related to the Application of Accident Prediction Models for the Computation of Accident Risk on Transportation Networks." *Transportation Research Record*, Vol. 1784, pp 17-26.

Lyon, C., J. Oh, B. Persaud, S. Washington, J. Bared (2003). "Empirical Investigation of Interactive Highway Safety Design Model Accident Prediction Algorithm Rural Intersections." *Transportation Research Record,* Vol. 1840, pp 78-86.

Maher, M. J., and I. Summersgill (1996). "A Comprehensive Methodology for the Fitting of Predictive Accident Models." *Accident Analysis and Prevention*, Vol. 28, No. 3, pp 281-296.

Mensah, A, and E. Hauer (1998). "Two Problems of Averaging Arising in the Estimation of the Relationship between Accidents and Traffic Flow." *Transportation Research Record,* Vol. 1635, pp 37-43.

Miaou, S. P., P. S. Hu, T. Wright, A. K. Rathi, and S. C. Davis (1992). "Relationships between truck accidents and highway geometric design: a Poisson regression approach." *Transportation Research Record*, Vol. 1376, pp 10-18.

Miaou, S. P. and H. Lum (1993). "Modeling Vehicle Accidents and Highway Geometric Design Relationships." *Accident Analysis and Prevention*, Vol. 25, No. 6, pp 689-709.

Miaou, S. P (1994). "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions." 73rd Annual Meeting Transportation Research Board, Washington, D. C.

Milton, J. and F. Mannering (1998). "The Relationship Among Highway Geometrics, Traffic-Related Elements and Motor-Vehicle Accident Frequencies." *Transportation,* No. 25, pp 395-413.

National Highway Traffic Safety Administration (NHTSA) (2004). "Traffic Safety Facts 2002." U.S. Department of Transportation, Washington, DC.

Noland, R. B., and M. A. Quddus (2004). "A spatially disaggregate analysis of road casualties in England." *Accident Analysis and Prevention,* Vol. 36, No. 6, pp 973-984,

Persaud, B, and K. Mucsi (1995). "Microscopic Accident Potential Models for Two-Lane Rural Roads." *Transportation Research Record*, Vol. 1485, pp 134-139.

Poch, M. and F. Mannering (1996). "Negative binomial analysis of intersection-accident frequencies." *Journal of Transportation Engineering*, Vol. 122, No. 2, pp 105–113.

Qin, X., J. N. Ivan, and N. Ravishanker (2003). "Selecting Exposure Measures in Crash Rate Prediction for Two-lane Highway Segments." *Accident Analysis and Prevention*, Vol. 938, pp 1-9.

Shankar, V., F. Mannering, and W. Barfield (1995). "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies." *Accident Analysis and Prevention,* Vol. 27, No. 3, pp 371-389.

Vogt, A. and J. G. Bared (1998). "Accident Models for Two-Lane Rural Road: Segments and Intersections." Federal Highway Administration Report Number FHWA-RD-98-133. Washington, D.C.

Wikipedia, the free encyclopedia (2006). *www.wikipedia.org/wiki/Poisson_distribution*

Wilson, E. H., J. D. Hurd, D. L. Civco, M. P. Prisloe, and C. Arnold (2003). "Development of a geospatial model to quantify, describe and map urban growth." *Remote Sensing of Environment,* Vol. 86, pp 275-285.

Zegeer, C. V., R. C. Deen, and J. G. Mayes (1981). "Effect of lane width and shoulder widths on accident reduction on rural, two-lane roads." *Transportation Research Record,* Vol. 806, pp 33-43.

This Appendix lists the steps to follow in preparing link-based population and employment data for either re-calculation of predicted accidents using the existing prediction equations or for re-estimation of accident prediction models. Note that these instructions are specifically for use of the CRCOG data. Some specific steps will be different for other data sets (in particular the need or the size of the buffer).

I. Get the updated population and employment data for each zone as a GIS layer and add them to a map document (.mxd document) along with the layer containing all the links. Clear any previous selections existing in the map document using the 'Selection' item on the main menu:
Selection → Clear Selected Features

II. If the digitization of the zone boundaries and links are not 100% coincident (assume they are not unless you know for certain that they are), then use ArcToolbox to create a buffer around each TAZ (Figure 19). The size of the buffer depends on how much the coordinates on the zone and link layers differ from one another. For the CRCOG database, 200 ft was necessary for the buffer; it is recommended to start with 200 ft and then spot check several locations to identify how many mis-identification errors occur.

ArcToolbox → Analysis Tools → Proximity → Buffer

Browse the 'Input Feature' or the layer which need to be buffered (TAZ layer in this case). Note the directory where the 'Output Feature Class' is being saved. Enter 200 feet as the 'Linear unit' and click OK. New buffer TAZ layer is automatically added to the map.

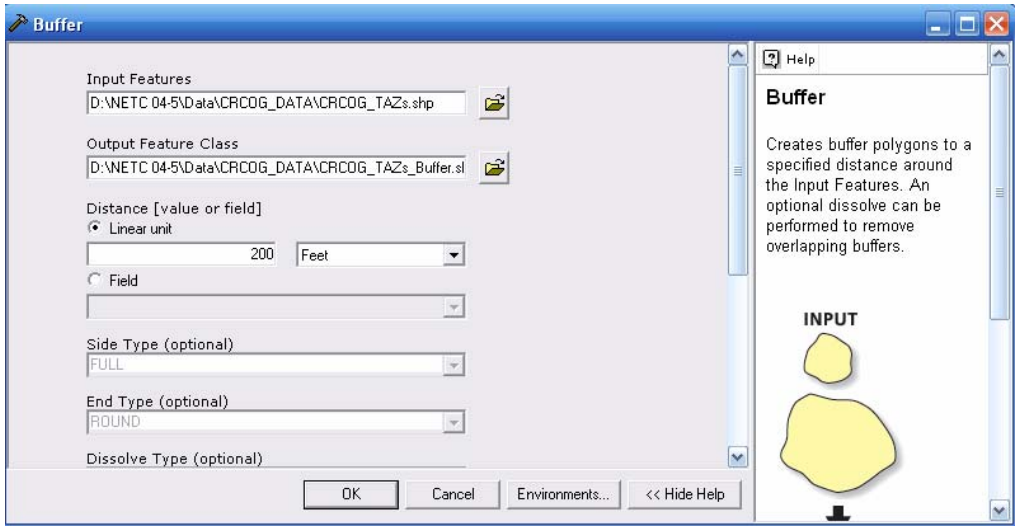


Figure 19: Buffer tool from ArcToolbox

III. Using ArcToolbox and the new buffered TAZ layer cut the links to create partial links such that each link is divided among different TAZs to which it belongs or forms the boundary of (Figure 20). This is done so that links which carry traffic from different TAZs can be assigned population and employment data from each of those TAZs based on their partial length.

ArcToolbox → Analysis Tools → Overlay → Intersect

From the 'Input Features' drop down box select the buffered TAZ layer first followed by the link layer. Note the directory where the 'Output Feature Class' is being saved and click OK. New intersected link layer is automatically added to the map.

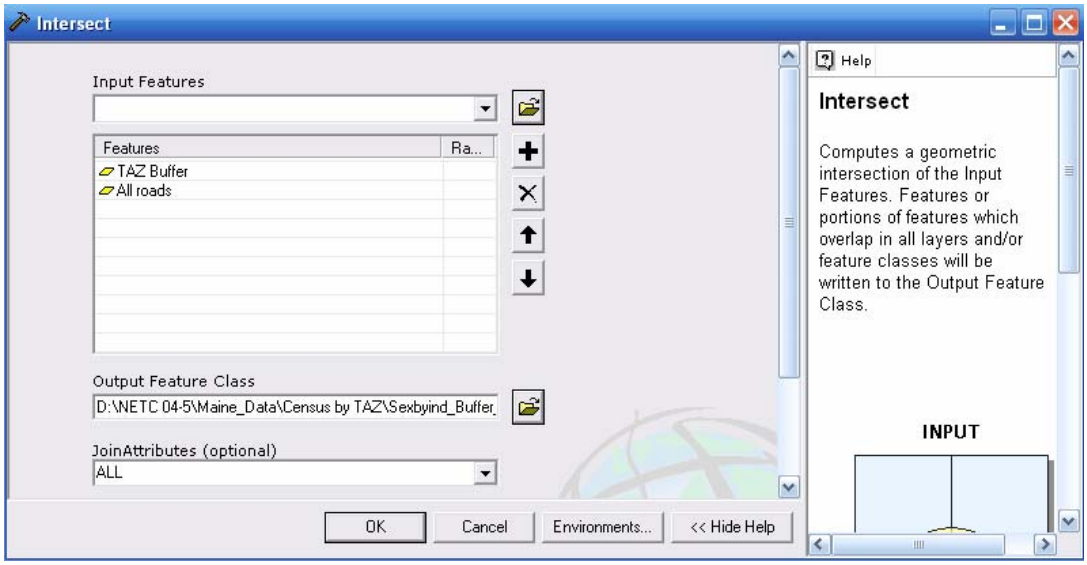


Figure 20: Intersect tool from ArcToolbox

This feature also assigns the TAZ number and all the other attributes of the TAZ layer to each link. Thus, if the TAZ layer contains the population and employment data, the split segments obtained from this procedure will contain this data along with the TAZ number to which it belongs. It is important to note that the number of links in the new layer can increase many times since each link crossing a TAZ boundary will produce at least two new split segments.

IV. This step is to calculate the length of new split segments since the population and employment data is proportional to the length of the segment in a zone. The ***units*** of this split segment length will be the units that your features are stored in, not the map units or display units of the data frame you are currently working with. So if your data is stored in feet, the calculated values will be in feet. If you want the calculated data to be in different units than the data's units, you could either add a conversion into the calculation expression, or (more simply) project your data into a coordinate system that uses the units you want the values to be in, and then perform the calculation.

Following are the steps to calculate this partial length (VBA code for ArcGIS ver 9.1 users):

a. Right-click the layer you want to edit and click 'Open Attribute Table';
b. Right-click the field heading for length and click 'Calculate Values.' If there is no field for length values, you can add a new field for length by clicking the 'Options' button and selecting 'Add Field';
c. Click 'Calculate Values';
d. Check 'Advanced';
e. Type the following VBA statement in the first text box:

```
Dim dblLength as double
Dim pCurve as ICurve
Set pCurve = [shape]
dblLength = pCurve.Length
```

f. Type the variable dblLength in the text box directly under the length field name;
g. Click OK.

ArcGIS ver. 9.2 users have an inbuilt function to accomplish this task as shown in Figure 21. By using the 'Calculate Geometry' feature one can calculate the length in miles (or any other unit) after adding the field 'partial_length' to the attribute table.

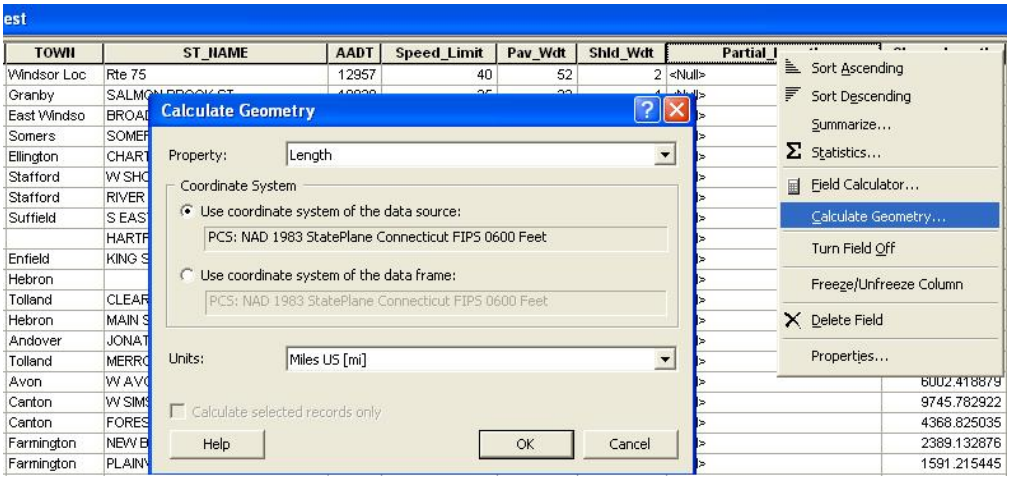


Figure 21: Calculation partial length in ArcGIS ver. 9.2

V. The attribute table of this new layer with split segments is then exported to a database (Microsoft Access®). This is done using the 'Export' option in the attribute table (Figure 22). Remember to save the table in a personal Geodatabase (Microsoft Access®) using the 'Save as type' dropdown box. Let's assume this new table name as 'split_links' from here onwards.

