2012

Identifying Traffic Safety Needs – A Systematic Approach: Research Report and User Manual

Andrew P. Tarko
Purdue University, tarko@purdue.edu

Shafiul Md Azam
Purdue University, sazam@purdue.edu

Jose Thomaz
Purdue University, jthomaz@purdue.edu

Mario Romero
Purdue University, maromero@purdue.edu

JOINT TRANSPORTATION RESEARCH PROGRAM

IDENTIFYING TRAFFIC SAFETY NEEDS: A SYSTEMATIC APPROACH, VOLUME 1: RESEARCH REPORT

Andrew P. Tarko
Professor of Civil Engineering
Purdue University
Corresponding Author

Shafiul Md Azam
Graduate Research Assistant
School of Civil Engineering
Purdue University

Jose Thomaz
Data Analyst
Center for Road Safety
Purdue University

Mario Romero
Graduate Research Assistant
School of Civil Engineering
Purdue University

This page intentionally left blank.

RECOMMENDED CITATION

CORRESPONDING AUTHORS

Prof. Andrew P. Tarko
School of Civil Engineering
Purdue University
(765) 494-5027
tarko@purdue.edu

JOINT TRANSPORTATION RESEARCH PROGRAM

NOTICE

This page intentionally left blank.

1. Report No. FHWA/IN/JTRP-2012/02	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle **Identifying Traffic Safety Needs – A Systematic Approach: Research Report and User Manual**		5. Report Date 2012	
		6. Performing Organization Code	
7. Author(s) Andrew P. Tarko , Shafiul Md Azam , Jose Thomaz , Mario Romero		8. Performing Organization Report No. FHWA/IN/JTRP-2012/02	
9. Performing Organization Name and Address Joint Transportation Research Program Purdue University 550 Stadium Mall Drive West Lafayette, IN 47907-2051		10. Work Unit No.	
		11. Contract or Grant No. SPR-3315	
12. Sponsoring Agency Name and Address Indiana Department of Transportation State Office Building 100 North Senate Avenue Indianapolis, IN 46204		13. Type of Report and Period Covered	
		14. Sponsoring Agency Code	

15. Supplementary Notes

16. Abstract

The Indiana Department of Transportation (INDOT) manages road safety in Indiana through safety emphasis areas, identification of safety needs within these areas, and development and implementation of transportation interventions that address the safety needs. Indiana applies system-wide performance screening to state roads using the somewhat limited, non-integrated tools available to date. A site-specific countermeasure development and evaluation process (scoring) is being used to determine essential project intent and merit or cost-effectiveness. This projects addresses the need for a unifying approach to data-driven identification of safety emphasis areas and safety improvement projects via a systematic evaluation of safety needs in the Indiana road network. Such a method must utilize a database that integrates the Indiana state network, traffic, road inventory, and crash data.

A method for safety screening of state roads in Indiana was developed with the following screening tasks in mind: (1) Identification of high crash locations (segments, intersection, ramps and bridges), corridors, and areas in terms of high crash frequency, crash rate, or proportion of a particular crash type. (2) Facilitate program-based screening (e.g., shoulder widening, median improvement, etc.). (3) Facilitate special programs and projects (e.g., five percent report)

In order to ensure the versatility of screening tasks and the data maintenance, four components are included in the screening tool: (1) data management, (2) standard screening, (3) roads clustering and special studies, (4) results presentation. The data management includes updating the existing data by utilizing renewable sources for data, reformatting the source data to meet the standards of the master database, integrating these data into tables that meet certain specifications, and replacing the existing data. The data management procedures include ArcGIS geo-processing and VBA-implemented and Model Builder codes that are not packaged as a single module but which are used separately as needed to maintain flexibility of the data management process.

A window-based user interface, query editor, facilitates selection of criteria for safety screening. To help the user build a query, the selection criteria have been grouped into four categories: geographical scope (state, county, etc.), element (segment, township, etc.), crash criteria, and road criteria. The query is translated to SQL and sent to the screening engine. The results include safety performance measures such as crash frequencies, rates, and proportion together with basic statistical indicators. Road segments and intersections that exhibit an excessive number of crashes may be concentrated along longer road sections. Clustering these elements may reveal large scale safety issues that otherwise might be overlooked if the screening analysis is focus on individual spots. The geo-coded results can be sorted in a table or displayed on GIS maps with Google Earth or ArcGIS to visualize the spatial distribution of the identified high-crash roads.

17. Key Words road safety, roads screening, safety needs, safety management, computer tool		18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 125	22. Price

Form DOT F 1700.7 (8-69)

This page intentionally left blank.

EXECUTIVE SUMMARY

IDENTIFYING TRAFFIC SAFETY NEEDS – A SYSTEMATIC APPROACH

Introduction

The Indiana Department of Transportation (INDOT) manages road safety in Indiana through safety emphasis areas, identification of safety needs within these areas, and development and implementation of transportation interventions that address the safety needs. Indiana applies system-wide performance screening to state roads using the somewhat limited, non-integrated tools available to date. A site-specific countermeasure development and evaluation process (scoring) is being used to determine essential project intent and merit or cost-effectiveness.

This project addresses the need for a unifying approach to data-driven identification of safety emphasis areas and safety improvement projects via a systematic evaluation of safety needs in the Indiana road network. Such a method must utilize a database that integrates the Indiana state network, traffic, road inventory, and crash data.

Findings

A method for safety screening of state roads in Indiana has been developed with the following screening tasks in mind:

(1) Identification of high crash locations (segments, intersection, ramps and bridges), corridors, and areas in terms of high crash frequency, crash rate, or proportion of a particular crash type.
(2) Facilitate program-based screening (e.g., shoulder widening, median improvement etc.).
(3) Facilitate special programs and projects (e.g., five percent report)

In order to ensure the versatility of screening tasks and the data maintenance, four components are included in the screening tool: (1) Data management, (2) Standard screening, (3) Roads clustering and special studies, (4) Results presentation.

The data management includes updating the existing data by reaching to renewable sources for data, reformatting the source data to meet the standards of the master database, integrating these data into tables that meet certain specifications, and replacing the existing data. The data management procedures include ArcGIS geo-processing and VBA-implemented and Model Builder codes that are not packaged as a single module but which are used separately as needed to maintain flexibility of the data management process.

A window-based user interface, query editor, facilitates selection of criteria for safety screening. To help the user build a query, the selection criteria have been grouped into four categories: geographical scope (state, county, etc.), element (segment, township, etc.), crash criteria, and road criteria. The query is translated to SQL and sent to the screening engine. The results include safety performance measures such as crash frequencies, rates, and proportion together with basic statistical indicators.

Road segments and intersections that exhibit an excessive number of crashes may be concentrated along longer road sections. Clustering these elements may reveal large scale safety issues that otherwise might be overlooked if the screening analysis is focus on individual spots.

The geo-coded results can be sorted in a table or displayed on GIS maps with Google Earth or ArcGIS to visualize the spatial distribution of the identified high-crash roads.

Implementation

The developed tool includes two major parts: data management component which is a collection of procedures and tools to facilitate data renewal and integration. Data management is complex but it is repeated only once a year or two years. The second part includes the screening and clustering tools that have a convenient user interface and are fed by datasets prepared in the data management process and optimized for screening. Screening and clustering have convenient user interfaces as they are expected to be run more frequently than the data management component.

A research report describes the concept and statistical fundamentals of the screening and clustering operations. A user manual developed as the part of the project includes an example study and is meant to help train the INDOT personnel. A sequence of hand-on workshop are helpful in facilitate the implementation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background

In 2010, the commissioner of the Indiana Department of Transportation (INDOT), signed Indiana's revised *Strategic Highway Safety Plan* (SHSP). SHSP, with INDOT as the lead agency, was developed cooperatively with several other state and federal agencies with a comprehensive approach to reducing traffic crash injuries and deaths through coordinated engineering, education, enforcement, and emergency response (the 4E's). The mission expressed in the document's preamble reads: "*The quality of crash data and our ability to analyze what it reveals has improved, providing a better picture of where and how to focus our traffic safety efforts to achieve maximum effect… Safety is one of the three core values at the heart of planning, designing, building and operating a transportation system.*"

Strategies corresponding to the emphasis areas were identified and agencies like INDOT, Indiana Criminal Justice Institute, Indiana State Police were appointed to lead the efforts in their areas. The plan emphasizes the importance of an integrated approach to traffic safety, data analyses, application of the latest research, and best practices from across the U.S. as a means of generating a sound basis and tools for safety management decisions.

INDOT has a primary role in addressing transportation infrastructure as it relates to traffic safety, mindful of the Highway Safety Improvement Program (HSIP) in its approach to addressing traffic safety performance and, correspondingly, adoption of emphasis areas for analysis and project selection. Population of the safety program with projects by the Traffic Safety Office is currently in process with system-wide performance screening using the somewhat limited, non-integrated tools available to date. A site-specific countermeasure development and evaluation process (scoring) is being used to determine essential project intent and merit or cost-effectiveness.

This method and associated (computer) applications does not readily and efficiently permit development of a truly systematic and optimized traffic safety program capable of automating the linkage of actual crash records/data to INDOT's road inventory to traffic volume assignments to performance/prediction models to site-specific analysis (of relative risk and positive countermeasures), all in a geo-spatial format. A unifying approach and a tool for determining the most relevant safety emphasis areas are needed that would lead to selection of the most cost-effective population of projects having the most favorable effect on statewide traffic safety performance. SPR-3315 was thus proposed to develop a comprehensive method of identifying road safety needs.

Although the project includes a strong practical component and a well-defined product, it also addresses certain fundamental questions, including identification of specific crash problems and problem areas related to both roads and drivers.

1.2. Literature Review

The development of strategic highway safety plans and associated research efforts have been performed in many states. The Iowa Department of Transportation developed their Strategic Highway Safety Plan in 1999. Following this strategic plan, there have been many research projects for the Iowa DOT to systematically identify high crash locations. A similar study *(1)* identifies hazardous locations by integrating a Geographic Information Management System (GIMS) and a Geographic Information System Accident Analysis and Location System (GIS-ALAS) database. This particular study focused mainly on five crash types, namely, horizontal curve, fixed object, rural four lane expressway intersection, and urban four-lane undivided corridor crashes. Since both databases were based on a GIS system, it was straightforward process for them to analyze the situation on a case by case basis. The crash ranking was based on the summation of the frequency rank, crash rate rank, and dollar cost rank; and the lowest value was considered the best overall crash ranking. Special emphasis was given to a horizontal curve detection technique utilizing a GIS application.

The latest federal transportation act, SAFETEA-LU, created the Highway Safety Improvement Program (HSIP) for the purpose of reducing traffic fatalities and serious injuries on public roads. Prior to SAFETEA-LU, traffic safety was not a "core" program, but merely a set-aside of the Surface Transportation Program (STP). Unlike almost all other funding programs, SAFETEA-LU mandates that every state submit the following annual reports to FHWA: (a) HSIP report, High Risk Rural Roads (HRRR) report, Section 130 report (related to the use of rail crossing safety funds), and a report describing not less than five percent of their highway locations exhibiting the most severe safety needs. The HSIP, HRRR, and Section 130 reports must report not only on the preceding year's project-by-project federal-aid expenditures, but also present before-and-after effectiveness reports (three years pre- and post-construction; again, project by project) in terms of reduced crashes, notably fatal and severe injury events. INDOT's reports are prepared by the Traffic Safety Office. In particular, the Five Percent Report that must be prepared annually starting from 2006 required a new scope for development and implementation of state-wide screening techniques. Two examples of such efforts are mentioned below.

In Indiana, crash screening and ranking has been conducted for the last few years as a joint effort of INDOT and the Center for Road Safety (CRS) of Purdue University *(2)*. Crash assignment methodologies for INDOT crash screening are based on a GIS technique, and the assignment is performed in ArcGIS. Separate tools were developed for network splitting, link and intersection identification, crash database processing, and, finally, for crash assignment. Since 2008, local roads have been considered in the screening process and relatively complicated interchange configurations have

been brought into the safety screening process. Separate crash rankings based on road location (segment vs. intersection), jurisdiction (state vs. local), and types have been considered in the Five Percent Report for the last three years. The ranking of state and local roads was performed by computing the Crash Cost Index (I), which is the difference between actual crash cost and typical crash cost divided by a standard deviation of these two parameters. The typical crash cost is based on the expected value of safety performance functions (SPFs) which consider only exposure variables in their models. For state roads, these SPFs were developed by a negative binomial model. For local roads, due to the lack of an important predictor variable, land-use exposure variables were considered to develop the typical crash frequency or crash rate and a data-mining technique was used.

The Delaware Department of Transportation (DDOT), in their 2007 Five Percent Report, used a linear reference milepost system to identify high crash locations (3). Only the state roadway systems are investigated in their five percent report because the suburban and municipal roadway systems were not included in their linear reference system. The roadway classifications include urban/rural, two-lane, and multilane divided/undivided categories. A total of 29 different roadway types were detected and a crash rate (CR) for each category was calculated. The Critical Crash Rate (CCR), which actually represents the statistical test of significance of the calculated crash rate, was estimated for each 0.30 mile segment of roadway. The critical ratio, which is the ratio of the CR and the CCR, if greater than or equal to 1, signifies that the segment is a candidate location. The report also contains the potential or already considered improvements and their estimated cost.

The use of GIS in identifying high crash locations has become a more promising method, and GIS methods have been successfully implemented in many INDOT projects. Some tools have been invented for easy manipulation and integration of a crash and GIS database (e.g., PASS-INPASS (4). Although GIS techniques have been used for a decade in crash identification and screening, researchers are trying to explore a more sophisticated use of GIS in solving certain safety problems. An important study (5) presented a macroscopic and microscopic case study of certain problem-solving processes using GIS techniques. The macroscopic model showed the use of a grid-based system in identifying potentially hazardous zones (e.g., areas where children are traveling to school are potentially vulnerable). A similar approach could be used to study crashes with alcohol involvement. Therefore, a GIS tool could be used for adopting interventions like education or patrolling and enforcement in specific regions. In another study, the applicability of grid-based and network-based techniques in screening and clustering of high crash locations was studied (6).

Some local government agencies have actively been considering GIS-based hazardous location identification systems. One such example is the South East Michigan Council of Government (SEMCOG), which has online GIS crash mapping for high crash locations. GIS methods have also been studied to identify clusters of crashes, where the Simple and Kernel Methods were studied to estimate the density of crashes (7). The Simple Method uses a preset number of cells in which the simple ratio of the crash number and a typical circular search area represent the cell density. The Kernel Method uses a method whereby a circle is drawn around each crash and kernel values are calculated using the quadratic kernel function available in ArcGIS. The sum of the overlapping kernel values represents the cell density. A specific cluster of pedestrian hot spots were calculated by a composite method which uses crash rates based on pedestrian volume and population in the area, as well as computed crash density based on the Simple or Kernel Methods.

In addition to crashes caused by the roadway geometry or an operational deficiency, the identification of zones having driver behavioral issues have been considered in many past research projects. Tarko et al. (8) proposed a method based jointly on the statistical significance and the excess of crashes above a typical number estimated with macroscopic safety performance functions (8). Use of the safety performance index (SPI) calculated for a geographical region, such as a county or city, was considered in a study for the state of Michigan (9). SPI is the sum of the weighted frequency of PDO, fatal, and injury crashes, with special focus on crash frequency (i.e., crashes due to alcohol intoxication) with a composite rate of all crash types, with population, registered vehicles, VMT, etc. as the exposure variables.

The identification process is non-trivial and the methods for selecting candidate road locations are numerous. Apart from advanced techniques for crash assignment, the screening technique also requires the statistically significant locations having safety issues which need to be addressed on a priority basis under the constraints of a limited budget. For this reason, statistical control techniques have been suggested for identifying hazardous locations (10), whereby a particular location was considered hazardous if the number of crashes at the location was higher than the upper control limit. A similar concept was proposed by Rudy (11) and Morin (12) for crash rates. Nakmoon et al. (13) proposed a quality control method based on negative binomial distribution. The distribution is determined from the variability of crash counts in the sample of similar type locations.

Estimating the expected number of crashes at a location is the central interest of the majority of statistical methods. The multivariate regression model proposed by Jorgensen (14) calculates the expected numbers of crashes at locations. Abbess et al. (15) and Maher (16) found that negative binomial distribution can be used to describe crash variability across locations. A loglinear regression model to calculate the expected number of crashes was proposed by Maycock and Maher (17).

Jorgensen *(14)* proposed to prioritize locations according to the difference between the reported and expected numbers of crashes divided by the square root of the expected frequency of crashes.

The introduction of data mining and similar approaches for crash prediction such as fuzzy logic, neural network, and classification trees *(18, 19, 20)* opened a wide spectrum of methodologies to handle complicated and vast amounts of data in a relatively easy way. Since the local road network is the largest part of a road network for any public road system, which often lacks important exposure variables like ADT, developing appropriate safety performance functions is a difficult process. A set of surrogate variables like land-use and demographic data and network characteristics have been proven to be a good set of variables for safety performance function *(21)*. This study also showed that an advanced data mining application performs well over the conventional negative binomial model.

AASHTO *(22)* first addition of the Highway Safety Manual (HSM) provides the factual information and tools to facilitate roadway planning, design, operations, and maintenance decisions based on consideration of their safety consequences. The primary focus of the HSM is the introduction and development of analytical tools for predicting the impact of transportation project and program decisions on road safety. The manual's biggest contribution is promoting analytical and data-based decision making in safety management. There still a great need for developing safety-management tools and processes for specific regions and jurisdictions that fit well local conditions and data availability. The HSM is quite useful in regions where local tools and methods and lacking.

1.3. Objective and Scope of Work

The primary objective of this research project was to develop a comprehensive safety identification tool to address statewide INDOT safety investment needs. A detailed list of the research objectives and scope of work are shown below:

- Identify INDOT-administered roads with safety needs based on an excessive number of crashes in certain categories (severe, off-road, head-on, night, etc.). The scope of this specific objective is state roadway elements like segments, intersections, ramps, and bridges.
- Develop a method of screening the road network for road deficiencies. Again, state roadway elements (e.g., segments, intersections, ramps, and bridges) are considered for screening for roadway deficiencies.
- Develop a method of identifying area-wide safety problems, such as aggressive driving, speeding, motorcycles, young drivers, etc. Certain geographic units (e.g., county, township, and city) were identified as the possible scope for implementing area-wide screening.
- Develop a user interface for safety screening and implementing the tool for screening as well as training INDOT personnel to use the tool.

1.4. Organization of the report

Chapter 1 introduces the study with a summary of past research and the specific objectives and scope of the study. Chapter 2 discusses the research approach and the safety screening tool concept. Chapter 3 discusses the efforts of creating and updating the master database as a part of data management. Chapter 4 discusses the components of the screening tool, including the selection criteria, statistical concepts, and methodology used in ranking road elements. Chapter 5 describes the clustering module and an example of special studies (five-percent report). Chapter 6 provides basic information regarding results presentation using the standard features of ArcGIS. Chapter 7 summarizes the results.

Appendix A describes the data structure and definitions that are included in the database utilized in the road network screening and accessed via the developed user interface. A separate Volume 2 provides a user manual for the developed database management process, the developed road network screening tool, and the data display with standard ArcGIS features.

2. RESEARCH APPROACH AND CONCEPT OF THE TOOL

2.1. Research Approach

Safety screening tools have been developed by the Center for Road Safety (CRS) for INDOT as part of past safety research, which have been used for a number of different research purposes. Some notable tools/methods developed and already implemented in Indiana are:

- Road Hazardous Analysis Tool (RoadHAT) for safety analysis of individual road site improvements;
- INPASS/PASS for corridor and regional analysis of the effect of specified safety countermeasures; and
- A GIS-based method for identifying roads with safety needs for use in Five Percent Report.

The first method provides a criterion for evaluating the safety level at individual road facilities. The second and third methods utilize GIS tools (TransCAD and ArcGIS) to assign crashes to roads and to predict crash frequencies with advanced econometric models (NB regression and decision tree). Although the three methods are successfully implemented in Indiana, they have been developed for different purposes.

One of the important research approaches for this project is integrating various components of safety data management and screening tools. Safety data management is a top priority issue for any agency. Without preparing and updating a comprehensive database, it is not possible to facilitate the necessary operations for safety screening. Therefore, the very first step for this project is the integration of the different pieces of datasets.

The research approach for this project can be summarized as follow:

1. Development of the overall concept of the systematic tool for identifying safety needs and predicting safety performance.

2. Identification of the basic data components and data collection from different sources.
3. Preprocessing of data components: adjusting ADT for different years, preprocessing of layers, and locating signalized intersections from crash databases.
4. Development of a master database, which includes the many different operations listed below:

   - bridge integration
   - splitting segments
   - network integration
   - updating segment database
   - updating intersection point database
   - crash assignment
   - calculating exposure variables like demographic and land-use features
   - preparation of final master database for interface

5. Development of specific concepts for screening/ranking and clustering techniques.
6. Development of "screening user interface."
7. Development of post-processing components for clustering.

2.2. Concept of the Tool

The tool for safety screening has been developed with the following screening tasks in mind:

- Identification of high crash locations (segments, intersection, ramps and bridges), corridors, and areas in terms of high crash frequency, crash rate, or proportion of a particular crash type.
- Facilitate program-based screening (e.g., shoulder widening, median improvement etc.)
- Facilitate special programs and projects (e.g., five percent report)

In order to facilitate the versatility of screening tasks and the data maintenance, four components are included in the screening tool:

1. Data management
2. Standard screening
3. Roads clustering and special studies
4. Results presentation

Figure 2-1 shows the components of the screening tool and the functional connections between them. The main components of this tool are described below.

Data Management Component

The data management includes updating the existing data by reaching to sources for new data, reformatting them to meet the standards of the Center for Road Safety (CRS) database (called also master database), integrating these data into tables that meet the master database specifications, and replacing the existing data. These new formatted and integrated data are then post-processed to make them ready for use by the network safety screening tool. The data maintenance is facilitated by a suite of procedures developed by the CRS or available in ArcGIS. The data updating may be performed annually or when a major change of data

at any of the data sources occurs to reflect these changes to the screening process.

The data management component facilitates the updating of the GIS and non-GIS data in a convenient and short time. The data management procedures include ArcGIS geo-processing and VBA-implemented and Model Builder codes that are not packaged as a single module but which are used separately as needed to maintain flexibility of the data management process. These procedures are presented in more detail in Chapter 3.

The data management process can be divided into the three subcomponents listed below in the order of their execution:

i. Data pre-processing
ii. Data integration
iii. Data post-processing

Data pre-processing is the preliminary operation on raw data to make it ready to integrate with the existing record. Two important tasks conducted to pre-process the database are:

i. Data acquiring from the sources
ii. Data preparation and formatting

Data integration places the formatted new data in the master database. Some specific tasks include integration of the road inventory and bridge data with the road network layers, linking crashes with road elements (crash assignment), creating intersection points, integrating the bridge and ramp layer, etc.

The road network layer is extracted from the TIGER network and simplified to meet the requirements of the Indiana safety management program. Originally, the TIGER layer served the GIS-based linking of crashes with road segments for the first five percent report. Then, a layer of intersections (also called intersection points) was created by identifying network nodes where at least three road segments meet. Next, the intersection information has been derived from the data associated with the connected segments and the corresponding data saved as intersection data. These two basic layers of road segments and intersections points and other GIS layers with geometric, AADT, and demographic data were integrated to form a database necessary for road network safety screening.

There are three distinct integration cases of the GIS attribute layers with the base road network layers:

i. Integration of *point* features; for example, crash assignment to the road network based on GIS coordinates.
ii. Integration of the *polyline* feature; for example, integration of the INDOT road inventory with the road network or integration of the bridges with the road network.
iii. Integration of the *polygon or regional* features; for example, integration of land-use and census data with the segment and intersection layers.

Data post-processing includes the specialized operations needed to prepare the integrated pieces of data for
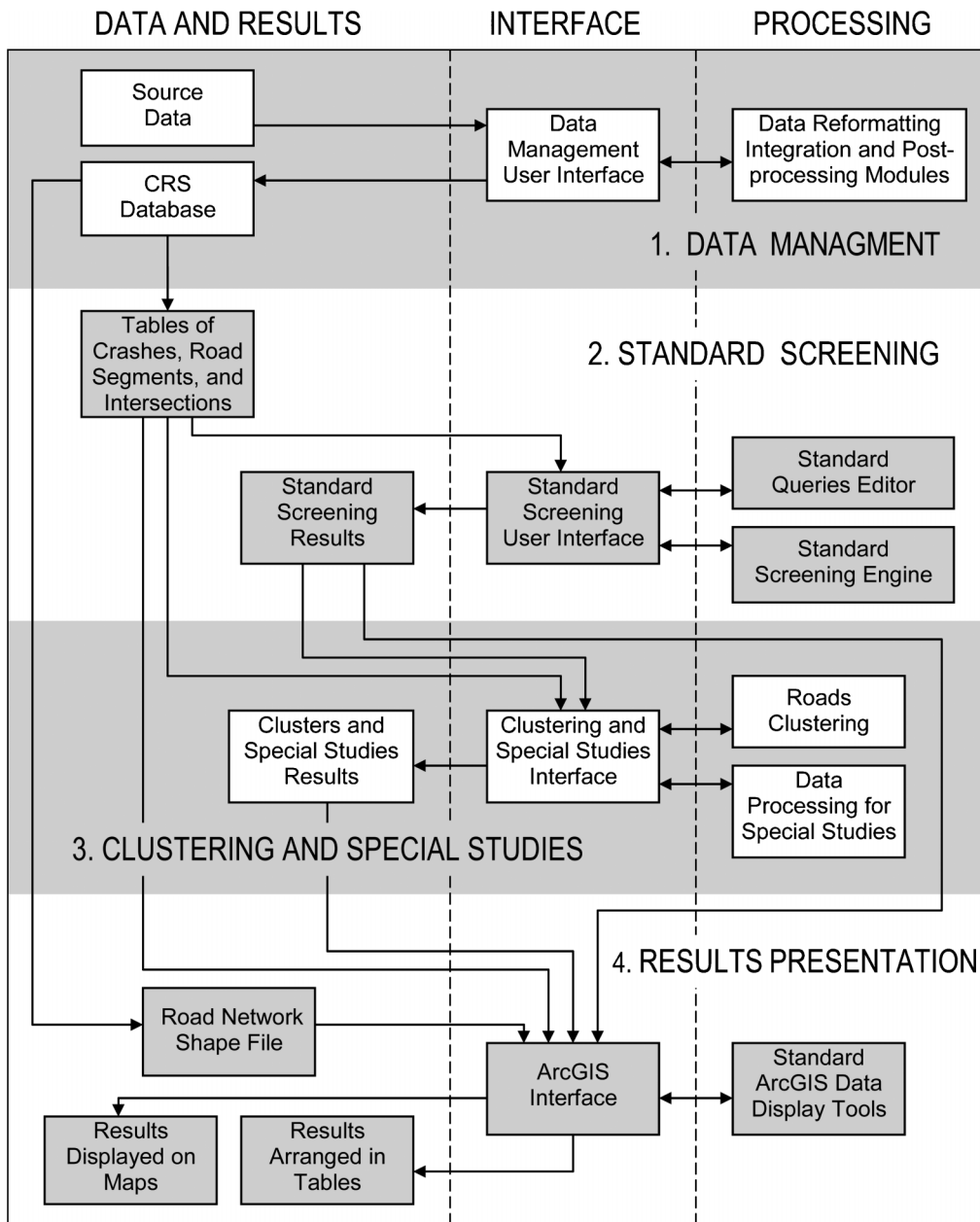
Figure 2-1   Components of the Network Safety Screening Tool – A Concept

the screening tool user interface. There are two distinct objectives for post-processing:

i. Post-processing increases the efficiency of network screening because it reduces the number of repetitive operations.
ii. Road element segments and intersection points are combined to form conjugate network components such as **single intersections** (a single intersection point with connected road segments) and **multiple intersections** (several closely spaced intersection points with adjacent road segments).

Standard Screening Tool

   A standard screening tool facilitates building queries for identifying crashes and road elements that meet certain criteria in order to perform a specific screening task. For example, the user may need a list of rural road segments with narrow shoulders with a considerable number of severe single-vehicle crashes to identify locations where widening shoulders might be justified. Two subcomponents of the standard screening interface are used to execute a standard screening task:

i. Query editor
ii. Screening engine

   Query editor facilitates selection of a set of criteria to be used to select the crashes and road elements. To help the user build a query, the selection criteria have been grouped into four categories: screening scope (entire

state, certain county, city, etc.), screened road elements (road segments, intersections, townships, etc.), crash criteria (single-vehicle crashes, incapacitating injury crashes, etc.), and road criteria (two-lane cross-section, shoulder narrower than 2 feet, etc.). For example, if the screening task is to rank road segments in a county by fatal crashes, the scope is *county*, the element is *road segment*, the crash criterion is *fatality*, and a road criterion is not specified (all segments). The concepts of screening scope, element, and other selection criteria are further described in Chapter 4.

Screening engine is the internal structure of the Standard Screening Tool that executes a certain query based on user-defined criteria. The screening engine receives the query task from the query editor (translated automatically into SQL format) and executes the screening task by using the host program's internal engine. It also transforms the raw query into a user-friendly output and computes specific parameters (e.g., mean and variance of population and confidence value).

Clustering and Special Studies

Road segments and intersections that exhibit an excessive number of crashes may be concentrated along longer road sections. Clustering these elements can reveal large scale safety issues that otherwise might be overlooked if the screening analysis is focus on individual spots. For example, clustering segments with excessive numbers of rear-end crashes may reveal a spill-over safety effect that originated at a signalized intersection with shortage of capacity or where traffic signals are poorly coordinated. Similarly, clustering smaller geographic units, such as townships, with a particular safety problem (e.g., speeding) can help identify larger areas where police enforcement or campaigning might be beneficial.

Clustering state road segments and intersections along state routes can help INDOT identify parts of corridors that require certain road improvements from a safety standpoint. These clusters might be found useful in scoping such projects.

For a special study, such as the five percent report, the user needs to run three queries: one for each severity type. The resultant tables include sufficient information to update the average crash costs, develop statistical models of crash frequency per crash severity, and calculate safety performance measures (confidence level and I index). These calculations are facilitated with a special study tool: the five percent report. Chapter 5 provides more information about the clustering and special studies modules.

Results Presentation

The standard screening tool presents the results of a query in the tabular format convenient for clustering and for additional processing that may be needed in special studies. The final results may also be displayed on GIS maps to visualize the spatial distribution of the identified roads. Such visualization is beneficial in presenting the results to decision-makers and to identify spatial pattern not detectable otherwise. Since the identified road components are geo-coded with the respective latitude and longitude, they can be visualized with the display features offered by Google Earth and ArcGIS. Chapter 6 provides more information.

3. DATA MANAGEMENT
3.1. Data sources

As discussed earlier, the **road network** representation used as a "spatial backbone" of the crash, road, and land use data is a refined network file from the TIGER dataset. The TIGER data are available to download from the U.S. Bureau of Census (http://www.census. gov/cgi-bin/geo/shapefiles2010/main) and from the Indiana GIS map repository (link: http://inmap. indiana.edu/dload_page/infrastructure. html). The data consist of links that represent road segments. The intersection table was added through processing the initial network shape file and associated data. This network is used for five percent report studies and its road segments and intersections are classified in accordance with the five percent report.

Indiana road inventory provides geometric data, including the number and width of lanes, the type and width of shoulders, the type and width of medians, and other pieces of information. The Indiana road inventory forms a separate GIS layer and its integration with the road network layer is necessary.

Average Daily Traffic (AADT) information is available in the Indiana road inventory dataset. The data from 2007, 2008, and 2009 have been integrated with the master database.

Ramp data are present in both the TIGER dataset and the Indiana road inventory. The ramp data found in the Indiana road inventory have been used due to its better quality and particularly due to its higher level of completeness.

Bridge data are extracted from the National Bridge Inventory (NBI) online database (http://www.fhwa.dot. gov/bridge/nbi/ascii.cfm). The dataset is in a tabular format and includes the coordinates for the bridge center points.

Weather data are extracted from National Climatic Data Center (NCDC) website (http://cdo.ncdc. noaa.gov/pls/plclimprod/poemain.accessrouter? datasetabbv=SOD). The weather measurements include daily rainfall, snow, and temperature values. These values need to be aggregated to the corresponding monthly values and then integrated with other data in the master database by associating the network elements (segments and intersections) with weather data from the nearest weather station.

County, township, and city/town boundary shape files are available at the Indiana Map website (http://www. indianamap.org). The Indiana Map is a collaborative effort of the Indiana Geographic Information Officer

(GIO), the Indiana Geographic Information Council (IGIC), the Indiana Department of Transportation (INDOT), the Indiana Geological Survey (IGS), the University Information Technology Services (UITS) of Indiana University, and other federal, state, and local partners. However, these databases required additional information from other sources. Land-use, demographics, and employment data are integrated by spatially joining the Traffic Analysis Zone (TAZ) level information, which is available from the Modeling and Forecasting Section, of INDOT, with the township and county information. This type of integration from smaller to larger aggregation levels help in the integration of the necessary exposure variables for crash screening. For example, the number of registered vehicles was an exposure condition to conduct crash rates based on safety screening. Since this information was available in the TAZ data, it was aggregated to a higher geographic unit, city or township, by spatial join. A list of all important basic data collected for building the master record sets is shown in Table 3-1.

All the sources with GIS coordinates should have a unique coordinate system to maintain the consistency of the spatial relations. In this case, the Universal Transverse Mercator (UTM) North American Datum (NAD) 1983 zone 16 was the coordination system for the network layer and hence all other shape files. Any shape files in different coordinate systems were converted to this particular datum.

3.2. Data preprocessing

Adjusting for AADT by year

AADT information is available in the Indiana road inventory, which must be updated annually. A separate AADT table with a reference to the Indiana road inventory has been created. Individual AADTs for 2007, 2008, and 2009 are available in the master database. Any missing values must be interpolated between the known values. The AADT values are included in the datasets used for screening networks.

In cases when the AADT values are already adjusted in the source data, only the respective growth factor for a specific year and the functional class are multiplied to the AADT. For the unadjusted ADT or traffic count values obtained from MPOs, all three adjustment factors were considered.

Preprocessing of the road network layer

As discussed earlier, the TIGER layer does not contain detailed information about road geometry or traffic information so it was necessary to obtain this information from other sources. Also, it was important to consider generic classifications that are already utilized by INDOT. The five percent project already established a classification for roadway elements and established safety performance functions. Therefore,

TABLE 3-1
Data Sources for Safety Screening Tool

Component Dataset	Contents	Source	Derivative Database
Road Network	Roadway network	TIGER Network refined for 5% report project Can be downloaded from U.S. Bureau of Census http://www.census.gov/cgi-bin/geo/shapefiles2010/main) and also from Indiana GIS map repository (link: http://inmap.indiana.edu/dload_page/infrastructure.html).	Intersection points layer
Indiana Road Inventory	Geometric Information like lanes, median, shoulder; pavement data like roughness or rutting and ADT (adjusted /total counts)	INDOT	Ramp layer
Traffic Counts	Traffic counts (not adjusted)	MPOs or local transportation agencies (collected as part of other CRS projects)	
Bridge Data	Bridge geometry information; condition indices; construction and maintenance data like year of construction, maintenance date, etc.	1. National Bridge Inventory (NBI) online source (http://www.fhwa.dot.gov/bridge/nbi/ascii.cfm)	Bridge GIS layer
Weather Data	Rainfall/snowfall, temperature, heating degree days (HDD), cooling degree days (CDD), etc.	National Climatic Data Center, NCDC online systems: http://cdo.ncdc.noaa.gov/pls/plclimprod/poemain.accessrouter?datasetabbv=SOD	Weather station point layer
Land-use and Demographic Data in Traffic Analysis Zones (TAZ)	Employment by types, Industry profile, demographic profile, and population (from 2000 census)	Modeling and Forecasting Section, INDOT	Aggregated information into city, townships, and county levels
County, Township & City Boundary Layer	Name, geographic extent, and IDs	Indiana Maps (*http://www.indianamap.org/*)	
Crash Data (2007-2009)	Collision, unit, operator information, etc.	Indiana State Police	n/a

the network used for the five percent report was considered as the base network.

In the earlier phase of the project, another layer provided by NAVTEQ, a subsidiary of NOKIA Corporation, was considered to be a plausible layer for this project. However, the network representation in the NAVTEQ layer was very detailed and complex; and the simplifications were not feasible under the current scope of this project. Also, in light of the ongoing five percent project, the current network can be very useful for compatibility reasons. Although the screening tool considered splitting the network into much smaller lengths than did the five percent report project, the geographic positions of the networks remain the same for ease of cross-referencing by spatial joining.

Preprocessing of a network involves certain operations on the network to make it ready for data integration. The following specific tasks are considered for processing:

i. Establishing a feasible boundary for state roads
ii. Establishing a generic classification for screening purpose

Only state roads were considered for development of the safety screening tool. Therefore, it was important to establish a feasible boundary for state roads to maintain network continuity and correct topology for all state segments and intersections. Figure 2 shows the method utilized to build the basic road (state road) network used in the screening tool. Table 3-2 shows the basic roadway segments and intersection types.

Locating signalized intersections from the crash database

Since information about the intersection control (signalized, unsignalized) is not available, codes for the control types included in the crash records must be used to identify which intersections are signalized. The steps to determine the signalization at intersections are explained below:
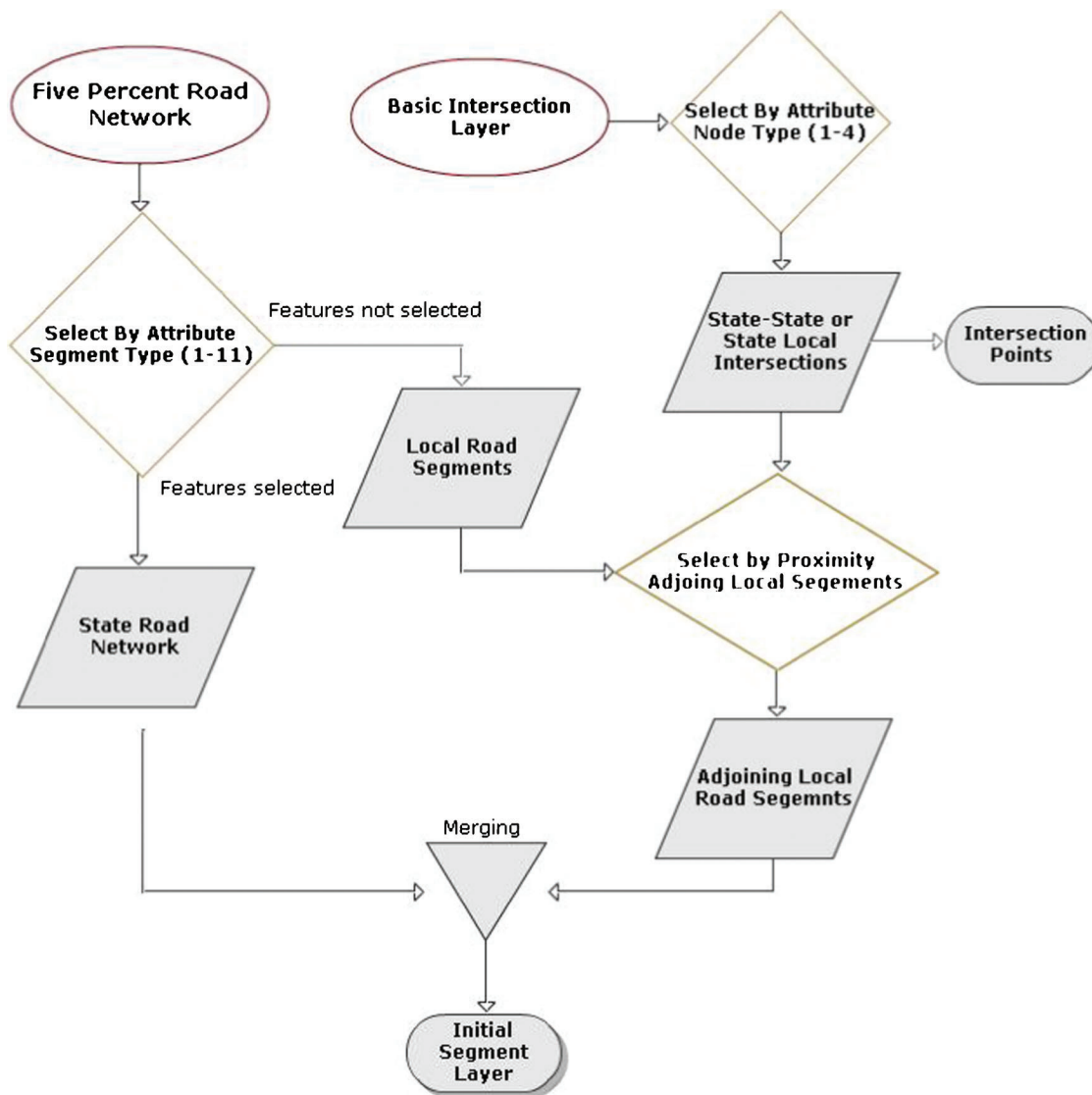

Figure 3-1   Building Basic Road Network (Segments and Intersections)

TABLE 3-2
Basic Segment and Intersection Types

Segment Type	Segment Code	Intersection Type	Intersection Code
Rural two-lane	1	Rural state intersection	1
Rural multilane	2	Rural state-local intersection	2
Rural interstate	3	Urban state intersection	3
Urban multilane	4	Urban state-local intersection	4
Urban two-lane	5		
Urban freeway	6		
Urban one-way	7		
Rural interchange freeway	8		
Rural interchange non-freeway	9		
Urban interchange freeway	10		
Urban interchange non-freeway	11		
Adjacent local road segment	12		

From the crash records linked to the intersections on state roads, rear-end and right-angle crashes within 250 feet from the center of the intersections were of interest. Therefore, the selection criteria were as follows:

- Crash year: 2007-2009
- Crash Type: rear-end (type= '01') and right angle (type = '07')
- Distance to intersection <=250 ft

The crashes reported in the three years are counted at each intersection by the intersection control type entered in the crash reports. Two numbers per each intersection were obtained:

1. S = number of crashes with "signalization control" (Control Type = 05), and
2. U= number of crashes with other type of control (Control Type<>05).

The following decisions were made:

1. $S \geq 2$ and $S \geq U \Rightarrow$ intersection is signalized,
2. $U \geq 2$ and $U > S \Rightarrow$ intersection is not signalized,
3. $S < 2$ and $U < 2 \Rightarrow$ few crashes, intersection is not signalized,
4. $S \geq 2$ and $U \geq 2 \Rightarrow$ inconclusive, intersection control information is contradicting.

Table 3-3 is a summary of the signal coding that was obtained after the above decision rule was implemented:

Results validation

A randomly selected 50 intersections were checked for verification using Google Street View. The results are shown in Table 3-4.

Nearly an entire set of "signalized intersection" was found to be correct. Only one intersection was found to be incorrect (unsignalized). Also, for "unsignalized intersection," all randomly selected intersections were found to be correct.

For local roads, a randomly selected 10 intersections were verified, five of which were signalized and the remaining five were unsignalized. The results were also correct for the whole sample. Figure 3-2 shows the signal control layout for the Lafayette/West Lafayette area as determined from the crash data.

3.3. Preparing the road network

Splitting segments

Splitting long segments into short element segments of uniform length is needed for a more precise identification of locations with safety issues along long segments. Before the splitting process was performed, the network needed to be "un-split" by using the GIS geo-processing tool that avoids combining segments of different names and different types. This step ensured reasonable homogeneity of the combined segments and avoided segment identification problems. Then, the combined segments, most of which start and end at intersection points, were split into shorter element segments using the following rules. Let D be the length of a long segment and x be the desired length of a short element segment. The number of short element segments n is:

TABLE 3-3
Summary of Traffic Control Coding from Crash Database

Intersection Type	Signalized	Unsignalized	Inconclusive	Total
Rural state intersection	95	435	116	646
Rural state-local intersection	300	26,892	1,879	29,071
Urban state intersection	127	81	29	237
Urban state-local intersection	1,211	6,477	1,431	9,119
All	**1,733**	**33,885**	**3,455**	**39,073**

TABLE 3-4
Results of Cross Verification from Google Street View (State Roads)

State/ Local	Signalized			Unsignalized		
	Total	Correct	Incorrect	Total	Correct	Incorrect
State	24	23	1	16	16	0
Local	5	5	0	5	5	0

$$n = \text{round}(D/x) \qquad (1)$$

and the length of the new short segments d is:

$$d = D/n \qquad (2)$$

The *round* operation is the standard rounding rule applied to zero decimal digits after rounding. For example, if the original length is D=892 ft and the desired uniform length is x=528 ft, then n=round (892/528)=round (1.69)=2 and the new length is d=892/2=446 ft.

The only exception from the above rule arose when the original segment length between the two intersections was shorter than 250 ft. In this case, the segment was not split irrespective of the desired uniform split length, x. Intersection points separated by less than 250 ft probably influence each other and were grouped as a multiple intersection. If the length of the segments between two intersections is greater than 250 ft, then Equations 1 and 2 were applied. The segment was split according to the result if n>1; otherwise, the segment had to be split in half.

3.4. Data integration

The road network layer is "an organizing agent," the backbone of the roadway, traffic, and land use data that are spatially related to the road network through GIS coordinates. This section of the chapter explains the data integration process that is needed and then the data update required for road safety screening.

Linking Inventory Data

The data integration algorithm for linking the Indiana inventory data is shown in Figure 3-3. The Indiana road inventory was converted to a route system by using a unique feature ID (i.e. OBjectID). Once the inventory layer was converted to a route system, the element segments from the master database were assigned to the corresponding routes if they were inside the buffer areas of these routes. Once assigned to routes, they were linked with the corresponding inventory segment using the linear referencing toolbox available in ArcGIS. A relatively small number of element segments were located outside the route buffer
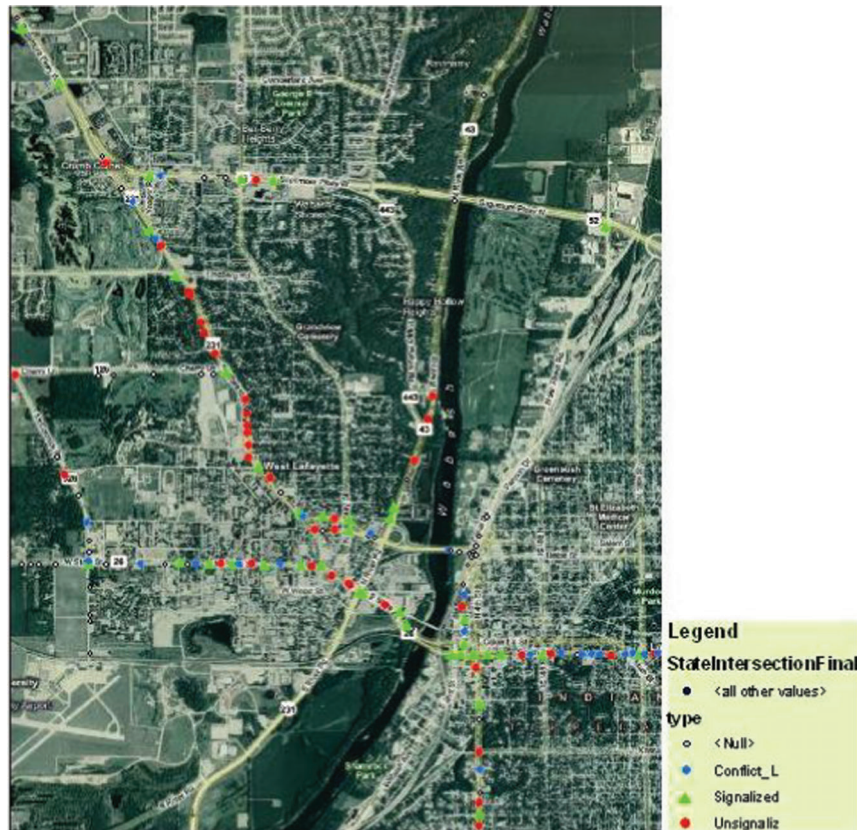

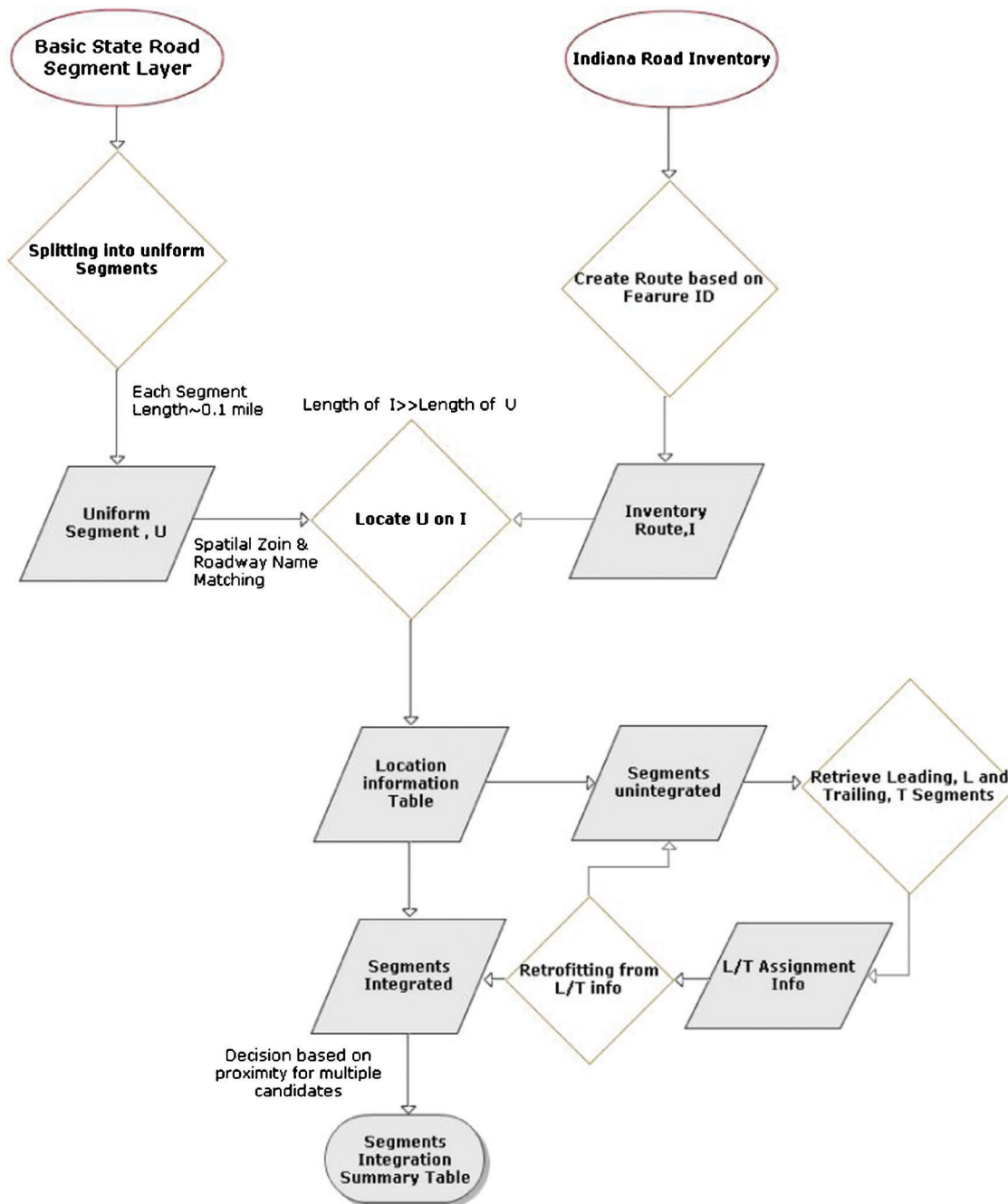Figure 3-2 Lafayette/West Lafayette Signal Control (Example)

Figure 3-3 Network Integration Algorithm

areas, which were left unassigned in the first attempt. They might be assigned in the next attempt if both of the neighboring segments are linked to the same inventory route. This criterion can be relaxed if the neighboring segments are assigned to different routes. In such a case, the element segment can be assigned to the closest route.

In some cases, both neighbor segments may be unassigned. In that case, the process is repeated for the next closest segments along the route, both ways, until a valid assignment is found. With a few iterations, most of the unassigned segments were integrated in this way.

Out of 132,699 state road uniform segments, 132,065 segments were integrated with the Indiana road inventory, which is 99.5% of the links. The adjoining local road segments were not considered for integration purposes because of the unavailability of AADT information.

Linking geographic areas

Each network element was linked to several geographic areas such as the county, township, or city/ town, based on its inclusion inside the geographic area.

All elements were connected to a specific township or county, and only those elements within the geographic jurisdiction boundaries of a city/town were connected to the corresponding city/town.

Updating intersection points layer

The original intersection point table contains segment names and ID information. However, while splitting the segments, ArcGIS assigned new field IDs to segments and hence the information needed to be updated. After integrating the network with the inventory, many attributes related to roadway geometry and traffic counts, such as median, shoulder, or ADT information, also were brought in as attributes of the segments. Therefore, it was necessary to update or bring new information to the intersection point layer. Intersection link information was obtained by using the ArcGIS geo-processing tool "generate near table." A very small buffer radius was used (e.g., <1 ft) for the geo-processing tool so that only the adjoining segments were considered to form the near table. After running this operation, each intersection point ID was referenced to the adjoining segment IDs. Once the near table was generated, each intersection could be associated with the adjoining segments to receive the required information. Major and minor road ADTs were determined based on averaging the link ADT values for a particular roadway. Also, information about medians, shoulder s, etc. was integrated with the intersection. Finally, the signalization information could be obtained from the crash data, as discussed in Section 3.2.2.

The following attributes were updated or added as part of this operation:

- Link Names
- Link IDs
- Major Road ADT
- Minor Road ADT
- Control Type (Signalized or not)
- Number of legs

Bridge integration

The bridge center point was located on the corresponding road based on its latitude and longitude and matching the facility carrier name with the road network. Then the start and end points of the bridge structure were located based on the segment length along the roadway, and a separate bridge layer was created.

In many cases, the presence of a bridge structure on the road does not allow for splitting long segments into element segments of uniform lengths. To avoid this problem, the bridge layer was kept separate and bridges were not included in the road network as an additional type of road segment. Crashes were assigned to road segments, ignoring the presence of bridges. To allow evaluation of safety on bridges, the crash assignment was repeated only for the bridge layer without roads. Figure 3-4 shows the algorithm of integrating the bridge layer with the basic segment layer.

Crash assignment

The overall crash assignment process was performed by considering the crash attributes and the proximity to particular segments and intersections. The crash assignment strategy is discussed below.

Convert crash records to a shape file

The latitude and longitude values of crash locations found in the electronic crash records were imported to ArcGIS. Using the ArcGIS geo-processing tool called "Display XY data," all crashes having valid coordinates were displayed on an ArcGIS map. The point display file was then exported to a shape file. The new shape file was then projected against the segment and intersection shape file. After this operation, crash shape files for three years (2007, 2008. and 2009) containing the collision information were generated.

Crash file aggregation

Three years of crash records in GIS format were aggregated in this operation. Although crash assignment can be performed year by year, this aggregation can eliminate some repetitive works. The ArcGIS geo-processing operation "merge" was used for this operation. For any additional year's crash data the "merge" tool feature can also be used to add it to the previous aggregated datasets. For example, if the 2010 year crash data are available, it should be first converted to a shape file and then aggregated with the combined (2007-2009) datasets by using the "merge" tool.

Getting Subsets of Candidate Crash Records

The aggregated crash records were filtered to generate subsets of crashes by using some roadway element specific selection criteria. For example, if crashes are to be assigned to the intersection points , only crashes between two conflicting movements in an intersection were considered (e.g., right angle, left/right turn, left turn, right turn ,etc.). As discussed earlier, the four basic elements to receive crashes are basic segments, intersection points, ramps, and bridges. The selection criteria for crashes within these four elements are shown in Table 3-5.

Assigning crashes based on proximity

Subsets of crashes were assigned to respective infrastructure elements based on proximity in terms of geo-coordinates. Crash point features were linked to the nearest segments, nodes, bridges, or ramps considering a buffer distance of 250 ft. The entire process was automated by using a set of ArcGIS geo-processing tools components in model builder. Volume 2 of this report, *User Manual,* explains the details of the model builder operation. Figure 6 shows a snapshot of the model builder tool developed for crash assignment.

The detailed descriptive statistics of crashes assigned to different roadway elements are shown is Table 3-6.

The total numbers of crashes on the state road systems were 63,664; 64,095 and 59,006 for the years 2007, 2008, and 2009, respectively. Therefore, the proportions of assigned crashes for these three years are approximately 91.6%, 98.3%, and 97.9%, respectively. Although a certain portion of the assigned crashes also include the adjacent local roads, it can be concluded that the crash assignment method has significantly linked a very high percentage of crashes.

3.5. Calculating exposure variables

The exposure values were calculated at levels of aggregation corresponding to the various elements as defined in the screening process (element segments, intersection points, counties, townships, etc.). For
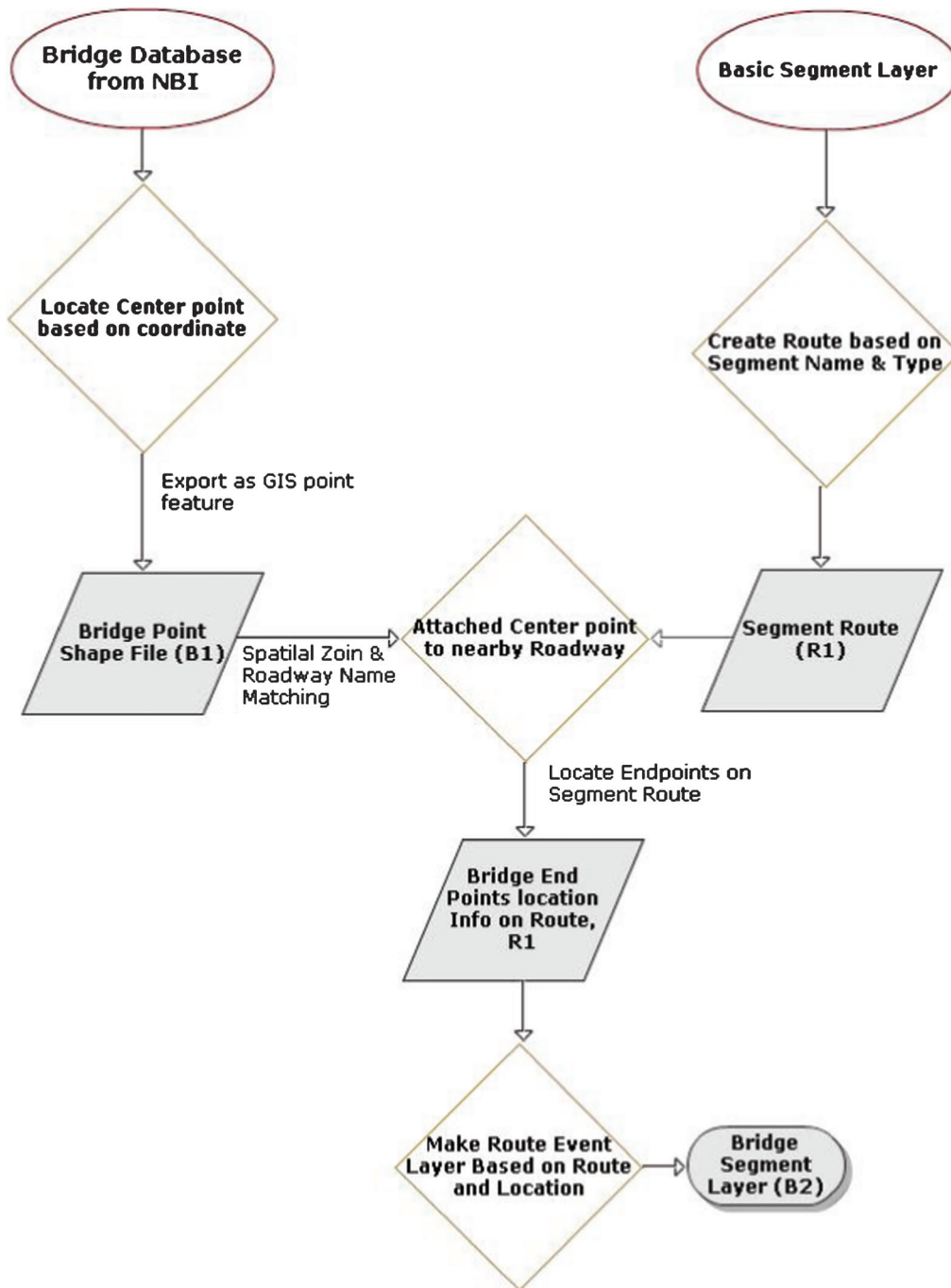

Figure 3-4   Algorithm for Bridge Layer Integration

TABLE 3-5
Element-Specific Selection Criteria for Generating Subsets of Crashes

Element Type	Selection Criteria for sub-setting candidate crashes
Segments	*Roadway Junction:* No Junction Involved
Segments linked to intersections	*Roadway Junction:* Four-Way Intersection, T-Intersection, Y-Intersection, Traffic Circle or Roundabout, Five Point or More *Manner of collision:* Rear End, Head On, Rear to Rear, Same Direction Sideswipe, Opposite Direction Sideswipe, Ran off Road, Backing, Other, Non-Collision
Intersection Points	*Roadway Junction:* Four-Way Intersection, T-Intersection, Y-Intersection, Traffic Circle or Roundabout, Five Point or More *Manner of collision:* Right Angle, Left Turn, Right Turn, Left/Right Turn
Ramps	*Roadway Junction:* Ramps/Interchange
Bridges	*Roadway Junction:* No Junction Involved, Ramps/Interchange *Manner of collision:* Rear End, Head On, Rear to Rear, Same Direction Sideswipe, Opposite Direction Sideswipe, Ran off Road, Backing, Other, Non-Collision

geographic units, the exposure variables are population, vehicle miles traveled (VMT), area, and number of registered vehicles. The VMT information in a network was aggregated to the geographic unit levels. It is important to mention that the calculated VMT was based on the ADT on the state road only. Since the ADT data were adjusted for each crash year, the individual VMT is available in the database. For roadway segments, ramps, and bridges, the exposure variables were ADT, VMT, and length. For intersection points, the exposure variables were the major road ADT and the minor road ADT. It is important to note that the population data available for the geographic unit were based on the 2000 census. Since the recent census data for 2010 are not available yet, a population adjustment for a crash year could not be made. The population exposure information can be updated once the 2010 census data are available.

3.6. Intersection points, single intersections and groups of intersections

An *intersection point* is the common area of two crossing roads. An intersection point and adjacent segments form a *single intersection*. Two (or more) intersection points separated from each other less than 250 ft form together with adjacent segments a *group of intersections*.

A single *state-state intersection* is formed by state-administered roads (US or SR). It includes the intersection area and the 500 feet long road segments adjacent to the intersection area. If another intersection is at the distance shorter than 1000 ft then the adjacent segment is half of the distance. Other adjacent segments remain 500 ft long. If another intersection is at the distance shorter than 250 ft, then the two intersections should be analyzed jointly as a *group of intersections*.

A *single state-local intersection* is formed by a state-administered road (US or SR) and a county or city

road. It includes the intersection area, the 500 feet long state road segments and 250 feet long local road segments adjacent to the intersection area. If another intersection located on the state-administered road is at the distance shorter than 1000 ft then the adjacent state segment is half of the distance. The other adjacent state segment remains 500 ft long. If another intersection on the state-administered road is at the distance shorter than 250 ft, then the two intersections should be analyzed jointly as a group of intersections. A *group of intersections* facility is provided by the RoadHAT to allow this analysis. Close intersections on local roads are ignored and the 250 ft long adjacent local road segment is always assumed.

Intersections along state-administered (major) roads separated from each other by less than 250 feet are grouped together and analyzed jointly. If all roads that cross the major road are local (county or city roads), then such a group is called: *group of state-local intersections* (can be rural or urban). Traffic volume are typically known only for the major road and the expected number of crashes are estimated bade on this single volume and other road inputs such as the length of the major road segment, number of intersections with traffic signals, number of non-major intersection legs, etc.

Intersections along state-administered (major) roads separated from each other by less than 250 feet are grouped together and analyzed jointly. If some of roads that cross the major road are state-administered (US and SR) and other roads are local (county or city), then such a group is called: *group of state-state and state-local intersections* (can be rural or urban). AADTs are typically known only for the state-administered roads and the expected number of crashes is estimated based on two AADT values: the major road AADT and the average AADT on crossing state-administered roads. Other road inputs include the length of the major road segment, number of
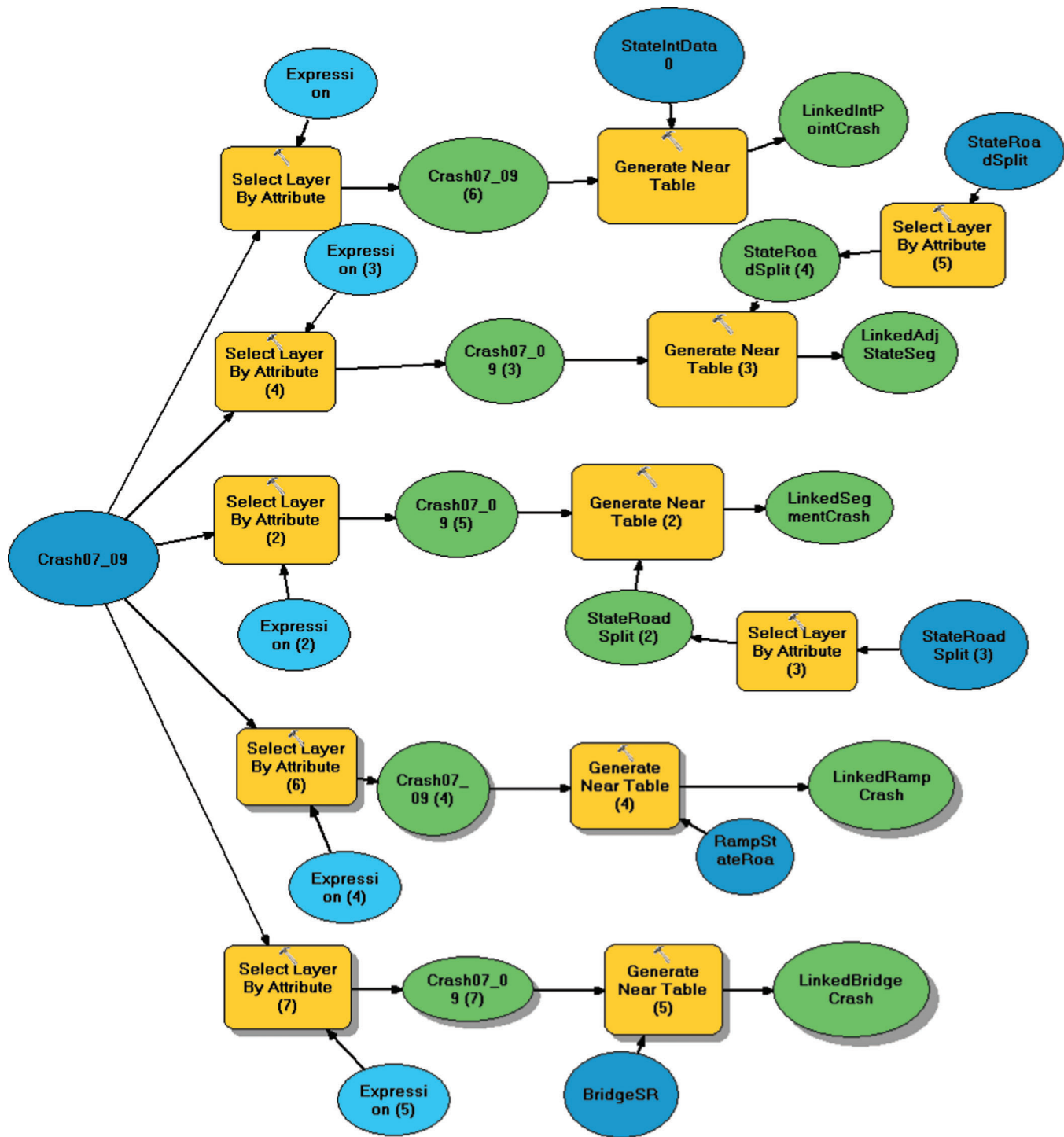
Figure 3-5   ArcGIS Model Builder Application for Crash Assignment

intersections with traffic signals, number of non-major intersection legs, etc.

3.7. Preparing Final Dataset for Interface

After all data components were prepared and integrated, the data for the intersection points, element

segments, and intersections (including intersection clusters) were generated, which is the main input to the standard screening tool. The relational database structure of the master record datasets is shown in Figure 7.

Appendix B shows the definition and code values of variables included in the master database.

TABLE 3-6
Crash Assignment Summary for State Road System

Elements	Year 2007	Year 2008	Year 2009
Segment	47,999	52,270	48,058
Intersection Points	7,184	7,079	6,333
Ramps	1,803	2,103	1,947
Bridge Crashes (common to segments)	450	517	511
Bridge Crashes (not include as segment crash)	1,353	1,586	1,436
Total Assigned Crashes	58,339	63,038	57,774
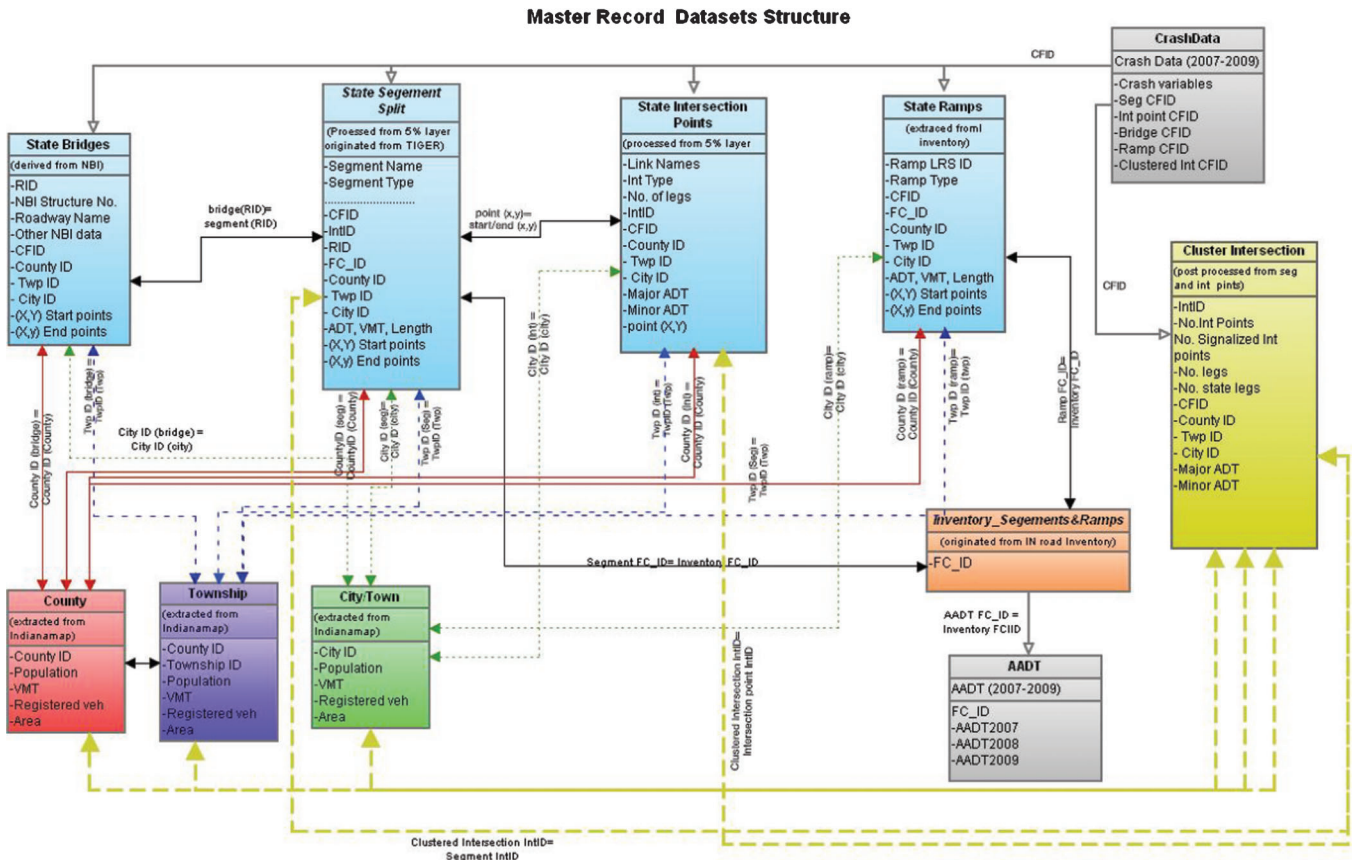
Figure 3-6   Dataset structure for the Standard Screening Tool

4. SCREENING CONCEPTS AND RANKING CRITERIA

4.1. Identification method

The safety identification method includes components designed to fulfill the agency's need to systematically investigate a particular problem. The following items are the core components necessary for a successful identification method:

- Defines scopes, elements, and selection criteria for identification
- Defines identification metric
- Defines exposures
- Statistical evaluation for ranking
- Clustering

4.2. Defining scopes, elements, and selection criteria for identification

Scope

Scope or domain is the geographic unit in which the user is willing to conduct the screening. In the safety screening tool, three scopes have been defined: state, county, and city/town. The scope can be limited to a particular county/township or multiple counties/townships, but should always be greater than the elements in geographic extent.

Element

Element is the smallest unit of aggregation level that a user wishes to investigate. Elements can be the facility

type (e.g., segments, intersection points, intersection, ramps, or bridges) or can be a smaller geographic unit within the scope. Therefore, the scope or domain is the group of elements an agency wishes to investigate.

Selection Criteria

After defining the scopes and elements, it is important to define the selection criteria. Selection criteria basically facilitate obtaining a subset of the elements within the scope. Within the conceptual framework of safety screening, which was discussed in Chapter 2, the "Screen Rule Editor" is used to define the selection criteria.

Selection criteria can be of two types:

- Crash Selection Criteria
- Element Selection Criteria

Crash Selection Criteria

Crash selection criteria are considered in order to investigate a specific type of crash. For example, an agency might be interested in only fatal or incapacitating injury types of crashes or only nighttime crashes. An example might be obtaining only alcohol-related crash locations for targeted enforcement purposes. The crash selection criteria are mainly dependent on the crash variables and their availability.

Element Selection Criteria

Element selection criteria also have a very specific purpose. Since the Indiana road inventory is embedded in the master record sets, a user might be interested in the crash propensity for a specific design condition (e.g., a particular roadway with a specific median type/width). Combining the crash and element selection criteria can serve as a great tool for choosing candidates for a specific program. Figure 8 shows the interaction among the scope, element, and selection criteria in the overall safety screening process.

4.3. Safety performance measures

After a user defines the scope, element, and selection criteria, it is important to define the unit of identification. The identification unit is analogous to the ***measures of safety*** which can have three basic types:

- *Crash Frequency:* Crash frequency is the crash counts of total crashes or a specific subset of crashes as determined by the user.
- *Crash Rate:* Crash frequency/exposure. Exposure can vary based on the type of elements selected.

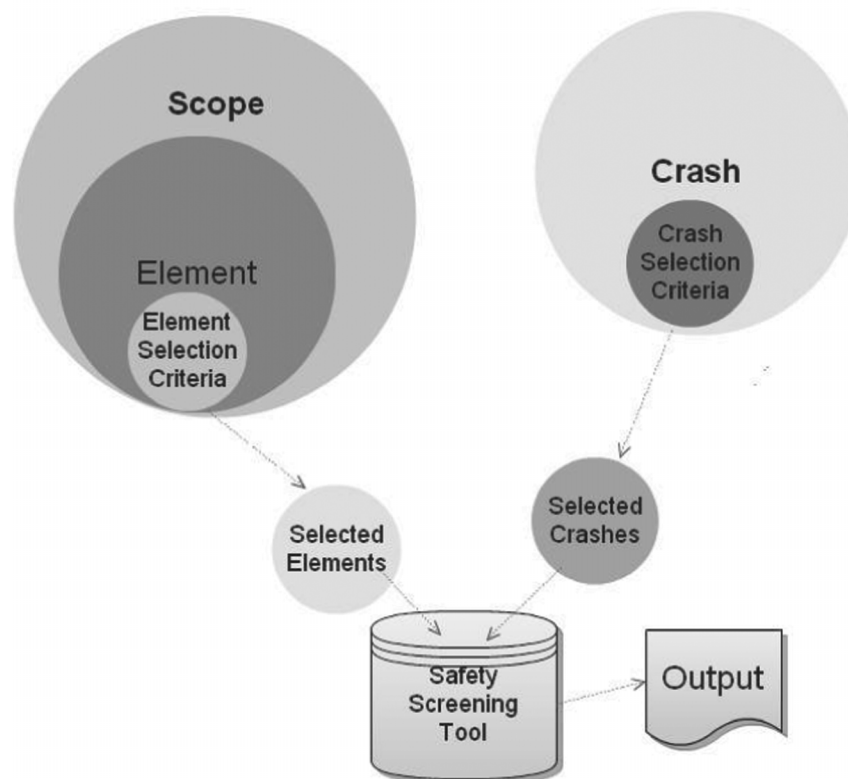Interrelationship of Scope, Element & Selection Criteria in Safety Screening

Figure 4-1   Scope, Element, and Selection Criteria for Safety Screening

- *Proportions of Crash:* Proportion is the ratio of two different crash counts with the criteria, which is: denominator of crash counts >= numerator of crash counts; for example, the proportion of rear-end crashes to the total number of crashes.

4.4 Exposure measures

Exposures are used to estimate crash rates. They can be included as variables in safety performance functions and can be ADT, VMT, population, etc., depending on the element under investigation (see Table 4-1).

4.5 Statistical Evaluation for Ranking

Notation:
t = number of study years,
h = threshold crash rate set by policy,
m = number of crashes expected for the evaluated element under average safety,
v = variance of the m estimate.
Evaluated element (segment, intersection, or area) during t years:
c = number of studied crashes,
r = number of reference crashes,
e = exposure.
All road elements in the query scope:
n = number of elements,
s = total number of studied crashes,
d = total number of reference crashes,
w = exposure.

Concept

Let c be the recorded number of crashes of a certain category used to evaluate a road element's safety during t years. An agency wants to know if this number of crashes indicates that there is a safety problem related to the considered crash category in the considered road element. The agency may have various criteria of evaluation that include the following options:

1. The expected number of crashes in the evaluated road element is higher than a typical number of crashes in the considered scope. The typical number of crashes is estimated with a Safety Performance Function (SPF) or as the average count among all the road elements in the scope.

TABLE 4-1
Exposure Measures for Different Road Elements

Element of Investigation	Exposure
County	Population, VMT, Registered vehicle, Area
City	Population, VMT, Registered vehicle, Area
Township	Population, VMT, Registered vehicle, Area
State Segment	Link volume (ADT, VMT), Length
State-State Intersection	Total approach volume (ADT, VMT)
State-Local Intersection	State (major) Road Volume
Ramp	Link volume (ADT, VMT), Length
Bridge	Link volume (ADT, VMT), Length

2. The expected crash rate in the evaluated road element is higher than the expected crash rate in the scope.
3. The expected crash proportion in the road element is higher than the expected crash proportion in the considered scope. The crash proportion is the ratio of the number of studied crashes and the number of reference crashes. The studied crashes are a subset of the reference crashes.
4. The expected number of crashes on the road element is higher than a certain threshold number set by policy (e.g., five crashes per year).
5. The expected crash rate on the road element is higher than a certain threshold rate set by policy (e.g., five crashes per 100 million VMT).

In the remainder of this document, Criteria 1-5 will be discussed. Criteria 1-3 criteria use the expected values that are not known and can only be estimated. The estimates are random variables with certain distributions, which means that the safety evaluation may only yield the probability that a criterion is met (estimated under certain assumptions). This probability will be called *confidence F*. An agency may decide what confidence level is sufficient to conclude that the evaluated network component needs attention. Another simplified measure is *index I*, which is the difference between the estimated safety and the threshold safety measured with the standard deviation of the difference estimate. This index may be considered more intuitive, but is less accurate than the confidence level.