Attribute Table → Options → Export

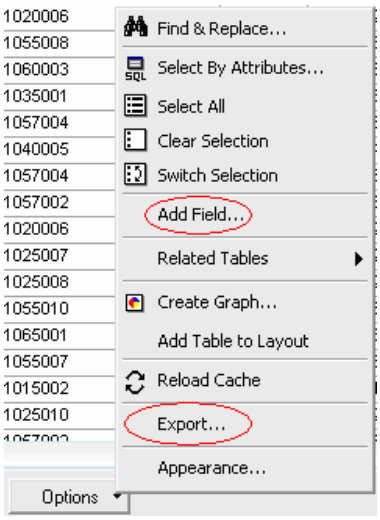


Figure 22: Export attribute table

VI. In many cases a link does not align perfectly with the boundary of a TAZ and crosses it at several places (Figure 23). The intersection of links with TAZ buffer boundary in step III will thus break this link into several pieces. For example, in the figure below both parts 1 and 2 of the segments belong to TAZ 1. These pieces have to be aggregated in order to create a compact database for distributed of land use in a TAZ among various links.

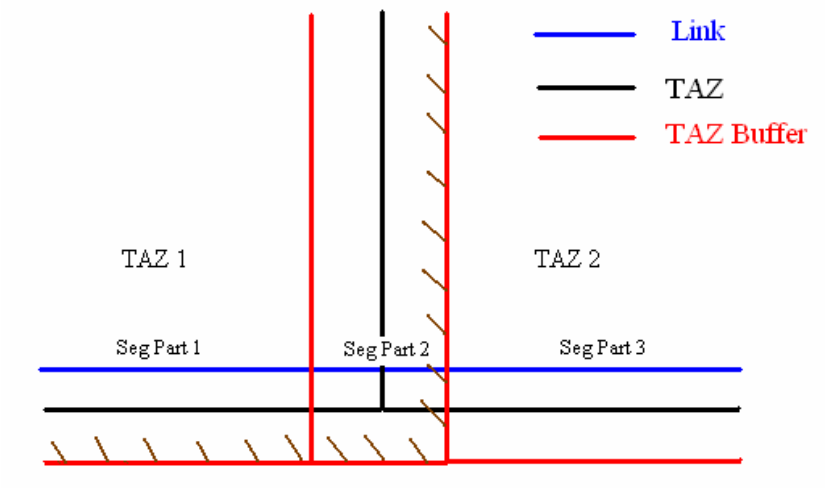


Figure 23: Breaking of links into parts by the TAZ boundary

Open the Microsoft Access database containing the table exported in step V and create the 'Make-Table Query' shown in Figure 24. This query is created on the table obtained from step III (split_links in Figure 24) and collects the pieces of the same segment in a TAZ and adds the partial length of each part. Microsoft Access displays the **Total** row in the design grid after clicking **Totals** Σ on the toolbar. Let's assume this new table name as 'step1' from here onwards.

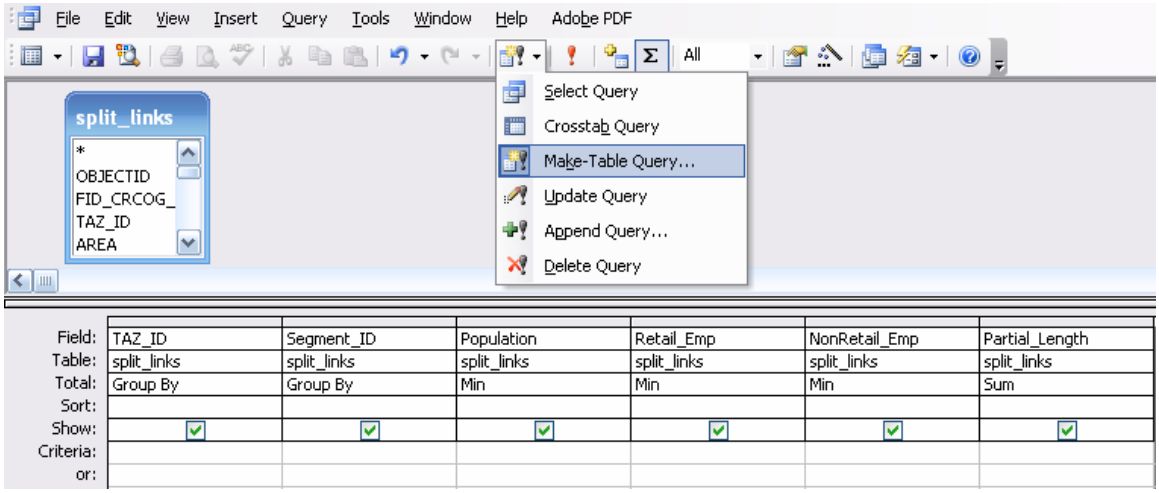


Figure 24: First Query in Access

VII. Since the population and employment data of a TAZ is distributed among all the links on the basis of their length in that zone, it is important to know the total length of all the segments in a particular TAZ. Create another 'Make-Table Query' shown in Figure 25 on the table obtained from step VI (step1 in Figure 25) to achieve this. Let's assume this new table name as 'step2' from here onwards.

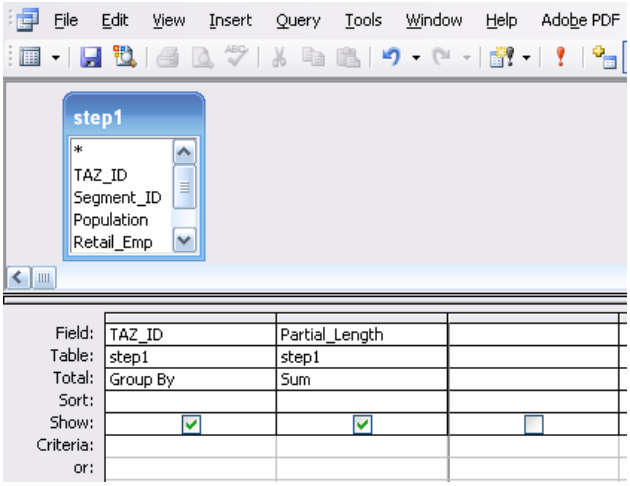


Figure 25: Obtaining total segment length in a TAZ

VIII. Tables 'step1' and 'step2' are now combined to create one table. Create a 'Make-Table Query' shown in Figure 26 to achieve this. Let's assume this new table name as 'step3' from here onwards.

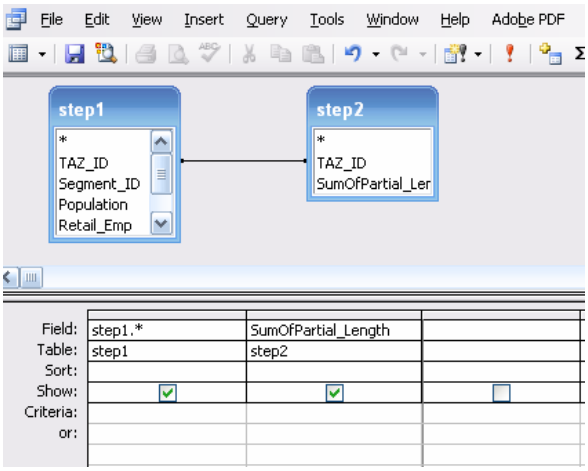


Figure 26: Combining tables

IX. Using the 'Design View' symbol on the top left hand corner of the table ( ) add three columns to table 'step3': weighted population (WPopulation), weighted retail employment (WRetail_Emp), and weighted non-retail employment (WNonRetail_Emp) of 'number' data type. Using an 'Update Query' shown in Figure 27, update these three fields such that:

$$WPopulation = Population \times \frac{Partial\_Segment\_Length}{Total\_Seg\_Length\_in\_TAZ}$$

Use similar format for retail and non-retail employment also. The 'Update Query' is denoted by the symbol in Microsoft Access and can be selected from the drop down box in the menu. The statement in 'Update To' field can be developed using the 'Build' tool and the 'Expression Builder' shown in the figure below.

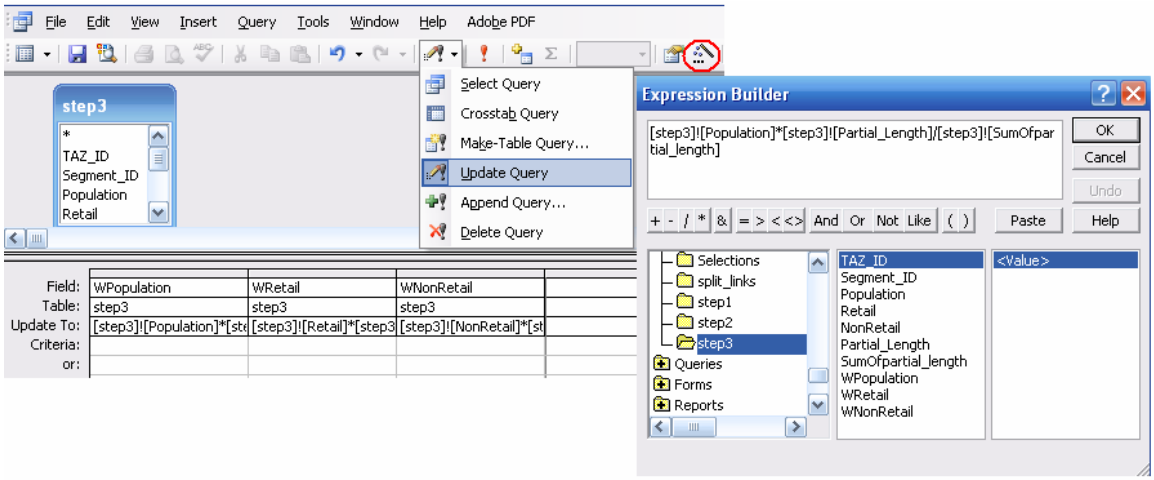


Figure 27: Updating fields

X. The final step is to aggregate pieces of the same segment in different TAZs in order to create one segment with the total population and employment data assigned to it from these TAZs. This is done by creating a 'Make-Table Query' shown in Figure 28 on updated table 'step3' where all the land use variables are aggregated by segment. After this step we will end up with exactly the same number of links with which we started in step I.

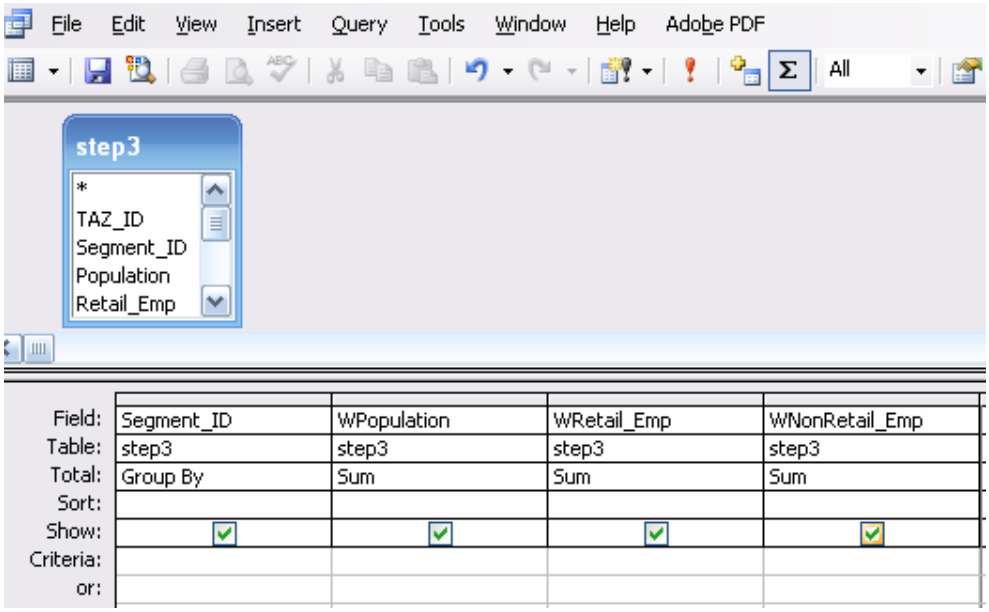


Figure 28: Final aggregation of links

This population and employment data, along with other predictors (AADT, Speed limit, etc.), can then be used to re-estimate the segment-intersection and segment-related crashes for all three road types using the models presented in chapter 5 and any computing application (e.g. Microsoft Excel or Access). It is assumed that the number of observed accidents on each link is already available as a column in this table. In order to use the GIS interface presented in appendix C this modified file has to be imported to the link shapefile or layer in GIS. To accomplish this export the table from the computing application (if using Excel) as a database file (.dbf) or an Access table. Use the following steps in the GIS map to join this .dbf (or Access table) file to the link shapefile or layer.

a. Right-click the link layer you want to join, point to Joins and Relates, and click Join.

b. Click the first dropdown arrow and click Join attributes from a table.

c. Click the second dropdown arrow and click the field name in the layer on which the join will be based (Segment_ID in our case).

d. Click the Browse button near the third dropdown arrow to search for new dbf (or Access table) file (with observed and predicted accidents) created above.

e. Click the fourth dropdown arrow and click the field in the table on which to base the join (Segment_ID in our case).

f. Click OK. The attributes of the table are appended to the layer's attribute table.

The new fields in the attribute table will be in the format 'tablename.fieldname'. Using the 'Add Field' option shown in Figure 22, create two fields named 'Tot_Count' and 'EstimatedAccidents' with type as 'Long Integer' (Figure 29). These fields will be added to the GIS layer, and not to the imported table, thus the column heading for the new fields will be: LinkLayerName.Tot_Count and LinkLayerName.EstimatedAccidents.

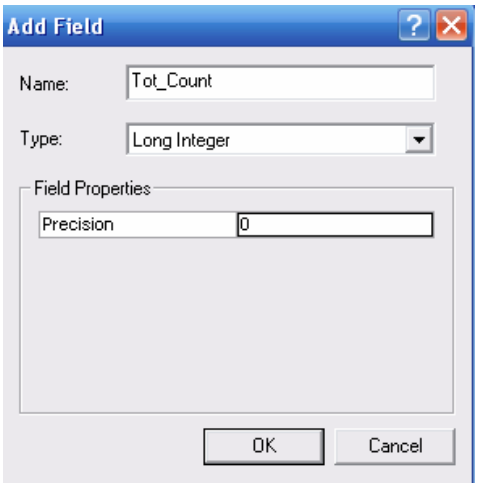


Figure 29: Creating field in GIS

Right-click on the column headings and choose the 'Calculate Values' option (Figure 30). Set the 'Tot_Count' and 'EstimatedAccidents' fields equal to the observed and predicted accidents respectively from the imported table. Remove the join by: Right-clicking the link layer to which the table was joined, point to Joins and Relates, point to Remove Join(s) and click 'Remove All Joins.' The layer is now ready with the updated predicted accidents. The use of GIS interface, developed as a part of this project, for comparison of observed and predicted accidents is presented in appendix C.

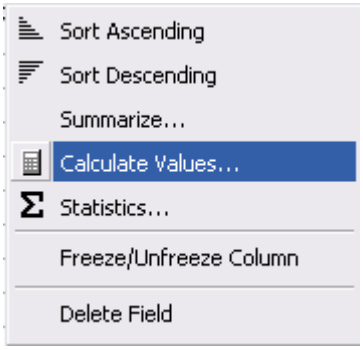


Figure 30: Calculate field values

<u>Important Note</u>: The users should be familiar with the use of geodatabases for storing the shapefiles, feature classes, and tables. The other (and older) ways of storing data as individual layers and shapefiles have severe limitations on adding fields and the length of the field name. With that the user WILL NOT be able to create the required field name 'EstimatedAccidents' for using the GIS interface presented in appendix C. Thus it is important that the user familiarizes himself with the use of geodatabases and these are used to store the shapefiles (see article below).

http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?ID=1459&TopicName=An%20overview%20of%20building%20geodatabases

Category 1: Intersection Crash Models

Table B-1 – Three and Four Leg Intersections

Parameter	Three-leg Estimate (SE)	Four-leg Estimate (SE)
Intercept	1.139 (0.529)*	***0.044 (0.428)***
Ln(Major AADT)	0.585 (0.145)	0.686 (0.122)
Ln(Minor AADT)	0.201 (0.114)	0.572 (0.110)
LU_Urban	***0.274 (0.244)*** **	***0.136 (0.120)***
LU_SubUrban	0.000 (0.000)	0.000 (0.000)
LU_Rural	-0.290 (0.147)	-0.342 (0.165)
Major Lanes = 2	-0.280 (0.168)	
Major Lanes = 3	0.000 (0.000)	
Major Lanes >= 4	***0.068 (0.166)***	
Major Speed <= 30	0.382 (0.180)	
Major Speed > 30	0.000 (0.000)	
Maj. Control = None		0.783 (0.352)
Maj. Control = Stop		0.000 (0.000)
Maj. Control = Signal		***0.081 (0.308)***
	0.183 (0.040)	0.160 (0.029)

* Estimate (Standard Error)

** Bold Italic formatting represents statistically insignificant variables

Category 2: Segment-Intersection Crash Models
Table B-2 – Rural – Two-lane – Undivided (N = 319)

Parameter	*Without Land Use*		*With Land use/Trips/Driveway*					
	Basic Model		**Land Use**		**Trips**		**Driveway**	
	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder
Intercept	1.193 (0.366)*	1.236 (0.358)	1.011 (0.369)	1.074 (0.361)	-4.289 (1.309)	-4.156 (1.305)	0.985 (0.393)	1.043 (0.385)
Ln(AADT)	0.700 (0.117)	0.707 (0.114)	0.669 (0.116)	0.672 (0.113)	0.647 (0.115)	0.654 (0.112)	0.629 (0.119)	0.630 (0.117)
Ln(Length)	0.572 (0.073)	0.573 (0.073)	0.463 (0.085)	0.471 (0.086)	0.276 (0.097)	0.287 (0.097)	0.418 (0.093)	0.430 (0.093)
Ln(Trips)					0.387 (0.089)	0.381 (0.089)		
LU_Pop / 1000			***0.527 (0.482)***	***0.459 (0.474)***				
LU_Retl / 1000			21.41 (8.868)	21.14 (8.904)				
LU_Nretl / 1000			***-0.519 (0.927)***	***-0.369 (0.931)***				
Res_Driway							0.014 (0.009)	***0.014 (0.009)***
Retl_Driway							0.089 (0.030)	0.093 (0.030)
Unsignal_Inter							***0.011 (0.040)***	***0.005 (0.041)***
Posted Speed < 40	***0.058 (0.176)*** **	***0.096 (0.178)***	***0.013 (0.177)***	***0.056 (0.178)***	***-0.031 (0.174)***	***0.013 (0.175)***	***-0.086 (0.182)***	***-0.031 (0.182)***
Posted Speed = 40	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Posted Speed > 40	-0.668 (0.164)	-0.688 (0.166)	-0.664 (0.166)	-0.681 (0.169)	-0.687 (0.161)	-0.723 (0.163)	-0.647 (0.162)	-0.667 (0.164)
Pav < 30 ft	***0.202 (0.190***		***0.205 (0.187)***		***0.205 (0.186)***		***0.240 (0.187)***	
Pav >= 30 to < 40 ft	0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)	
Pav >= 40 ft	0.448 (0.177)		0.459 (0.175)		0.496 (0.173)		0.503 (0.175)	
Shld < 3 ft		***0.118 (0.184)***		***0.117 (0.182)***		***0.113 (0.180)***		***0.176 (0.183)***
Shld >= 3 to < 6 ft		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)
Shld >= 6 ft		***0.244 (0.170)***		***0.258 (0.169)***		0.308 (0.168)		0.346 (0.171)
Dispersion	1.118 (0.092)	1.135 (0.094)	1.091 (0.091)	1.108 (0.092)	1.058 (0.088)	1.076 (0.090)	1.076 (0.090)	1.092 (0.091)
AIC	64.18	59.25	66.42	61.24	80.86	75.40	70.38	65.60

* Estimate (Standard Error) ** Bold Italic formatting represents statistically insignificant variables

Table B-3 – Suburban and Urban – Two-lane – Undivided (N = 573)

Parameter	*Without Land Use*		*With Land use/Trips/Driveway*					
	Basic Model		**Land Use**		**Trips**		**Driveway**	
	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder
Intercept	***-0.116 (0.314)****	***-0.339 (0.291)***	***-0.208 (0.319)***	***-0.399 (0.299)***	-3.576 (0.739)	-3.803 (0.712)	***-0.302 (0.350)***	***-0.503 (0.332)***
Ln(AADT)	1.215 (0.094)	1.267 (0.087)	1.110 (0.097)	1.153 (0.092)	1.074 (0.097)	1.103 (0.092)	1.137 (0.096)	1.183 (0.091)
Ln(Length)	0.578 (0.049)	0.570 (0.049)	0.425 (0.061)	0.421 (0.060)	0.381 (0.062)	0.373 (0.061)	0.432 (0.068)	0.424 (0.068)
Ln(Trips)					0.253 (0.049)	0.258 (0.049)		
LU_Pop / 1000			0.783 (0.267)	0.752 (0.267)				
LU_Retl / 1000			3.497 (1.721)	3.664 (1.737)				
LU_Nretl / 1000			***0.523 (0.351)***	***0.530 (0.349)***				
Res_Driway							***0.001 (0.005)***	***0.002 (0.005)***
Retl_Driway							0.042 (0.013)	0.042 (0.013)
Unsignal_Inter							0.076 (0.027)	0.075 (0.027)
Posted Speed < 35	0.168 (0.100)	***0.132 (0.101)***	0.192 (0.098)	***0.163 (0.099)***	***0.151 (0.098)***	***0.119 (0.099)***	***0.126 (0.099)***	***0.105 (0.099)***
Posted Speed = 35	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Posted Speed > 35	-0.345 (0.080)	-0.336 (0.080)	-0.281 (0.081)	-0.269 (0.081)	-0.306 (0.079)	-0.294 (0.077)	-0.292 (0.079)	-0.283 (0.079)
Pav < 30 ft	***-0.152 (0.113)***		***-0.124 (0.111)***		***-0.086 (0.112)***		***-0.090 (0.112)***	
Pav >= 30 to < 40 ft	0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)	
Pav >= 40 ft	***0.100 (0.092)****		***0.057 (0.091)***		***0.051 (0.091)***		***0.064 (0.091)***	
Shld < 3 ft		***0.027 (0.088)***		***0.041 (0.088)***		***0.061 (0.087)***		***0.064 (0.087)***
Shld >= 3 to < 6 ft		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)
Shld >= 6 ft		0.209 (0.097)		0.171 (0.096)		0.184 (0.095)		***0.142 (0.097)***
Dispersion	0.638 (0.041)	0.636 (0.041)	0.616 (0.040)	0.614 (0.039)	0.609 (0.039)	0.606 (0.039)	0.615 (0.040)	0.613 (0.039)
AIC	113.44	114.91	126.86	128.30	136.64	139.57	127.63	128.74

* Estimate (Standard Error) ** Bold Italic formatting represents statistically insignificant variables

Table B-4 – Suburban and Urban – Four-lane – Undivided (N = 229)