Comparing the estimates of counts, rates, and proportions for the evaluated component and the population network could be done with the assumption that these estimates are Gamma-distributed. Unfortunately, comparing two Gamma variables is not easy. Instead, an approach based on Negative Binomial distribution is used, which yields the same results. In this approach, the comparison of Gamma estimates of expected counts, rates, and proportions is done through a single Negative Binomial crash count with the mean and variance set in a way that insures valid F and I calculations.

Criterion 1

It is assumed that the number of crashes in consecutive periods and in the same network component follow the Poisson distribution. The expected number of crashes, estimated based on a single period, is a Gamma variable with its mean and variance equal to the crash count in the period c. The two corresponding parameters of the Gamma distribution are $\alpha = 1/c$ and $\beta = c$.

The second expected value is the number of crashes estimated with the SPF or calculated as the average number among all of the road elements included in the considered scope. In the first case, SPF yields u and a, thus $m = u \cdot t$. In the second case, v for the scope of elements has to be calculated first and then, $m = s/n$ and $a = (v-m)/m^2$. All the involved variables are explained at the beginning of the section. The estimate of the typical number of crashes m in t years is Gamma-distributed with parameters $\alpha = 1/a$ and $\beta = am$.

It has been confirmed through extensive simulation that the probability of Gamma variable $G(1/c, c)$ being higher than $G(1/a, a \cdot m)$ is equal to the probability that c is higher than the Negative Gamma variable $NG(1/a, 1/(1+a \cdot m))$:

$$F = S_{x=0..c-1} \text{NegBinomDist}(x, 1/a, 1/(1+a \cdot m)). \quad (3)$$

The following useful equivalence is used to evaluate the cumulative Negative Binomial distribution:

$$S_{x=0..c-1} \text{NegBinomDist}(x, y, z) = \text{BetaDist}(z, y, c), \quad (4)$$

and specifically:

$$F = \text{BetaDist}(1/(1+a \cdot m), 1/a, c). \quad (5)$$

The corresponding calculation of the index I is:

$$I = (c-m)/(c+a \cdot m^2)^{1/2}. \quad (6)$$

The simulated relationship between the index I and the significance F is shown in Figure 4-2. It can be concluded that an I lower than 1.25 indicates weak statistical evidence of a safety problem ($F < 0.90$), an I between 1.25 and 1.5 indicates considerable evidence (F between 0.90 and 0.95), an I between 1.5 and 2 indicates strong evidence (F higher than 0.95), and an I larger than 2 indicates very strong evidence.

Criterion 2

This criterion can be converted to the form of Criterion 1 in the following way:

1. Estimate the rate for the considered scope as s/w.
2. Calculate the threshold number of crashes for the evaluated element as $m = e \cdot s/w$.
3. Calculate the variance of the m estimate assuming that e and w exposures are fixed values and s has the Poisson variance: $v = (e/w)^2 s$.
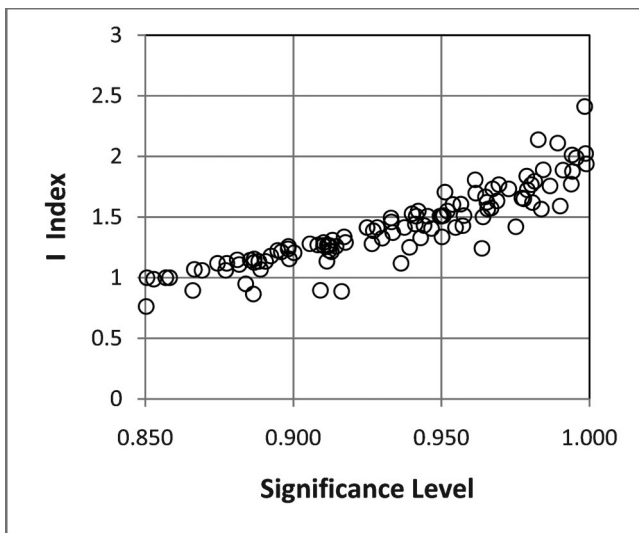
Figure 4-2 Relationship between Index of Frequency I and Significance Level F.

4. Calculate the over-dispersion as $a = v/m^2$, thus $a = 1/s$.
5. Calculate F and I using the same equations as in Criterion 3:

$$F = \text{BetaDist}(1/(1+am), 1/a, c) \text{ and } I = (c-m)/(c+a \cdot m^2)^{1/2}.$$

Criterion 3

This criterion can be converted to the form of Criterion 1 in the following way:

1. Estimate the proportion for the population network as s/d.
2. Calculate the threshold number of crashes for the evaluated network component as $m = r \cdot s/d$.
3. Calculate the variance of the m estimate assuming that r, s, and d have the Poisson variance: $v = (2 \cdot c \cdot r \cdot s \cdot d + r^2 s \cdot d + r \cdot s^2 d - 3r^2 s^2)/d^3$ (see comments about the derivation below).
4. Calculate the over-dispersion as $a = v/m^2$, thus $a = 2c/(r \cdot s) + 1/r + 1/s - 3/d$.
5. Calculate F and I using the same equations as in Criterion 3:

$$F = \text{BetaDist}(1/(1+a \cdot m), 1/a, c) \text{ and } I = (c-m)/(c+a \cdot m^2)^{1/2}.$$

The r, s, and d counts are not independent as explained in Figure 4-3.

The variance of $m = rs/d$ is calculated as the variance of $m = (c+c_1)(c+c_2)/(c+c_1+c_2+c_3)$ with four independent sources of Poisson variance: c, c_1, c_2, and c_3. The variance has been derived from the following equation:

$$\text{var } m(c,c_1,c_2,c_3) = \left(\frac{\partial m}{\partial c}\right)^2 \cdot c + \cdots + \left(\frac{\partial m}{\partial c_3}\right)^2 \cdot c_3 \quad (7)$$

The validity of the derived variance and of the assumption of Gamma distribution applied to this criterion has been evaluated using simulation of 10,000 values of the m estimates for two distinct sets of values of c, r, s, and d. The simulated distribution of the m estimates and corresponding Gamma distributions with the parameters calculated in steps 2, 3, and 4 are shown in Figure 4-4 (for (c=10, s=210, r=210, d=510, m=45.3, v=18.1) and Figure 4-5 for (c=1, s=6, r=3, d=18, m=0.44 , v=1.0). The simulation-based evaluation confirms validity of the method for estimating right-hand distribution tails of m estimates.

This criterion is similar to Criterion 1 with the difference that the threshold u (annual frequency) and corresponding $m = u \cdot t$ are not estimated but are set by policy as deterministic value (a=0). This difference allows calculating the significance F using the Poisson distribution $F = \text{Poisson}(c-1, m, 1)$ and the expression for factor I is $I = (c-m)/c^{1/2}$.

Criterion 5

This criterion is nearly identical to Criterion 4 with the difference of calculating m based on the assumed by policy threshold crash rate h, $m = e \cdot h$ (a=0). The significance F is calculated with the Poisson distribution $F = \text{Poisson}(c-1, m, 1)$ and the expression for factor I is $I = (c-m)/c^{1/2}$.
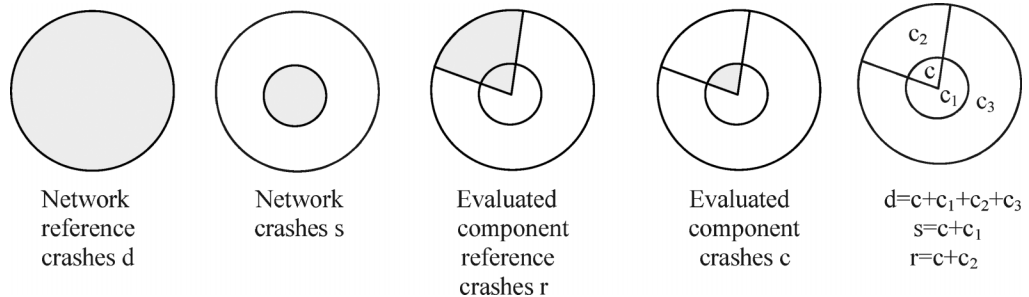
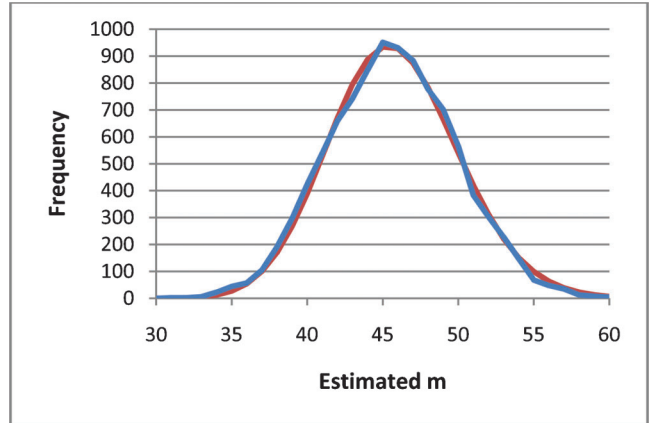Figure 4-3   Dependence between Crash Counts r, s, and d

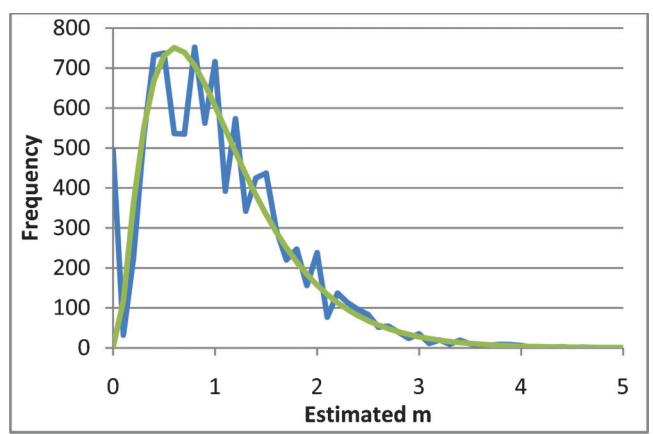Figure 4-4   Case 1 - Simulated (blue) vs. Calculated (red) Distributions of m.

Figure 4-5   Case 2 - Simulated (blue) vs. Calculated (green) Distributions of m.

Computations

Table 4-2 provides equations useful in calculating various safety performance indicators for the five cases.

5. CLUSTERING

The screening tool identifies which road elements experience an excessive number of crashes. Clustering these elements into longer road sections may reveal useful spatial regularities that may be useful to INDOT engineers in scoping corridor improvement projects and other safety-oriented programs.

It is important to note that elements with safety needs should be clustered based on the safety performance measures in order to obtain relevant road clusters from the safety management point of view. The following text describes the statistical basis of clustering and the clustering method itself.

5.1. Statistical Basis

There are three basic safety measures that can be used to identify road elements with excessive numbers of certain categories of crashes: crash frequency, crash rate, and crash proportion. Crashes are subject to a strong random fluctuation over time and two safety performance indices: Confidence F and Index I are proposed to estimate the level of statistical confidence indicated in the detected excessive number of crashes as a systematic issue rather than the effect of random fluctuation.

Significance F is the probability of a safety level equal to or better than the one observed during the period of analysis if the expected safety level in the long run is average for the type of location and under the given exposure. The higher significance F is, the stronger the evidence is that the location experiences a real safety

TABLE 4-2
Calculating m, a, F, and I for Each of the Five Safety Criteria

Safety Measure	Reference	Mean	Over-dispersion	Significance F	Index I
Crash Frequency	Typical annual frequency u	$m=s/n$	$v=\Sigma_i(c_i - m)^2/(n-1)$, $i=1\ldots n$, $a=(v-m)/m^2$, $a\geq0$.	BetaDist$(1/(1+am), 1/a, c)$	$(c-m)/(c+am^2)^{1/2}$
Crash Rate	Population rate	$m=es/w$	$a=1/s$	BetaDist$(1/(1+am), 1/a, c)$	$(c-m)/(c+am^2)^{1/2}$
Crash Proportion	Population proportion	$m=rs/d$	$a=2c/(rs)+1/r+1/s-3/d$	BetaDist$(1/(1+am), 1/a, c)$	$(c-m)/(c+am^2)^{1/2}$
Crash Frequency	Annual frequency u by policy	$m=s/n$	$a=0$	Poisson$(c-1,m,1)$	$(c-m)/c^{1/2}$
Crash Rate	Rate h by policy	$m=eh$	$a=0$	Poisson$(c-1,m,1)$	$(c-m)/c^{1/2}$

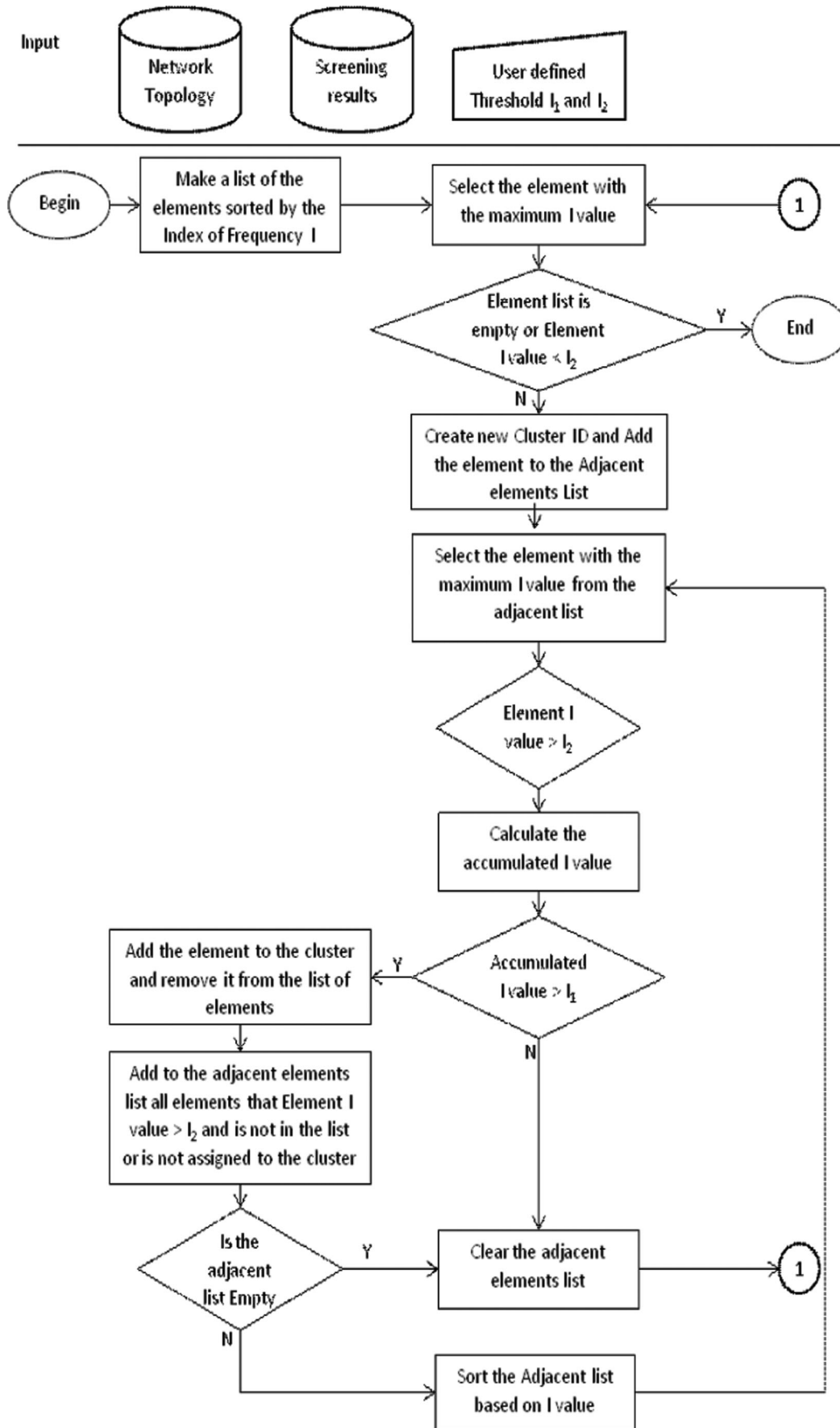Figure 5-1   Clustering Algorithm Flowchart

problem. The values of F=0.90 and higher are typically used.

Index I is the difference between the safety observed during the period of analysis and the safety expected given the location type and exposure divided by the standard deviation of the difference estimate. It is a simplified measure of Significance F. Values of I=1.5 and higher provide sufficient evidence that the location experiences a real safety problem.

Significance F is calculated as BetaDist(1/(1+am), 1/a, c), while Index I is calculated as: $(c-m)/(c+am^2)^{1/2}$, where: a is the over-dispersion parameter, m is the average crash count in a long run, and c is the actual crash count in the period of analysis. Equations for calculating the values of m and a for different safety measures are shown in Table 4-2.

5.2. Clustering Method

One of the important operations of clustering road elements is evaluation of the safety level in current clusters to ensure that the obtained clusters experience excessive number of crashes. A practical method of updating safety evaluations in clusters is aggregation of the safety measures of individual network elements included in the cluster. The exact method based on Significance F is statistically and computationally troublesome because summing two Gamma variables does not yield a Gamma variable, and

the convenient equivalency between Negative Binomial and Beta distribution cannot be used. Therefore, Index I, which is easy to update for clusters, is calculated instead. The following equation is used to calculate the Index I for clusters of multiple road elements:

$$I = \frac{\sum_i (c-m)_i}{\sqrt{\sum_i (c+am^2)_i}} \qquad (8)$$

where values of $(c-m)_i$ and $(c+am^2)_i$ are known for any road element i. The clustering algorithm is shown in Figure 5-1. It is important to note that the clustering process is controlled by two user-selected threshold values: I_1 and I_2. The recommended ranges are: (1.25–2) for I_1 and (0–1.25) for I_2 with the recommendation that $I_1 > I_2$. The user can restrict the clusters' building only along the same routes to follow the common practice in scoping road projects. Other restrictions may be added to the algorithm as needed. A list of clusters and their elements is obtained based on the screening results, the network topology, and the parameters set by the user.

6. RESULTS PRESENTATION

Result presentation is the final step of the screening process. In this phase, the user can visualize the results obtained in the standard screening or clusters/special studies. ArcGIS provides many visualization tools that
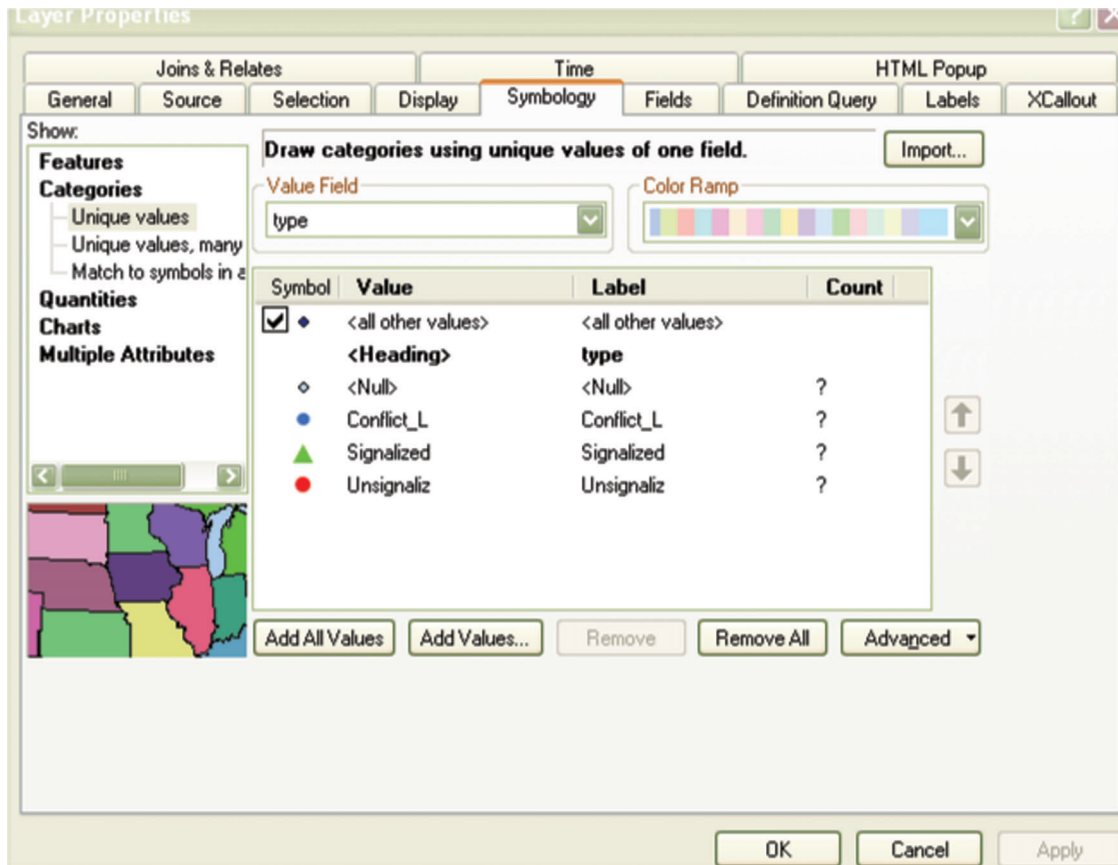

Figure 6-1   Symbology Window in ArcGIS

Figure 6-2   Labeling Feature (left- no labeling, right-labeling)
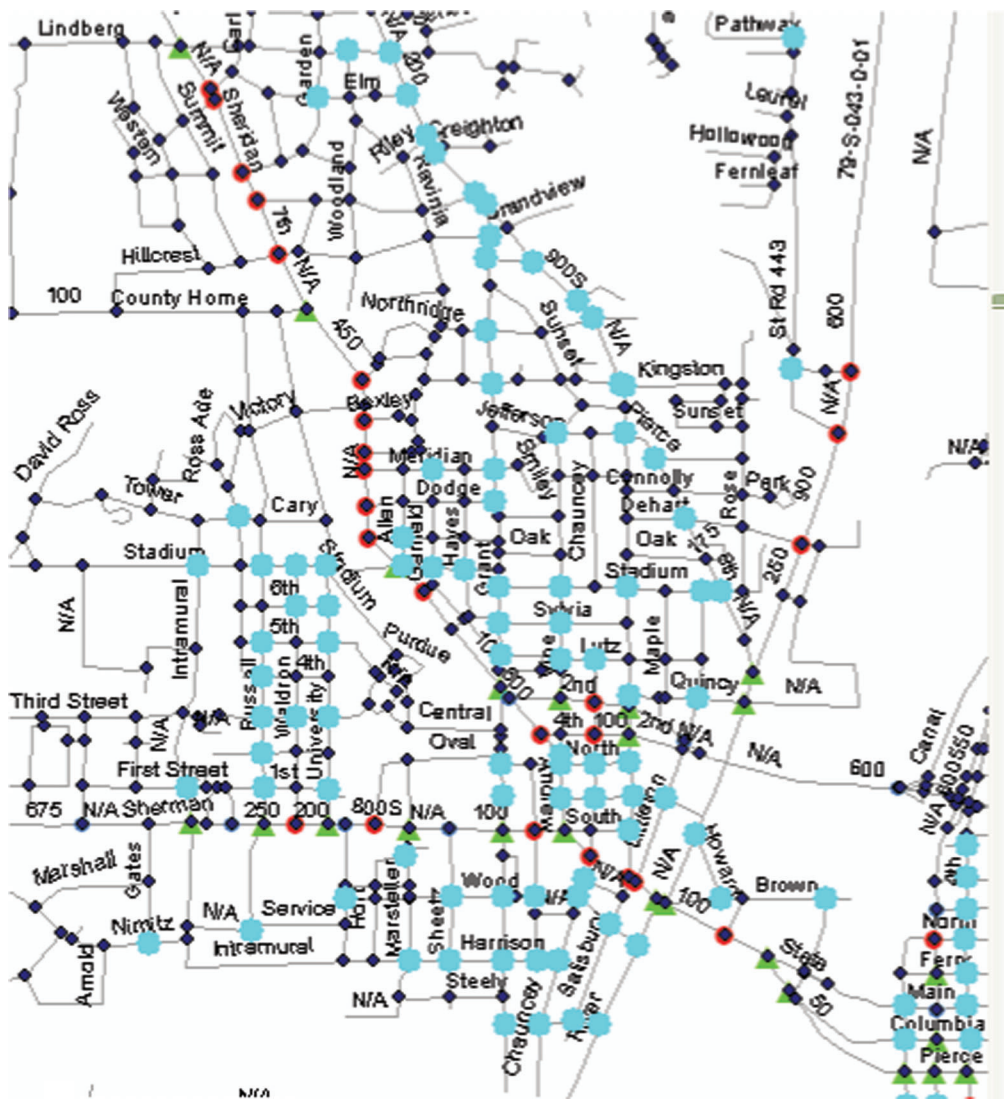

Figure 6-3   Selection by Attribute

can accomplish this job. Among various features available in ArcGIS , the following three visualization tools are widely used:

- Symbology
- Labels
- Selection by Attributes

Symbology refers to visualization of a feature (i.e. a single element), categories of elements, quantities, etc., by colors or symbols. Symbology has a special procedure for preparing charts like bar charts or pie charts to be shown as part of the individual element. Figure 6-1 shows a sample Symbology selection window.

Label is useful in displaying a name or a value of a particular attribute on a map; for example, individual roads on a map can be labeled with their names to enhance visualization or to help identify a specific feature on the map. Figure 6-2 shows a network layer with and without labels.

Selection by attribute can highlight particular elements of interest on a map. For example, intersections having more than 10 crashes per year can be easily selected and marked. Figure 6-3 shows an example of visualization made by "select by attribute tool" in which the highlighted local roads were found to have signalization. The user can easily scan through the map once the features are selected.

Details about results display and visualization are not discussed in the User Manual.

7. SUMMARY

- The conceptual framework of the safety screening tool was developed in order to cope with the complexity of the data management and safety screening operations.
- Three years (2007-2009) of crash data were assigned to fairly uniform-sized segment of 0.1 mile. Crashes involving intersections between two different roadways close to intersections were assigned to intersection points, whereas the links between intersections receive segment - related crashes. Closely spaced intersections were grouped together as clustered or multipoint intersections. Also, an isolated intersection was created by combining the intersection points and adjoining links within roughly 250 ft.
- As part of SPR 3315, the safety screening tool was developed to facilitate the overall screening for the state road elements using user-defined crash selection criteria. The user needs to specify the scope, element, and selection criteria for a particular screening. Apart from the crash selection criteria, the network selection criteria can facilitate screening for roadway deficiencies. Screening criteria was also implemented on geographic units, such as counties, townships, and cities/towns.
- Clustering roads with the same safety problems facilitated the grouping of roadway elements to attain a sizeable project scope.
- The draft user interface is being delivered to INDOT as part of this report. A workshop for training INDOT personnel will be arranged in the near future in order to get INDOT's feedback.

APPENDIX A
A DATA DEFINITION FOR MASTER DATABASE ATTRIBUTES

TABLE A-1
List of variables in table 'State Small Segments'

Variable Name	Definition
OID/FID	Object ID (unique IDs assigned by ArcGIS)
SEGNAME	Segment Name
SEGTYPE	Type of Segment as per 5% report:
	1-Rural two-lane
	2-Rural multilane
	3-Rural interstate
	4-Urban multilane
	5-Urban two-lane
	6-Urban freeway
	7-Urban one-way
	8-Rural interchange freeway
	9-Rural interchange non-freeway
	10-Urban interchange freeway
	11-Urban interchange non-freeway
	12-Adjacent Local road segment
RID	Route ID (Route established by roadway name and type)
StateLocal	State/local road indicator (1=State, 0=local)
Length	Length (in ft)
Split	Used for data management purpose
Start_Sig	Distance (in ft) to nearest upstream signalized intersection (-1 = too far or no signalized intersection present)
Start_Sig	Distance (in ft) to nearest upstream signalized intersection (-1 = too far or no signalized intersection present)
Start_Int	Distance (in ft) to nearest upstream intersection (-1 = too far or no intersection present)
End_Int	Distance (in ft) to nearest downstream intersection (-1 = too far or no intersection present)
AdjSeg	Indicator if the segment in linked with intersection point (1=linked with intersection point, 0= not linked)
CFID	Crash Feature ID (an indicator to relate with Seg CFID present in crash table)
ID	Original ID present in 5% segment layer
COUNTYID	County Code
TWPID	Township Code
CITYID	City Code
COUNTY	County Name
TOWNSHIP	Township Name
CITY	City/Town Name
ADT07, ADT08, ADT09	Adjusted ADT for 2007, 2008 and 2009
VMT07, VMT08, VMT09	Vehicle Miles Traveled /year calculated from ADT
X_Start	X-coordinate of start point in decimal degree
Y_Start	Y-coordinate of start point in decimal degree
X_End	X-coordinate of end point in decimal degree
Y_End	Y-coordinate of end point in decimal degree
IntID	Intersection ID of the corresponding isolated or clustered intersections (-1 = the segment is not a part of any intersection)

Variable Name	Definition
OID/FID	Object ID (unique IDs assigned by ArcGIS)
ID	Original ID present in 5% layer
H12LINK, H13LINK, H14LINK, H15LINK, H16LINK, H17LINK	Segment name from link #1 to link #6
NSEGS	Number of links connected to the intersection point
INTYPE	Type of intersection points as per 5% report
	1-Rural state intersection
	2-Rural state-local intersection
	3-Urban state intersection
	4-Urban state-local intersection
	5-Local intersection
ControlTyp	Type of signal control obtained from crash data
	-signal, -stop, -unsignalized
CFID	Crash Feature ID (an indicator to relate with Int. Point CFID present in crash table)
COUNTYID	County code
TWPID	Township code
CITYID	City code
COUNTY	County name
TOWNSHIP	Township name
CITY	City/town name
OID, TARGET FID	Used for linking purpose
Majseg	Name of major segment
majADT07, majADT08, majADT09	Adjusted ADT for 2007, 2008 and 2009 for major segment
Minseg	Name of minor segment
minADT07, minADT08, minADT09	Adjusted ADT for 2007, 2008 and 2009 for minor segment (0-no ADT available)
X_point	X-coordinate of the intersection point in decimal degree
Y_point	Y-coordinate of the intersection point in decimal degree
IntID	Intersection ID of the corresponding isolated or clustered intersections (-1 = the intersection is not a part of any intersection)

TABLE A-3
List of variables in 'Intersection (Isolated or Clustered)'

Variable Name	Definition
IntID	Intersection ID (unique for an isolated or clustered intersection)
NumIntPoints	Number of Intersection Points
NumSigIntPoints	Number of signalized intersection points
NumNonSignalIntPoints	Number of non-signalized Intersection points
majADT07, majADT08 majADT09	Adjusted major ADT in 2007, 2008, 2009
minADT07, minADT08, minADT09	Adjusted minor ADT in 2007, 2008, 2009
Urban	Urban Indicator (1= urban, 0=rural)
COUNTYID	County ID
TWPID	Township ID
CITYID	City/town ID
COUNTY	County name
TOWNSHIP	Township name
CITY	City/town name
NumStStInter	Number of state-State Intersections
NumStLocInter	Number of state-local intersections
TotalLegs	Total number of legs
NumStateLegs	Total number of state legs
NumLocalLegs	Tota number of local legs
NumLegsWithMed	Number of legs with median
NumLegsNOMed	Number of legs without median

Variable Name	Definition
OID/FID	Object ID (unique IDs assigned by ArcGIS)
LRS_ID	Linear Reference System ID present in Indiana Road Inventory Database
TYPE1	Type of Ramp as per 5% report:
	1-Ramp
	2-Loop
	3-Other
P	Ramp Letter
	Code-Northbound Road-Eastbound Road
	A-SE Ramp -SW Ramp
	B-NE Ramp-SE Ramp
	C-NW Ramp-NE Ramp
	D-SW Ramp-NW Ramp
	Code-Northbound Road-Eastbound Road
	E-SE Loop-SW Loop
	F-NE Loop-SE Loop
	G-NW Loop-NE Loop
	H-SW Loop-NW Loop
	Code-Northbound Road/Eastbound Road
	J-Connector from A Ramp
	K-Connector from B Ramp
	L-Connector from C Ramp
	M-Connector from D Ramp
	P-Loop Connector from E to F
	R-Loop Connector from G to H
	S-Loop Connector from H to E
	T-Loop Connector from F to G
	Other non-standard connectors N, Q, and U through Z
Length	Length (in feet)
FC_ID	Linker to Indiana Road Inventory
RampID	Interchange numbers are assigned according to the milepost of priority roads
CFID	Crash Feature ID (an indicator to relate with Seg CFID present in crash table)
COUNTYID	County code
TWPID	Township code
CITYID	City code
COUNTY	County name
TOWNSHIP	Township name
CITY	City/town name
ADT07, ADT08, ADT09	Adjusted ADT for 2007, 2008, and 2009
VMT07, VMT08, VMT09	Vehicle miles traveled /year calculated from ADT
X_Start	X-coordinate of start point in decimal degree
Y_Start	Y-coordinate of start point in decimal degree
X_End	X-coordinate of end point in decimal degree
Y_End	Y-coordinate of end point in decimal degree

TABLE A-5
List of variables in table 'State Bridges'[1]

Variable Name	Definition
OID/FID	Object ID (unique IDs assigned by ArcGIS)
RID	Route ID for referencing to Segments (Route established by roadway name and type)
TYPE1	Type of Ramp as per five percent report:
	1-Ramp
	2-Loop
	3-Other
StructureN	NBI structure number
InventoryR	The inventory route (The inventory route is a 9-digit code composed of 5 segments)
FeatsInter	Description of the features intersected by the structure and a critical facility indicator
FaciCarr	The facility being carried by the structure.
Loc	Narrative description of the bridge location (e.g. ,3.5 km south of SR 69
BypassLeng	The bypass/detour length: represents the total additional travel for a vehicle which would result from closing of the bridge
Toll	The toll status of the structure.
	1-toll bridge, 2-on toll road, 3-on free road, 4-on interstate segment under secretarial agreement , 5-toll bridge is a segment under secretarial agreement.
Owner	Actual name of the owner (federal, state, etc.)
FuncClass	Functional Classification of Inventory Route.
	Rural
	01- Principal Arterial - Interstate
	02 -Principal arterial - other
	06 -Minor arterial
	07 -Major collector
	08 -Minor collector
	09 -Local
	Urban
	11 -Principal arterial - interstate
	12 -Principal arterial - other freeways or expressways
	14 -Other principal arterial
	16 -Minor arterial
	17 -Collector
	19 -Local
YrCons	The year of construction of the structure (4 digit)
LanedOn_Un	The number of lanes being carried by the structure
ADT	Average daily traffic
YearADT	Year of ADT
DesignLoad	Live load for which structure was designed
ApprRdwayW	Approach roadway width in meter
BridgeMedi	Bridge Median Code:
	0- No median
	1 -Open median
	2 -Closed median (no barrier)
	3 -Closed median with non-mountable barriers
Skew	Angle between the centerline of a pier and a line normal to the roadway centerline
TrafSafety	Bridge inspection shall include the recording of information on the following traffic safety features so that the evaluation of their adequacy can be made. (0-inspected features does not meet criteria; 1-meets criteria; N-N/A)
	Items of Information:
	A- Bridge railings 1 digit
	B- Transitions 1 digit
	C- Approach guardrail 1 digit
	D- Approach guardrail ends 1 digit
HisSign	Historical Significance (scale 1–5)
NavContrl	Indicate for this item whether or not navigation control (a bridge permit for navigation) is required.
	N-n/a, 0-no navigation control, 1-navoigation control is present
NavHorCle	Navigation horizontal clearance measure in meter

Variable Name	Definition
TypeServic	Type of service under or over bridge (first digit on bridge, second digit under bridge) 1- Highway 2- Railroad 3 -Pedestrian-bicycle 4 Highway-railroad 5 -Highway-pedestrian 6 -Overpass structure at an interchange or second level of a multilevel interchange 7 -Third level (Interchange) 8 -Fourth level (Interchange) 9 -Building or plaza 0 -Other
Strype_Main	Structure type main (material type and design)
NoSpanMain	Number of spans in main unit
InvTotalHo	Inventory road total horizontal clearance (in meter)
LenggOfMaxS	Length of maximum span (in meters)
StrLength	Length of bridge measured(in meters) back to back of back walls of abutments or from paving notch to paving notch
CurbWidth	Curb of sidewalk width
BridgeRdway	Bridge roadway width in meter s(curb to curb)
DeckWidth_	Deck width (out to out) in meter
MinVerClea	A 4-digit number should be coded to represent the clearance to the nearest hundredth of a meter
……	……
FMEAS	Start point on RID (meter)
TMEAS	End point on RID (meter)
RteLen	Length of route (meter)
SEGTYPE	Segment type (as per 5% segment type)
StateLocal	1=State, 0=Local
CFID	Crash feature ID (an indicator to relate with Seg CFID present in Crash Table)
COUNTYID	County Code
TWPID	Township Code
CITYID	City Code
COUNTY	County Name
TOWNSHIP	Township Name
CITY	City/Town Name
X_Start	X-coordinate of start point in decimal degree
Y_Start	Y-coordinate of start point in decimal degree
X_End	X-coordinate of end point in decimal degree
Y_End	Y-coordinate of end point in decimal degree

[1]NBI Data Dictionary contains the complete data definition

TABLE A-6
List of variables in table 'County'

OID/FID	Object ID (unique IDs assigned by ArcGIS)
COUNTY	Name of County
CTYFIPS	County FIPS (Federal Information Processing Standard) Code
STNAME, STATE, STFIPS	State name, state abbreviation, state FIPS
AreaSgMile	Area in square miles
SUM_VMT	Summation of township's VMT aggregated over county
POP2000	Population per Census 2000
SUM_Townsh	Total vehicles: summation of township's vehicles aggregated over County
VMT07, VMT08, VMT09	Vehicle miles traveled /year calculated from ADT

Variable Name	Definition
OID/FID	Object ID (unique IDs assigned by ArcGIS)
COUNTY, Cnty_Name	County code, county name
COUSUBFP, MCD_Name	Township code, township name
SUM_AGG_VE	Total number of vehicles (found by aggregate TAZs over township)
SUM_A_AGFO	Number of agricultural/forestry employment
SUM_B_MINI	Number of mining employment
SUM_C_CONS	Number of construction employment
SUM_D_MANU	Number of manufacturing employment
SUM_E_TRAN	Number of transportation related employment
SUM_F_WHOL	Number of wholesale business employment
SUM_G_RETA	Number of retail business employment
SUM_I_SVCS	Number of. service employment
SUM_J_PUBA	No. public administration employment
SUM_TOT_EMP	Number of total employment
SUM_EDUCATION	Number of education employment
AggHHInc	Total household
Ruralper	Percentage of rural area
UrbanPer	Percentage of urban area
Suburbper	Percentage of suburban area
StationID	Weather Station ID
Station Name	Station
SUM_POP2000	Population per census 2000
SUM_WHITE....SUM_HISPAN	Population by race per census 2000
SUM_MALE, SUM_FEMALE	Population by gender per census 2000
SUM_AGE_UN5 … SUM_AGE_65	Population by age group per census 2000
SUM_HOUSOLD… SUM_RENTER	Number of household s and other household variables
STNAME, STATE, STFIPS	State name, state abbreviation, state FIPS
AreaSgMile	Area in square miles
VMT07, VMT08, VMT09	Vehicle miles traveled /year calculated from ADT

Variable Name	Definition
OID/FID	Object ID (unique IDs assigned by ArcGIS)
PLC	Place Code from U.S. Census Bureau
PLACEFP	County and place FIPS code of place from U.S .Census Bureau
NAME	Name of the city
LSAD	Place type code (57-census designated place, 58-city, 59-city, Indianapolis, 60-town)
LSAD_TRANS	Place type literal meaning
STFID	State, county and place FIPS code
POP2000	The year 2000 population for the palce
WHITE, BLACK…HISPANIC	Population by different race
MALES, FEMALES	Population by gender
AGE_UNDER5 to AGE_65_UP	Population by age
MED_AGE, MED_AGE_M, MED_AGE_F	Median age, median age for male, median age for females
HOUSEHOLDS	Total number of households
AVG_HH_SZ	Average household size
HSEHLD_1_M …RENTER_OCC	Different household and household unit variables
AreaSgMile	Area in sq mile
SUM_PropVeh	Total number of registered vehicle (aggregated over cities/towns from TAZs)
VMT07, VMT08, VMT09	Vehicle miles traveled per year calculated from ADT

DATA MANAGEMENT

INTRODUCTION

Data management in this tool includes updating the existing data by contacting the sources for new data, reformatting them to meet the standards of the CRS database (called also master database), integrating these data into tables that meet the master database specifications, and replacing the existing data. These new formatted and integrated data are then post-processed to make them available for use by the network safety screening tool. The data maintenance is facilitated by a suite of procedures developed by the CRS or available in ArcGIS. The data updating may be performed annually or when a major change of data at any of the data sources occurs to reflect these changes to the screening process. See Figure B-1 for the components associated with the data management user interface.