Parameter	*Without Land Use*		*With land use/trips/driveway*					
	Basic Model		**Land Use**		**Trips**		**Driveway**	
	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder
Intercept	***0.606 (0.882)****	***1.015 (0.807)***	***0.858 (0.877)***	1.417 (0.832)	***-0.798 (1.508)***	***-0.052 (1.422)***	***0.195 (0.945)***	***0.613 (0.864)***
Ln(AADT)	1.155 (0.223)**	1.009 (0.203)	0.990 (0.221)	0.829 (0.211)	1.112 (0.226)	0.965 (0.208)	1.144 (0.222)	0.989 (0.201)
Ln(Length)	0.553 (0.087)	0.565 (0.088)	0.438 (0.100)	0.490 (0.101)	0.467 (0.114)	0.503 (0.111)	0.416 (0.115)	0.416 (0.115)
Ln(Trips)					***0.096 (0.084)***	***0.077 (0.085)***		
LU_Pop / 1000			***0.247 (0.546)***	***0.443 (0.556)***				
LU_Retl / 1000			8.339 (2.530)	6.216 (2.465)				
LU_Nretl / 1000			***-0.430 (0.323)***	***-0.468 (0.328)***				
Res_Driway							***-0.011 (0.019)***	***-0.010 (0.019)***
Retl_Driway							0.036 (0.014)	0.039 (0.014)
Unsignal_Inter							***0.027 (0.061)***	***0.036 (0.062)***
Posted Speed < 35	***0.176 (0.173)***	***0.215 (0.174)***	***0.172 (0.171)***	***0.208 (0.173)***	***0.155 (0.173)***	***0.203 (0.174)***	***0.214 (0.172)***	***0.232 (0.174)***
Posted Speed = 35	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Posted Speed > 35	-0.521 (0.143)	-0.459 (0.141)	-0.443 (0.142)	-0.385 (0.142)	-0.469 (0.149)	-0.420 (0.147)	-0.460 (0.142)	-0.393 (0.141)
Pav < 50 ft	-0.342 (0.193)		***-0.247 (0.191)***		***-0.310 (0.194)***		***-0.250 (0.193)***	
Pav >= 50 to < 60 ft	0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)	
Pav >= 60 ft	-0.333 (0.151)		-0.477 (0.151)		-0.352 (0.152)		-0.336 (0.147)	
Shld < 2 ft		***0.195 (0.170)***		***0.143 (0.168)***		***0.171 (0.171)***		***0.199 (0.167)***
Shld = 2 ft		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)
Shld > 2 ft		***-0.104 (0.143)***		***-0.114 (0.140)***		***-0.110 (0.143)***		***-0.126 (0.141)***
Dispersion	0.821 (0.079)	0.834 (0.080)	0.771 (0.075)	0.799 (0.077)	0.816 (0.078)	0.831 (0.080)	0.787 (0.076)	0.794 (0.077)
AIC	45.97	42.38	53.79	45.89	45.28	41.20	49.43	47.30

* Estimate (Standard Error) ** Bold Italic formatting represents statistically insignificant variables

Category 3: Segment-Related Crash Models
Table B-5 – Rural – Two-lane – Undivided (N = 319)

Parameter	*Without Land Use*		*With land use/trips/driveway*					
	Basic Model		**Land Use**		**Trips**		**Driveway**	
	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder
Intercept	1.992 (0.254)*	2.015 (0.248)	1.996 (0.254)	2.024 (0.249)	***1.276 (0.923)***	***1.137 (0.924)***	2.159 (0.265)	2.190 (0.259)
Ln(AADT)	0.393 (0.079)**	0.388 (0.077)	0.392 (0.080)	0.385 (0.079)	0.380 (0.081)	0.371 (0.079)	0.401 (0.081)	0.392 (0.079)
Ln(Length)	0.889 (0.053)	0.893 (0.053)	0.899 (0.053)	0.903 (0.053)	0.897 (0.054)	0.904 (0.054)	0.874 (0.055)	0.876 (0.055)
Ln(Trips) / Len					***0.051 (0.064)***	***0.063 (0.064)***		
LU_Pop/1000*Len			***0.202 (0.224)***	***0.187 (0.222)***				
LU_Retl/1000*Len			***3.757 (2.399)***	***3.823 (2.403)***				
LU_Nretl/1000*Len			-0.871 (0.386)	-0.786 (0.383)				
Res_Driway / Len							-0.009 (0.004)	-0.009 (0.004)
Retl_Driway / Len							***-0.010 (0.008)***	***-0.009 (0.008)***
Unsignal_Inter / Len							***-0.021 (0.015)***	***-0.023 (0.015)***
Posted Speed < 40	***-0.027 (0.126)***	***-0.019 (0.126)***	***-0.061 (0.126)***	***-0.051 (0.126)***	***-0.031 (0.126)***	***-0.023 (0.126)***	***0.052 (0.129)***	***0.057 (0.129)***
Posted Speed = 40	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Posted Speed > 40	-0.466 (0.111)	-0.474 (0.111)	-0.498 (0.112)	-0.501 (0.112)	-0.461 (0.111)	-0.468 (0.111)	-0.457 (0.110)	-0.463 (0.110)
Pav < 30 ft	***0.118 (0.117)***		***0.104 (0.117)***		***0.115 (0.117)***		***0.124 (0.118)***	
Pav >= 30 to < 40 ft	0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)	
Pav >= 40 ft	0.296 (0.128)		0.310 (0.128)		0.299 (0.128)		0.220 (0.129)	
Shld < 3 ft		***0.098 (0.113)***		***0.080 (0.113)***		***0.096 (0.113)***		***0.105 (0.114)***
Shld >= 3 to < 6 ft		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)
Shld >= 6 ft		0.270 (0.123)		0.264 (0.122)		0.282 (0.123)		***0.194 (0.125)***
Dispersion	0.468 (0.047)	0.466 (0.047)	0.459 (0.046)	0.460 (0.046)	0.467 (0.047)	0.467 (0.047)	0.459 (0.046)	0.459 (0.046)
AIC	210.51	209.71	209.51	208.08	209.16	208.67	212.40	211.64

* Estimate (Standard Error) ** Bold Italic formatting represents statistically insignificant variables

Table B-6 – Suburban and Urban – Two-lane – Undivided (N = 573)

Parameter	*Without Land Use*		*With land use/trips/driveway*					
	Basic Model		**Land Use**		**Trips**		**Driveway**	
	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder
Intercept	2.187 (0.281)*	2.141 (0.261)	2.196 (0.280)	2.178 (0.264)	***0.773 (0.655)***	***0.784 (0.625)***	2.122 (0.286)	2.044 (0.269)
Ln(AADT)	0.306 (0.084)**	0.305 (0.078)	0.250 (0.085)	0.242 (0.081)	0.237 (0.088)	0.227 (0.084)	0.303 (0.084)	0.311 (0.078)
Ln(Length)	0.800 (0.047)	0.799 (0.046)	0.824 (0.047)	0.825 (0.046)	0.825 (0.048)	0.826 (0.047)	0.848 (0.049)	0.844 (0.048)
Ln(Trips) / Len					0.106 (0.044)	0.104 (0.044)		
LU_Pop/1000*Len			0.272 (0.085)	0.263 (0.084)				
LU_Retl/1000*Len			0.865 (0.489)	0.883 (0.484)				
LU_Nretl/1000*Len			***0.009 (0.105)***	***0.011 (0.103)***				
Res_Driway / Len							***1E-4 (0.002)***	***3E-4 (0.002)***
Retl_Driway / Len							***0.001 (0.003)***	***3E-4 (0.003)***
Unsignal_Inter / Len							0.019 (0.005)	0.018 (0.005)
Posted Speed < 35	0.224 (0.090)	0.218 (0.090)	0.250 (0.089)	0.245 (0.089)	0.222 (0.090)	0.216 (0.089)	0.214 (0.091)	0.216 (0.090)
Posted Speed = 35	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Posted Speed > 35	-0.231 (0.072)	-0.216 (0.072)	-0.181 (0.072)	-0.167 (0.072)	-0.213 (0.072)	-0.199 (0.072)	-0.217 (0.072)	-0.204 (0.072)
Pav < 30 ft	***0.130 (0.098)***		***0.151 (0.096)***		***0.158 (0.098)***		***0.146 (0.097)***	
Pav >= 30 to < 40 ft	0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)	
Pav >= 40 ft	0.145 (0.085)		***0.127 (0.085)***		***0.133 (0.085)***		0.148 (0.085)	
Shld < 3 ft		0.174 (0.079)		0.183 (0.078)		0.182 (0.079)		0.194 (0.079)
Shld >= 3 to < 6 ft		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)
Shld >= 6 ft		0.247 (0.087)		0.235 (0.086)		0.247 (0.086)		0.232 (0.087)
Dispersion	0.438 (0.034)	0.431 (0.034)	0.421 (0.033)	0.415 (0.033)	0.432 (0.034)	0.425 (0.033)	0.427 (0.033)	0.422 (0.033)
AIC	239.30	245.73	249.79	256.15	242.90	249.32	245.36	251.40

* Estimate (Standard Error) ** Bold Italic formatting represents statistically insignificant variables

Table B-7 – Suburban and Urban – Four-lane – Undivided (N = 229)

Parameter	*Without Land Use*		*With Land use/Trips/Driveway*					
	Basic Model		**Land Use**		**Trips**		**Driveway**	
	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder	Pavement	Shoulder
Intercept	2.304 (0.715)*	2.150 (0.670)	2.306 (0.721)	2.047 (0.690)	***0.095 (1.325)***	***-0.137 (1.290)***	1.870 (0.750)	1.750 (0.705)
Ln(AADT)	0.416 (0.182)**	0.431 (0.173)	0.366 (0.183)	0.414 (0.177)	0.364 (0.182)	0.364 (0.174)	0.454 (0.182)	0.454 (0.173)
Ln(Length)	0.776 (0.079)	0.772 (0.079)	0.789 (0.077)	0.790 (0.078)	0.800 (0.079)	0.808 (0.081)	0.733 (0.078)	0.729 (0.079)
Ln(Trips) / Len					0.149 (0.076)	0.159 (0.077)		
LU_Pop/1000*Len			0.193 (0.101)	0.225 (0.105)				
LU_Retl/1000*Len			***0.413 (0.387)***	***0.344 (0.395)***				
LU_Nretl/1000*Len			***0.112 (0.070)***	***0.091 (0.072)***				
Res_Driway / Len							***0.004 (0.003)***	***0.004 (0.003)***
Retl_Driway / Len							0.009 (0.003)	0.010 (0.003)
Unsignal_Inter / Len							***-0.011 (0.009)***	***-0.010 (0.009)***
Posted Speed < 35	0.360 (0.151)	0.395 (0.156)	0.204 (0.152)	0.254 (0.156)	0.311 (0.150)	0.349 (0.155)	0.373 (0.147)	0.394 (0.152)
Posted Speed = 35	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Posted Speed > 35	-0.377 (0.123)	-0.361 (0.124)	-0.282 (0.123)	-0.269 (0.125)	-0.296 (0.128)	-0.278 (0.130)	-0.291 (0.124)	-0.261 (0.126)
Pav < 50 ft	-0.461 (0.165)		-0.468 (0.162)		-0.423 (0.165)		-0.375 (0.164)	
Pav >= 50 to < 60 ft	0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)	
Pav >= 60 ft	***-0.189 (0.128)***		***-0.205 (0.128)***		-0.218 (0.128)		-0.206 (0.125)	
Shld < 2 ft		***0.024 (0.150)***		***-0.053 (0.147)***		***-0.021 (0.149)***		***0.043 (0.146)***
Shld = 2 ft		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)		0.000 (0.000)
Shld > 2 ft		***-0.097 (0.125)***		***-0.110 (0.123)***		***-0.113 (0.124)***		***-0.102 (0.123)***
Dispersion	0.519 (0.062)	0.543 (0.064)	0.484 (0.060)	0.510 (0.062)	0.506 (0.061)	0.529 (0.063)	0.489 (0.060)	0.506 (0.061)
AIC	86.70	79.10	91.08	82.55	88.51	81.28	91.31	85.77

* Estimate (Standard Error) ** Bold Italic formatting represents statistically insignificant variables

I. Overview

Accident Model User Interface (AMUI) is a Geographic Information System (GIS) -integrated interface developed to provide users with a tool that presents crash model results on a map. AMUI is a map document that contains two customized modules: Link Accident Module and Intersection Accidents Module.

AMUI is not a stand-alone application. Therefore, ***ArcGIS must be available to the users***. AMUI was programmed using ArcObjects with Visual Basic for Application (VBA). ArcObjects is a set of objects specifically designed for programming with ArcGIS Desktop Applications. The VBA is a simplified version of Visual Basic and is designed to be embedded within applications. That is, ArcGIS comes with VBA so that programmers can customize it by putting frequently-used tools into one or more modules. Although AMUI cannot be used outside ArcGIS, users can enjoy full built-in functionality of ArcGIS in addition to customized tools.