DATA RETRIEVAL FROM SOURCE DATA

The various components of the source data are:

Basic Road Network
Indiana Road Inventory
Traffic Counts Data
Bridge Data
Weather Data
Land-use and Demographic Data
County, Township, and City Boundary
Crash Data

Following is a brief description of the process to retrieve data from these sources:

Step 1- Obtaining Network Data from Source

There are two designated data sources for the network data TIGER that is used as the base layer for the screening tool. This shape file is public and can be downloaded for free.

Source1:
The first source is the TIGER census website, which is available at the following link of the U.S. Census Bureau. http://www.census.

gov/cgi-bin/geo/shapefiles2010/main. This source provides the most updated TIGER file.

Choose "Road" as the layer type and submit. The next link will appear.

To obtain all roads select "Indiana" in the "All Roads" Category and then submit. User may also select the "Primary and Secondary" category in order to access the state road layer.

For the "all roads" category, the current download scheme is to download the county shapes files one by one. The user needs to use the ArcGIS software utility to merge these files. However, for "Primary and Secondary" or to start the road category the user needs can update the whole shape file for Indiana one time. Once the submit button is pressed, the download option will pop-up and the user needs to save the files in any designated space in user's computer. Once the files are downloaded, they can be added to the ArcGIS map by using the Add Data feature in ArcGIS.

Source 2:
Another source of downloading the TIGER file is from Indianamap (http://www.indianamap.org). The Indiana Map is a collaborative effort of the Indiana Geographic Information Officer (GIO), the Indiana Geographic Information Council (IGIC), the Indiana Department of Transportation (INDOT), the Indiana Geological Survey (IGS), the University Information Technology Services (UITS) of Indiana University, and other federal, state, and local partners.

The following link provides access to download the TIGER file. http://inmap.indiana.edu/dload_page/infrastructure.html

However, the TIGER line shape file available in this source are not as recent as those available from the U.S.Census Bureau.

A snapshot of the TIGER line database for all roads (which was originally used for the five percent report) is shown in Figure B-5.

The left map shows the complete network and the right map shows the state roads only.

The basic TIGER file does not contain detailed geometry information. The important variables present in the TIGER database are listed in Table B-1.

The original TIGER map (which was downloaded from Indianamap) was modified for the five percent report and this
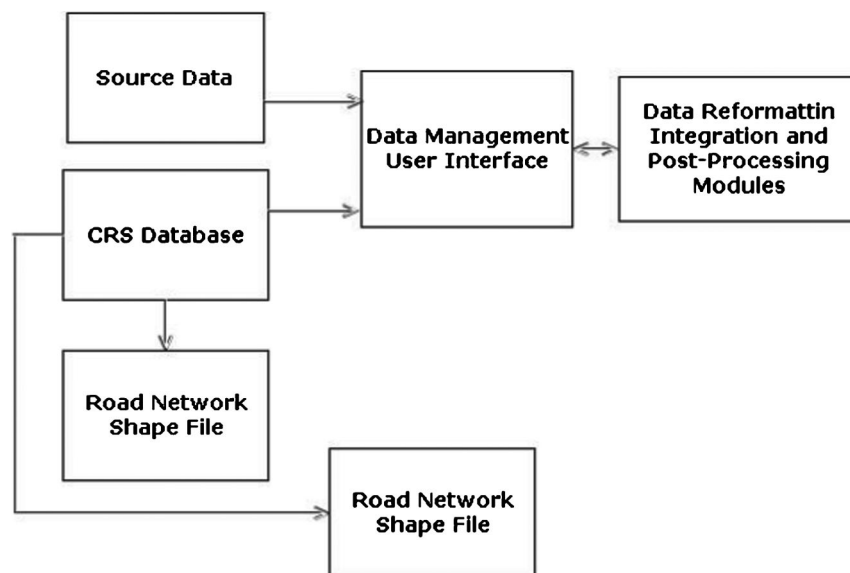
Figure B-1   Data Management User Interface Components
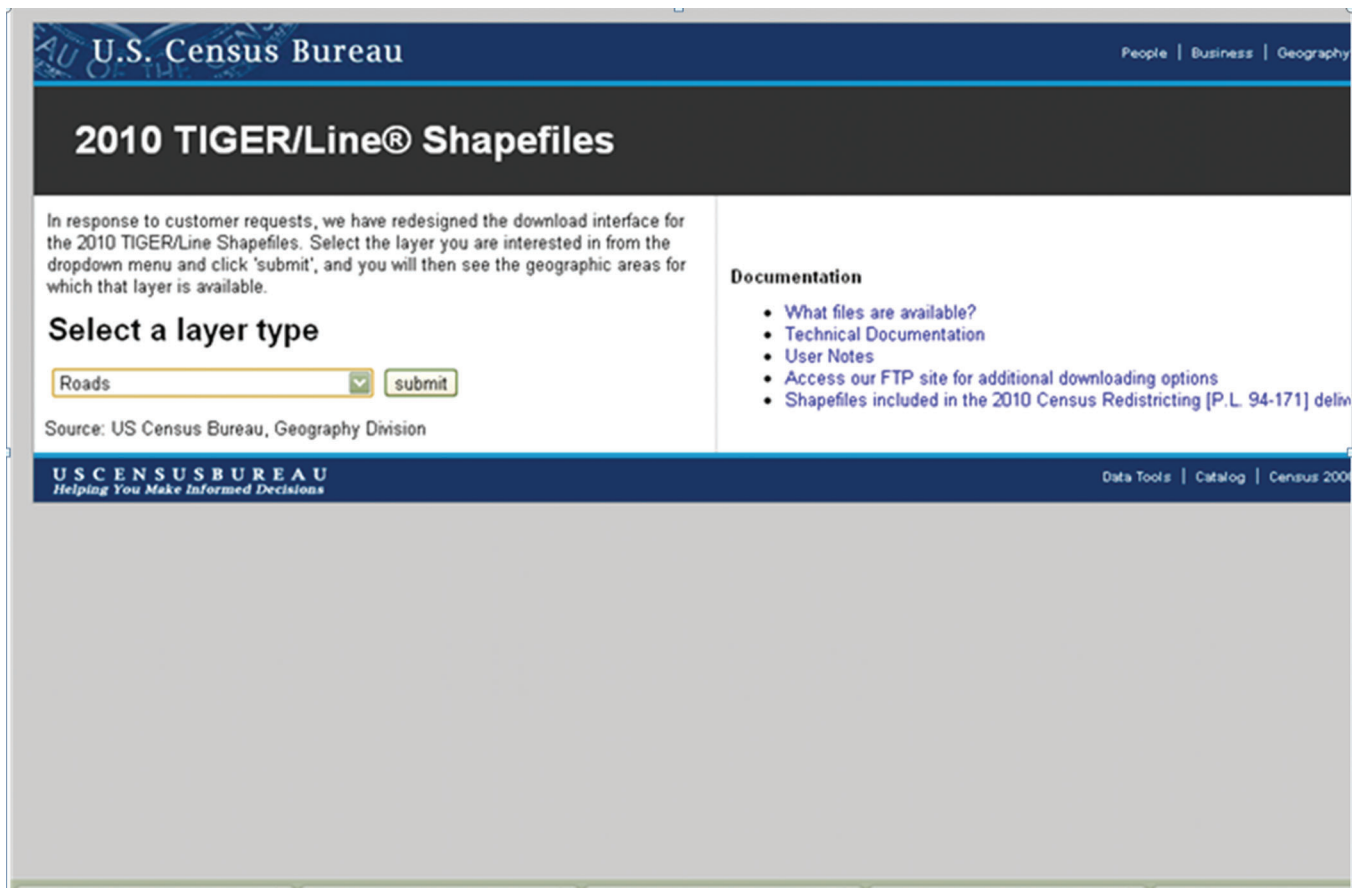
Figure B-2   Downloading TIGER Road Shape file for Census Bureau: Choosing Road Option
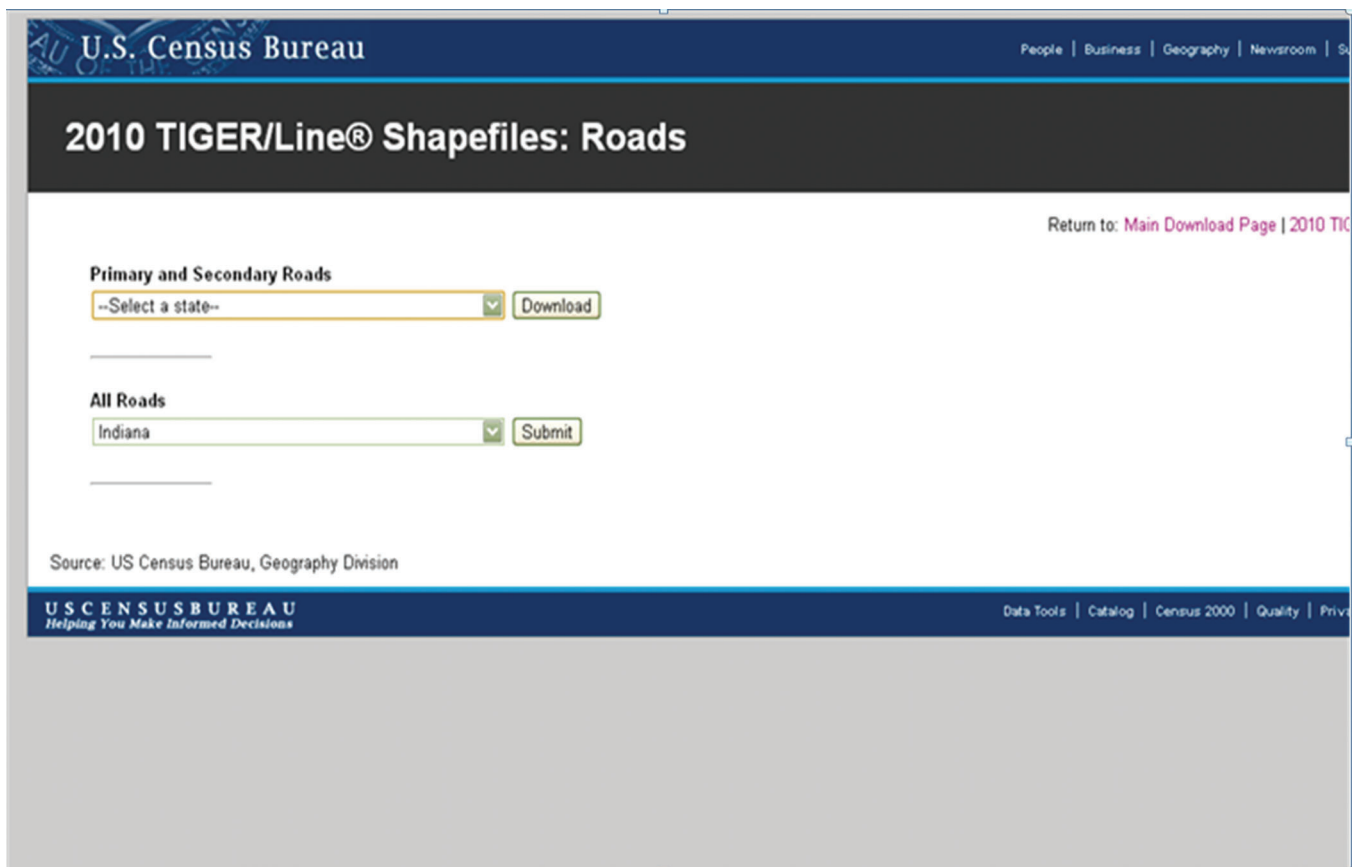
Figure B-3   Downloading TIGER Road Shape file: Selecting State

Figure B-4   Downloading TIGER Road Shape file from Indiana Map
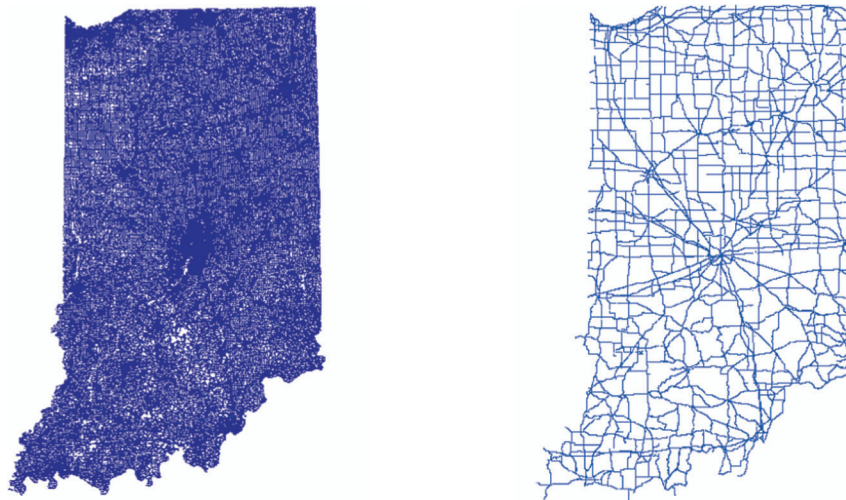
Figure B-5   All Roads (left) and State Roads Only (right) for TIGER Road Shape File

FID_	Object ID
FENAME	Feature Name
FETYPE	Feature Type
FEDIRS	Feature Direction
CFCC	Census Feature Classification Code (road classification used by the U.S. Bureau of the Census)
FRADDL	Start Address (left)
TOADDL	End Address (Left)
FRADDR	Start Address (right)
TOADDR	End Address (right)
ZIPL	Zip Code (left)
ZIPR	Zip Code (right)

modified layer was used for the screening tool. The following modifications were made after the file was downloaded:

Ramp layer was removed from the map to make it simple.

Each double line feature for divided roadways was made a single line. Intersections were created at the intersections of two poly-lines with the exception of bridge nodes. The modified layer then had 673,315 segments. Around 530,031 nodes were created from the layer based on the intersecting segments.

The segments were then classified into 12 segment categories based on the CFCC classification.

It is recommended that the user refer to the five percent research reports submitted to INDOT from 2007 through 2010 to gain a detailed understanding of the transition from the TIGER layer to the more refined five percent layer.

Step 2- Obtaining Road Inventory Data

The Indiana road inventory is available from INDOT. The inventory has already been converted to the shape file and therefore can be visualized in ArcGIS. There is no public domain to download the inventory data. The Indiana road inventory
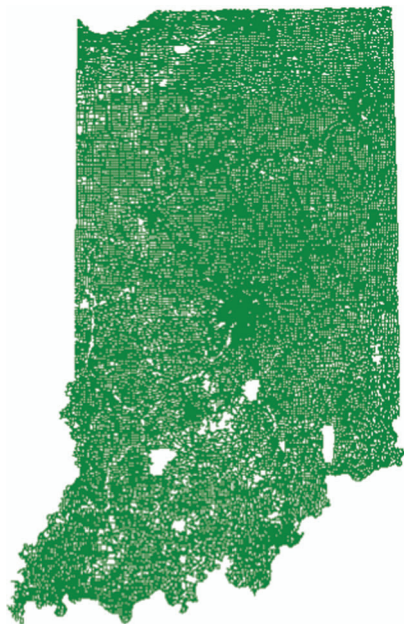
Figure B-6   Indiana Road Inventory

provides a wide range of information for building a comprehensive road datasets and has 177,402 segments, which is longer than the segment length in the TIGER file. Figure B-6 shows a snapshot view of the Indiana road inventory.

Step 3-Obtaining Bridge Data

Bridge data must be extracted from the National Bridge Inventory (NBI) online database. NBI is a collection of information (datasets) covering just under 600,000 bridges in the U.S.that are located on public roads, including interstate highways, U.S. highways, state and county roads, and publicly-accessible bridges on federal lands. NBI provides a state by state summary analysis of the number, location, and general condition of highway bridges within each state.

The NBI ASCII file can be accessed at http://www.fhwa.dot.gov/bridge/nbi/ascii.cfm. Figure B-7 shows that different years of NBI ASCII data are available. For updating purposes, the most recent year data should be chosen. The above link will lead to another page having the list of all states. Choose "Indiana" to proceed to the next step (Figure B-8)

A disclaimer page will pop-up. The user needs to right click the "proceed on link" (Figure B-9) and save the data in csv. format.

Since the final bridge data collection with the process will be in a non-GIS (.csv or .txt) format, this file will later require conversion to a GIS format based on the coordinate (latitude and longitude) information. This process will be done in the data reformatting module and will be discussed in that particular section.

Step 4- Obtaining Weather Data

Weather data need to be extracted from the National Climatic Data Center (NCDC) website. Weather stations collect daily rainfall, snow, and temperature values which are required to be aggregated to monthly values in order to integrate with the master record sets. If the user's domain is gov, mil, edu, us, or k12; free access is available to the NCDC online systems. See the following instructions to download data from NCDC.

Follow the link: http://cdo.ncdc.noaa.gov/pls/plclimprod/poemain.accessrouter?datasetabbv =SOD and continue with the advanced option (Figure B-10).

The user will reach the page shown in Figure B-11. Select state as :Indiana" and "Entire State" and send output to "Web form."

Continue then to the following page where the user will reach the webpage as shown in Figure B-12.

Select the date range based on the user-defined period. Here, we have selected January 2005 to December 2009. Then, continue to the next page and select the check-box "Inventory Review" and enter your e-mail address on the form shown in Figure B-13. Then submit the request.

User will receive FTP links to download data to your designated e-mail address. Depending on size, it may take several minutes to hours to process the request. The links sent to users will be similar to that shown below:

Links 1 and 2 contain similar data. The user needs to use Link 2 if he wants to further process data in an Excel worksheet. Link 3 shows the station information and Link 4 contains the inventory list.

Step 5-Extracting Ramp Data

The ramp network is present in both TIGER and the Indiana road inventory. However, the Indiana road inventory contains more complete ramp data and is thus preferred over TIGER. The ramps therefore were extracted from the Indiana Road inventory for this

NBI ASCII Files

- Recording and Coding Guide for the Structure Inventory and Appraisal of the Nations Bridges (pdf, 0.8 Mb) (doc, 2.
  - Errata Sheet for Coding Guide (03/2004)
- Record Format

- 1992 Data
- 1993 Data
- 1994 Data
- 1995 Data
- 1996 Data
- 1997 Data
- 1998 Data
- 1999 Data
- 2000 Data
- 2001 Data
- 2002 Data
- 2003 Data
- 2004 Data
- 2005 Data
- 2006 Data
- 2007 Data
- 2008 Data

Figure B-7   Options to Download a Particular Year of NBI Data

Download all states as a ZIP file (51 mb).

State	5a=1	5a=2	5a=A to Z
Alabama	16,038	1,149	257
Alaska	1,136	113	0
Arizona	7,643	1,276	56
Arkansas	12,589	513	1,119
California	24,571	6,328	2,733
Colorado	8,522	1,002	27
Connecticut	4,191	952	316
Delaware	861	211	86
District of Columbia	249	73	0
Florida	11,941	2,834	1,133
Georgia	14,672	1,938	507
Hawaii	1,169	48	0
Idaho	4,134	433	14
Illinois	26,353	3,754	294
Indiana	18,552	1,690	177
Iowa	24,731	1,149	128
Kansas	25,542	1,257	50
Kentucky	13,858	1,796	330
Louisiana	13,435	595	265
Maine	2,393	440	1
Maryland	5,208	1,070	120
Massachusetts	5,130	79	536
Michigan	10,928	3,293	9

Figure B-8   Choosing Indiana from NBI Database for a Particular Year

tool. Also, as mentioned previously, the ramp layer from the TIGER file was removed to simplify the network.

Open ArcMap and add Inventory layer. Click "Selection->Select by Feature" and add the following query (i.e. "RAM" <> "). The query will select only those features which have a RampID (Figure B-15).

Upon submitting this query, the ramps will be selected. Export the ramp file as a new layer (e.g., Ramplayer.shp).

The new ramp layer can be added to the ArcGIS map window for further processing. Once added to the map, the ramp layer can be distinguished from the rest of the inventory layer. See Figure B-17 where the ramp layer can be easily distinguished from the inventory layer.

Step 6- Obtaining TAZ, County, Township and City Boundary Data

The Traffic Analysis Zone (TAZ) contains the land-use and demographic information, which can be obtained from different sources. However, for the comprehensive planning data required for travel demand forecasting, the INDOT Modeling and Forecasting Section is a useful source. This layer is not available for download from a public domain.

There is also a free source available at the following link: http://inmap.indiana.edu/dload_page/ infrastructure.html.

In order to obtain the county, township, and city boundary layer, the following steps should be followed:

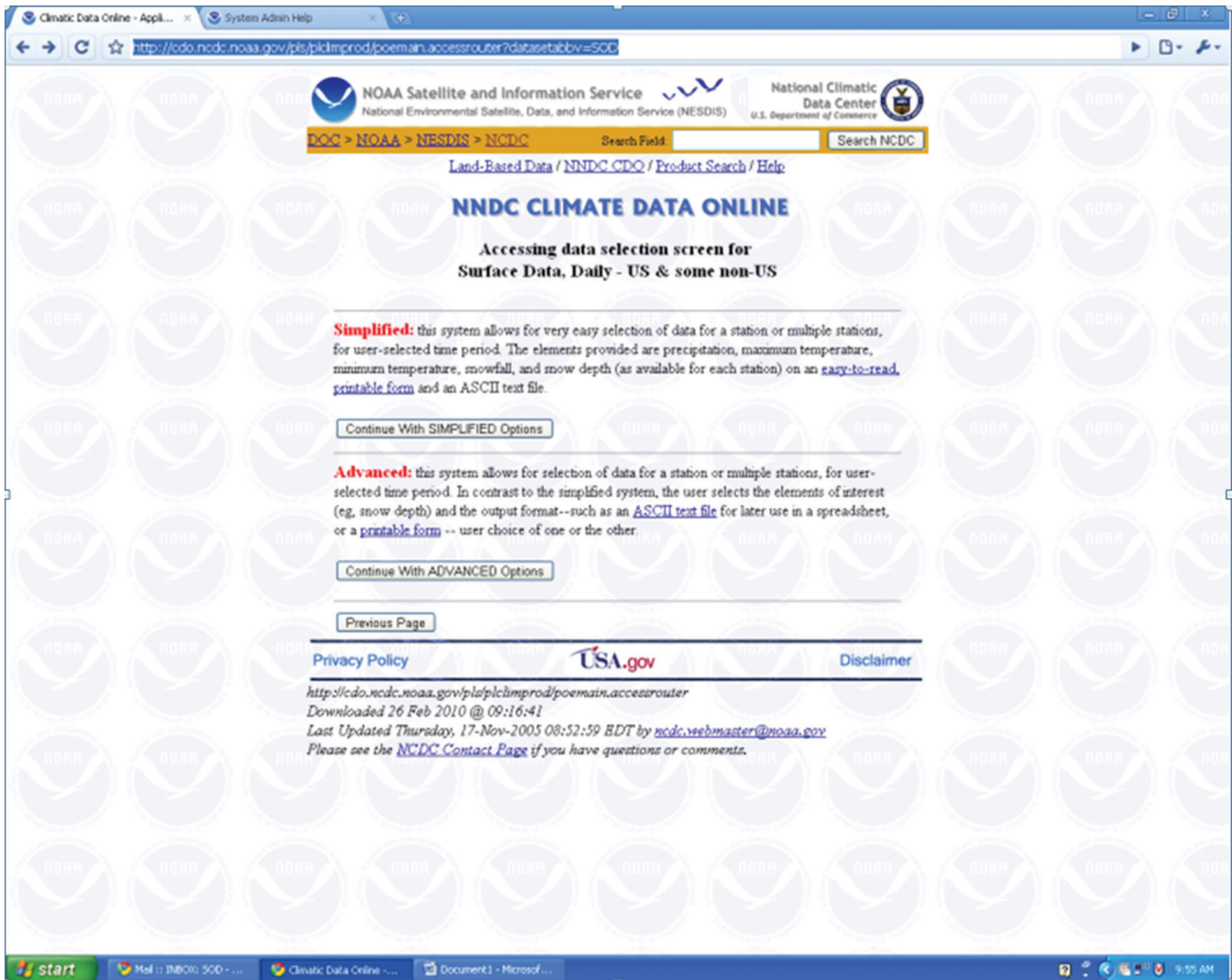Figure B-9   Saving File from Disclaimer Page from NBI Database

Figure B-10   Accessing Data Selection Screen from NCDC

Figure B-11   Selections of Spatial Range and File Type for Data

Go to the following link:
http://inmap.indiana.edu/dload_page/demographics.html

Download the three shape files: a) county, b) populated areas, and c) minor civil divisions (townships)

Once the features are downloaded, add the layers in ArcGIS database in order to visualize and check for correctness. The four layers should resemble the below maps:

All four shape files are polygon data. There are a total 92 counties, 1,011 townships, 601 cities or populated areas, and 4,720 internal (i.e., those within Indiana) and surrounding TAZs in Indiana.

Step 7- Obtaining Crash Data

Crash data is provided by the Indiana State Police.

The ARIES website was developed and is maintained by Open Portal Solutions to serve as a portal to the State of Indiana's repository for the traffic collision and surveillance reports completed by Indiana law enforcement agencies. The online repository link to ARIES is: http://aries.in.gov/Public/Home.aspx

The code values for the variables present in crash data can be explained by using the data dictionary found at http://www.nhtsatsis.net/

crashforms/Pages/state/in/IN_ARIES_2009_CrashDataDictionary_sub3_30_10.pdf.

DATA PREPROCESSING

Step-1 Adjusting for AADT by year

AADT information should be updated for each year. For each additional AADT, add a new column to the new table. Adjustment factors are available in the form of weekday factors, seasonal adjustment factors, and annual growth rate factors by functional class.

A useful link for adjustment factor information is: http://www.in.gov/indot/files/INDOT _2009_AADT_Adjustment_Factors.pdf.

Equations to adjust AADT:

From already adjusted AADT from previous year: AADT (adjusted) = Previous year AADT (adjusted) = Previous year AADT*GrowthFactor

From unadjusted ADT from previous year: AADT (adjusted) = Previous year ADT(unadjusted)*Weekday Factor*SeasonalFactor *GrowthFactor
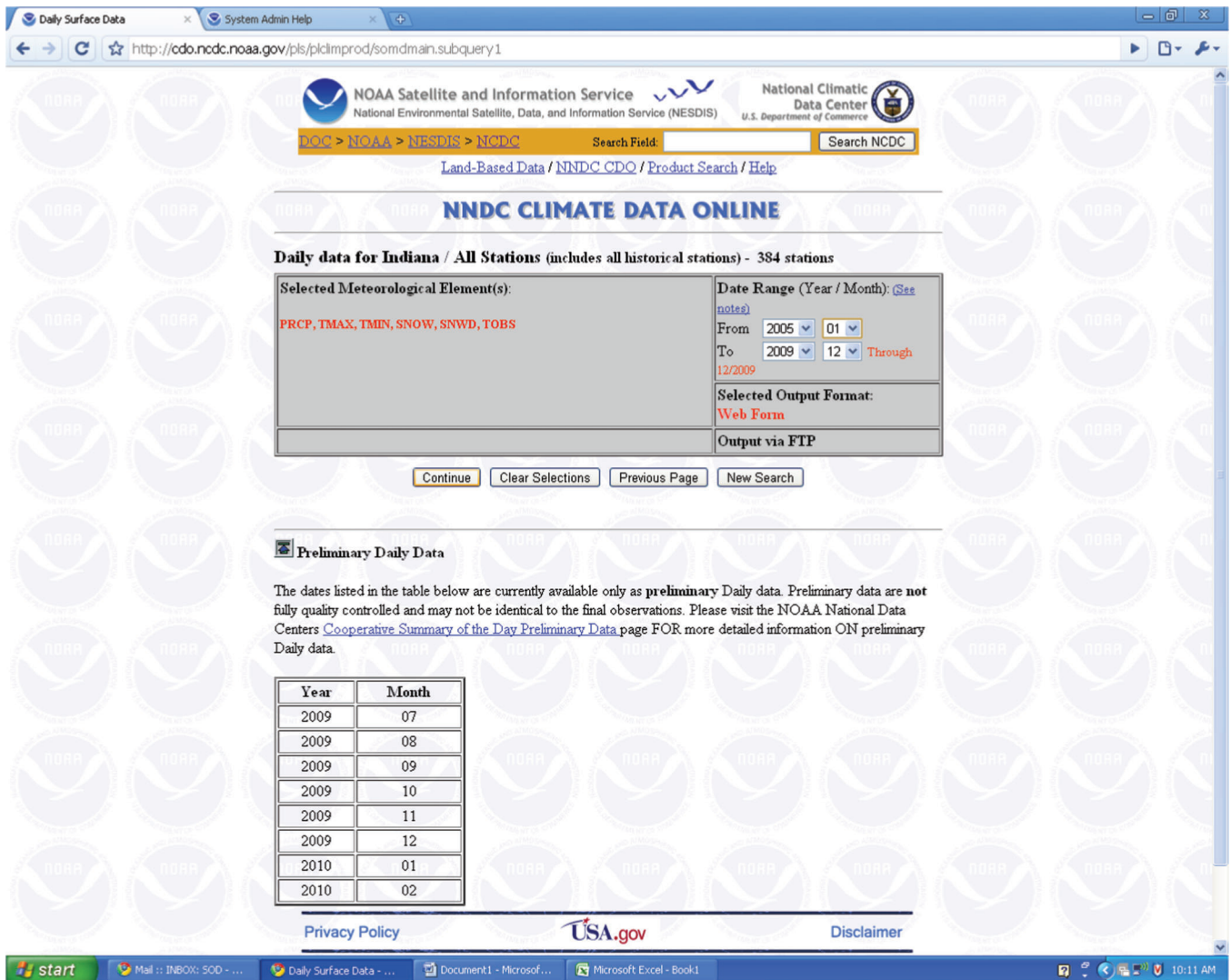
Figure B-12 Selection of Data Range

From traffic counts from previous year: AADT (adjusted) = Previous year Traffic counts *dailyfactor*Weekday Factor*Seasonal Factor *GrowthFactor

BRIDGE LAYER INTEGRATION

In this process, the bridge center point is located on the corresponding road based on its latitude and longitude, matching the facility carrier name with the road network. Then the start and end points of the bridge structure are located based on the segment length along the roadway and a separate bridge layer is created.

Step 1- Making Shape File from Previously Downloaded NBI Datasets (.csv or .txt format)

In order to create a shape file from the downloaded NBI datasets, add the .csv (or .xls) file in ArcGIS map window.

Right click the non-GIS (here .xls) file in the Table of Contents window in ArcGIS and click "display XY data."

Select the latitude value (in decimal degree) as the Y-value and the longitude value (in decimal degrees) as the X-value.

The bridge points will be populated in the map. The new displayed feature is known as "XY Event Source" in ArcGIS terminology, which is a temporary file that needs to be saved as a permanent shape file. Right click the newly created XY Event Source" and export the data. Choose the file type as .shp (shape file). A similar GIS representation will be created which the user can use any time for future purposes.

Step 2- Create Route Based on Segment's Name and Type

In order to create a route based on a segment name and type, first add the segment shape file in ArcGIS and then click the "Create route" geoprocessing tool under the Linear Reference System Tools.

The 'Create Route' window will appear. User needs to specify: the input polyline feature (i.e. the segment layer), route identifier field and, the output route layer. In order to use the road name and type both as a identifier for the route ID, a new field needs to be created. The new field "Segment_Type," is created by concatenating the segment name and type. For example, a specific value for "Segment_Type" can be "StateRd6~7," which stands for the road name and type separated by a character '~'. After creating the route by using the criteria, it should combine small segments up to the locations until when the names and type do not change.
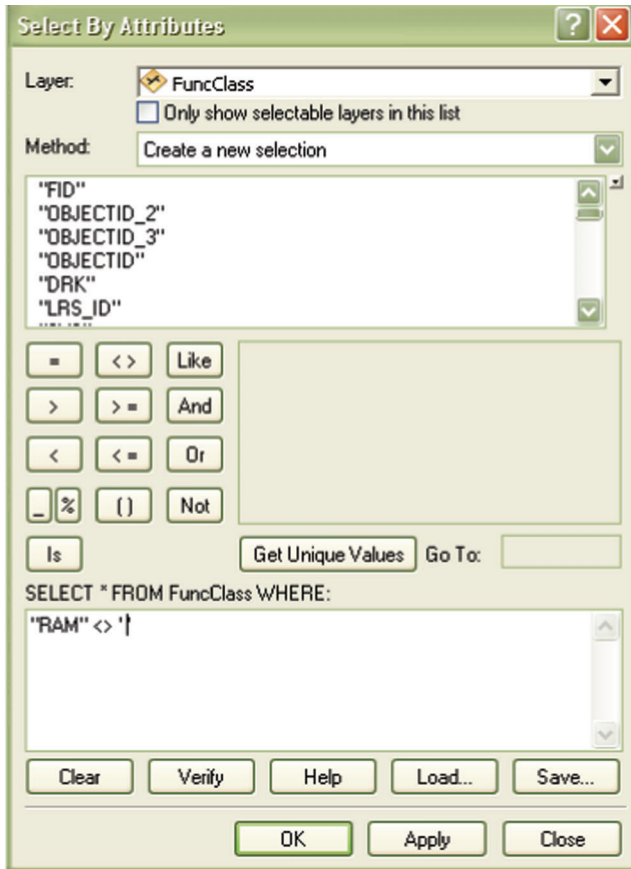
http://cdo.ncdc.noaa.gov/cgi-bin/cdo/cdoprod.pl

NOAA Satellite and Information Service
National Environmental Satellite, Data, and Information Service (NESDIS)

National Climatic Data Center
U.S. Department of Commerce

DOC > NOAA > NESDIS > NCDC   Search Field: [   ]   [Search NCDC]

Land-Based Data / NNDC CDO / Product Search / Help

NNDC CLIMATE DATA ONLINE

SOD - Surface Data, Daily - US & some non-US, Request Summary

Indiana / All Stations - includes 166 stations *(See selected stations below)*

Selected Meteorological Element(s): 'PRCP', 'TMAX', 'TMIN', 'SNOW', 'SNWD', 'TOBS'	Date Range (Year / Month): 2005/01 to 2009/12
	Selected Output Format: Web Form
	Selected Output Media: FTP
	Total Data Rows Available: 40631 - View Inventory
	Output File Size Estimate (bytes): 184320000

☑ Inventory Review: I have reviewed the Inventory File to see if the elements/dates desired are included *before* ordering. Some time periods or elements may be missing. For daily data, there are often missing days and/or missing elements (eg, temperature) in the data received.

IMPORTANT! Please enter a valid email address below so we can notify you when your request has finished processing.

E-mail Address: [xyz@purdue.edu   ]

[Submit Request]

Stations included:

COOPID	WBANID	Station Name	State	County	Clim. Div.	Period of Record
120096	99999	ALEXANDRIA 5 S	IN	MADISON	05	199304 to 200911
120177	99999	ANDERSON SEWAGE PLT	IN	MADISON	05	189505 to 200911
120182	99999	ANDERSON WTR WKS	IN	MADISON	05	191012 to 200911
120200	99999	ANGOLA	IN	STEUBEN	03	189301 to 200911
120331	99999	ATTICA 2E	IN	FOUNTAIN	04	199106 to 200911
120550	99999	BEDFORD 4 SW	IN	LAWRENCE	08	194701 to 200911
120676	99999	BERNE WWTP	IN	ADAMS	03	191001 to 200911

Figure B-13   Data Request Submission

Your SOD - Surface Data, Daily - US request has completed.....

The following files are available:

Link 1-

SOD- Surface Data, Daily - US - Web Form (4847093 bytes - compressed / 64804377 - uncompressed)

Link 2 –

SOD- Surface Data, Daily - US - ASCII Form (13671070 bytes - compressed / 64804377 - uncompressed)

Link 3 –

SOD- Surface Data, Daily - US - Stations included in your search (27726 bytes)

Link 4 –

SOD - Surface Data, Daily - US - Data inventory for your search (227995 bytes)

Figure B-14   Links Sent by NCDC to Download Weather Data

Figure B-15   Query to Extract Ramp Data

Step 3- Attaching Center Point To Nearby Roadway by Spatial Join

After creating the route, it needs to be linked to the bridge center point. The user needs to use 'Locate feature along route" under the linear referencing tool to spatially locate the bridge on the segments.

The user needs to select the Bridge layer as the input feature and the route created in setp-2 as the input route. A certain search radius (e.g., 150 ft) should be selected. The user also needs to make sure that the "Keep only closest route location" is checked so that a bridge center point can be referenced to more than one route. (see Figure B-26).
Figure B-28 shows the output after running step 3.

Step 4-Roadway Name Matching

Upon running step 3, a new table called the "bridgeNearTable" will be created. All routes within a certain distance to the bridge are identified and their names compared to the road name of the bridge. The route with the same name as the bridge road is assigned to this bridge. In the case of more than one routes having the same name as the bridge road, the closest route is assigned. An application has been developed in ArcGIS to perform the matching. Cick the tab "Bridge" in the data processing tool to perform the matching.