II. Before Opening the Interface

It has been observed that the newly released Internet Explorer 7 (IE7) interferes with ArcGIS and ArcToolbox possibly causing problems while opening a map document[1]. Therefore, it is recommended that the users install all available security patches and updates for GIS available at the ESRI Website[2].

III. Detailed Features of AMUI

1. Opening the Interface

Start ArcMap and open the map document, *NETC Accident Model Comparison Module.mxd.* Figure 31 shows the initial screen when the user interface is opened. On the bottom of the tool bar area (within a red ellipse in Figure 31) are two modules customized for AMUI. If they are not shown in the tool bar, use the right click button on the mouse to open the context menu, and check "*NETC UI Toolbar*." Then drag the buttons to where you want them on the tool bar.

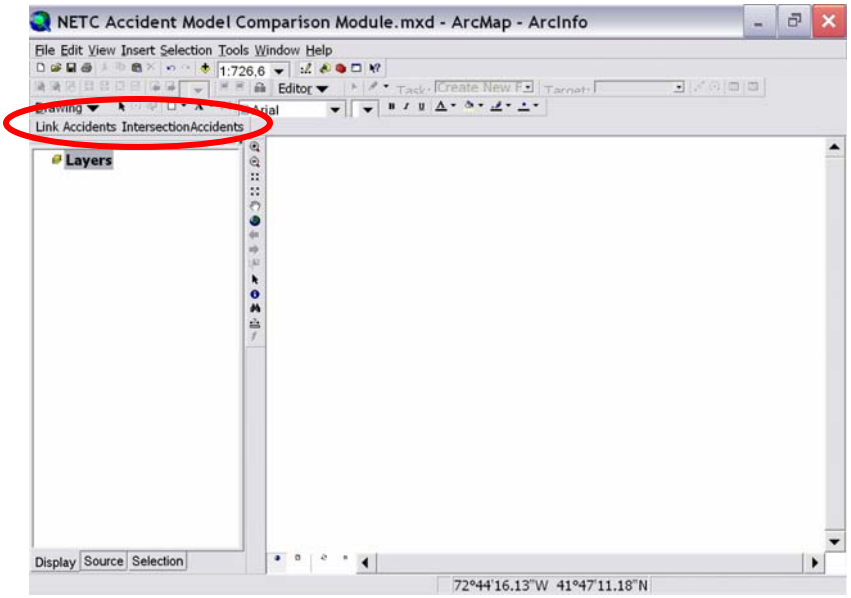


Figure 31: Initial Screen on the GIS Interface

[1] http://forums.esri.com/Thread.asp?c=93&f=1148&t=189707&mc=18
[2] http://support.esri.com/index.cfm?fa=downloads.patchesServicePacks.viewPatch&PID=43&MetaID=1150

2. Opening the Modules

Click '*Link Accidents*' or '*Intersection Accidents*' to open the modules shown in Figure 32. The '*Link Accident*' module can only read polyline layers or shape files; on the other hand, '*Intersection Accident*' module can read only point layers or shape files.

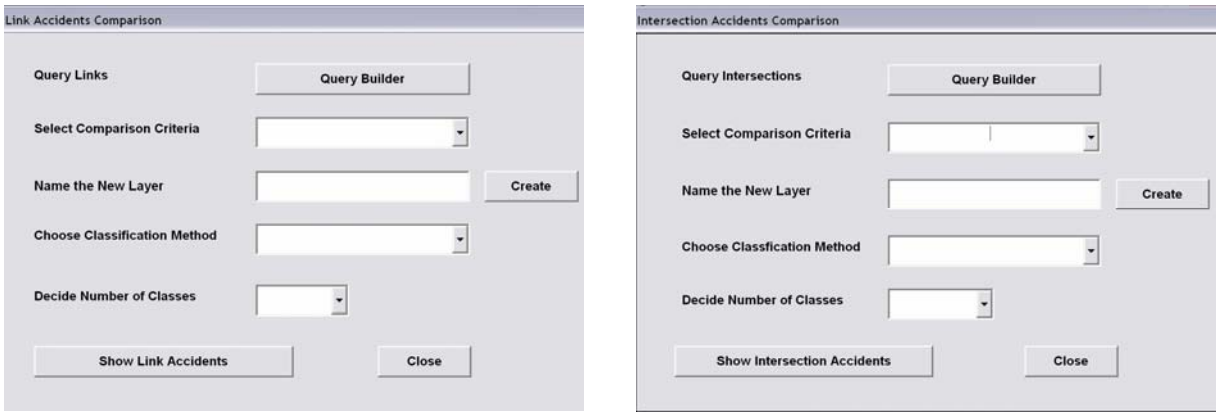


Figure 32: The Link and Intersection Accidents Modules

3. Detailed Features of Modules

The two modules have similar features; therefore, the detailed description of the Link Accidents Module is presented below as an example. The enlarged picture of Link Accident Module is shown in Figure 33.

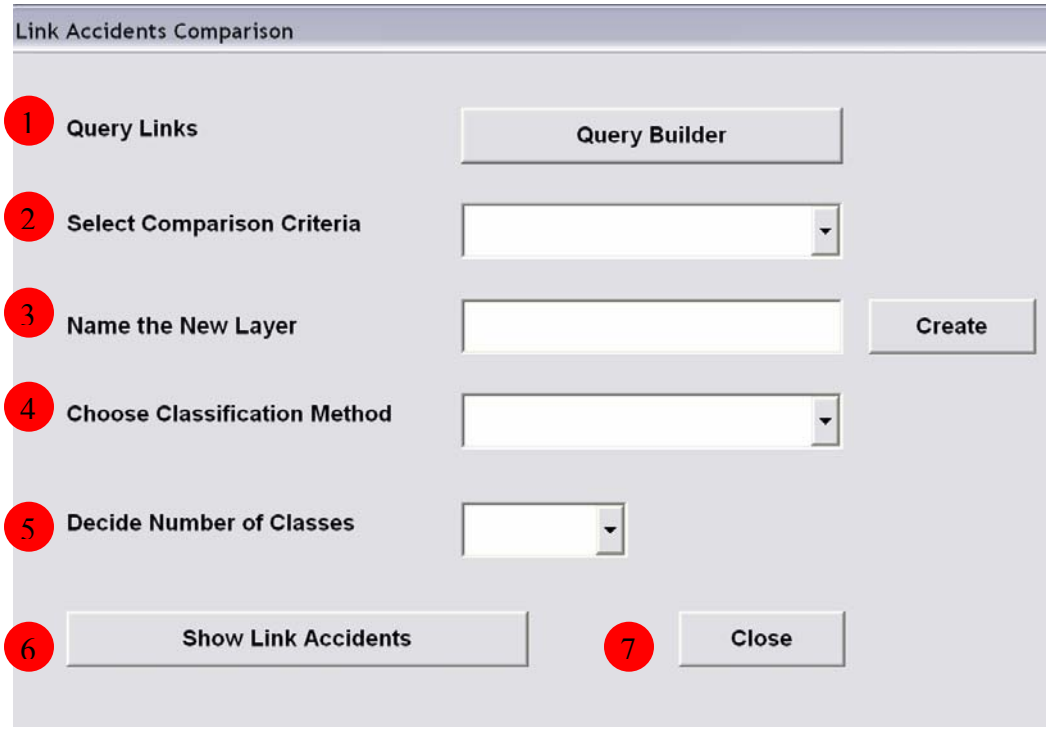


Figure 33: Accident Module

1 **Query Links** – This works similar to '*Select by Attributes*' tool in ArcGIS. Users can select specific links or intersections by defining a selection criterion. For example, users can select all the links in a town or a TAZ.

2 **Select Comparison Criteria** – Five comparison criteria are provided in the module. These comparison criteria use specific fields from the attribute table of the layer provided with this interface. A brief description of these criterions is provided below. Terms in parentheses are the corresponding field names which have to be present in the working layer of shape file.

(1) Observed (*Tot_Count*) – This presents the observed crash counts on the road segments.
(2) Predicted (*EstimatedAccidents*) – This presents the estimated crash counts from the statistical models.
(3) Difference (*Difference*) – This presents the difference between the predicted and observed crash count (i.e. difference = observed – predicted). A positive value means that there is an underestimation of the number of crashes. On the other hand, a negative value indicates overestimation at that location.
(4) Ratio (*Ratio*) – This is the ratio of predicted over observed (i.e. ratio = observed/predicted) crashes. A value over 1.0 means underestimation and a value under 1.0 indicate overestimation.

Note: The users must use the same field name for storing the observed and estimated crashes, and also the difference and the ratio of these values in the attribute tables of the feature layers or the shape file in order to use these comparison criteria. The following criterion can be used for any field.

(5) User Defined Field – Selecting this criterion enables the users to select a specific field from the layer or shape file attribute table. This is useful if the user does not have the values with the correct field names as specified above.

3 **Name the New Layer** – Provide a name for the new layer that will contain and display the selected links or intersections. Remember to click '*Create*' after entering the name.

4 **Choose Classification Method and** 5 **Decide Number of Classes**

Described here are five ways to classify the values in the fields into a given number of class breaks. Users can choose between 3 and 10 classes.

(1) Manual Class Breaks – This method divides the data into specified number of classes at the class break values defined by the user.
(2) Equal Interval – This method divides the classes such that each class has equal intervals (10-20, 20-30, etc.). The entire range of the data between minimum and maximum is divided equally into the specified number of classes.
(3) Quantile – This method classifies data into the specified number of classes such that each class had equal number of observations in it.
(4) Natural Breaks – This method identifies break points by picking the class breaks that best group similar values and maximize the differences between classes. The features are divided into classes whose boundaries are set where there are relatively big jumps in the data values.

6 **Show Link / Intersection Accidents** – Clicking this button displays the map with the new layer.

7 **Close** – This button ends the module.

IV. How to Use the Interface

This section provides the step-by-step procedure for the use of these modules. Examples 1 and 2 demonstrate the '*Link Accident Module*' and Examples 3 and 4 present the '*Intersection Accident Module*'. The link and intersection layers are based on the GIS layers provided by the Capitol Region Council of Governments (CRCOG).

Example 1. Using query builder

Using query builder enables the users to select specific links or intersections that satisfy a certain criterion. Following are the steps involved with using this feature:

a. Open *NETC Accident Model Comparison Module.mxd*

b. Add data to the map: Layers or shape files can be added to an empty map from the geodatabase using the File→Add Data option from the main menu. '*Predicted Accidents*' layer was used for this example. ***It should be noted that if there are multiple layers in the map document, the layer with the complete dataset (i.e. the working layer) should always be at the top in the table of contents (TOC) on the left hand side. This can be done by dragging the layer to the top of the TOC. In addition, any existing selections in the map must be cleared (Figure 34).***

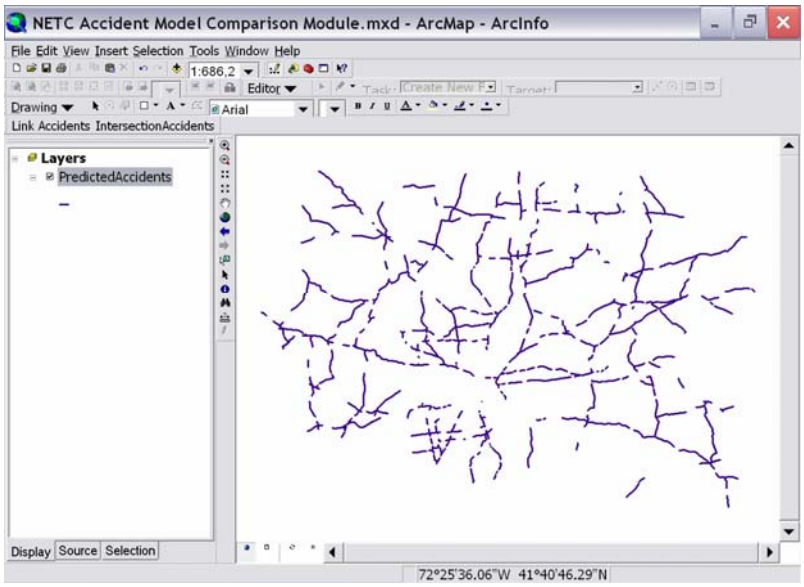


Figure 34: Initial screen with complete data

c. Click '*Link Accident*' on the toolbar and then click the '*Query Builder*' button. For this example, the segments with length greater than 1 mile were selected. After defining the selection criterion click '*OK*'. Figure 35 shows the query builder and the selected links which are highlighted.

Users should make sure that links from correct layer are selected and the selection method shows "Create a new selection" (Figure 35).

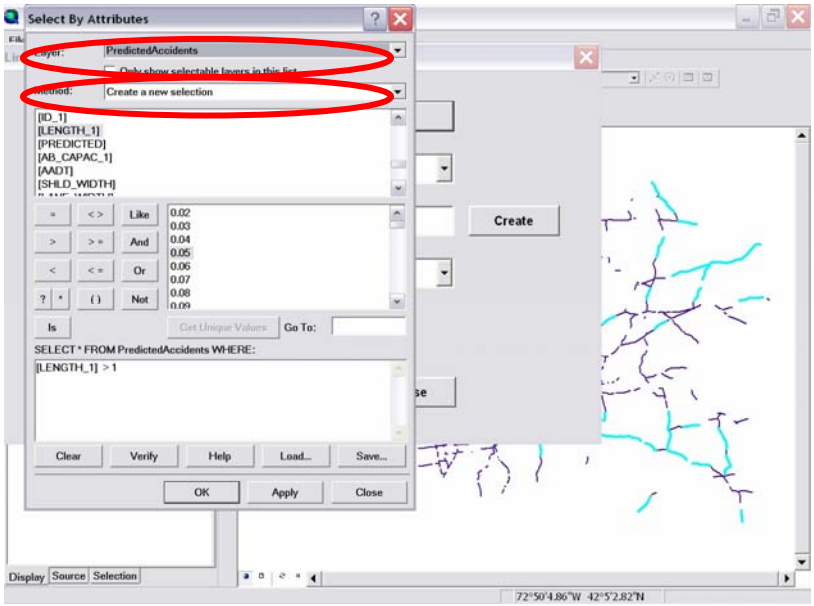


Figure 35: Query Builder

d. Next, a comparison criterion is selected. In this case we need to observe the number of crashes predicted by the statistical models thus '*Predicted*' was selected (Figure 36).

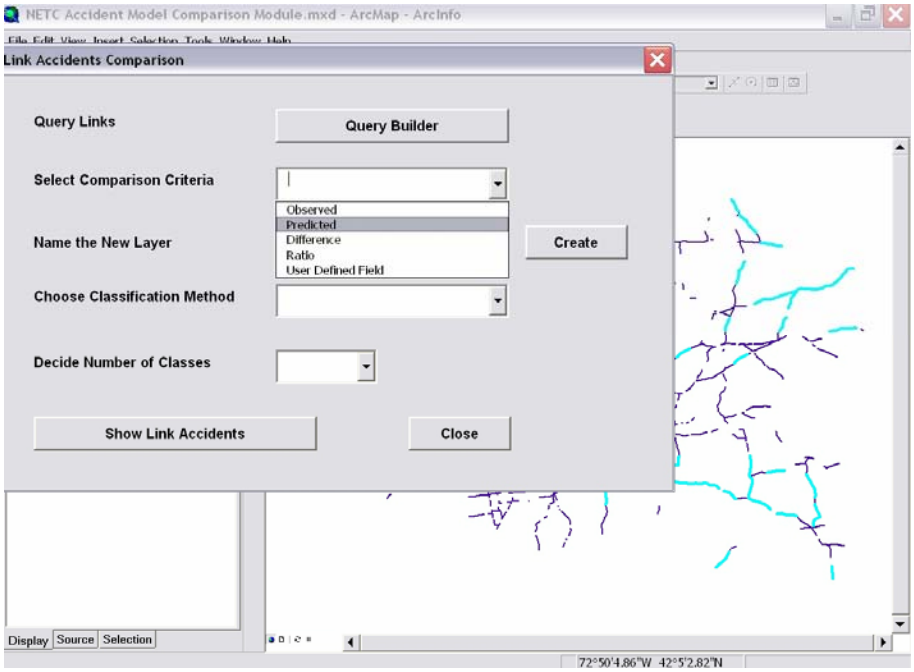


Figure 36: Selecting Comparison Criterion

e. Enter a new layer name and click '*Create*' button to create a new layer to store and display the crashes on the selected links. '*Query Link*' is the layer name provided for this example (Figure 37).

f. The next step is to select a classification method. The '*Manual Class Breaks*' method is used.

g. Once the number of classes has been selected, a new window pops up asking the users to provide the class break values. Three classes were selected for this example, thus two class break values (30 and 60) were provided (Figure 37). Click '*OK*' to go back to the module.

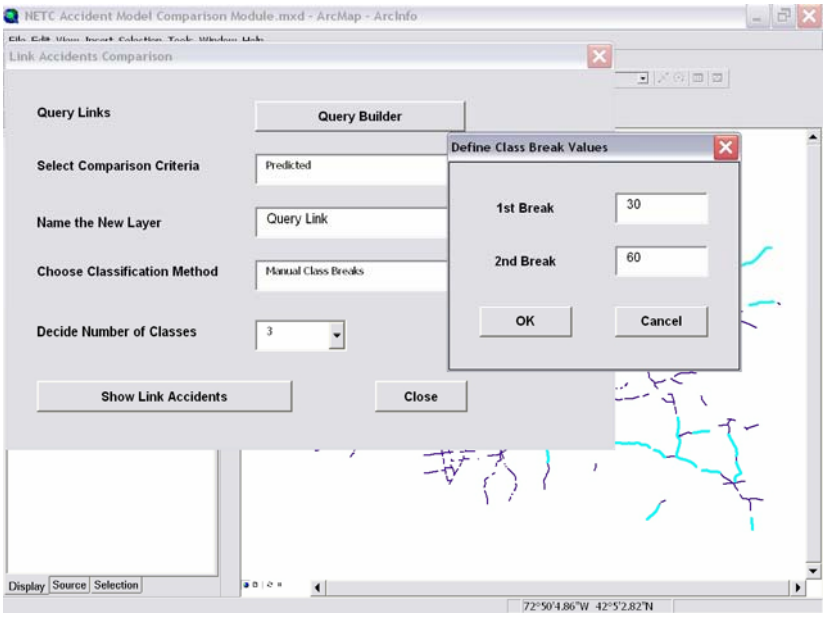


Figure 37: Defining Class Breaks

h. Click '*Show Link Accidents*' followed by '*Close*'. The resulting map is presented in Figure 38.

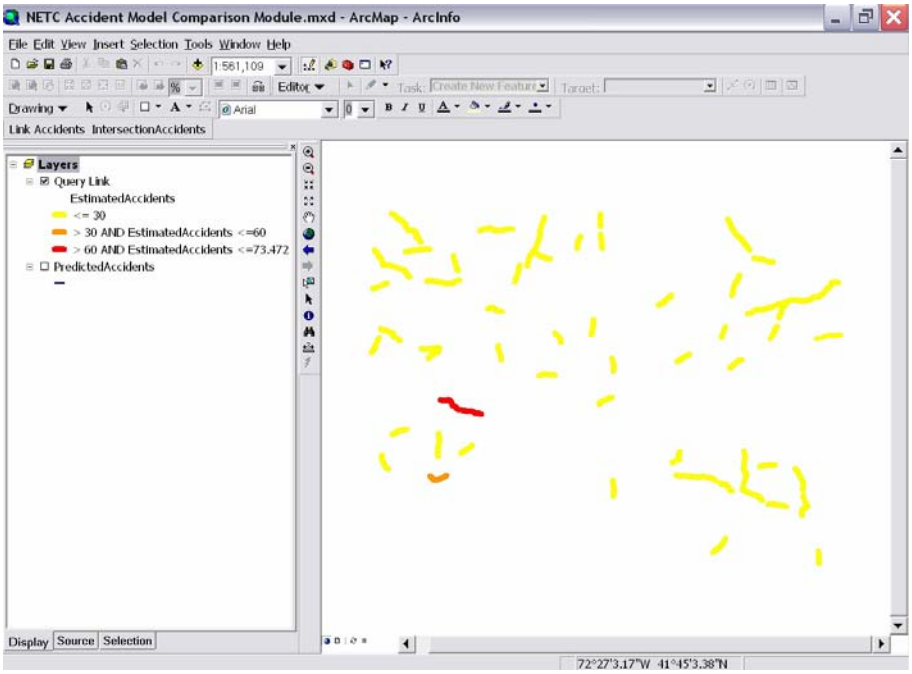


Figure 38: Final Map from Example 1

Example 2. Using rubber band

Users can select links or intersections using a selection tool (referred to as the rubber band) provided by ArcGIS.

a. Open *NETC Accident Model Comparison Module.*
Before going on to the next stage, make sure to clear any existing selection. In addition, if there are multiple layers in the map document, the working layer should always be at the top of the TOC. This can be done by dragging the layer to the top of the TOC. If necessary, click 'Full Extent ( )' button to see the entire map.

b. Now click '*Select Features ( )*' button and drag the mouse to select the area of interest. Figure 39 shows the selection result.

c. Select '*Difference*' as the comparison criterion.

d. Provide '*Rubber Link*' as the new layer name and click '*Create*' (Figure 40).

e. For this example, the '*Equal Interval*' method was selected with four classes.

f. Click '*Show Link Accidents*' and '*Close*'. The resulting map is shown in Figure 41.

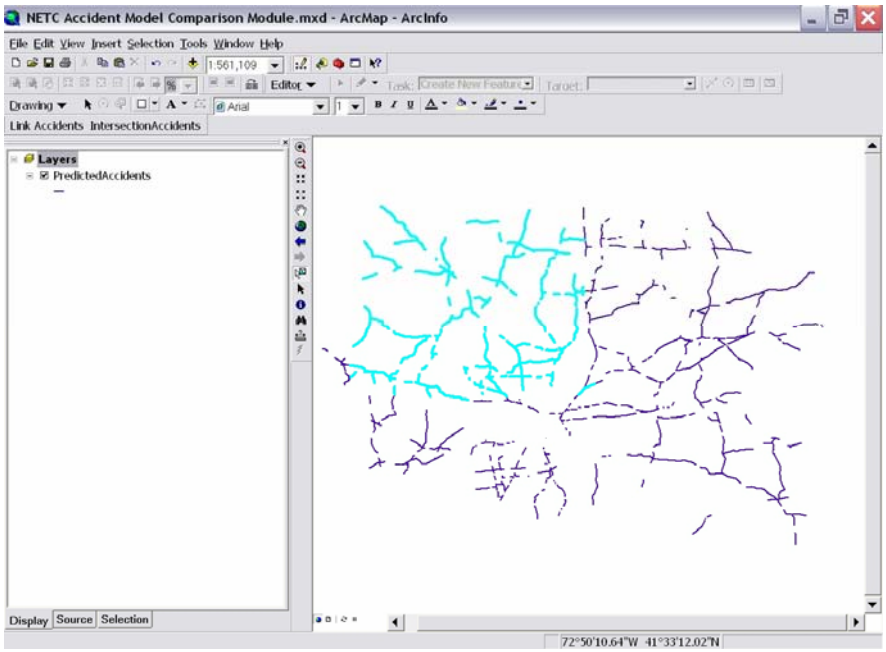


Figure 39: Selecting links using the 'Rubber Band'