Step 5-Creating Bridge End Points Locations on Roadway

Once the matching is completed, the start and end point locations should be calculated for the route. For this step, two new fields, FMEAS and TMEAS (from measurement and to measurement), need to be created and then a specific formula is used to calculate the start and end point location on the route (see Figure B-30).
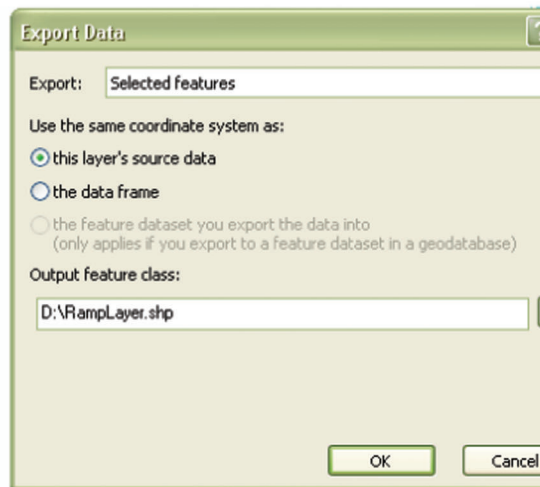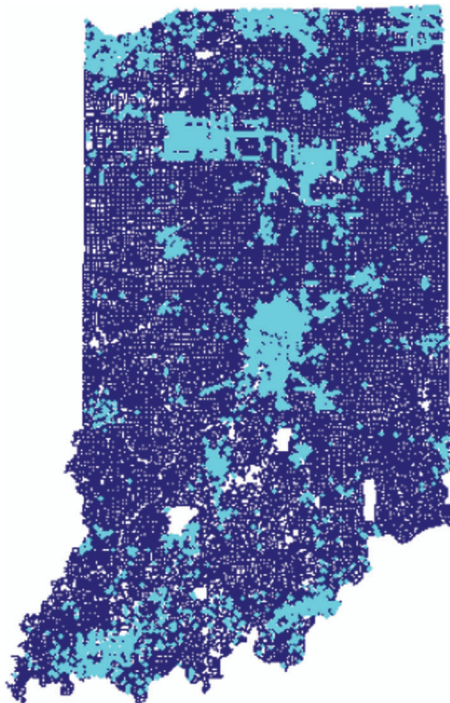
Figure B-16   Exporting the Ramp Layer
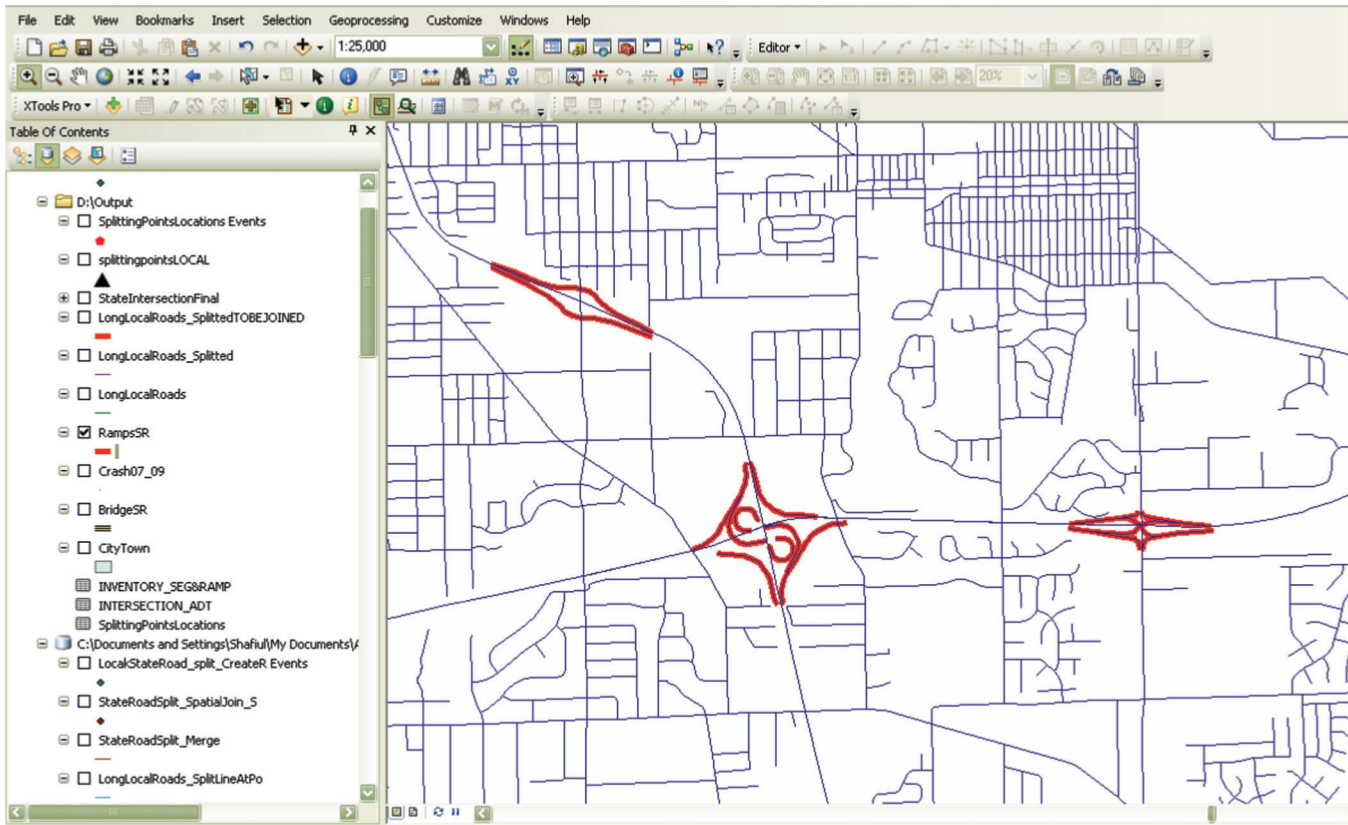
Figure B-17   Adding New Ramp Layer to Map Window

According to Figure B-30, 'LenofStruc' stands for a variable in the NBI bridge datasets, which represents a length of bridge structure (in 1/10 of a meter). MEAS is the measure of the bridge center point on the route in meters.

The ArcGIS routing system, by default, uses meters and the user needs to be cautious about the unit conversion.

Step 6- Create Bridge Segment Layer

After the FMEAS and TMEAS are calculated for the segments, the bridge can be created as a separate polyline feature. Use the tool "Make Route Event Layer" under "Linear Referencing Tools" in ArcGIS and provide inputs for 1) Route, 2) Route Identifier Field, and 3) Event Table. Here, for the event table, the input table will be the bridge near table. Also, the newly created fields in Step 5 (FMEAS and TMEAS) need to be specified.

After submitting the geo-processing "make route event layer" process, a new shape file will be created.

SPLITTING SEGMENTS

Splitting long segments into short element segments of uniform length is needed for a more precise identification of locations with safety issues along long segments. Before the splitting process is performed, the network needs to be "un-split" by using the GIS geo-processing tool that avoids combining segments of different names and different types. This ensures reasonable homogeneity of the combined segments and avoids the problem with the segment identification.

There are two different steps in splitting segments:

Step 1: Un-split the Segments

To un-split a segment, ArcGIS 10 has a new geo-processing tool available under the following link: data management> features> un-split line.

The un-split tool mainly has an option for the input features and output features directory location and name (see Figure B-33).

Step 2: Split Line by Split Tool in Data Management Interface

Use the splitting tool under the data management interface. Specify the maximum distance between two intersections which will not be split and the typical uniform length for splitting (in ft).

LINKING INVENTORY DATA

The data integration algorithm for linking the Indiana inventory data is shown in Figure B-35. The step-by-step approach for data integration are discussed below:

Step 1: Creating Inventory Route

First, from the Indiana road inventory, a route should be created and the segment needs to be located on the layer. This method is similar to the steps followed for "Bridge Integration."

In order to create a route based on the inventory layer, add the inventory shape file in ArcGIS and then click the "Create route" geoproecssing tool under Linear Reference System Tools. The 'Create Route' window will appear. User needs to specify: the input
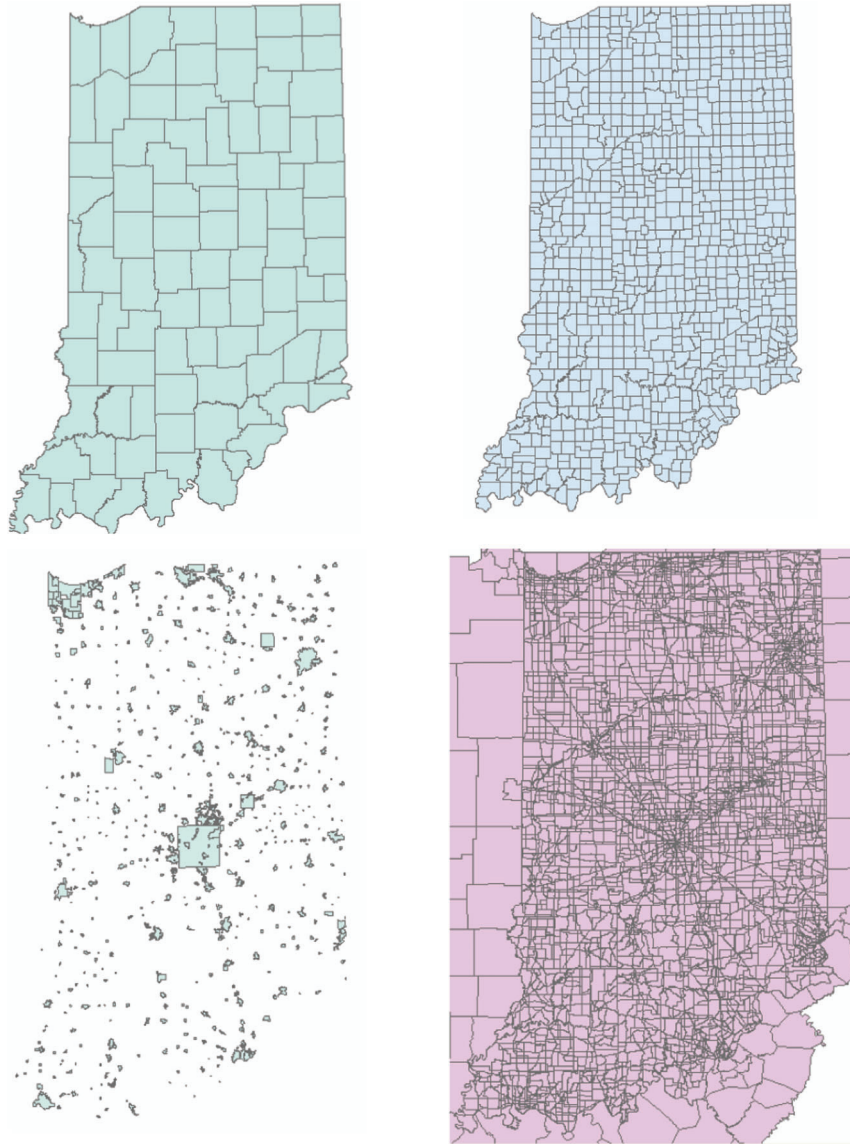
Figure B-18   County, Township, City, and TAZ Maps

polyline feature (i.e. the inventory layer), route identifier field and, the output route layer. The functional class route will be created.

Step 2- Locating Segment Layer To Nearby Inventory

After creating the inventory route, the segments need to be linked to the route. The user needs to use "Locate feature along route" in linear referencing tool to spatially locate the bridge on the segments. The input for this module area is input feature=StateSegment Layer, Input Route= Inventory Route, Route Identifier = ObjectID1. Also, the search radius can be set to 150 ft. After running the module, a table showing the route event will be produced.

Step 3: Create Point Layer for Start and End Points

Create a point layer for start and end points feature class by using "Feature Vertices To Points'" tool. Run this first for the "Start" point and then for the "End" Point. Two shape files, StateRoad_Split_Start and StateRoad_Split_End, will be created.

Step 4 Data Integration

The user needs to use the "Data Integration" tab under the data management tool to finish the integration. The inputs for these operations are the segment file, the point layers for the start and end points, and the output table created in Step 2.

After running Step 4, a shape file the integration results will be created as shown in Figure B-39.

CRASH ASSIGNMENT

Step 1: Convert Crash Records to Shape File

The crash records need to be converted to a shape file in order to perform the proximity-based crash assignment. Step 1 of the bridge layer integration shows a way to convert the NBI datasets into a shape file. Similar techniques will be used to convert the crash table (collision table only) into a shape file.

ANNUAL GROWTH FACTORS BY FUNCTIONAL CLASSIFICATION
2000 - 2009

YEAR TO	YEAR FROM									
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009

Urban - Interstate (11), Freeways and Expressways (12)

YEAR TO	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
2000	-	0.921	0.890	0.863	0.853	0.831	0.817	0.786	0.803	0.788
2001	1.086	-	0.966	0.937	0.926	0.903	0.887	0.853	0.872	0.855
2002	1.124	1.035	-	0.970	0.958	0.934	0.919	0.883	0.902	0.885
2003	1.159	1.067	1.031	-	0.988	0.963	0.947	0.911	0.930	0.913
2004	1.173	1.080	1.043	1.012	-	0.975	0.958	0.922	0.941	0.924
2005	1.203	1.108	1.070	1.038	1.026	-	0.983	0.945	0.966	0.948
2006	1.224	1.127	1.089	1.056	1.043	1.017	-	0.962	0.982	0.964
2007	1.273	1.172	1.132	1.098	1.085	1.058	1.040	-	1.021	1.002
2008	1.248	1.147	1.108	1.075	1.062	1.035	1.018	0.979	-	0.981
2009	1.270	1.169	1.130	1.096	1.083	1.055	1.038	0.998	1.019	-

Urban - Principal Arterials (14), Minor Arterials (16), Collectors (17), Local (19)

YEAR TO	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
2000	-	0.969	0.914	0.926	0.937	0.936	0.930	0.942	0.972	0.976
2001	1.032	-	0.943	0.956	0.967	0.966	0.960	0.972	1.003	1.008
2002	1.094	1.060	-	1.013	1.025	1.024	1.017	1.031	1.064	1.068
2003	1.080	1.046	0.987	-	1.012	1.011	1.004	1.017	1.050	1.054
2004	1.067	1.034	0.975	0.988	-	0.999	0.992	1.005	1.037	1.041
2005	1.068	1.035	0.976	0.989	1.001	-	0.993	1.006	1.038	1.042
2006	1.075	1.042	0.983	0.996	1.008	1.007	-	1.013	1.046	1.050
2007	1.061	1.028	0.970	0.983	0.995	0.994	0.987	-	1.032	1.036
2008	1.028	0.997	0.940	0.952	0.964	0.963	0.956	0.969	-	1.004
2009	1.024	0.993	0.936	0.949	0.960	0.959	0.953	0.965	0.996	-

Rural - Interstate (01)

YEAR TO	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
2000	-	0.968	0.923	0.918	0.906	0.902	0.895	0.888	0.903	0.911
2001	1.033	-	0.953	0.949	0.936	0.932	0.924	0.917	0.933	0.941
2002	1.084	1.049	-	0.995	0.982	0.977	0.970	0.962	0.979	0.987
2003	1.089	1.054	1.005	-	0.987	0.982	0.974	0.967	0.983	0.992
2004	1.103	1.068	1.018	1.013	-	0.995	0.987	0.979	0.996	1.005
2005	1.109	1.073	1.023	1.018	1.005	-	0.992	0.984	1.001	1.010
2006	1.118	1.082	1.031	1.026	1.013	1.008	-	0.992	1.009	1.018
2007	1.127	1.091	1.040	1.034	1.021	1.016	1.008	-	1.017	1.027
2008	1.107	1.072	1.022	1.017	1.004	0.999	0.991	0.983	-	1.009
2009	1.097	1.062	1.013	1.008	0.995	0.990	0.982	0.974	0.991	-

Rural - Principal Arterials (02), Minor Arterials (06)

YEAR TO	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
2000	-	0.978	0.966	0.988	0.961	0.962	0.953	0.953	1.002	1.007
2001	1.022	-	0.987	1.009	0.982	0.983	0.974	0.974	1.024	1.029
2002	1.035	1.013	-	1.022	0.995	0.996	0.987	0.987	1.036	1.043
2003	1.013	0.991	0.978	-	0.973	0.974	0.965	0.965	1.015	1.020
2004	1.041	1.018	1.005	1.028	-	1.001	0.992	0.992	1.043	1.048
2005	1.040	1.017	1.004	1.027	0.999	-	0.991	0.991	1.042	1.047
2006	1.049	1.027	1.013	1.036	1.008	1.009	-	1.000	1.052	1.057
2007	1.049	1.027	1.013	1.036	1.008	1.009	1.000	-	1.052	1.057
2008	0.998	0.976	0.964	0.985	0.959	0.960	0.951	0.951	-	1.005
2009	0.993	0.971	0.959	0.981	0.954	0.955	0.946	0.946	0.995	-

Rural - Major Collectors (07), Minor Collectors (08), Locals (09)

YEAR TO	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
2000	-	1.002	0.980	0.978	0.974	0.986	0.991	0.984	1.053	1.045
2001	0.998	-	0.978	0.976	0.972	0.984	0.989	0.983	1.051	1.042
2002	1.020	1.022	-	0.997	0.993	1.006	1.011	1.004	1.074	1.065
2003	1.023	1.025	1.003	-	0.996	1.009	1.014	1.007	1.077	1.069
2004	1.027	1.029	1.007	1.004	-	1.013	1.018	1.011	1.081	1.073
2005	1.014	1.016	0.994	0.991	0.987	-	1.005	0.998	1.067	1.059
2006	1.009	1.011	0.989	0.986	0.982	0.995	-	0.993	1.062	1.054
2007	1.016	1.018	0.996	0.993	0.989	1.002	1.007	-	1.070	1.061
2008	0.950	0.952	0.931	0.928	0.925	0.937	0.942	0.935	-	0.992
2009	0.957	0.959	0.939	0.936	0.932	0.944	0.949	0.942	1.008	-

Figure B-19   Adjustment Factors Downloaded from INDOT Website
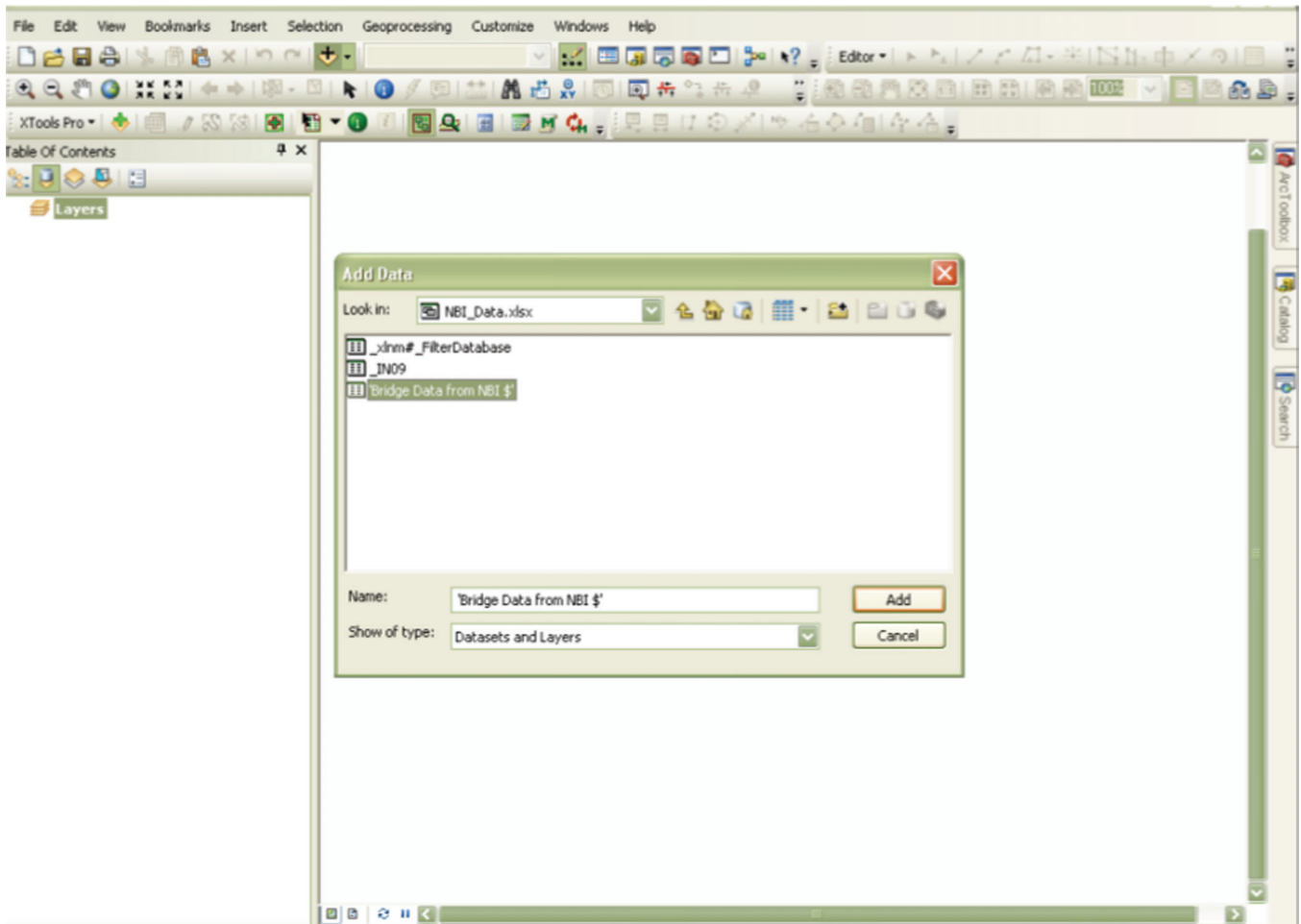
Figure B-20   Adding NBI Data in ArcMap

Step 2: Aggregate Crash Files

Although crash assignment can be performed year by year, the aggregation of different years' crash files can eliminate some repetitive work. The ArcGIS geo-processing operation "merge" needs is used for this operation. The merge tool is available in Data ManagementTools> General>Merge.

Once the merge tool window opens, the user needs to add each year's crash files (.shp) one by one in the Input Datasets list.

The "field map" options in this tool can help edit the final fields layout. It is required to have each crash file's field name and types identical to one another in order to perform the operation of this tool successfully.

Step 3: Use Model Builder to Assign Crash

The crash assignment model builder is a customizable tool developed at the Center for Road Safety (CRS) to assign the crashes. Any tool developed as a "Model builder" first needs to be opened in the ArcGIS environment. The user needs to right click on the ArcToolBox window and add the corresponding user-defined tool. After the whole list of user-defined tools are added to the ArcToolBox, the user needs to right click on the "CrashAssignmentFinal" tool and open the edit environment. The model builder's editing environment will appear.

In the editing environment, the user has the flexibility to change the criteria for crash assignment. Each component in "yellow"

indicates a particular geo-processing operation. Therefore, the model builder has mainly two different geo-processing operations: "Select by Attribute" and "Generate Near Table." Components in blue are inputs and components in green are output.

By double clicking the expression input for the very first component, the user will be able to see the query builder for crash selection. The user can change the query in order to obtain a different subset of crashes for a specific element.

By double clicking on the "Generate Near Table" component, the user can open the second geo-processing tool "Generate Near Table." It can give users the flexibility to choose the input feature (i.e., crash layer), the near feature (i.e., element layer) and the buffer distance.

Once the parameters are properly specified, clicking on the run button at the top will generate whole sets of crash assignment modules and one output table for each element. The output table will show the crash IDs and corresponding assigned element IDs.

Once the output tables have been created, they can be joined with the crash table by common field IDs (FID in output file with FID in crash file). See Figure B-45 (left).

Once the tables are joined, the user needs to create a specific field in the crash table to copy the assigned IDs for a specific element. For example, for a segment, the user may add a field SegID and for intersection, IntID, and so on. Finally, after copying the IDs, the user may remove the join between the two tables. After crash assignment, the user should have a table as shown in Figure B-44 (right).
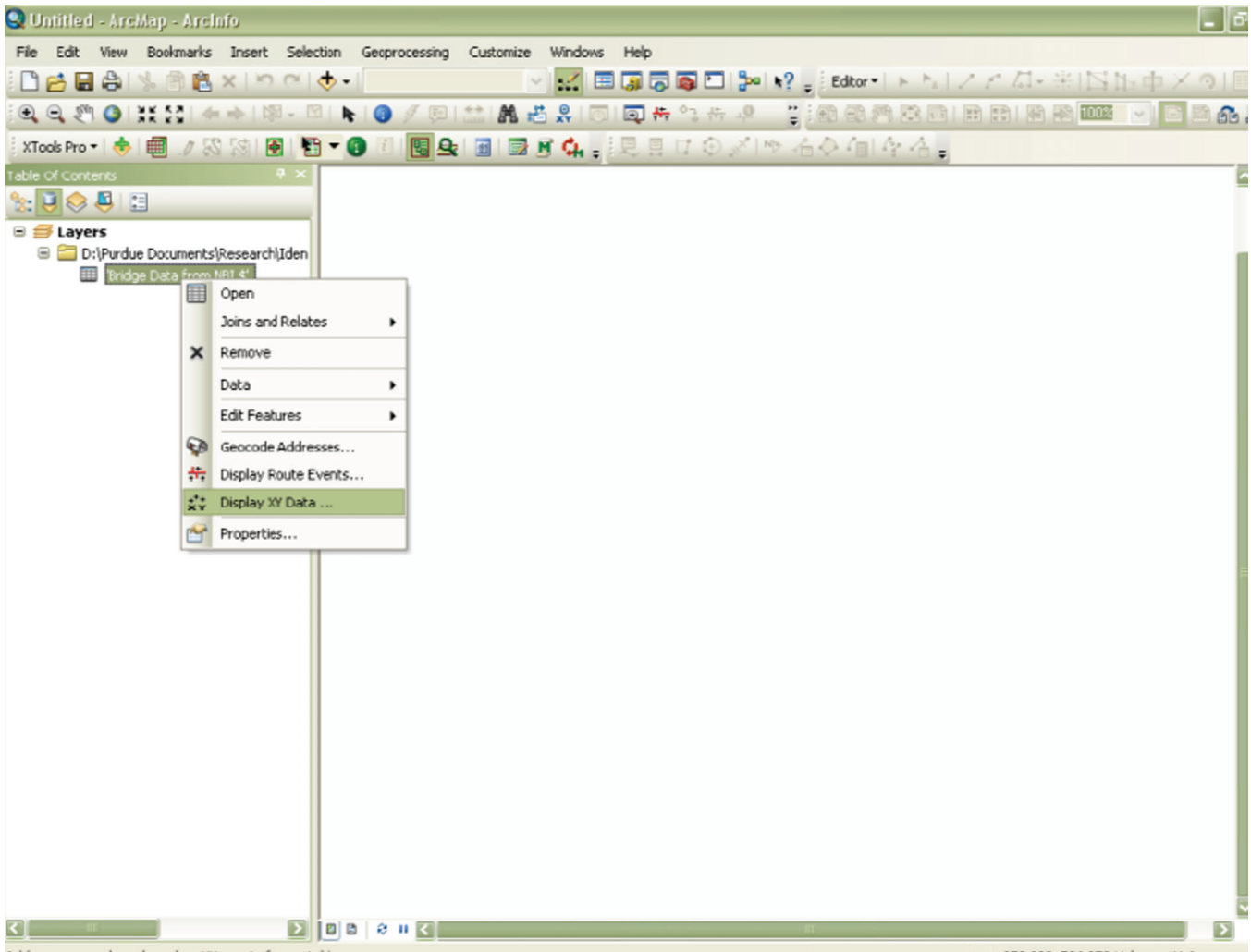
Figure B-21   Displaying the Bridge Center Points on Map
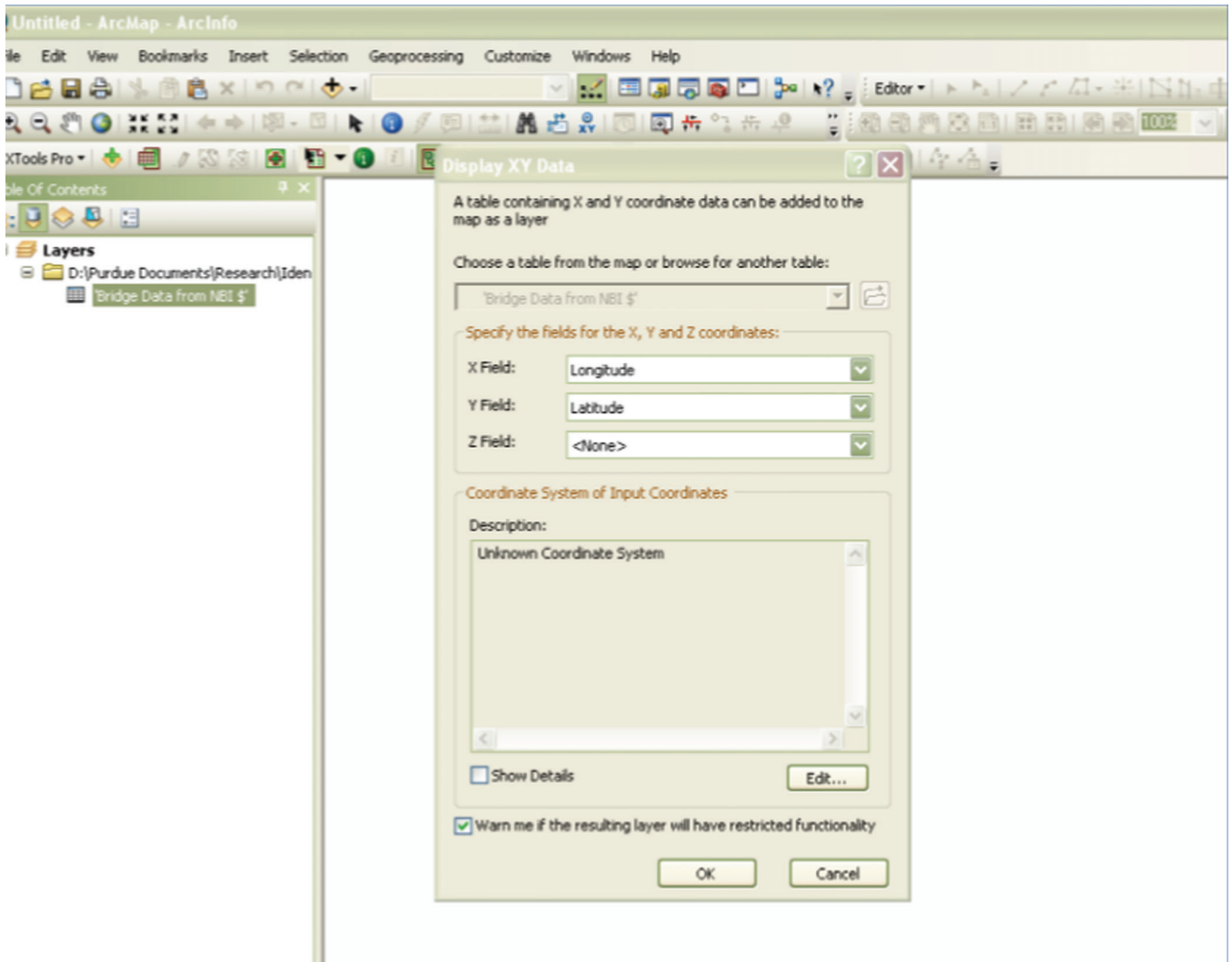
Figure B-22   Selecting the Inputs for Displaying Center Point of Bridge
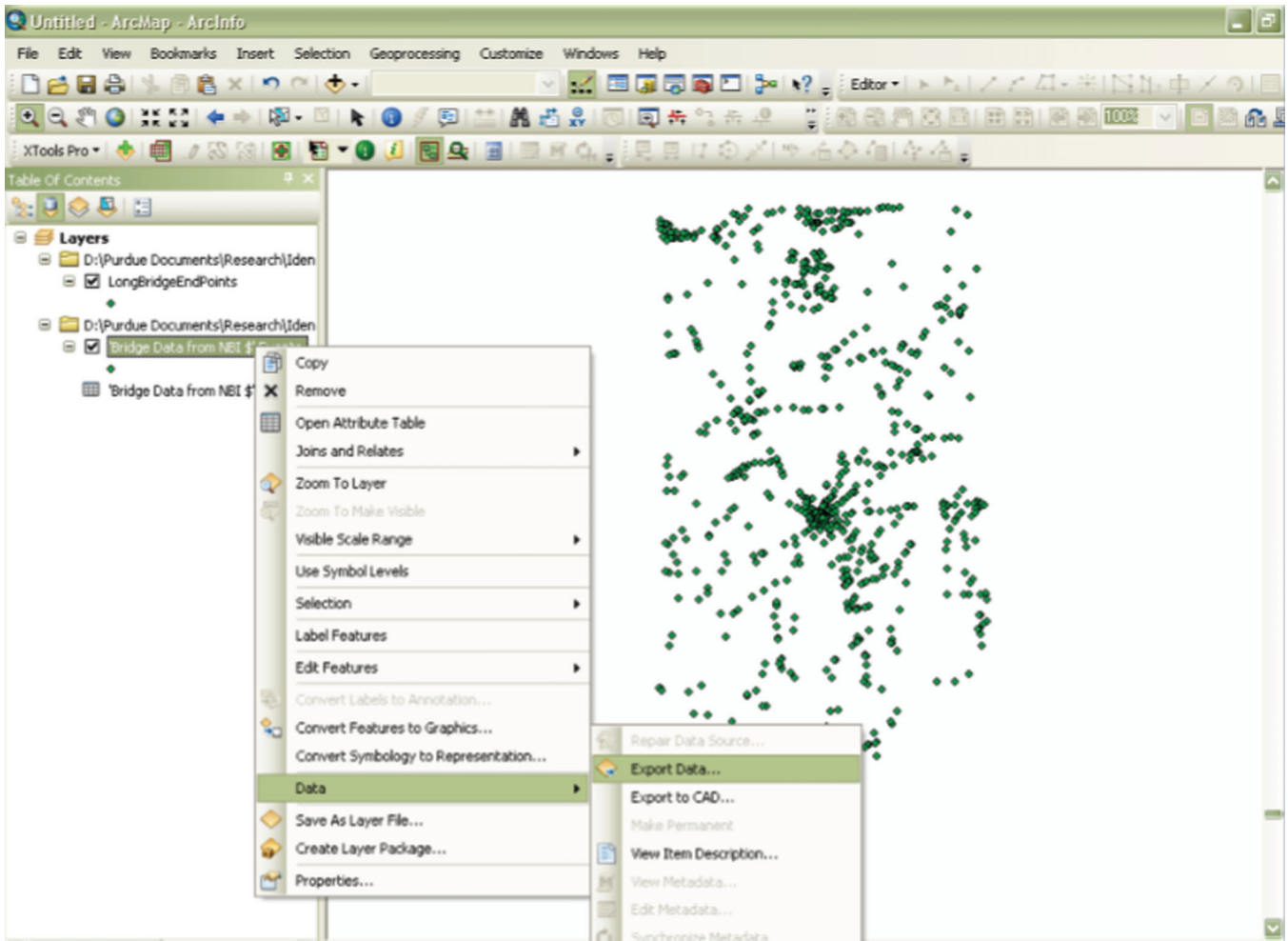
Figure B-23   Exporting the Bridge Center Point Display File to Shape File
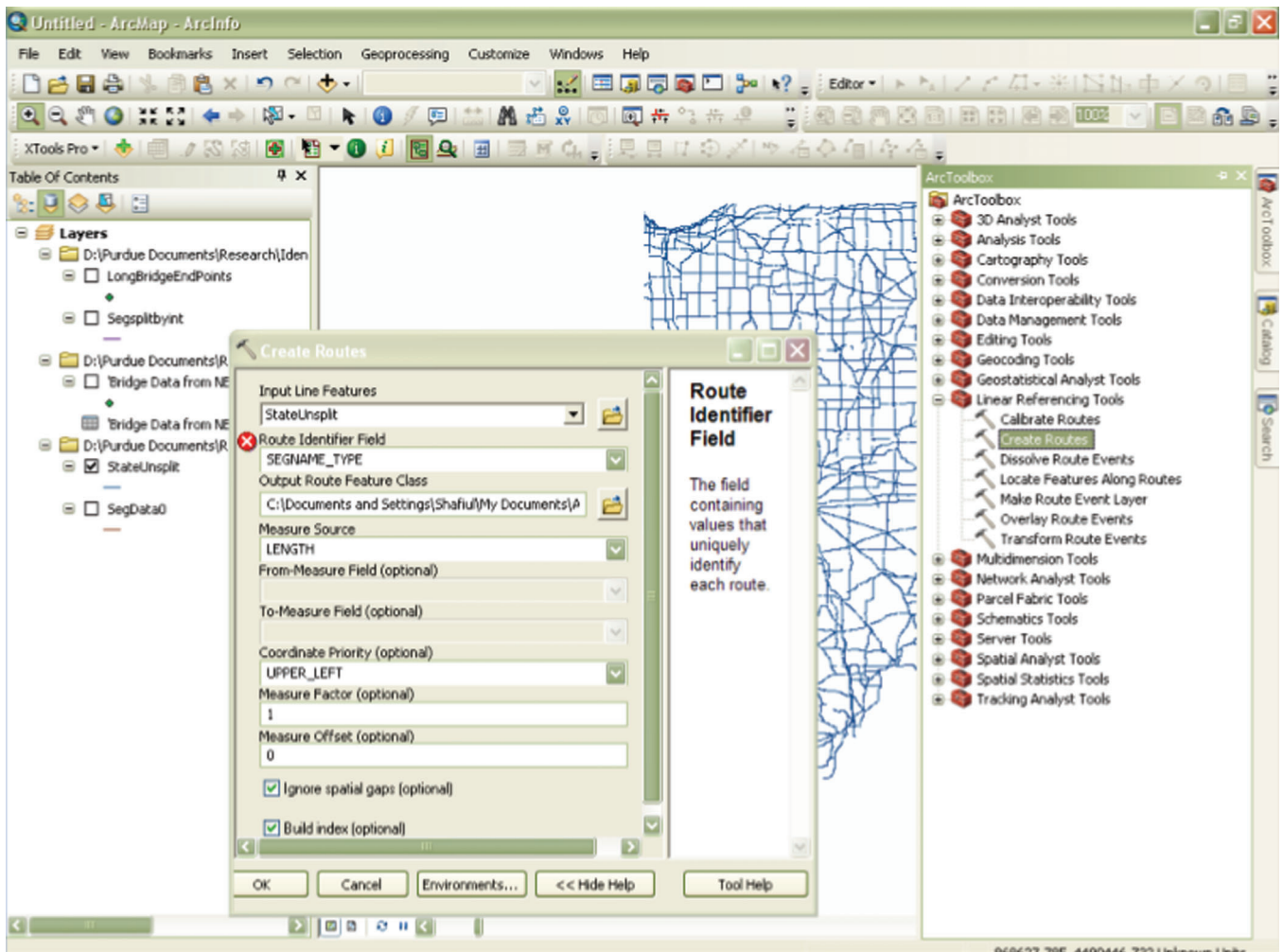
Figure B-24   Create Route Using LRS Tool in ArcGIS
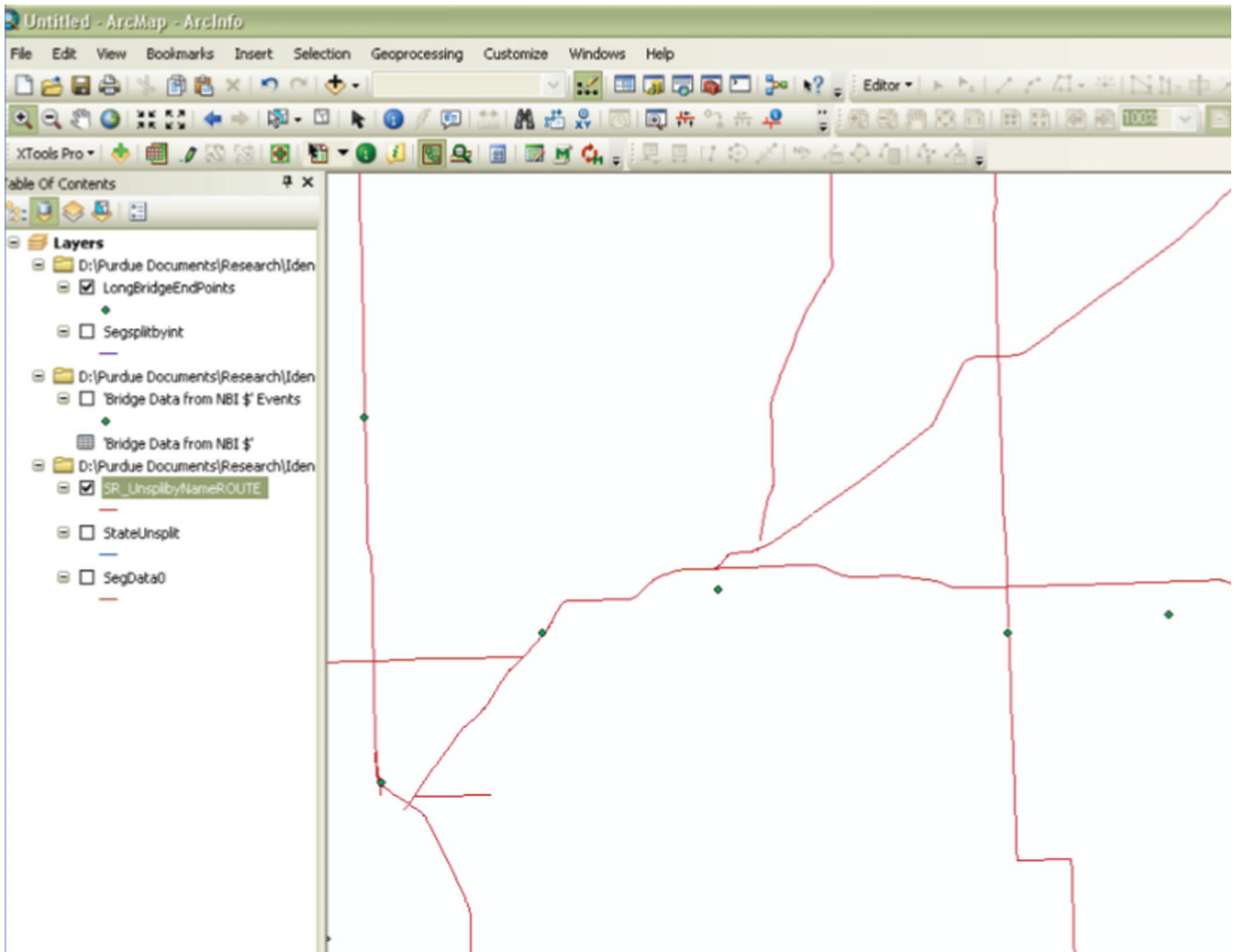
Figure B-25   Routes Created and Bridge Center Points
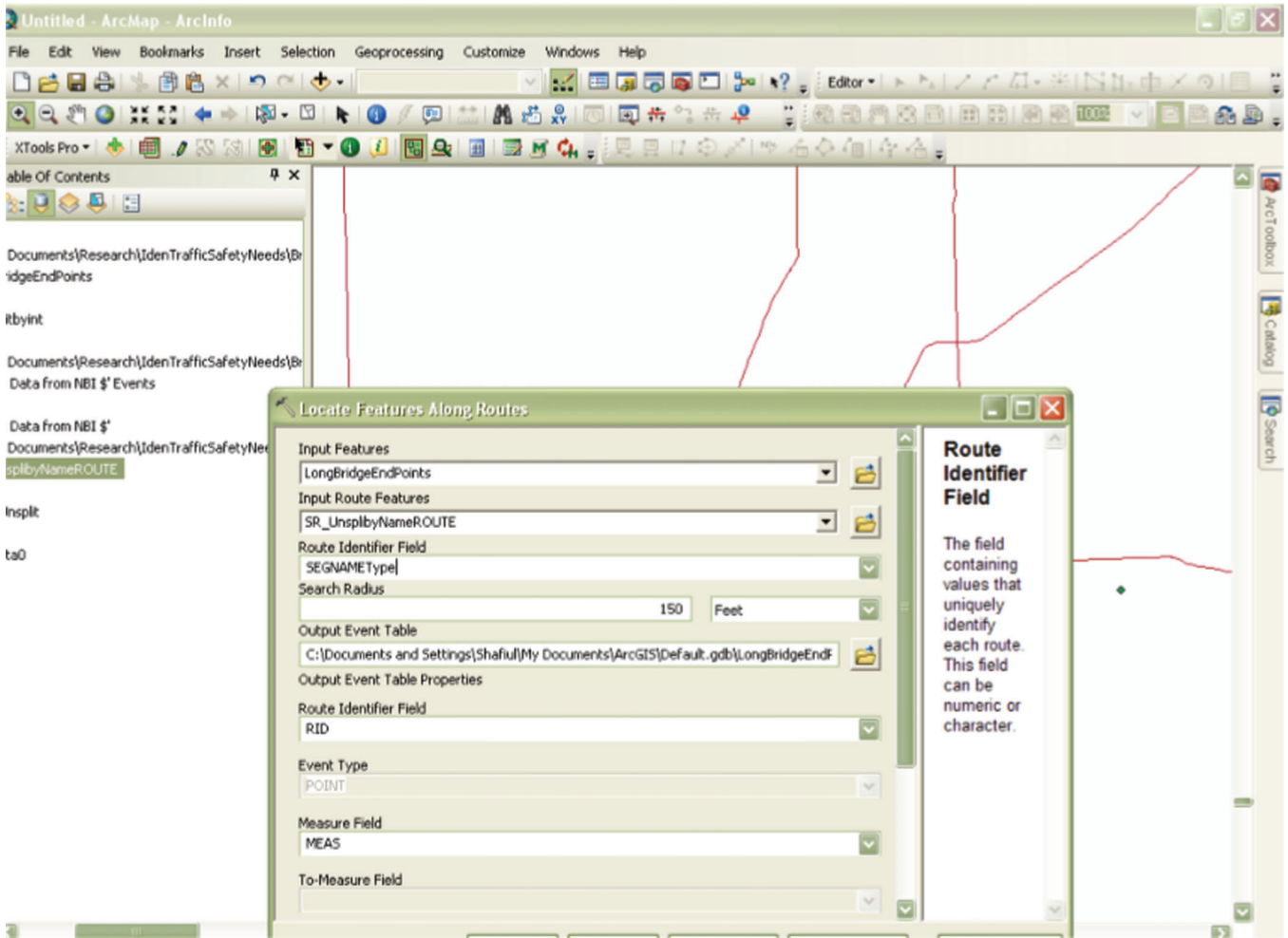
Figure B-26　Locating Bridge Point along the Route-1
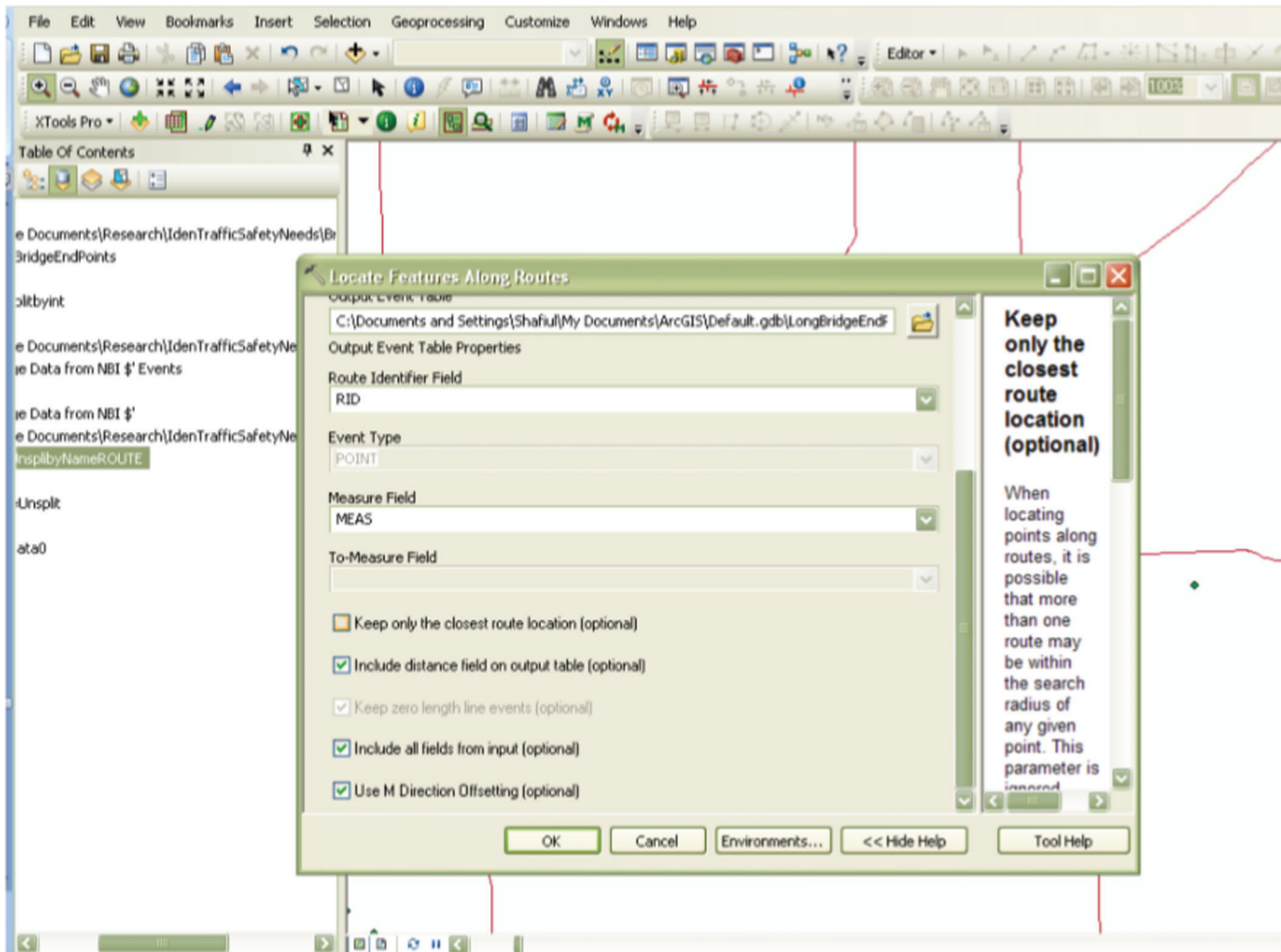
Figure B-27   Locating Bridge Point along the Route-2