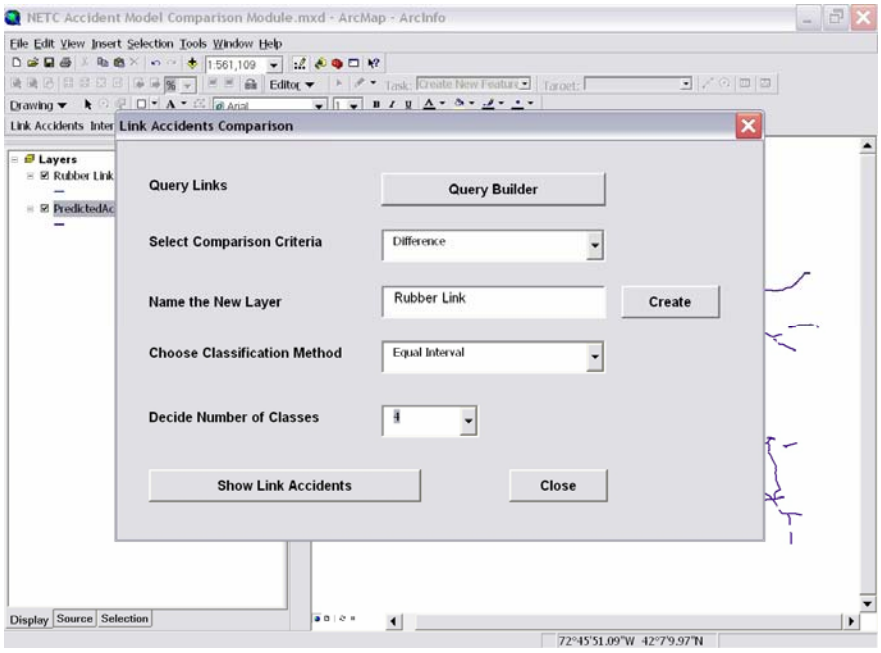


Figure 40: Full Link Accident Module

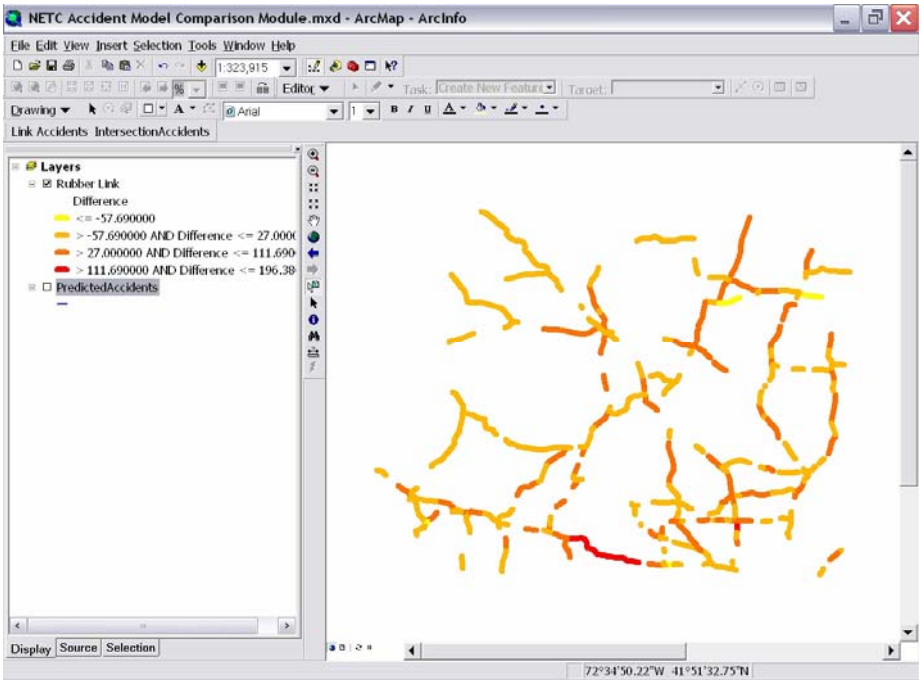


Figure 41: Final map from Example 2

Example 3. Select individual links or intersections by mouse click.

Users can select individual links or intersections using the '*Select Features ( )*' tool. To select multiple links or intersections, click individual links or intersections while holding the shift key on the key board.

a. Open *NETC Accident Model Comparison Module.*

Before going on to the next stage, make sure that there is no existing selection. In addition, if there are multiple layers in the map document, the working layer should always be at the top of the TOC. If necessary, click 'Full Extent ( )' button to see the entire map.

The '*Intersection Accidents*' module and '*Intersection_Nodes*' layer were used (Figure 42).

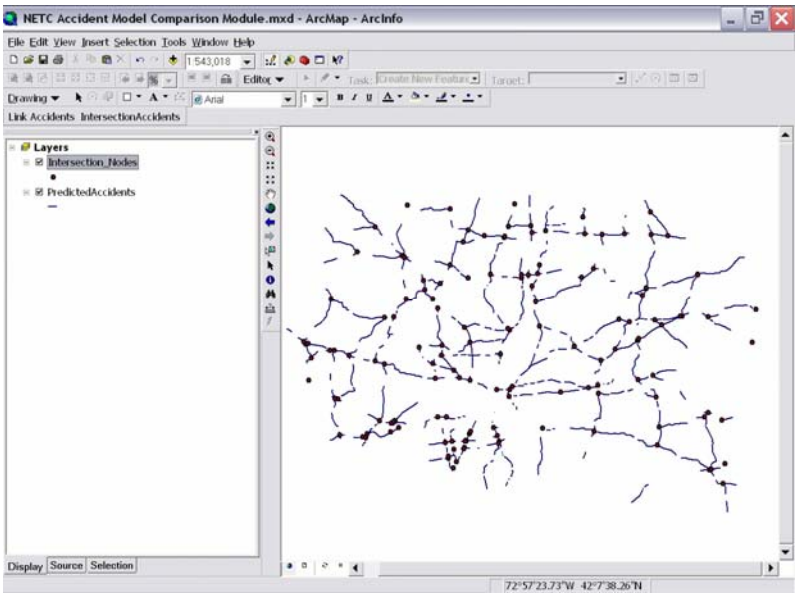


Figure 42: Maps showing links and intersections

b. Click '*Select Features (*  *)*' button and click intersections of interest while holding the shift key. Figure 43 shows the selection result.

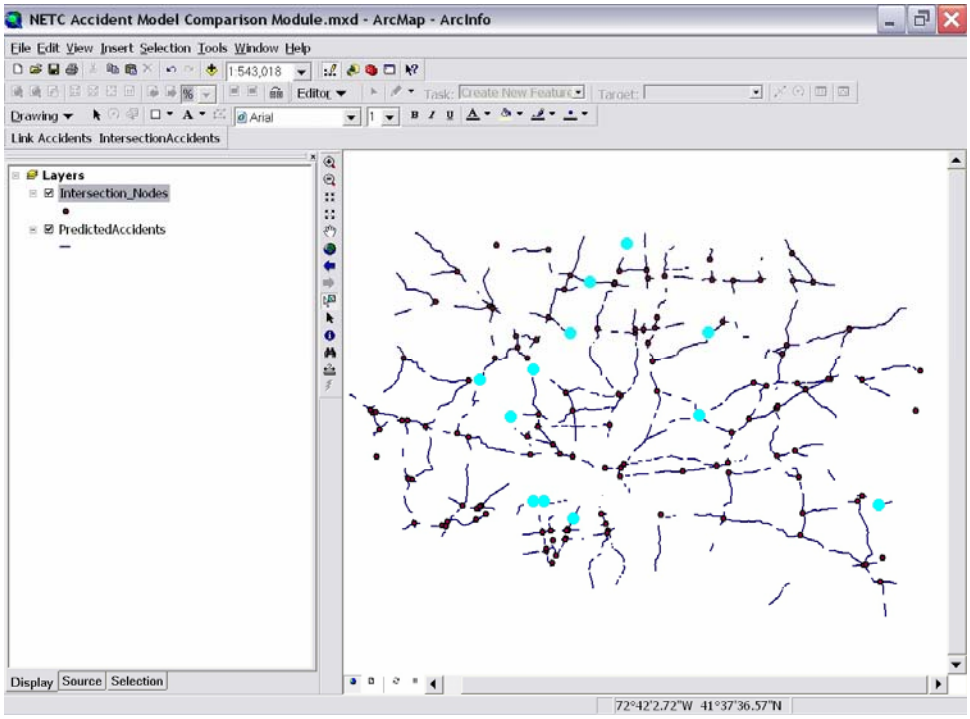


Figure 43: Selected intersections

c. Use '*Observed*' as the comparison criterion.
d. Use '*Mouse Click Intersection*' as the new layer name and click '*Create*'.
e. Use '*Quantile*' as classification method with four classes (Figure 44).

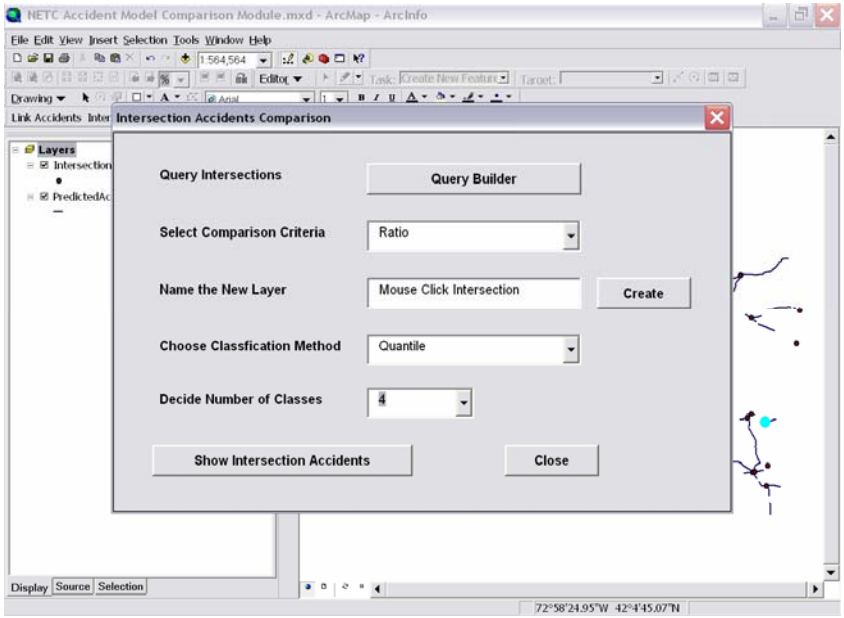


Figure 44: Intersection Accident Module

f. Click '*Show Intersection Accidents*' and '*Close*'. The resulting map is shown in Figure 45.

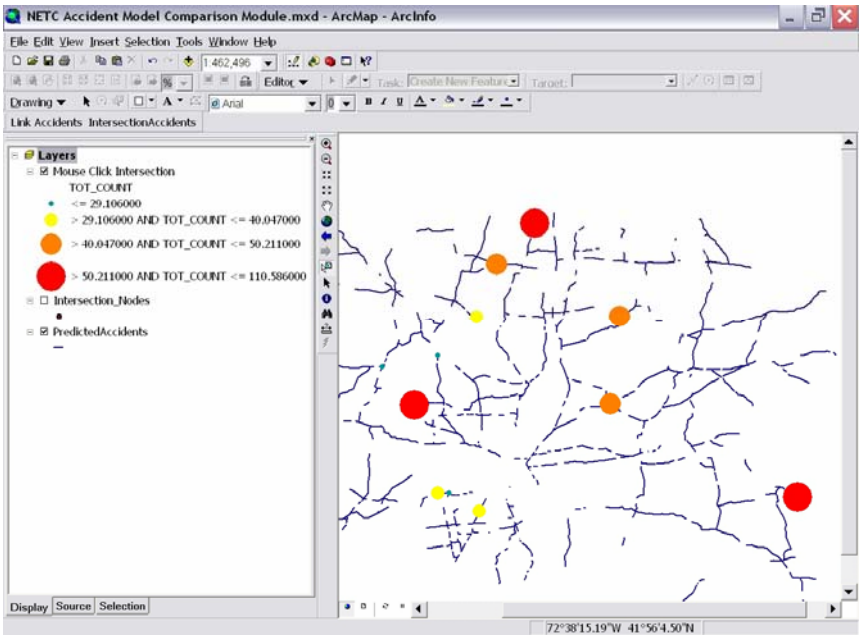


Figure 45: Final map from Example 3

Example 4. Select individual links/intersections, then use query builder to create a new selection from the current selection

Users can select individual links or intersections by using '*Select Features ( )*'. These selected links or intersection can be queried further using the query builder in the modules.

a. Open *NETC Accident Model Comparison Module.*
Before going on to the next stage, make sure to clear any existing selection. In addition, if there are multiple layers in the map document, the working layer should always be at the top of the TOC. If necessary, click "Full Extent ( )" button to see the entire map area.
The '*Intersection Accidents*' module and '*Intersection_Nodes*' layer were used.

b. Click '*Select Features ( )*' button, and then drag the mouse to select intersections. Figure 46 shows the selection result.

c. Open '*Intersection Accident*' module, and click '*Query Builder*'. Provide the query criterion. For example, intersections with four legs were selected from the existing selection set.
Users should make sure that the intersections from correct layer are selected and the selection method shows 'Select from current selection"(Figure 47).

d. Select '*User Defined Fields*' as comparison criterion. Enter the field name of into the new window. Figure 48 shows use of the field '*Test*' from the attribute table of the intersection layer. Remember this field should already exist in the attribute table of the working layer.