Figure B-28   Output File after Locating Bridge along Route
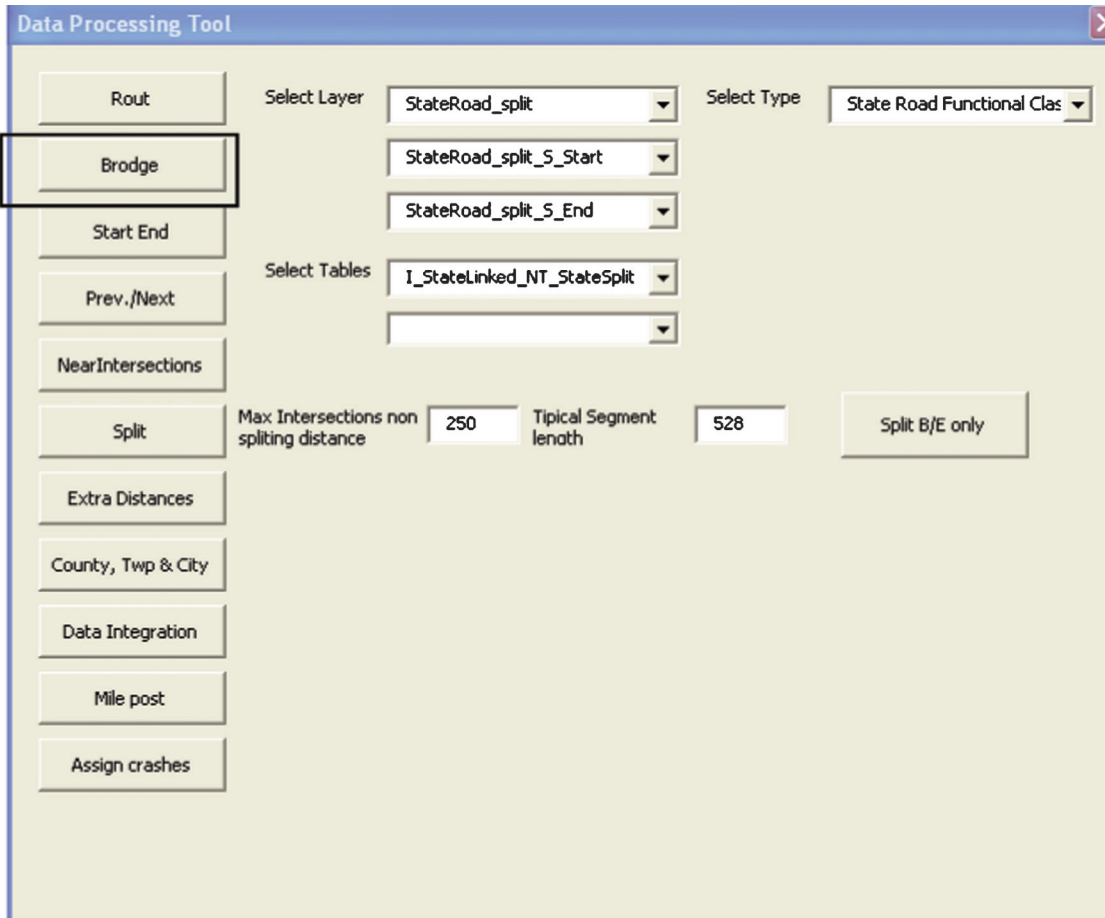
Figure B-29   Bridge Name Matching Tool under Data Management Interface

Figure B-30   Calculating Start and End Point Locations
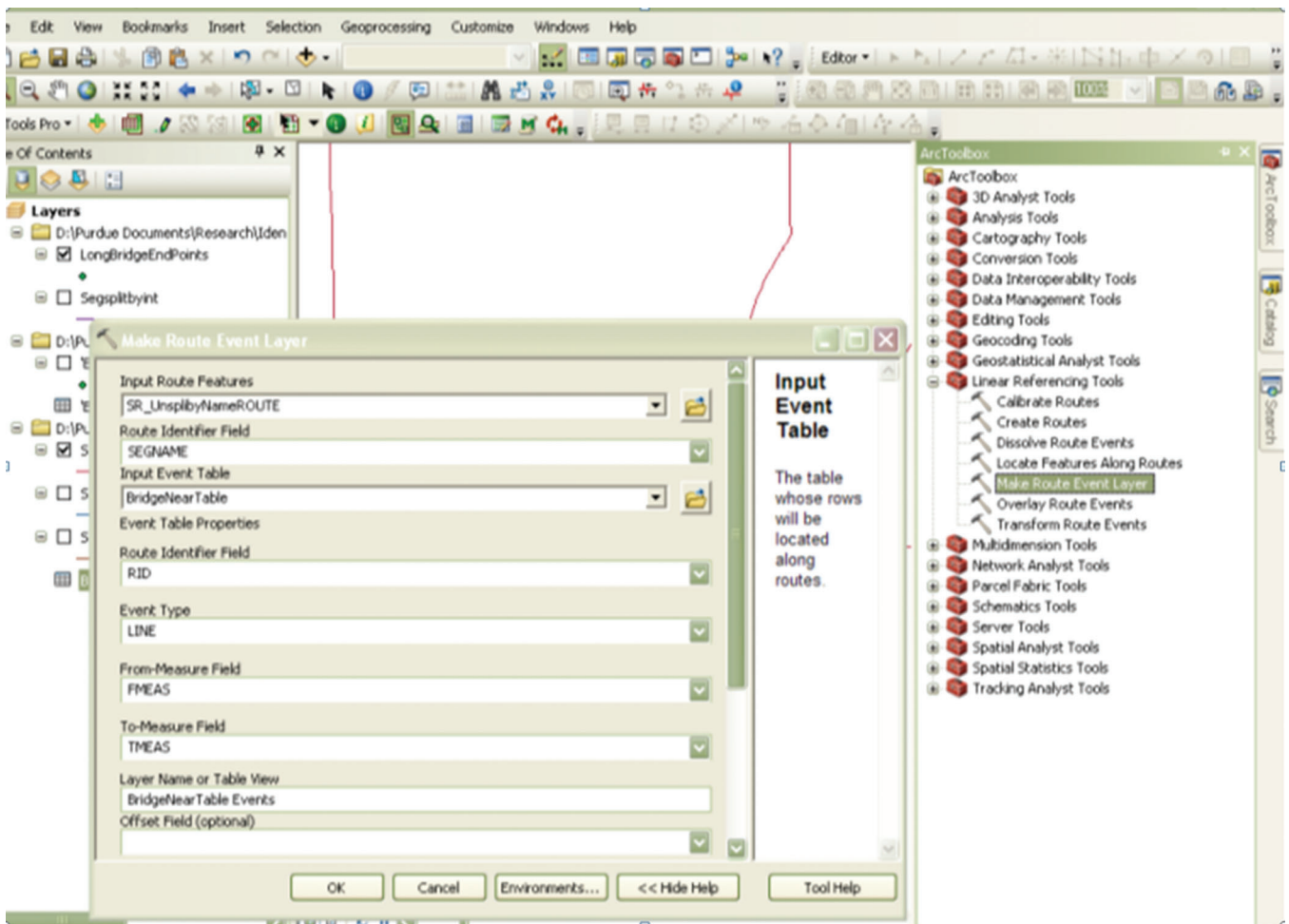
Figure B-31   Create Bridge Polyline Layer Input
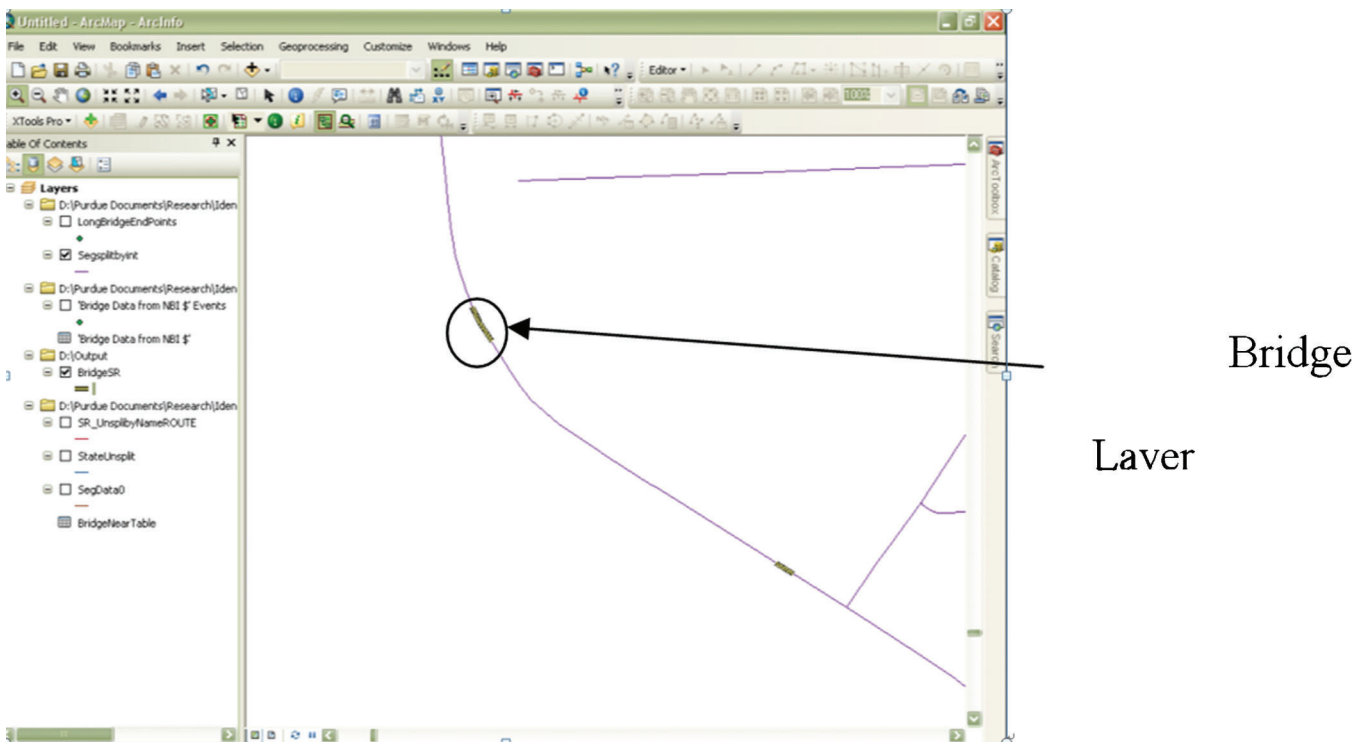
Figure B-32   Bridge Layer along with Segment Layer
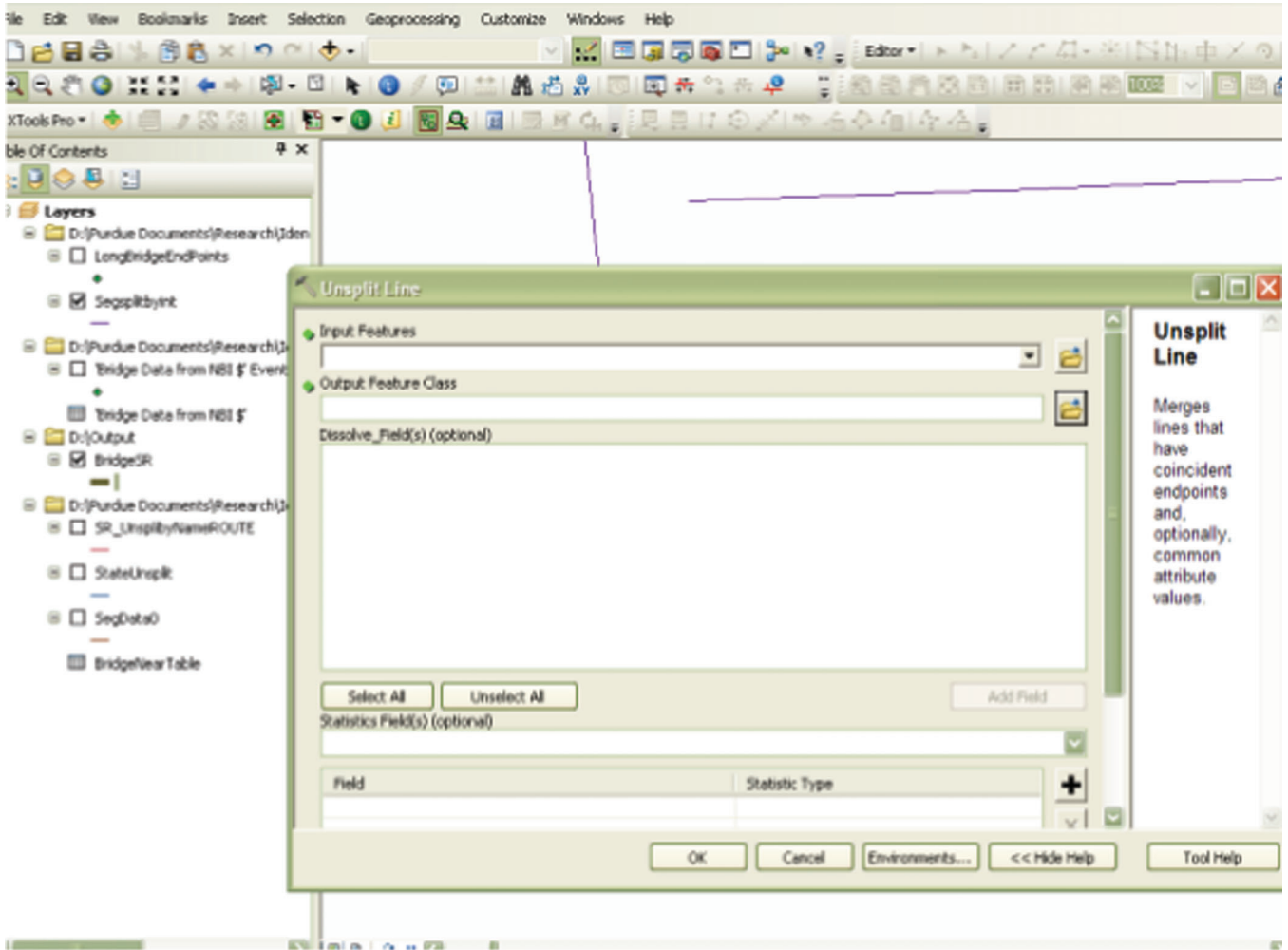
Figure B-33   Un-split Segments before Splitting (ArcGIS 10)

Figure B-34   Split Tool in Data Management Interface
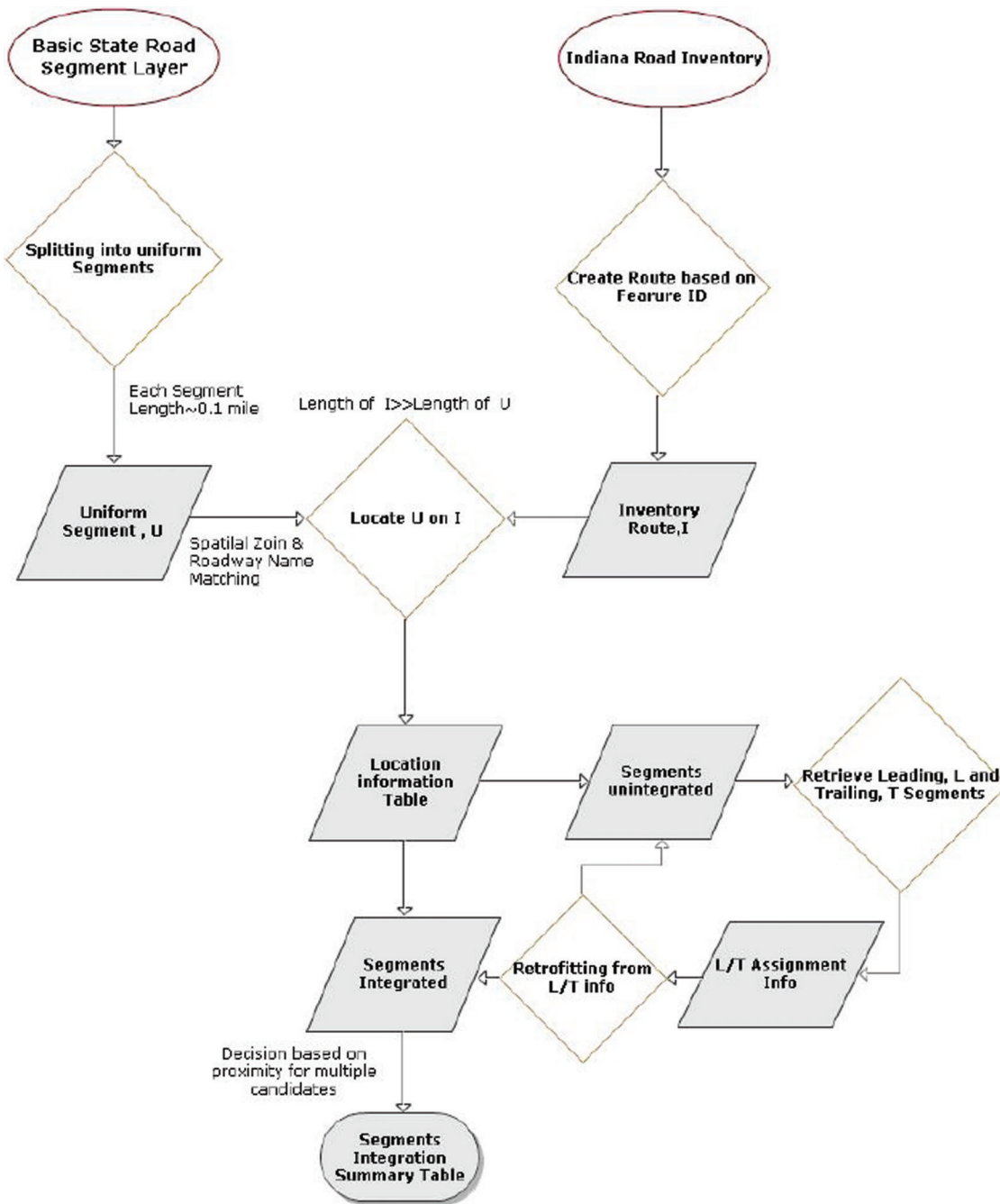
Figure B-35   Data Integration Algorithm
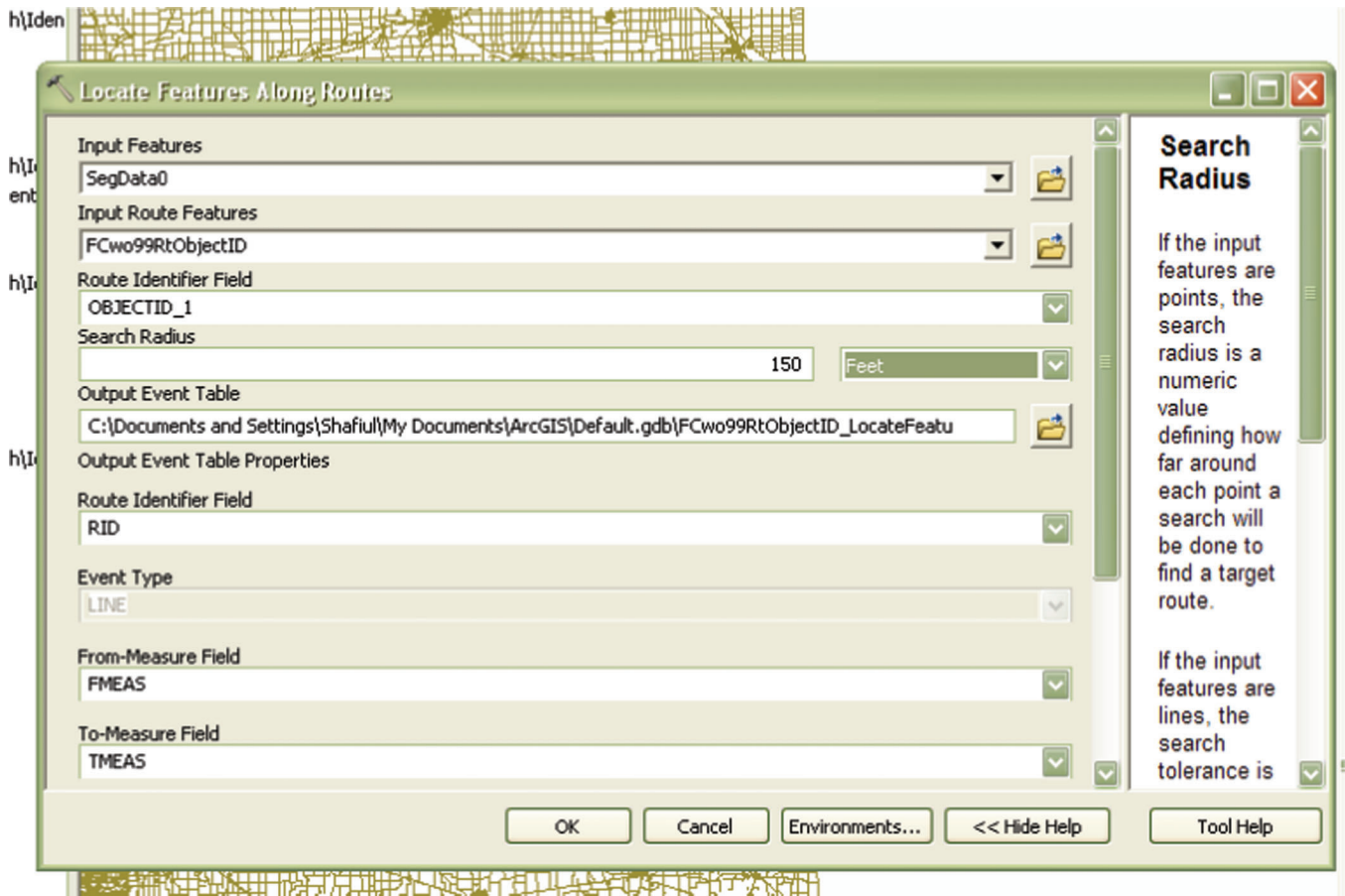
Figure B-36 Locating Segment Layer Along Inventory Route
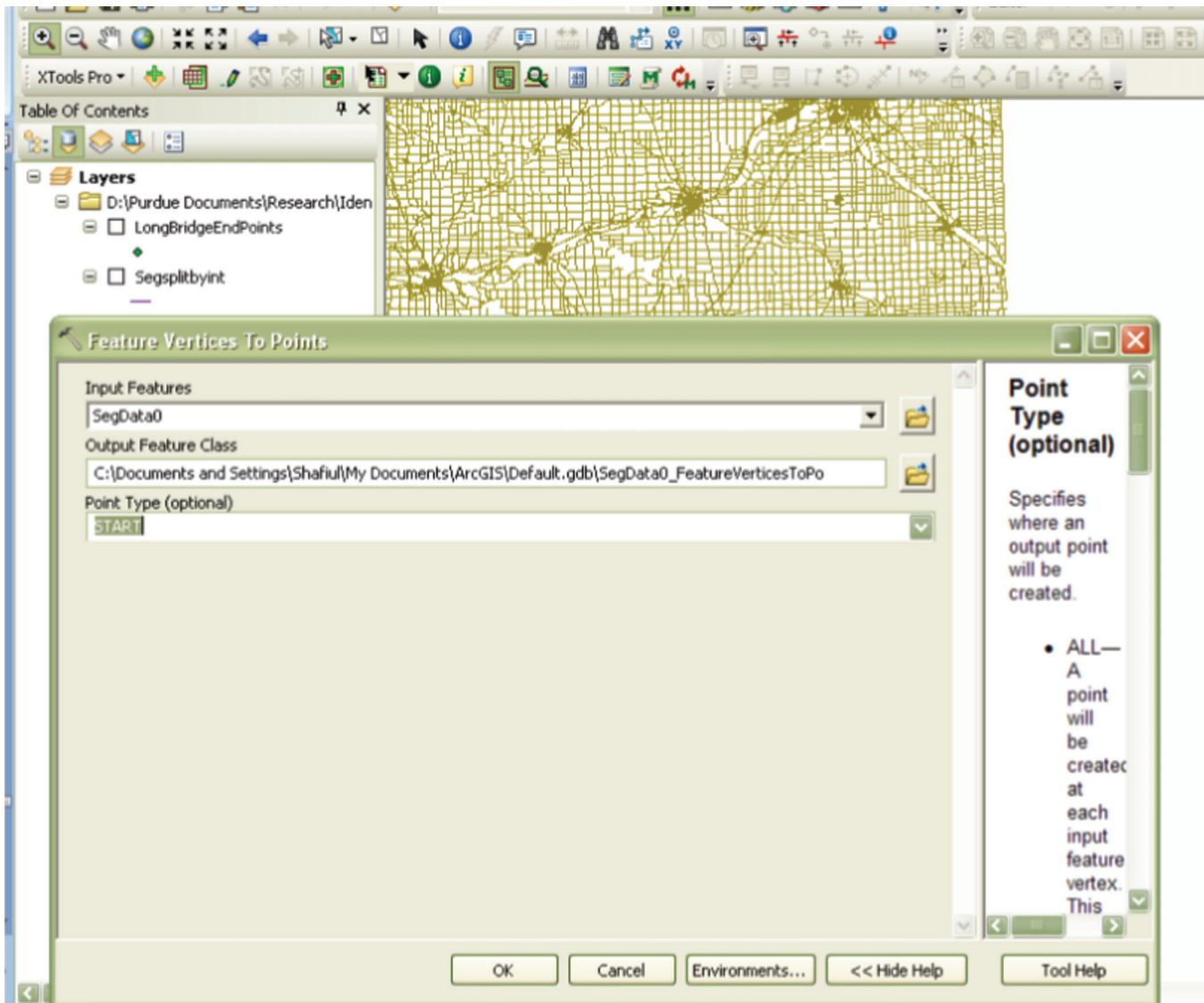
Figure B-37   Creating Point Layer for End Points
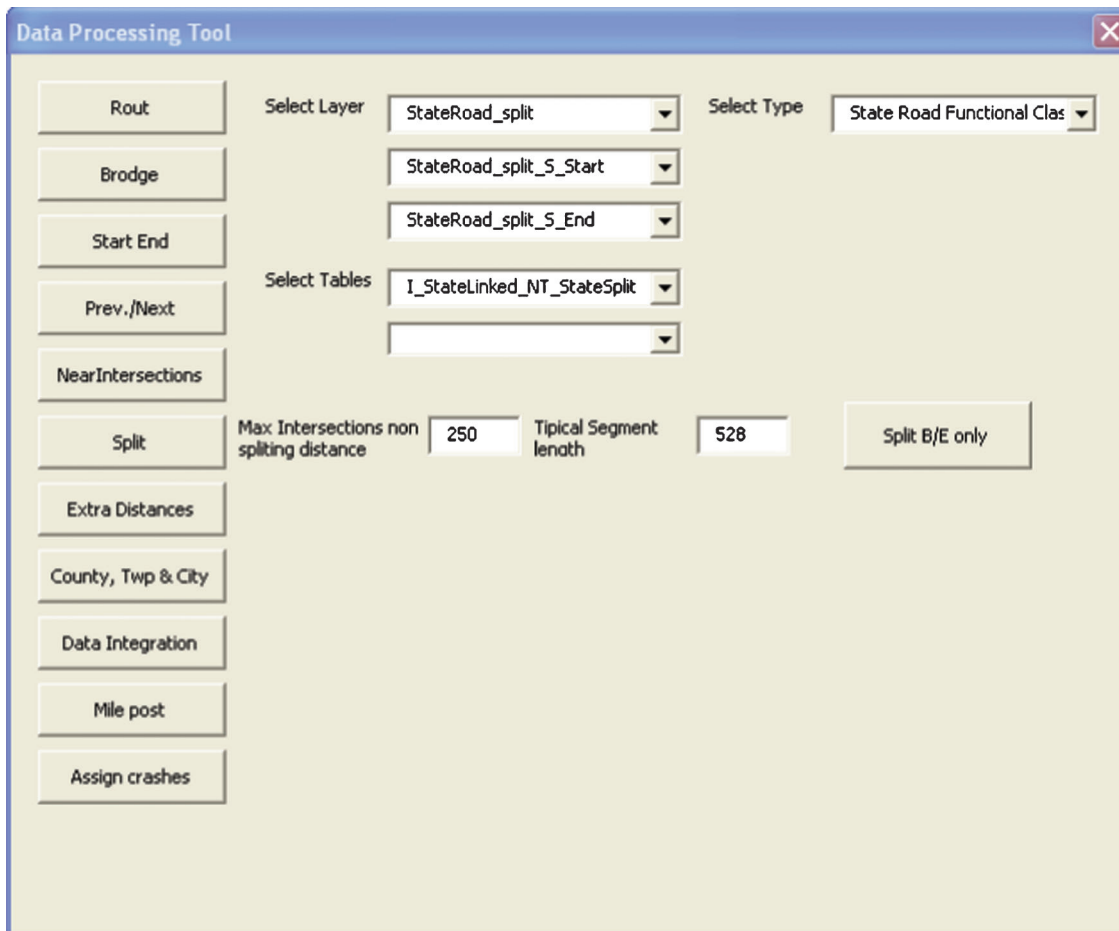
Figure B-38   Data Integration Tool in Data Management Interface

Contents

.ayers
D:\Purdue Docume
☐ LongBridgeEnd

☐ Segsplitbyint

D:\Purdue Docume
☐ 'Bridge Data fr

'Bridge Data fr
D:\Purdue Docume
☑ I_StateLinked

☐ SR_UnsplibyNa

☐ StateUnsplit

☐ SegData0

BridgeNearTab
D:\Purdue Docume
☑ FCwo99RtObje

D:\Output
☐ BridgeSR

I_StateLinked

FID	Shape	ID *	ASGNID	NEXTID	PREVID
21	Polyline M	135788	29841	-1	24378677
22	Polyline M	137202	3366	-1	24460994
23	Polyline M	412017	-2	-1	24445443
24	Polyline M	24365270	87242	24459437	24365323
25	Polyline M	24365272	49468	24365273	24365338
26	Polyline M	24365273	49468	24365274	24365272
27	Polyline M	24365274	49468	24365275	24365273
28	Polyline M	24365275	-49468	24365276	24365274
29	Polyline M	24365276	8856	24365277	24365275
30	Polyline M	24365277	8856	24365278	24365276
31	Polyline M	24365278	-8856	24365279	24365277
32	Polyline M	24365279	12560	24459439	24365278
33	Polyline M	24365281	73107	24365282	24459429
34	Polyline M	24365282	73107	24365281	24459430
35	Polyline M	24365284	73107	24365285	24459430
36	Polyline M	24365285	73107	-24459432	24365284
37	Polyline M	24365288	73107	24459431	24365289
38	Polyline M	24365289	73107	24365290	24365288
39	Polyline M	24365290	19354	24365289	24365291
40	Polyline M	24365291	19354	24365290	24365292
41	Polyline M	24365292	19354	24365293	24365291
42	Polyline M	24365293	19354	24365292	24365294
43	Polyline M	24365294	19354	24365318	24365293
44	Polyline M	24365295	11410	24365296	24459433
45	Polyline M	24365296	11410	24459434	24365295
46	Polyline M	24365300	57597	24365301	24459435
47	Polyline M	24365301	57597	24463767	24365300
48	Polyline M	24365302	36491	24365303	24365378
49	Polyline M	24365303	36491	24365304	24365302
50	Polyline M	24365304	36491	24365305	24365303
51	Polyline M	24365305	36491	24365306	24365304
52	Polyline M	24365306	36491	24365307	24365305

I◀ ◀ 0 ▶ ▶I (0 out of 47697 Selected)

I_StateLinked

Figure B-39 Output Table after Running the Data Integration Tool
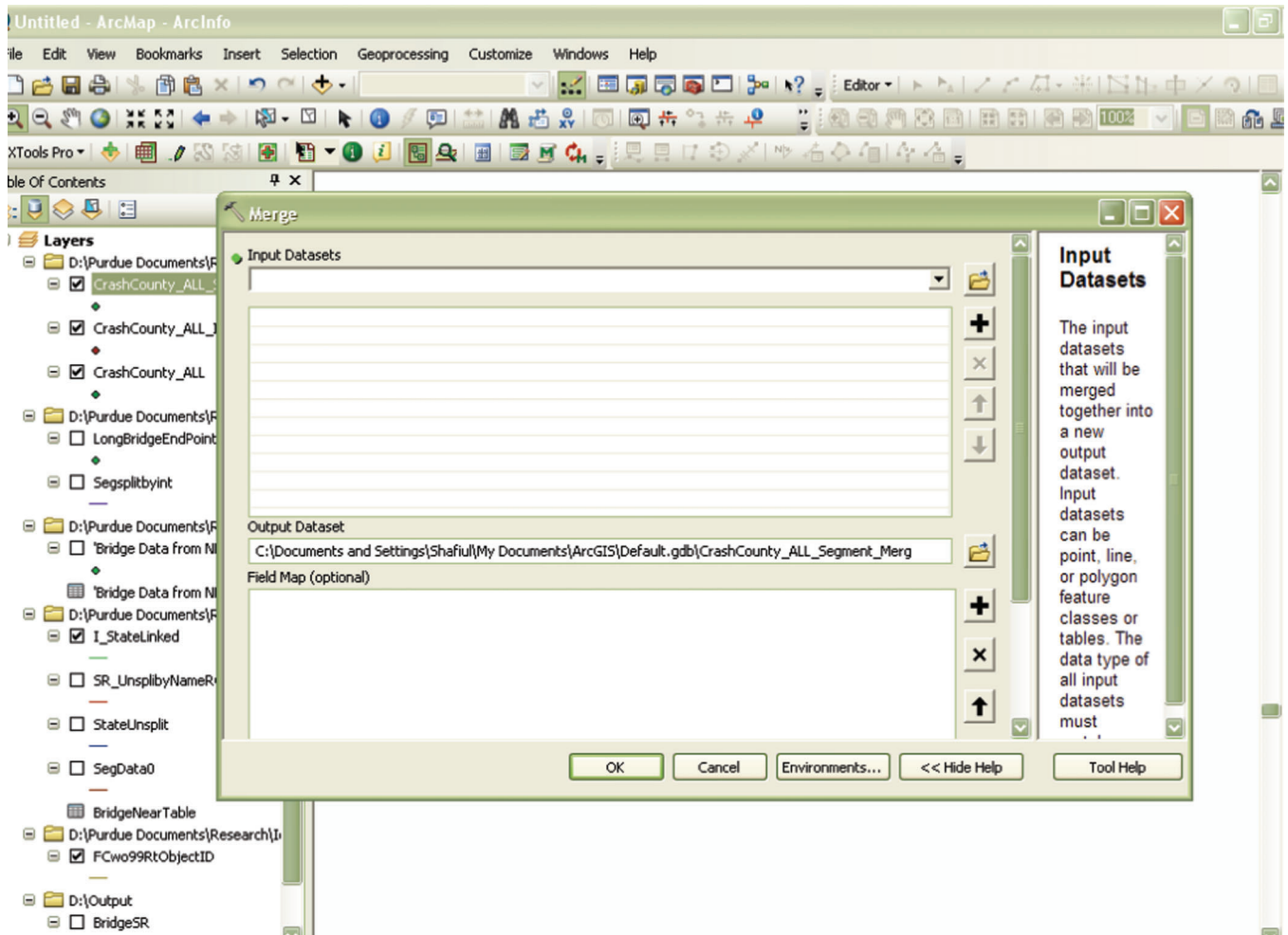
Figure B-40   Merging Multiple Year Crash Shape Files
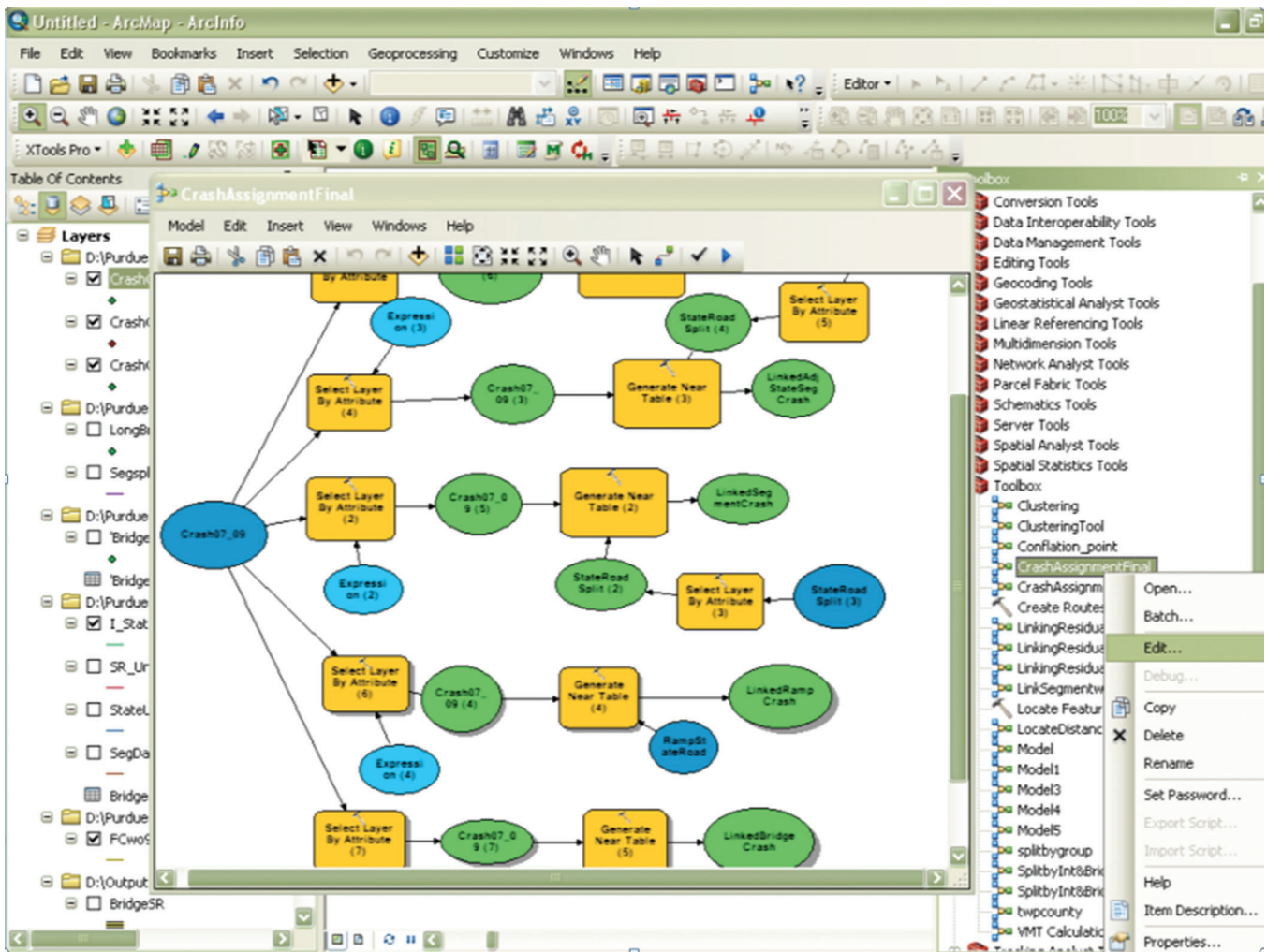
Figure B-41   Model Builder Editing Environment
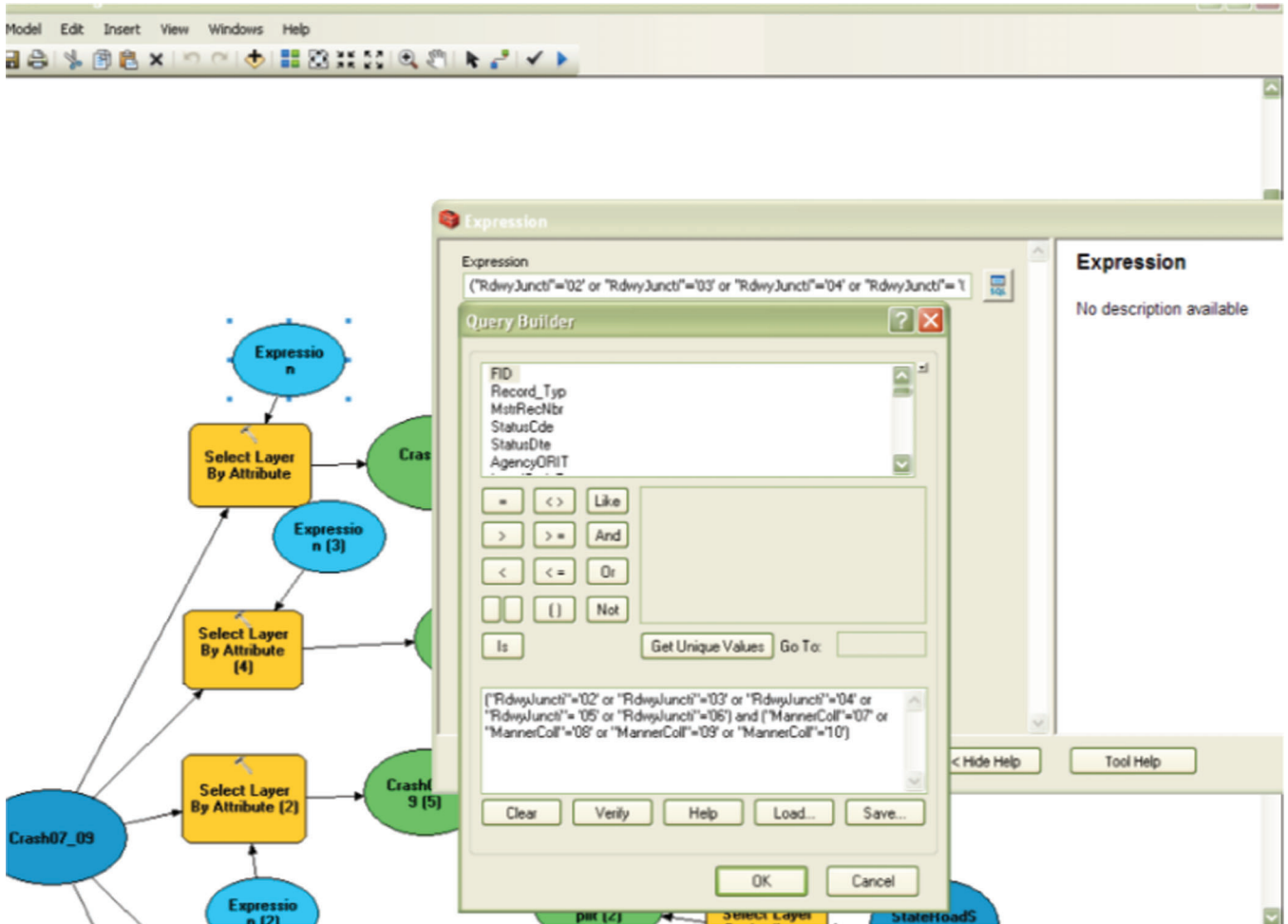
Figure B-42   Editing Model Builder Crash Selection Criteria for Subsets of Crashes
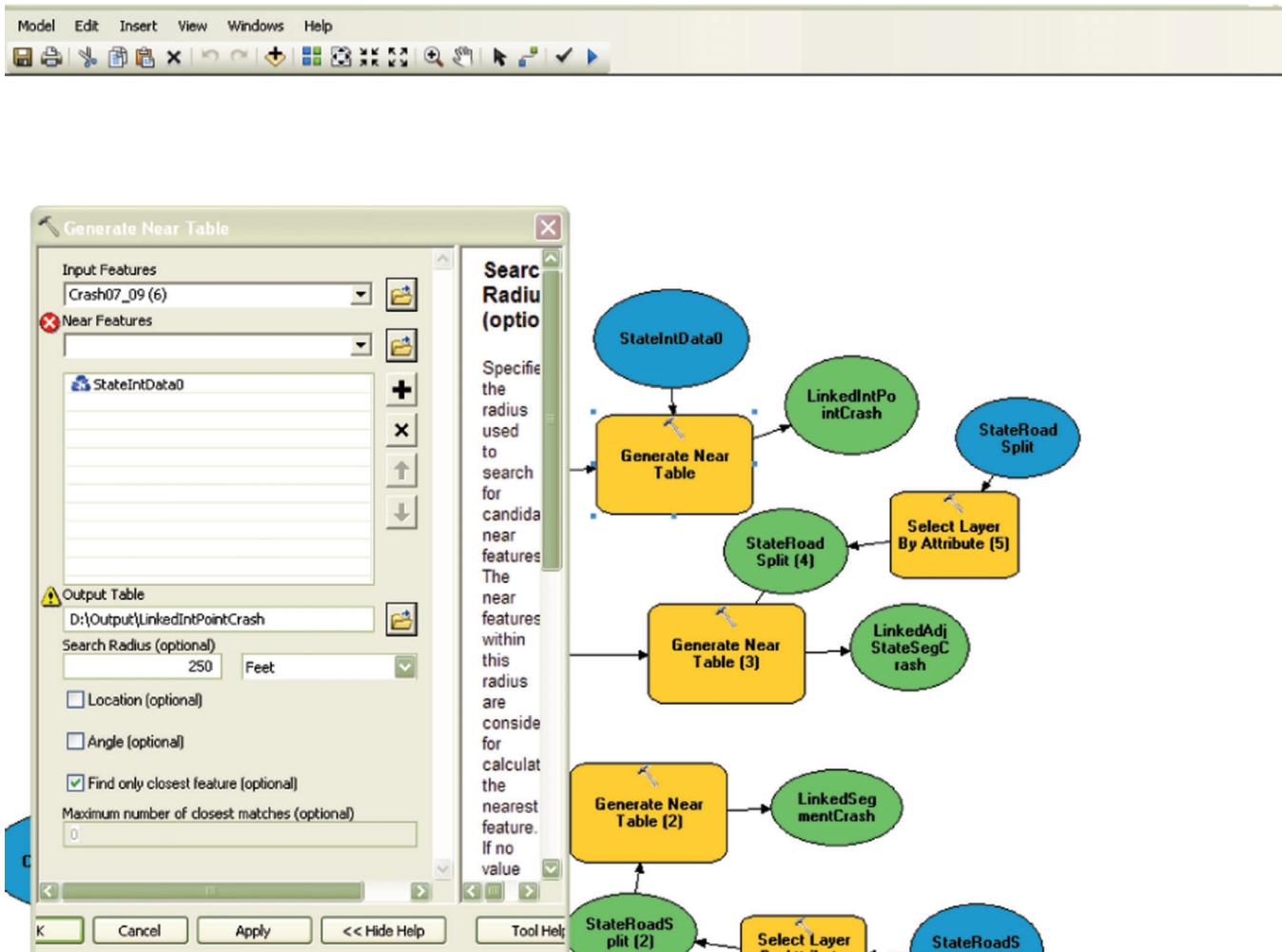
Figure B-43   Editing Generate Near Table Option in Crash Assignment Model Builder
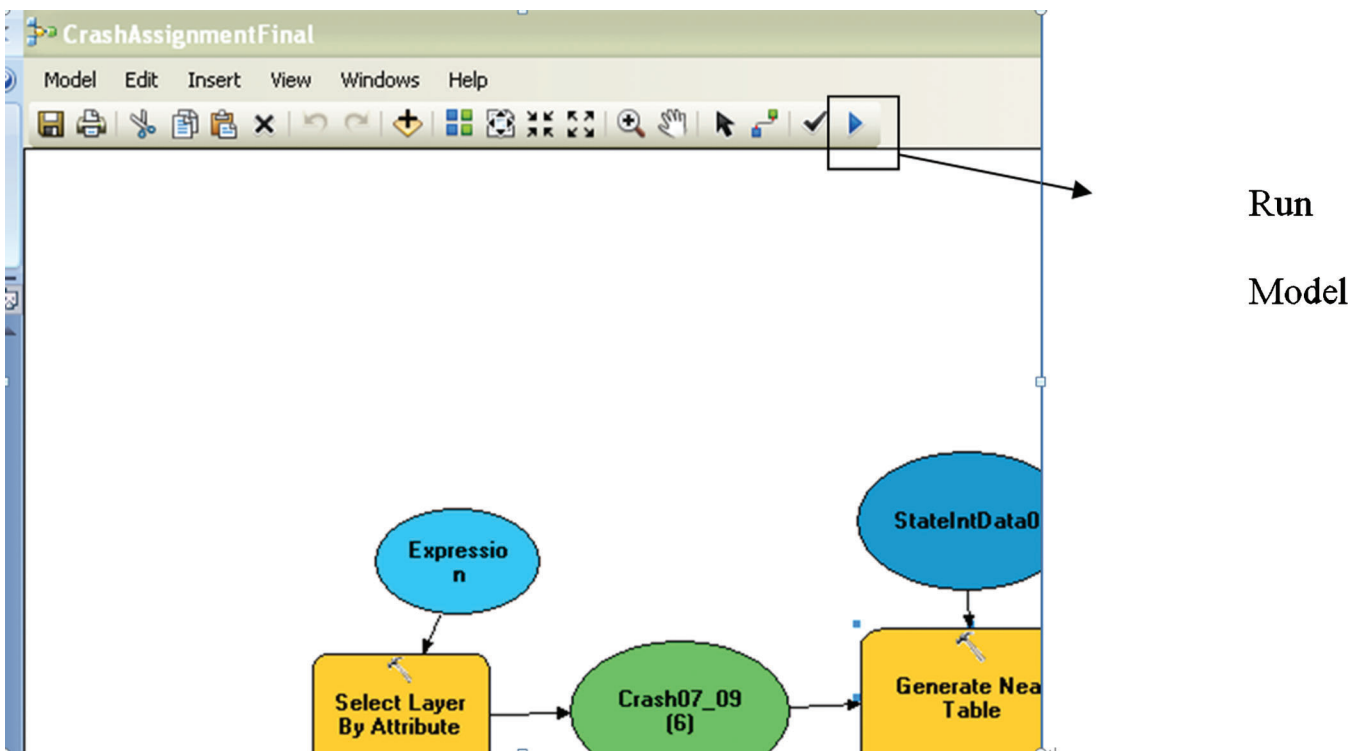
Figure B-44   Running Crash Selection Model Builder

UniqueLoca	year	NEAR_FID	NEAR_DIST
05US27NWINCHESTERST	2007	300963	20.11417
07PRIVATELANEAT4033NSALEMRRDSALEMRDRD	2007	-1	-1
12CR900NRDUS27	2007	-1	-1
07CR1100NCR200W	2007	342157	29.057804
052NDSTMONROEST	2007	341619	8.194714
07NHARRISONSTWWATERST	2007	300285	16.756429
12CR300US224E	2007	341619	8.194714
12CR500WRDSR218W	2007	-1	-1
05ADAMSSTUS27	2007	301015	4.499006
077THSTNUTTMANAVE	2007	341555	7.659227
12HARRISONSTMEIBERSST	2007	341590	22.94674
05NUTTMANAVEUS27	2007	301173	7.270593
12MARSHALLSTN3RDST	2007	341624	3.312051
08CLERMONTAVEPIEDMONTAVE	2007	341673	16.947732
12CR200WUS224W	2007	341762	4.144493
07ADAMSDRSTADIUMDR	2007	-1	-1
05CR400NRDUS27	2007	-1	-1
09PATTERSONSTUS27N	2007	301014	8.650735
12CR1150NRDCR200ERD	2007	-1	-1
0728THSTCR550N	2007	-1	-1
129THSTUS224E	2007	341534	16.97859
0513THSTADAMSST	2007	301015	15.78768
05CR450NUS27N	2007	300963	20.11417
05MONROESTUS27N	2007	301171	4.430368
051STSTMONROEST	2007	341621	13.117906
07CR850NRDUS27HWY	2007	301258	11.193401
12200NRD	2007	-1	-1
12CR000SCR500S	2007	300469	22.805623
052NDSTUS224E	2007	341619	8.194714
07EMONROESTPIQUARD	2007	341744	25.731941
06PIQUARDUS224E	2007	341663	9.074161

UniqueLoca	year	SegID	IntID	RampID	BridgeID
12US224HWY	2007	-1	-1	-1	-1
12US224HWY	2007	79022	-1	-1	-1
05US27NWINCHESTERST	2007	-1	34441	-1	-1
10JACKSONSTALYJACKSONSTALY	2007	-1	-1	-1	-1
12CR250WRDCR700SRD	2007	-1	-1	-1	-1
12S13THST	2007	49085	-1	-1	-1
12USHWY27HWY	2007	90704	-1	-1	-1
12CR300SR218	2007	-1	-1	-1	-1
07GRANTSTWINCHESTERST	2007	-1	-1	-1	-1
12650ERDSR124W	2007	-1	-1	-1	-1
12CR200US224	2007	493	-1	-1	-1
12N4THST	2007	82376	-1	-1	-1
0712THSTUS224E	2007	41870	-1	-1	-1
12PATTERSONST	2007	-1	-1	-1	-1
07PRIVATELANEAT4033NSALEMRRDSALEMRDRD	2007	-1	-1	-1	-1
12CR900NRDUS27	2007	-1	-1	-1	-1
12CR850SRDUS27S	2007	90670	-1	-1	-1
12CR700US27	2007	-1	-1	-1	-1
12POLKSTSR218	2007	-1	-1	-1	-1
12EMONROESTHARVESTERLN	2007	-1	-1	-1	-1
10CR850US27	2007	-1	-1	-1	-1
12S13THSTWINCHESTERST	2007	90278018	-1	-1	-1
12N17THST	2007	-1	-1	-1	-1
12BOLLMANST	2007	50286	-1	-1	-1
07CR1100NCR200W	2007	-1	-1	-1	-1
09CR300NCR300W	2007	-1	-1	-1	-1
12S13THST	2007	90735	-1	-1	-1
12US27NWINCHESTERRD	2007	90278018	-1	-1	-1
12CR200SCR500W	2007	-1	-1	-1	-1
12WINCHESTERST	2007	90278018	-1	-1	-1
12SALEMRDUS33N	2007	-1	-1	-1	-1

Figure B-45   Crash Assignment Results: (Left) - Output Joined with Crash Table, (Right) - Final Look of Crash Table

8. REFERENCES

1. IDOT. *A Systematic Identification of High Crash Location*. Iowa Department of Transportation Project TR-442. Center for Transportation Research and Education, Iowa State University, and Iowa Department of Transportation, 2001.

2. CRS and INDOT. *2008 Five Percent Report*. Indiana Department of Transportation and Center for Road Safety, Purdue University, 2008.

3. DDOT. *2007 Five Percent Report*. Delaware Department of Transportation, 2007.

4. Tarko, A. P., Iqbal, M. A., Inerowicz, M., Liang, H., Panicker, G. and Ramos J. *Safety Conscious Planning in Indiana: Predicting Safety Benefits in Corridor Studies, Volumes 1: Research Report, Volume 2: PASS and INPASS User Manual*. Publication FHWA/IN/JTRP-2006/21. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2007. DOI: 10.5703/1288284313366 and 10.5703/1288284313367.

5. Miller, J. S. *Geographic Information Systems – Unique Analytic Capabilities for the Traffic Safety Community*. Paper No. 00-0101. Transportation Research Record 1734, 2000.

6. Breyer, J. P. *Tools to Identify Safety Issues for Corridor Safety-Improvement Program*. Paper No. 00-0474. Transportation Research Record 1717, 2000.

7. Pulugurtha et al. New Methods to Identify and Rank Pedestrian Crash Zones: An Illustration. *Accident Analysis and Prevention*, Vol. 39, 2007, pp. 800–811.

8. Tarko, A., Sinha, K. C., and Farooq, O. Methodology for Identifying Highway Safety Problem Areas. Transportation Research Record, Transportation Research Board, No. 1542, 1996, pp. 49–53.

9. Kar, K., and Datta, T. K. *Development of a Safety Resource-Allocation Model in Michigan*. Transportation Research Record, No. 1865, 2004, pp. 64–71.

10. Norden, M. Orlansky, J., and Jacobs, H. *Application of Statistical Quality Control Technique to Analysis of Highway Accident Data*. HRB, No. 117. National Research Council, Washington D. C., 1956, pp. 17–32.

11. Rudy, B. M. *Operational Route Analysis*. Highway Research Record, No. 341. National Research Council, Washington D. C., 1962, pp. 1–29.

12. Morin, D. A. *Application of Statistical Concepts to Accident Data*. Highway Research Record 188. National Research Council, Washington D. C., 1967.

13. Nakmoon, S., Taylor, W. C., and Vincent, M. *Another Look at Identifying Hazardous Locations* (Based on negative binomial distribution). Proceedings of Transportation Research Board Annual Meeting, Washington D. C., 2003.

14. Jorgensen, N. O. *Statistical Detection of Accident Blackspots*. OTA-PIARC 11th International Study Week in Transportation and Safety, Brussels, 1972.

15. Abbess, C., Jarrett, D., and Wright, C. C. Accidents at Black Spots: Estimating the Effectiveness of Remedial Treat-ment with Special Reference to the Regression-to-Mean Effect. *Traffic Engineering and Control*, Vol. 10, 1981, pp. 535–542.

16. Maher, M. J. Fitting Probability Distributions to Accident Frequency Data. *Traffic Engineering and Control*, Vol. 28, No. 6, 1987. Pp. 356–357.

17. Maycock, G., and Maher, M. J. *Generalized Linear Models in Analysis of Road Accidents – Some Methodological Issues*. Conference Proceedings of Traffic Safety Theory and Research Methods, Session 4, Amsterdam, 1988.

18. Abdelwahab, H. T., Abdel-Aty, M. A. *Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accident at Signalized Intersection*. Transportation Research Record, No. 1746, 2001, pp. 6–13.

19. Delen, D., Sharda, R., and Besson, M. Identifying Significant Predictors of Injury Severity in Traffic Accidents Using a Series of Artificial Neural Networks. *Accident Analysis & Prevention*, Vol. 38, 2006, pp. 434–444.

20. Chong, M., Abraham, A., and Paprzcycki, M. *Traffic Accident Data Mining Using Machine Learning Paradigms*. Oklahoma State University Computer Science Department, 2004.

21. Tarko, A., and Azam, S. *Screening Road Networks for Hazardous Locations Addressing the Local Roads Issue*. AATT, Athens, Greece, 2008.

22. AASHTO. *Highway Safety Manual, First Edition*. American Association of State Highway and Transportation Officials, 2010.

JOINT TRANSPORTATION
RESEARCH PROGRAM

IDENTIFYING TRAFFIC SAFETY NEEDS:
A SYSTEMATIC APPROACH,
VOLUME 2: USER MANUAL

Andrew P. Tarko

Professor of Civil Engineering
Purdue University
Corresponding Author

Shafiul Md Azam

Graduate Research Assistant
School of Civil Engineering
Purdue University

Jose Thomaz

Data Analyst
Center for Road Safety
Purdue University

Mario Romero

Graduate Research Assistant
School of Civil Engineering
Purdue University

This page intentionally left blank.

RECOMMENDED CITATION

CORRESPONDING AUTHORS

Prof. Andrew P. Tarko
School of Civil Engineering
Purdue University
(765) 494-5027
tarko@purdue.edu

JOINT TRANSPORTATION RESEARCH PROGRAM

NOTICE

This page intentionally left blank.

PREFACE

The following User Manual describes a Safety Needs Identification Package (SNIP) for identifying roads and areas that experience and excessive number of crashes and may require special attention. The Manual includes four chapters. Chapters 1, 2, and 3 describe tool needed to perform the screening task and to present the results. Chapter 4 presents an example study with SNIP to illustrate its features. Additional information about the engineering basis of the operations described in this volume can be found in Volume 1: Identifying Traffic Safety Needs – A Systematic Approach – A Research Report. Volume 1, Appendix B, describes the data management procedures that must be performed annually or every other year to ensure that the data used by the SNIP are up to date.

The SNIP runs in the MS Windows XP/Vista/7 environment and requires the MS .NET Framework 3.5, MS Access (minimum 2007 version), and Google Earth. The MS .NET Framework is installed automatically during the first installation of SNIP if the computer does not have this component. An Internet connection must be turned on when installing the MS .NET Framework and when using the Google Earth.

To install the SNIP, the user should unzip the provided file into a folder of his/her choice, copy the **SNIP Files** folder to the MS Windows **My Documents** folder and run the installation process by clicking on the **setup.exe** file. The **readme.txt** file explains the installation steps.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 SAFETY NEEDS IDENTIFICATION PACKAGE - SNIP

The SNIP program has three main components:

1. Screening,
2. Clustering, and
3. Visualization.

1.1 Requirements

The SNIP is developed for MS Windows XP/Vista/7. To be able to run all SNIP components Microsoft Access 2007 or later, MS .Net Framework, and Google Earth must be installed. If Google Earth is not installed, it can be downloaded and installed at no charge from the following web page: http://www.google.com/earth/index.html. Furthermore, the installation of the MS .NET is executed by the SNIP installation application if the computer is connected to the Internet. An Internet connection is also needed to run the visualization tool via the Google Earth.

1.2 Installation

To install the SNIP, the user should unzip the provided file into a folder of his/her choice, copy the **SNIP Files** folder to the MS Windows **My Documents** folder, and run the installation process by clicking on the **setup.exe** file (Figure 1). The **readme.txt** file explains the installation steps.

Once you installed the program and made sure it is working, you may delete the unzipped files in the folder with the **setup.exe** file () to save disk space. Be sure to save the zipped/compressed file in case you need to reinstall the program.

DO NOT delete the files in the SNIP Files folder. Also, DO NOT delete the files in the folder to which the program was installed.

1.3 Running SNIP

The program can be run through the icon created by the installation process at the desktop or through the start menu were the SNIP program should be listed. The SNIP program allows you to access the three modules of the tool by selecting the appropriate tab.

Screening tool

The first tab allows running the screening tool by clicking on the "Start Screening" button (Figure 2). The screening tool identifies which road elements experience an excessive number of crashes. Chapter 2 explains how to use the screening tool.

Clustering tool

The clustering tool found under the second SNIP tab (Figure 3) allows clustering road elements identified with the screening tool. The clustering combines small road elements into larger parts of the road network based on the safety performance measures. The obtained clusters may be more suitable for certain safety interventions. Chapter 3 gives more details of the clustering tool.

Visualization tool

The third tab shown in Figure 4 allows creating KML files for visualization of results of the screening tool and of the clustering tool. KML is a file format used to display geographic data in a GIS-based browser such as Google Earth or Google Maps. Chapter 4 explains the visualization tool.
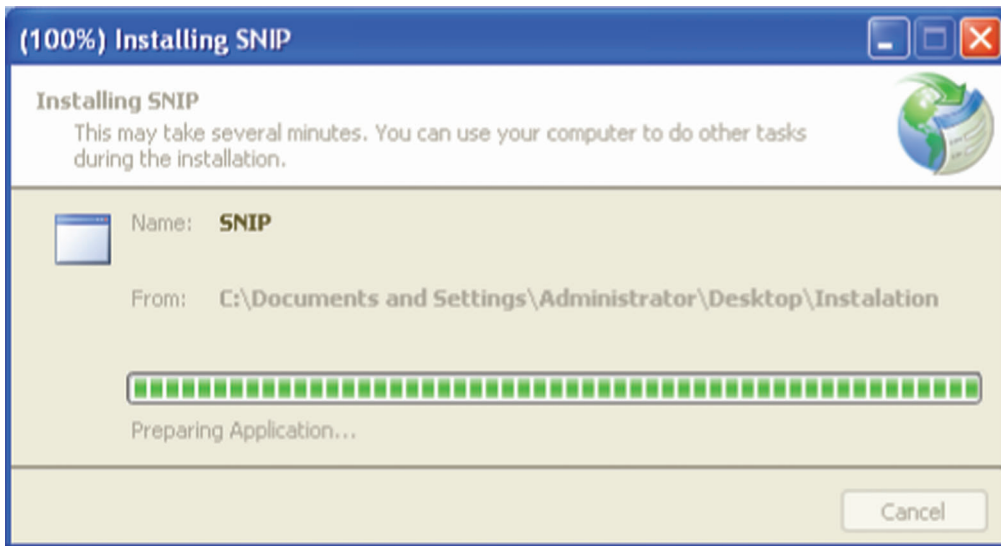

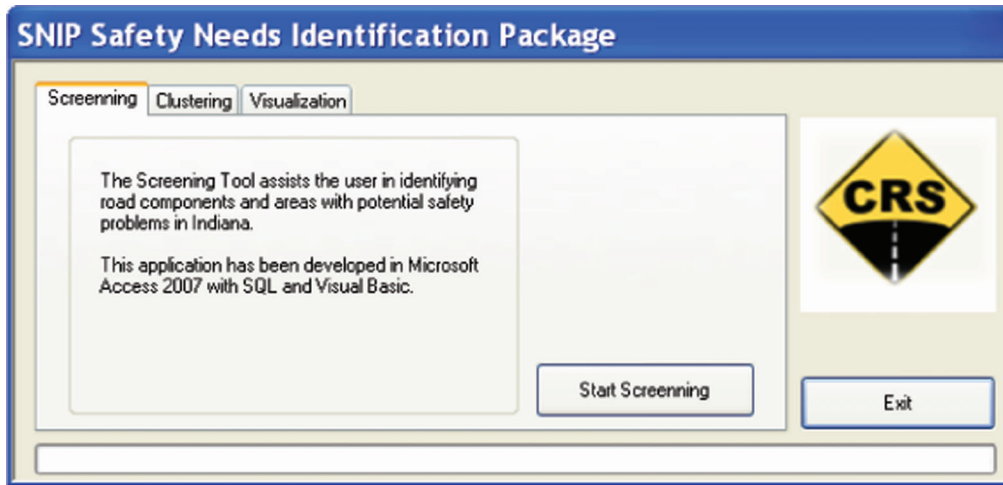Figure 1   SNIP Installation
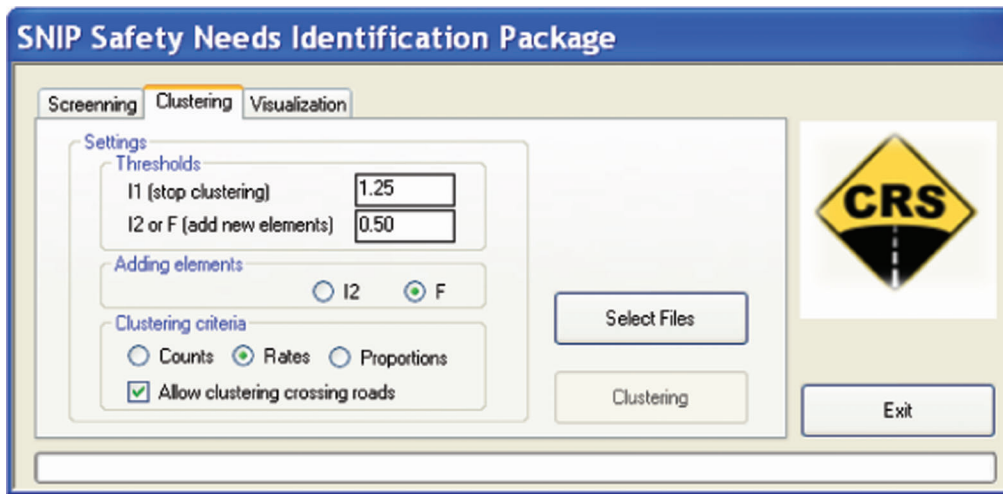
Figure 2   SNIP Screening Tab

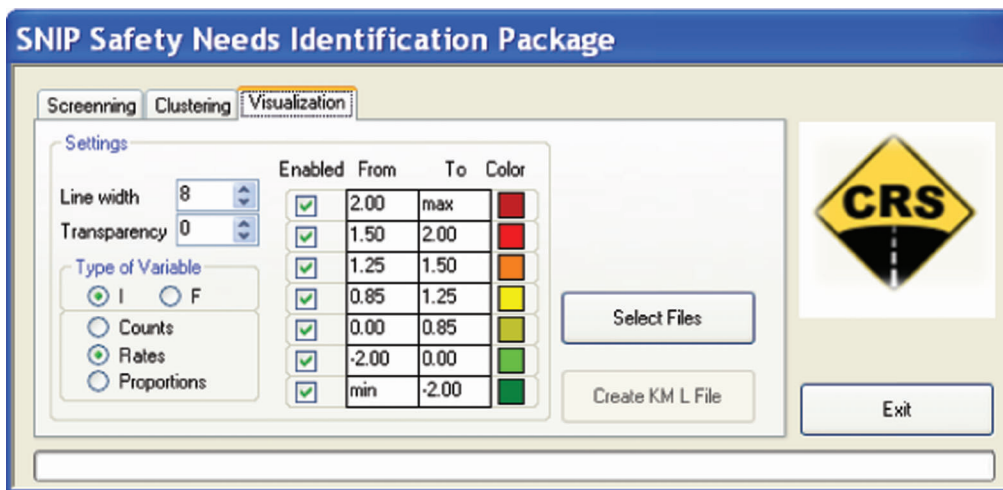Figure 3   SNIP Clustering Tab

Figure 4   SNIP Visualization Tab

2 SCREENING TOOL

2.1 Introduction

The Screening Tool assists the user in identifying areas and road elements (segments and intersections) with potential safety problems in the Indiana roadway system. The screening tool uses the Indiana crash data, a comprehensive inventory of the Indiana roadway system, AADT information, ramps, bridges, and census data. It was developed in Microsoft Access 2007, using SQL and Visual Basic for applications and is one of the components in a broader safety evaluation system.

The tool contains an interface to the query editor that helps the user build queries by offering a number of choices of crash and road element properties (Figure 5). The selection is then converted internally to SQL by the tool's screening engine and executed in Microsoft Access, to extract both crashes and road elements of interest. These road elements and their respective assigned crashes undergo a number of statistical analyses, and the final results are then presented to the user in tabular form. The results can then be used by the clustering and special studies module, or moved into the visualization tool.

To properly use the screening tool, the user should understand how the tool breaks down the screening process into the following steps:

1) Identify the scope of the problem
   The scope is the geographic unit where the user intends to conduct the screening. In the safety screening tool, four scopes have been defined; namely, state, county, township, and city. The scope can be limited to a particular county/township or multiple counties/townships, but it should always be greater than the elements in geographic extent.
2) Identify the family of elements that will be evaluated
   An element is the smallest unit of aggregation level for which the user would like to investigate safety performance. Elements can be of the infrastructure type, such

as segments, intersection points, intersections, ramps, or bridges, or can be a geographic area such as county and city within the scope. Therefore, the scope can be seen as a basket of elements, and an agency's main interest would be to investigate the safety performance of elements within the scope.