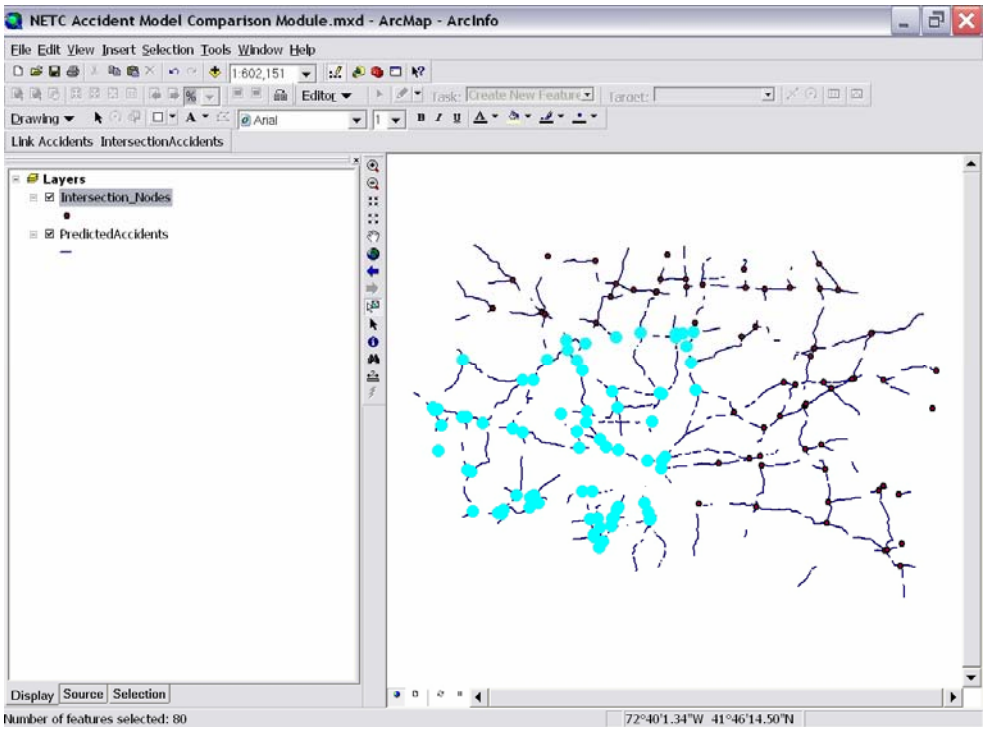


Figure 46: Selected Intersections

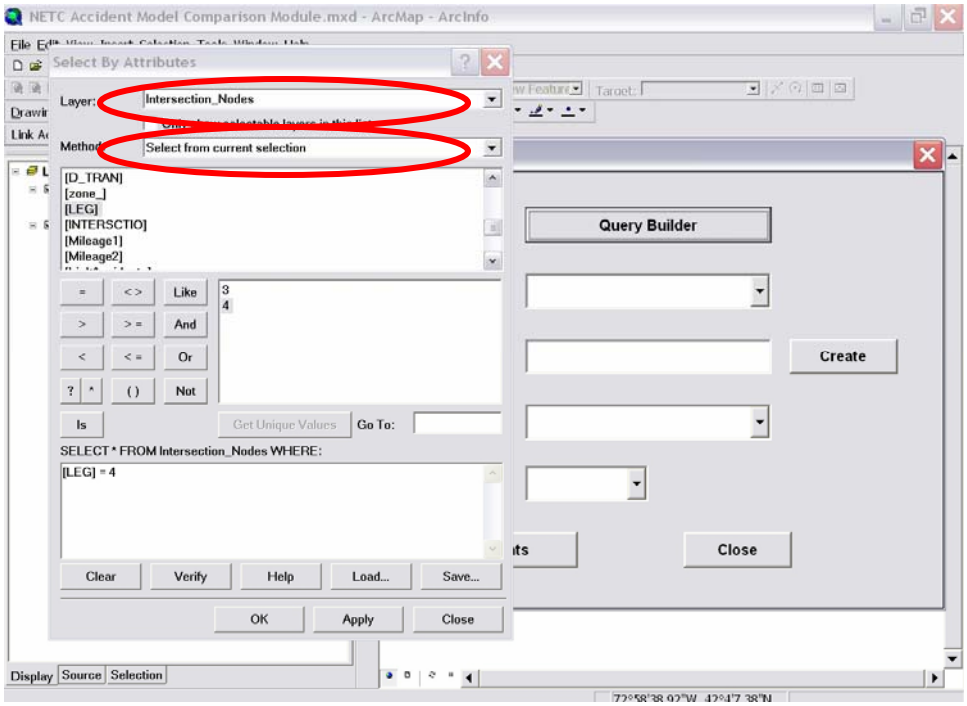


Figure 47: Query Builder

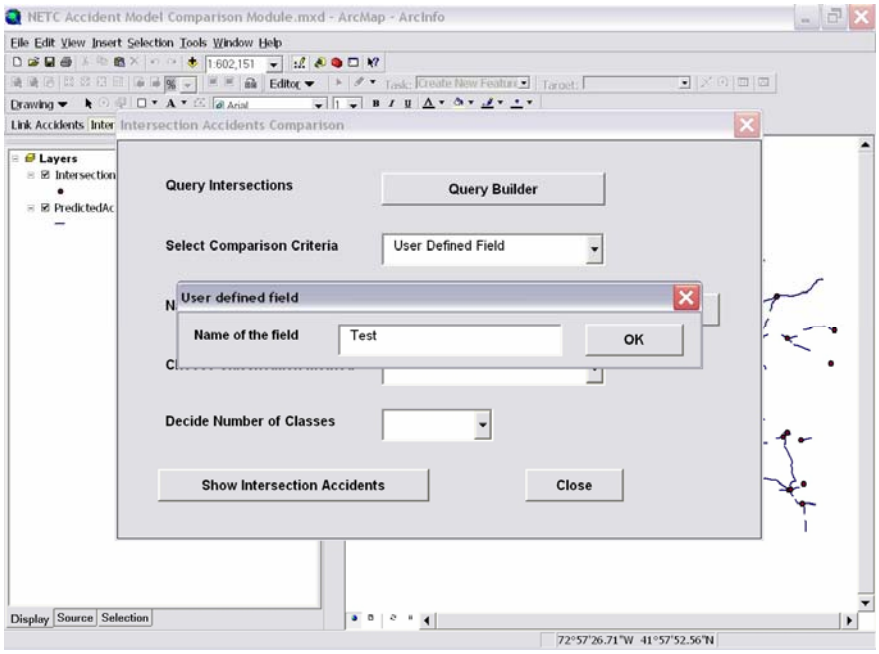


Figure 48: Using user defined fields

e. Use '*Rubber and Query*' as the new layer name and click '*Create*'.

f. For this example, '*Natural Breaks*' method was selected with four classes (Figure 49).

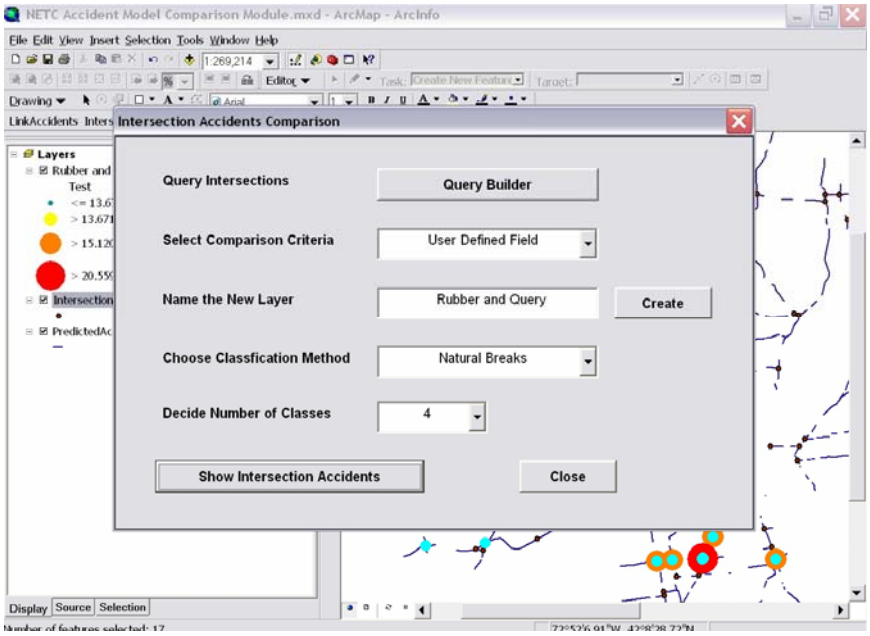


Figure 49: Defining comparison criteria and classification method

g. Click '*Show Intersection Accidents*' and '*Close*'. The resulting map is displayed in Figure 50.

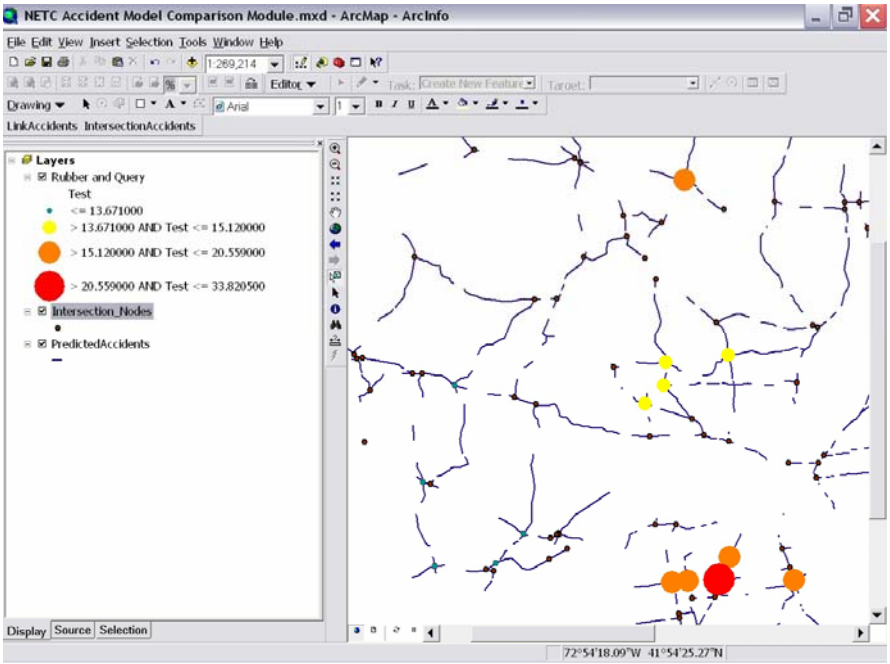


Figure 50: Final map from Example 4

V. Troubleshooting

This section lists a few common errors that users might encounter while using this interface and what to do when they arise.

1. User interface toolbar is not shown

When the map document is opened, the user interface is often not displayed on the toolbar. If this is the case, use the right click button on the mouse to open the context menu, and check '*NETC UI Toolbar*'. Then drag the buttons to where you want them on the tool bar.

2. 'Unsupported function' error message

Users may encounter the '*Unsupported function*' error message. This is an unexplained error message and arises when there is a conflict in executing some of the commands in VBA. It is suggested that you save your map document, exit ArcGIS, and reopen the map document to continue working.

3. 'Query Builder' does not respond

Sometimes the '*Query Builder*' on the module does not respond. If you encounter this unexplained error, please close the interface by clicking the close button () on the **interface**, **not on the query builder**. Then, save the current document and reopen it to continue working.