3) Specify the type of crashes of interest for the study
   After defining the scope and element of interest, it is important to define the crash type of interest. A crash type is defined by the corresponding crash selection criteria. For example, an agency might be interested in fatal or incapacitating injury crashes only or nighttime crashes only for some specific study. Another example is looking exclusively at the alcohol-related crashes for a targeted enforcement purpose. Crash selection criteria are dependent on the crash report variables and their availability.
4) Specify what conditions the elements of interest must satisfy
   Element (or road) selection criteria are also present in the tool. The Indiana road inventory is embedded in the master database, thus allowing a user to look at the crash propensity for specific design conditions. For example, users might investigate a particular roadway with a specific median type/width or intersections of a certain type with a specific number of legs. The combination of both crash and element selection criteria can be a great tool for choosing candidates for a specific program. Figure 6 shows the interaction among scope, element, and selection criteria in the overall safety screening process.
5) Select the metrics by which the elements will have their safety evaluated.
   After the user defines the scope, element, and selection criteria, it is important to define the way safety will be ascertained. There are three basic measures of safety available to the SNIP user:

   - *Crash Frequency:* Crash frequency represents the crash count of either the total crashes or a specific subset of crashes following some criteria determined by the user.
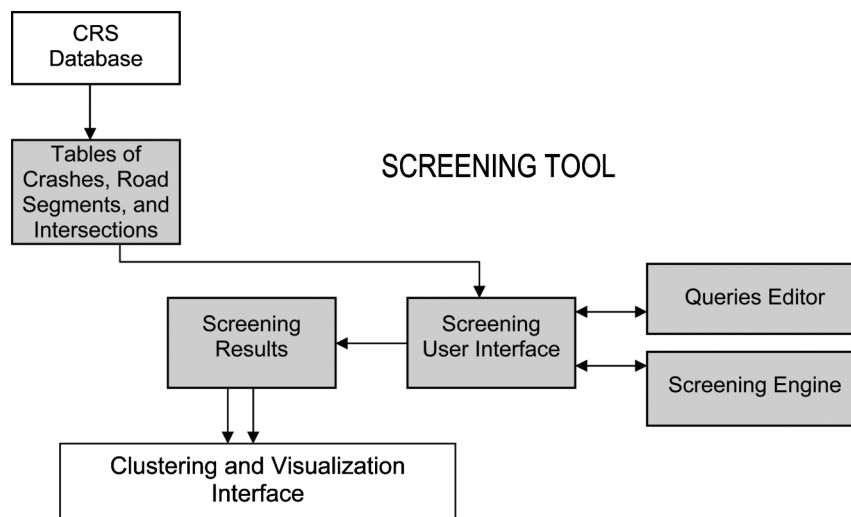

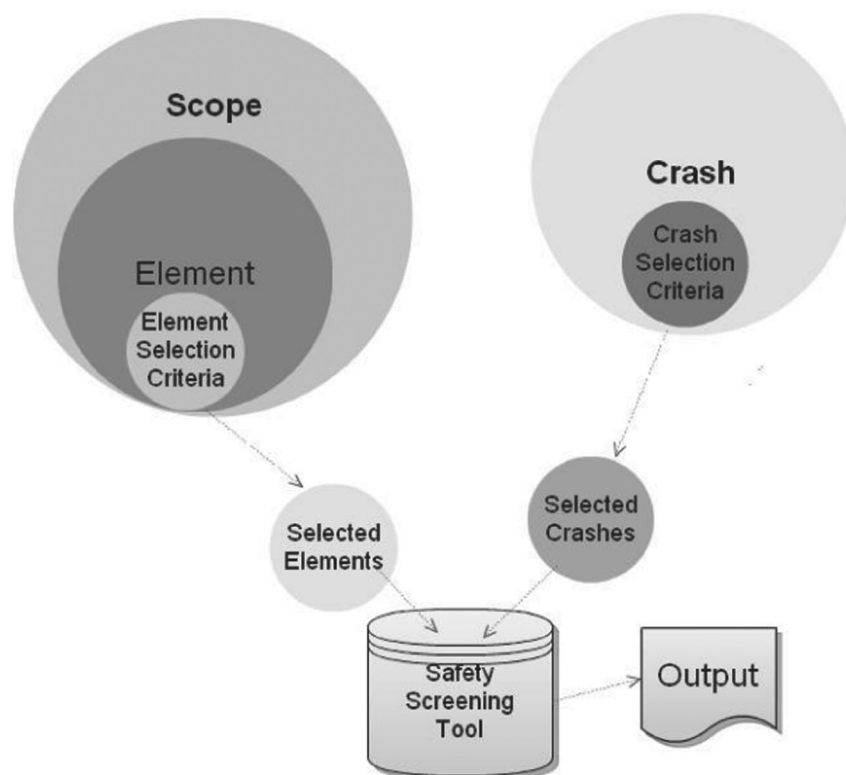Figure 5   Screening Tool Architecture

Figure 6   Scope, Elements, and Crashes, and their Selection Criteria

- *Crash Rate:* Crash rate is the frequency/exposure ratio. The measure of exposure can vary depending on the type of elements selected (Table 1).

- *Proportions of Crash:* Proportion is the ratio of two crash counts: (1) Studied crashes (numerator), (2) Reference crashes (denominator). The only condition is that the studied crashes are a subset of the reference crashes. For example, if the nighttime

crashes are studied, then all crashes (nighttime and daytime) could be a valid type of reference crashes.

6)   Evaluate the results.

The results of the queries are presented in a tabular form and they include a list of all elements matching the element selection criteria in the chosen scope, with their corresponding frequencies, rates, and/or proportions of the crashes that satisfy the crash selection criteria. The tables include basic information about the elements and their GIS coordinates, where applicable, to assist later visualization of the results. Measures of statistical confidence (confidence F and index I, as explained later) are included to determine if the element indeed experiences higher number of crashes than the average number in the scope. Table 2 displays the equations used in calculating the performance indicators used for each type of metrics available to the user.

Assuming that:

t = number of study years,
h = threshold crash rate set by policy,
m = number of crashes expected for the element under average safety,
v = variance of the m estimate.

Evaluated road element (segment, intersection, or area) during t years:

c = number of studied crashes,
r = number of reference crashes,
e = exposure.

TABLE 1
Exposure Measures **for Different Road Elements**

Element of Investigation	Exposure
County	Population, VMT*, Registered vehicles, Area
City	Population, VMT*, Registered vehicles, Area
Township	Population, VMT*, Registered vehicles, Area
State Segment	Daily VMT, Length
State-State Intersection	Intersection ADT
State-Local Intersection	State (major) ADT
Ramp	Daily VMT, Length
Bridge	Daily VMT, Length

Note * Note available in the current version

TABLE 2
Calculating m, a, F, and I for Each of the Three Safety Criteria

Safety Measure	Scope Mean	Over-dispersion	Confidence F	Index I
Crash Frequency	$m=s/n$	$v=\Sigma_i(c_i - m)^2/(n-1)$ $a=(v-m)/m^2$, $a \geq 0$	BetaDist$(1/(1+am), 1/a, c)$	$(c-m)/(c+am^2)^{1/2}$
Crash Rate	$m=es/w$	$a=1/s$	BetaDist$(1/(1+am), 1/a, c)$	$(c-m)/(c+am^2)^{1/2}$
Crash Proportion	$m=rs/d$	$a=2c/(rs)+1/r+1/s-3/d$	BetaDist$(1/(1+am), 1/a, c)$	$(c-m)/(c+am^2)^{1/2}$

All road elements in the query scope:
n = number of elements,
s = total number of studied crashes,
d = total number of reference crashes,
w = total exposure.

2.2 User Interface

The MS Access must be installed before the Screening Tool can be executed. The user may run the screening tool directly from the SNIP tab or may locate the **Standard** Screening Tool.accde **file in the** My Documents/SNIP Files **folder and double click on it to open. It will automatically bring up the MS Access in which the tool runs.**

When the user runs the screening tool for the first time, it is possible that MS Access display a warning alert saying that certain content in the database has been disabled (Figure 7). If that occurs, the user has the option to click on the options button in the warning and enable the content (Figure 8).

If the user enables the content as described above, the effect is temporary and only lasts during the execution of the program. Every time the user runs the tool, he/she will be greeted by the warning. To have a more permanent solution, the user could follow these steps:

1) Click on the Office Button, and then click on "Access Options" at the right bottom of the menu.
2) Click on Trust Center, select Trust Center Settings, and then click on Macro Settings.
3) Select "Enable all macros".

4) Click on the OK buttons until the user is back in Access. At this point the user should close the application and open the tool again. The Visual Basic code and initialization procedures will run immediately.

It is important to notice that macros are blocked by default to protect the user from running malicious code. If the user decides to enable all macros; he/she should only open databases from reliable sources. It is a trade between convenience and security.

The steps described in Section 2.1 are presented to the user in a tabbed interface. These tabs are arranged according to these steps. This type of layout allows the user to not only follow the sequence of steps in a natural fashion, but also to jump back to a particular previous step to make specific modifications before re-running the analysis. A description of the interface tabs and associated steps follow.

Steps 1 and 2 – Identification of Scope and Element

When the user opens the Screening Tool, it will initialize itself and all variables. After each analysis, the user is advised to press the "Reset analysis" button, to remove temporary tables and files and reinitialize the program.

The first tab takes the user to the page where the first two steps of the study are executed.

Identification of the Scope: Four options are included: State, County, Township, and City. All options, with the exception of State, will open a small window displaying a list of individual counties,
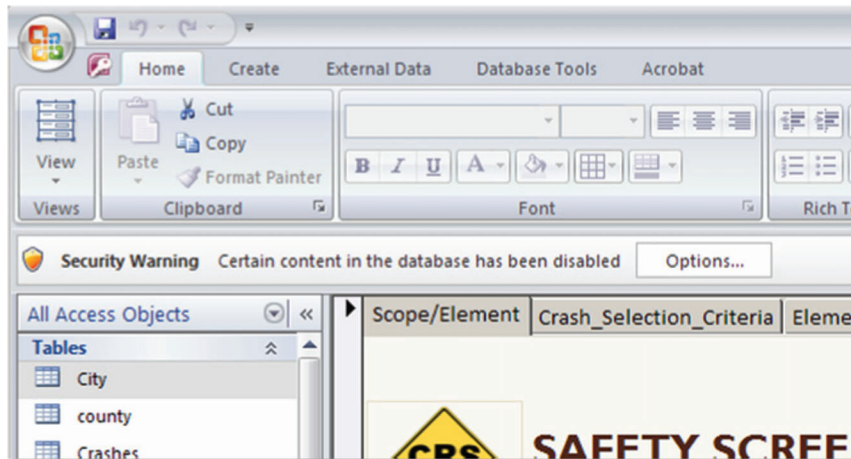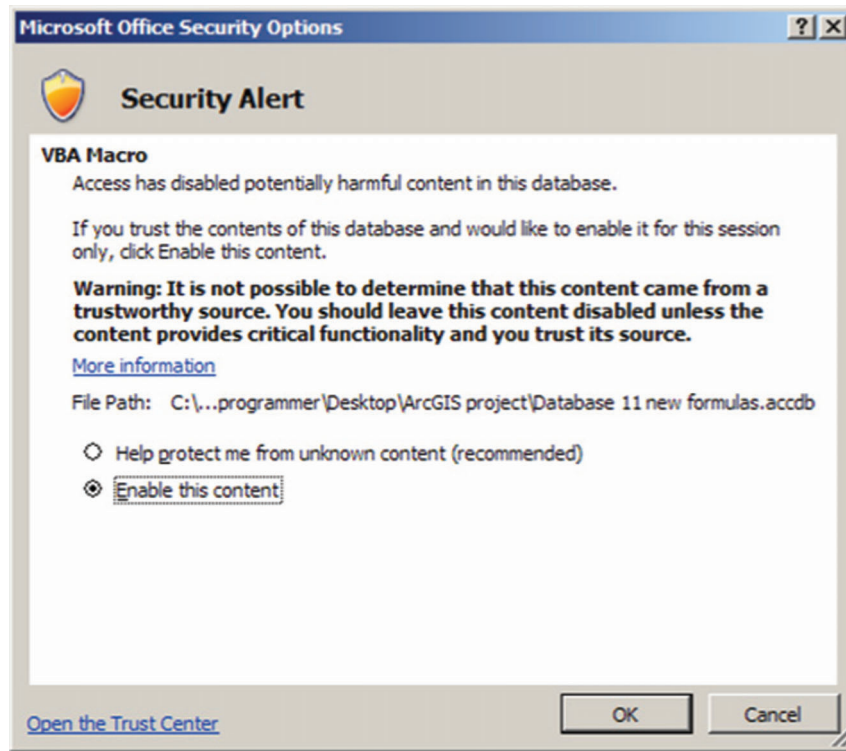

Figure 7   Macros Blocked Warning

Figure 8   Option to Enable Macros and VBA

townships, or cities for the user to select. This version of the tool only allows for the choice of one county, township, or city.

Identification of the Element: This tool offers different elements to be explored; four are of a geographical nature (state, county, township, and city)); and five are related to the transportation infrastructure (segments, intersections, intersection points, ramps, and bridges). Depending on the scope selected by the user, the ineligible elements are grayed out. For instance, if the user selected City for the scope, then the element options for State, County, and Township would be automatically unavailable to the user.

When an element is chosen by the user, the appropriate set of element selection criteria will also be highlighted in the element selection criteria tab.

Step 3 – Defining the Crash Selection Criteria

This tab permits the user to define what types of crashes are of interest for the study.

Presently, only a selected choice of commonly used crash selection criteria is offered to the user. The tool was developed in a way that allows for the easy addition of further criteria, if desired.

The user should select only one option in each of the seven framed categories presented. The option "All" is the default in all categories, meaning that no restrictions are being imposed in those categories. As the user selects different criteria, only the crashes that simultaneously satisfy all user-select criteria will be considered

in the safety study. For example, if the user marks "Truck Involved" and "Speeding," only crashes involving speeding and trucks will be selected for the analysis.

The "Time of Day" criteria let the user select up to three different time intervals when the crashes take place. If the user desires to specify time intervals of interest, values can be entered by selecting a time field by clicking on it, and then using either the mouse on the arrows or pressing the up/down keyboard keys. The time can be changed with a precision of minutes. After the desired time is displayed, clicking on it will change the time color to black, meaning it has been set. Fields where the "To" time period is left as 0:00 will be ignored. Also, the user should always start entering time intervals using the top two fields and moving downwards, as more time intervals are desired.

After the criteria for selection is established, the user should press the "Extract Crash Data" button. Once the extraction is completed, the button "View Extracted Crashes" will allow the user to view a table containing the crash data that satisfy the criteria.

Step 4 – Defining the Element Selection Criteria

This page permits the user to select specific conditions for the infrastructure elements being considered in the study. It follows the same concepts introduced in the Crash Selection Criteria tab.

As the content of this page serves different groups of elements, the group of constraints appropriate for the

Figure 9   Scope/Element Tab

Figure 10   Crash Selection Criteria Tab

Figure 11   Element Selection Criteria

elements selected in the "Scope/Element Tab" is high-lighted in blue. The constraints that are not applicable to the chosen elements will not be highlighted. Figure 11 illustrates the case where **segments** were chosen by the user. The available selection criteria for segments are framed in blue, whereas the ones with intersections and intersection points are not.

Similar to the "Crash Selection Criteria" page, the value "All" is the default for all selection categories, indicating that no restriction is being applied. Presently, the user can select only one value for each framed criteria group, and only elements that satisfy all user selected criteria will be extracted for analysis.

After the selection criteria are properly marked, the user should click on the "Extract Element Data" button. Like before, once the extraction is completed, the button "View Extracted Elements" will allow the user to view a table containing the element data that satisfy the criteria.

Step 5 – Selecting the Metrics – Frequencies, Rates, and Proportion Tabs

The Metrics tab takes the user to the selection of the safety performance measures. Frequencies, the first evaluation metric, are always calculated automatically, so there is no need for any user input for this metric. The program always calculates the annual frequency of crashes that satisfy the user- defined crash selection criteria for each element from the chosen scope that

satisfy the element selection criteria defined by the user. A measure of statistical confidence for the frequency is also indicated.

The Metrics tab takes the user directly to the second evaluation metric. The page will automatically present the user with the exposure measures available for the type of element chosen by the user in the "Scope/Element" tab. The inappropriate choices will be grayed out.

Similar to the Frequency tab, the program will calculate, for each element in the selected scope and satisfying the element selection criteria, the average annual ratio between the counts of crashes satisfying the crash selection criteria for that element and the chosen measure of exposure for those elements.

Elements with missing measures of exposure will have their rates left blank. A measure of statistical confidence for the rates is also indicated.

Inside the proportions section of the Metrics tab, the SNIP presents the user with a combined copy of the Crash Selection Criteria. This time, each crash criteria category is presented to the user only two possible values. The user has a choice to leave the chosen options as it is or select the option "All," in which case the reference crashes are expanded to all cases and no longer restricted to the previously selected category. Every time the user changes one of the previously restrictive criteria to "All," the number of reference crashes may increase. The frequency of crashes in this expanded reference crash type is the denominator in the crash proportion. In the example shown in Figure 14,

Figure 12   Metrics

Figure 13   Metrics – Rates

Figure 14 (Metrics - Proportions form)

Vehicles Pedestrians or Bicyclists Involved
- ⦿ All Cases
- ○ Bus Involved
- ○ Truck Involved
- ○ Motorcycle Involved
- ○ Pass. Vehicles Only Involved
- ○ Alcohol Involved
- ○ Pedestrian Involved
- ○ Bicyclists Involved

Severity
- ⦿ All
- ○ Fatal
- ○ Fatal + Incapacitating
- ○ Incapacitating Only
- ○ Not Fatal nor Incapacitating

Contributing Factor
- ⦿ All
- ○ Speeding
- ○ Driver Innatention
- ○ Aggressive Driving
- ○ Improper Lane Use

Collision Type
- ⦿ All
- ○ Frontal Crash
- ○ Side Swipe
- ○ Rear Ending
- ○ Roll Over

Light Conditions
- ⦿ All
- ○ Daylight
- ○ Down / Dusk
- ○ Dark / Lights on
- ○ Dark / Lights off

Days of Week
- ⦿ All
- ○ Weekend
- ○ Weekdays

Time of Day
- ⦿ All
- ○ Selection Below

From	To
0:00	0:00
or	
0:00	0:00
or	
0:00	0:00

Figure 14   Metrics - Proportions

Figure 15 (Output Tables)

Tabs: Scope/Element | Crash_Selection_Criteria | Element_Selection_Criteria | Metrics | Results

[Start Screening Calculations] [Export Results to Access File]

Scope: County -ALLEN   **Exposure:** VMT
Element: Segments

County	Township	City	Segment_Name	Segment_ID	Segment_Type	Length_Ft	VMT	X_Start	Y_Start	X_End
ALLEN	St. Joseph	Fort Wayne	02-S-930-0-01	40627005	Local Road	524.9		-85.10837	41.112048	-85.1071869
ALLEN	Washington	Fort Wayne	02-S-930-0-01	837	Urban Multilane	467.0	1445261	-85.1301275658	41.1182153606	-85.12
ALLEN	Washington		02-U-030-0-01	145161	Rural Multilane	516.6	77	-85.2131471172	41.125273064	-85.211
ALLEN	Eel River		McDuffee	64387384	Local Road	524.9		-85.297485	41.195249	-85.2973256
ALLEN	Washington		02-S-003-0-01	659	Urban Multilane	565.0	1358992	-85.1655570791	41.156697428	-85.165
ALLEN	Washington	Fort Wayne	N/A	80179633	Local Road	524.9		-85.181171	41.117748	-85.1798938
ALLEN	Wayne	Fort Wayne	N/A	77956	Local Road	386.1		-85.13597	41.055048	-85.13
ALLEN	Wayne	Fort Wayne	N/A	81630	Local Road	551.3		-85.14007	41.087748	-85.14
ALLEN	Wayne	Fort Wayne	N/A	81603	Local Road	82.7		-85.13187	41.046048	-85.13
ALLEN	Washington		02-S-003-0-01	105144	Rural Multilane	419.0	762871	-85.165972	41.175848	-85.1659720
ALLEN	Aboite		N/A	81495047	Local Road	524.9		-85.236402712	41.0730062771	-85.236
ALLEN	Eel River		02-U-033-0-01	105633	Rural Two Lane	628.3	501820	-85.2906217417	41.1826949393	-85.289
ALLEN	Cedar Creek		02-S-001-0-02	620	Urban TwoLane	240.3	172265	-85.07188	41.18093	-85.072
ALLEN	St. Joseph		N/A	78020	Local Road	116.0		-85.076269	41.178048	-85.076
ALLEN	Washington	Fort Wayne	02-S-930-0-01	105385	Urban Multilane	469.0	1189641	-85.1439714092	41.118108737	-85.142
ALLEN	Jefferson		02-U-030-0-01	1003	Rural Multilane	533.1	755602	-84.9104885134	41.0335708371	-84.912
ALLEN	Washington	Fort Wayne	02-S-930-0-01	105383	Urban Multilane	490.1	1243154	-85.1369488798	41.1182480138	-85.13
ALLEN	Aboite		02-U-024-0-01	875	Urban Multilane	464.5	602691	-85.2658911862	41.0243641692	-85.264
ALLEN	Cedar Creek	Leo-Cedarville	02-S-001-0-02	610	Rural Two Lane	195.0	145391	-85.021368	41.203349	-85.0220593
ALLEN	Perry	Huntertown	02-S-003-0-01	105116	Rural Multilane	492.1	476375	-85.1675720004	41.2358987773	-85.167
ALLEN	Springfield		02-S-037-0-01	105197	Rural Two Lane	497.0	241505	-84.927105	41.192273	-84.9285233
ALLEN	Monroe		02-S-101-0-01	105243	Rural Two Lane	528.5	50396	-84.862236	40.950044	-84.8621930
ALLEN	St. Joseph	Fort Wayne	02-S-930-0-01	105356	Urban Multilane	532.0	1135477	-85.089969	41.105048	-85.0896491
ALLEN	Washington		N/A	80209653	Local Road	524.9		-85.163964655	41.168578541	-85.165
ALLEN	Adams	New Haven	02-S-930-0-01	821	Urban Multilane	345.1	695756	-85.058668	41.069548	-85.0599180
ALLEN	Washington	Fort Wayne	02-S-003-0-01	105146	Urban Multilane	586.8	1411415	-85.163643	41.144265	-85.1631226
ALLEN	Jefferson		02-U-024-0-01	853	Rural Two Lane	649.5	497401	-84.985946	41.088669	-84.9842856

Record: I ◀ 1 of 3117 ▶ ▶I ▶⁕   No Filter   Search

Figure 15   Output Tables

the user has selected "only fatal and incapacitating injury" crashes with "truck involvement." The user is offered only the opportunity to change any of these two selections to the value of "All."

After the new selections take place, the user should click on the "Extract Reference Crashes" button. Like before, once the extraction is completed, the button "View Reference Crashes" will allow the user to view a table containing the reference crash data that satisfy the criteria.

Step 6 – Evaluation of Results

Arriving at the Results Tab, the user should press the button "Start Screening Calculations." At this point, all the three evaluation metrics will be calculated, as well as the indicators for their confidence level. This is also where the user finds the results of the screening process. The output table includes detailed information about all elements satisfying the element selection criteria in the scope.

The user is also able to find how many of the studied crashes have been assigned to each element, as well as the safety performance values and the corresponding confidence levels.

The element properties included in the results vary with the type of element chosen. In the case of geographic areas selected as elements, such as county, township, or city, information about the population, area, and number of registered vehicles are included. In the case of infrastructure elements such as road segments, intersection points, intersections, ramps, and bridges, the latitude and longitude coordinates will be provided. This information will permit the table to be used as input for visualization with ArcGIS or Google Earth. By pressing the "Export Results to Access" button, the user will be able to save the results to a file that can be imported to be used by the clustering and visualization tools.

3 CLUSTERING TOOL
3.1 Introduction

The screening tool identifies which road elements experience an excessive number of crashes; clustering these elements into longer road sections may reveal useful spatial regularities, that otherwise may go unnoticed, and may be helpful to INDOT engineers in scoping corridor improvement projects and other safety-oriented programs.

It is important to note that elements with safety needs should be clustered based on the safety performance measures in order to obtain road clusters relevant from the safety management point of view.

There are three basic safety measures that can be used to identify road elements with an excessive number of crashes of a certain category: crash frequency, crash rate, and crash proportion. These measures were introduced and discussed earlier in this report. Crashes are subject to a strong random fluctuation over time and two safety performance indices:

Confidence F and Index I, are proposed to estimate the level of statistical confidence that the detected excessive number of crashes indicates a systematic issue rather than just the effect of random fluctuation.

Confidence F is the probability of a safety level equal to or better than the one observed during the period of analysis if the expected safety level in a long run is average for the type of location and under the given exposure. The higher the confidence F is, the stronger is the evidence that the location experiences a real safety problem. The values of F=0.90 and higher are typically used.

Index I is the difference between the safety observed during the period of analysis and the safety expected given the location type and exposure, divided by the standard deviation of the difference estimate. It is a simplified measure of Confidence F. The values I=1.5 and higher provide sufficient evidence that the location experiences a real safety problem.

Confidence F is calculated as BetaDist(1/(1+am), 1/a, c), while Index I is calculated as $(c-m)/(c+am^2)^{1/2}$, where a is the over-dispersion parameter, m is the average crash count in a long run, and c is the actual crash count in the period of analysis.

One of the important operations of clustering road elements is evaluation of the safety level in current clusters to be able to claim that the obtained clusters experience excessive numbers of crashes. A practical method of updating safety evaluation in clusters is aggregation of safety measures. The exact method, based on Confidence F, is statistically and computationally troublesome because summation of Gamma variables does not yield a Gamma variable, and the convenient equivalency between Negative Binomial and Beta distribution cannot be used. That is why the Index I is calculated, which is easy to update for clusters. The following equation is used to calculate the Index I for clusters of multiple road elements:

$$I = \frac{\sum_i (c-m)_i}{\sqrt{\sum_i (c+am^2)_i}}$$

where values of $(c-m)_i$ and $(c+am^2)_i$ are known for any road element i and they are summed up for clusters. The clustering algorithm is shown in Figure 16. It is important to note that the clustering process is controlled by two user-selected threshold values: I_1 and I_2. The recommended ranges are: (1.25-2) for I_1 (0-1.25) for I_2 with the recommendation that $I_1 > I_2$. The user can restrict the clusters building only along the same routes to follow the common practice in scoping road projects. Other restrictions may be added to the algorithm as needed. A list of clusters and their elements is obtained based on the screening results, the network topology, and the parameters set by the user.

The clustering tool's requirements are presented below; also, a description of the interface is shown. It includes the data importation process, the user settings, clustering calculation, and results.
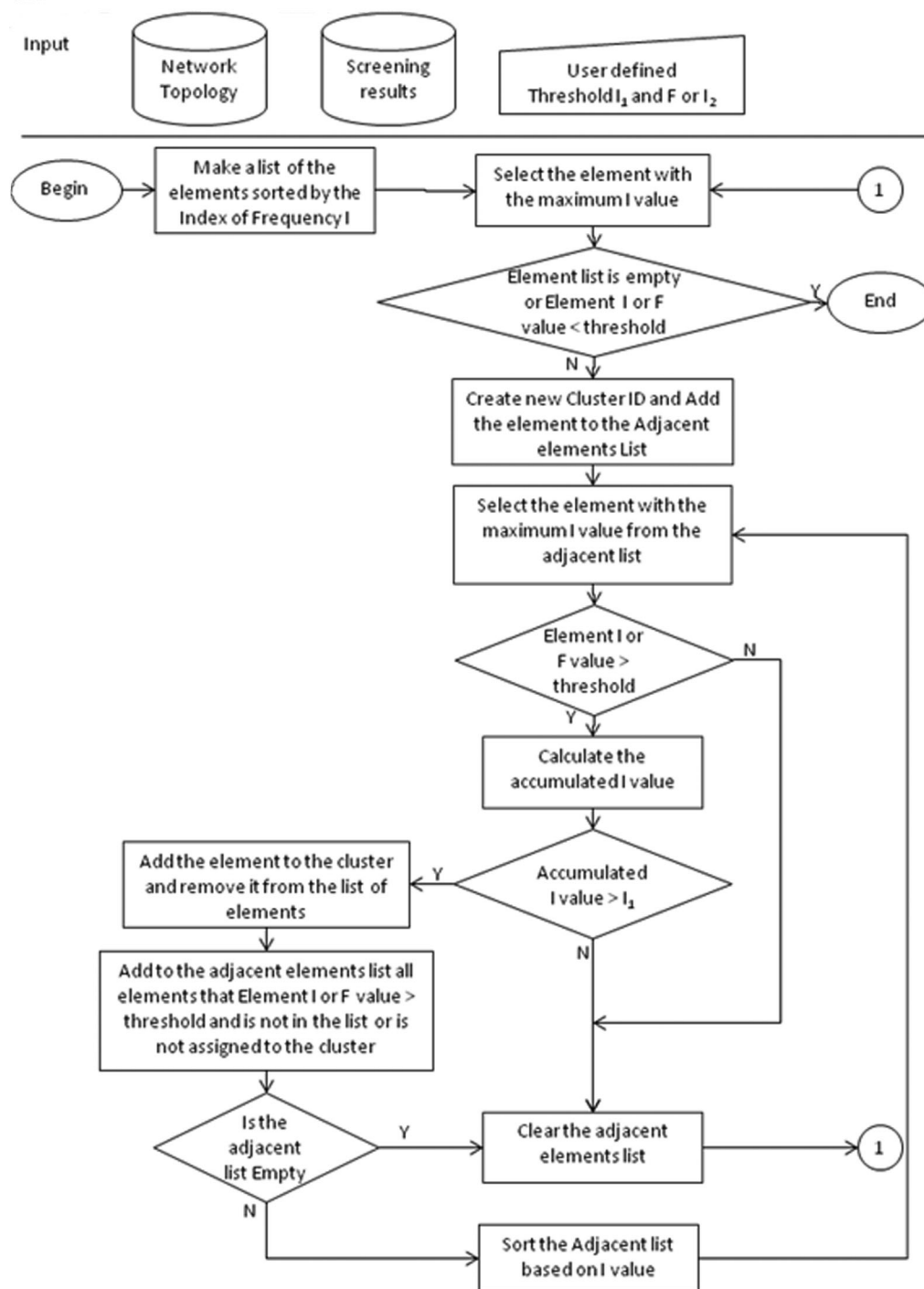
Figure 16   Clustering Algorithm Flowchart

3.2 Running the Tool

A description of each necessary step to run the process is shown below, which includes the user settings, the open data files, the clustering calculation, and the results.

Settings

The clustering process is controlled by two user-selected threshold values: index I1 and index I2 or threshold confidence F. I1 is used to control the end of the process and is the minimum I value for the final
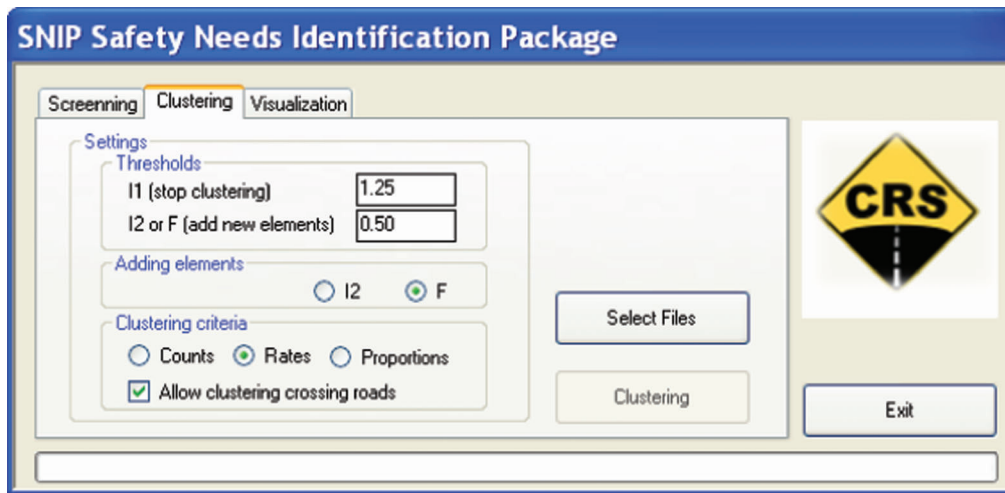
Figure 17   Settings

cluster. I2 and threshold F are used to control adding elements. The user can select either I2 or threshold F to control adding elements. The algorithm allows adding elements that have I values greater than threshold I2 or Confidence F value greater than threshold F. The recommended ranges are: (1.25-2) for I1; (0.5-0.7) for F; (0-1.25) for I2 with the recommendation that I1>I2.

Thus, before running the clustering tool, there are three basic settings needed: setting threshold I1, selecting new element criteria (I2 or threshold F), and allowing (or not) clustering crossing roads (Figure 17).

The user must also select the clustering variable criteria selecting counts, rates or proportions and allowing (or restricting) the clusters building along the same routes to follow the common practice in scoping road projects by selecting (deselecting) the "Allow Clustering Crossing Roads" check box. Allowing this option may lead to clusters that "turn" at an intersection to another state-administered route.

Selecting Files

Input to the clustering tool is included in the screening tool results. One or more files with the results must be selected. Clicking on the "Select Files" button shown in Figure 18 opens the Windows file selection window where the appropriate files can be selected. It is important to note that it is necessary to open all the at once by highlighting these using the control key and pressing button Open.

Clustering

Once the settings and the data opening are done, the "Clustering" button is enabled. By clicking on it, the clustering process begins. It is important to note that this process may take several minutes. After it ends, the user is asked to provide a file name and a folder to save the results.

At the end of the clustering process a window opens with a summary report that shows the total number of clusters found, the total segments included, and the total intersection points included, as shown in Figure 19.

Results Table

The clustering process creates two datasets (tables): the Segment table and the Intersection Points table.

Each table contains the Element ID and the Cluster ID (ID of the cluster to which this element belongs). These tables are used in the visualization process as explained in Chapter 4.

4 VISUALIZATION OF THE RESULTS

The user has two options of visualization of the results:

1. The visualization tool embedded in the SNIP package, or
2. External tools such as the ones provided by the ArcGIS.

Both the methods are explained in this chapter.

4.1 SNIP Visualization Tool

The results presentation is the final step of the Safety Needs Identification Package. In this phase, the user can visualize the results obtained in the screening or clustering steps. This SNIP module is shown in Figure 20 allows creating KML files for both the screening and the clustering tools. KML is a file format used to display geographic data in an Earth browser such as Google Earth or Google Maps.

Settings

Before running the clustering tool, there are four basic setting needed: color code settings: line width, transparency and variable using for color coding.
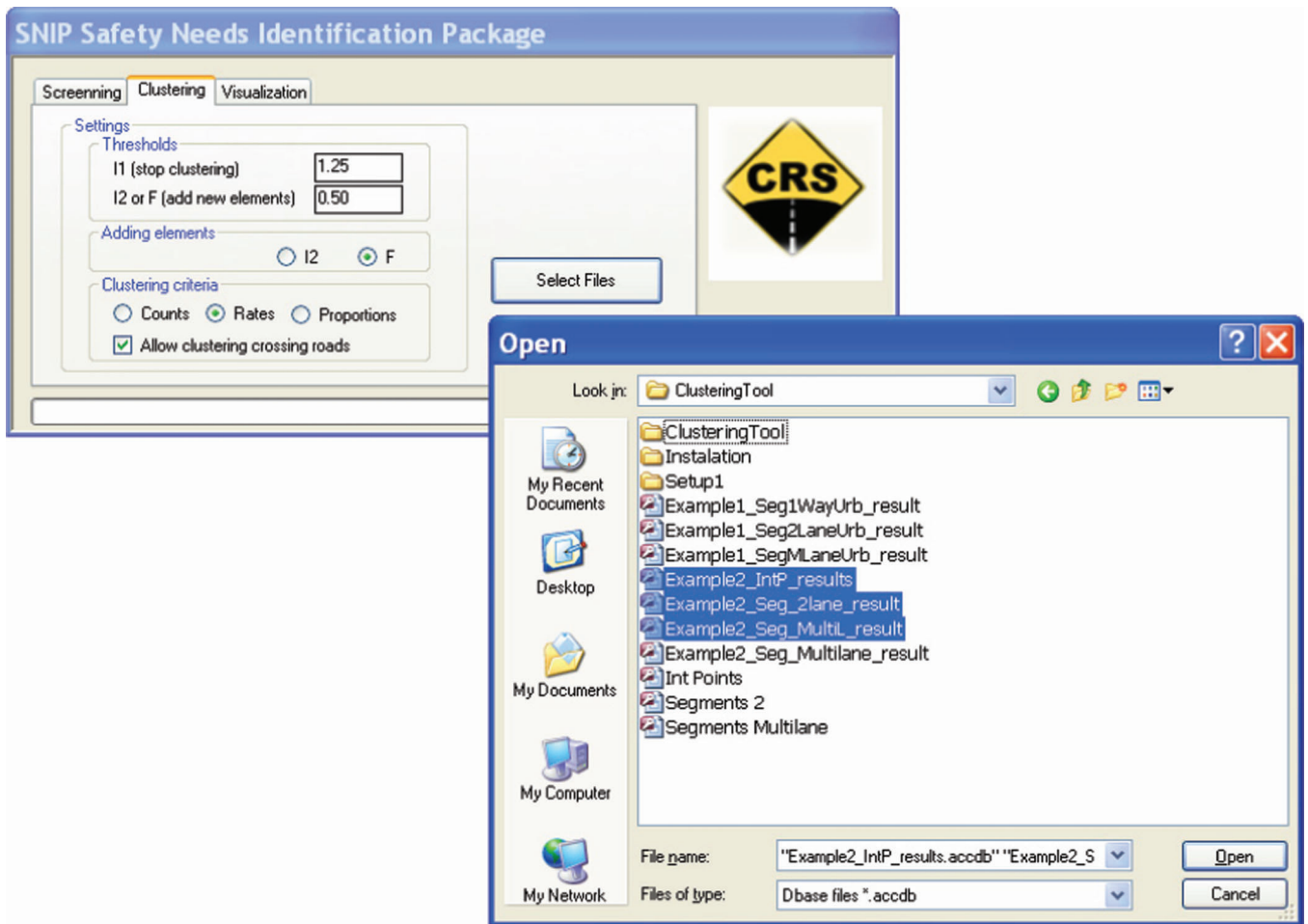
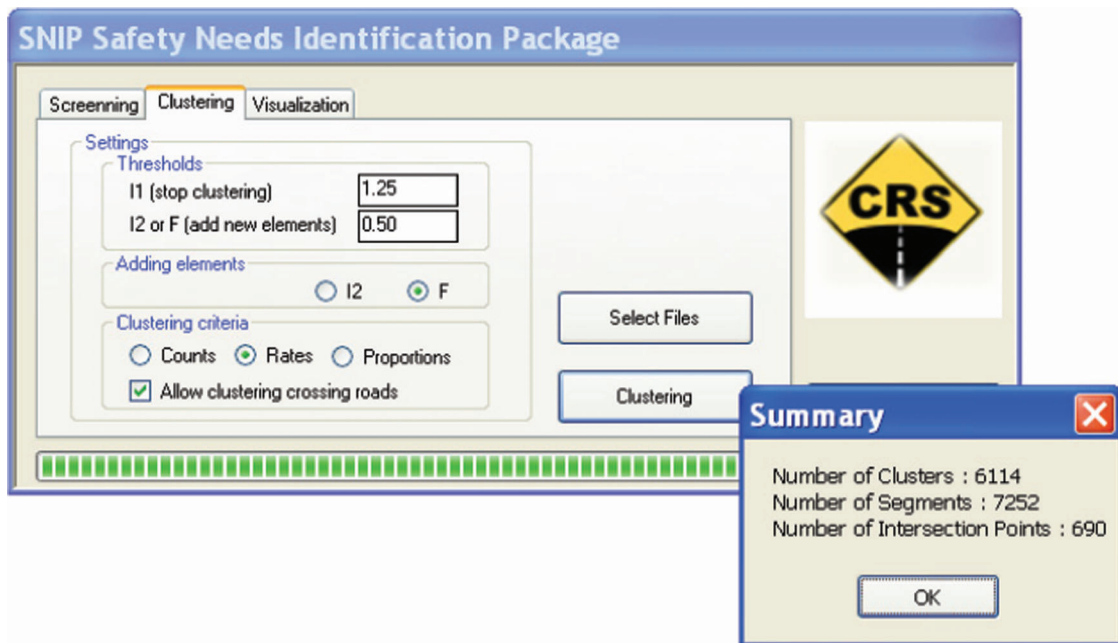Figure 18   Data selecting

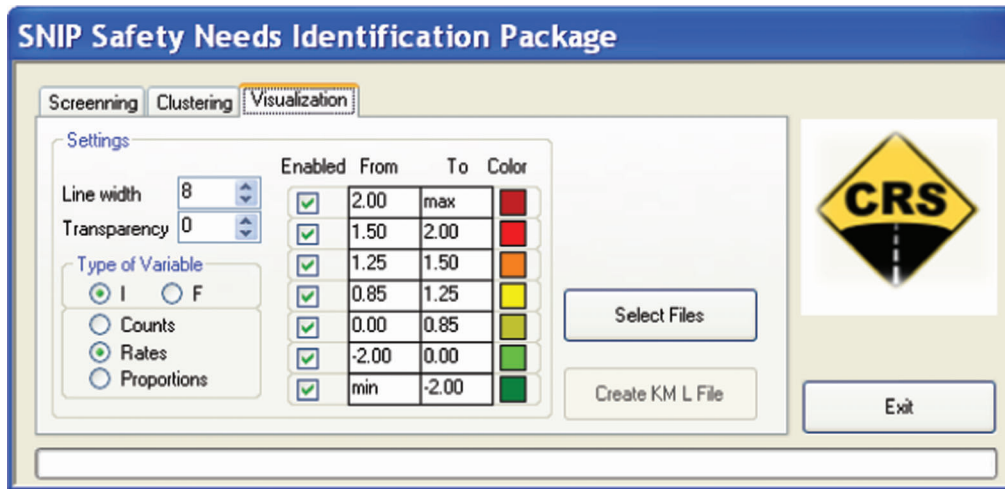Figure 19   Clustering Summary

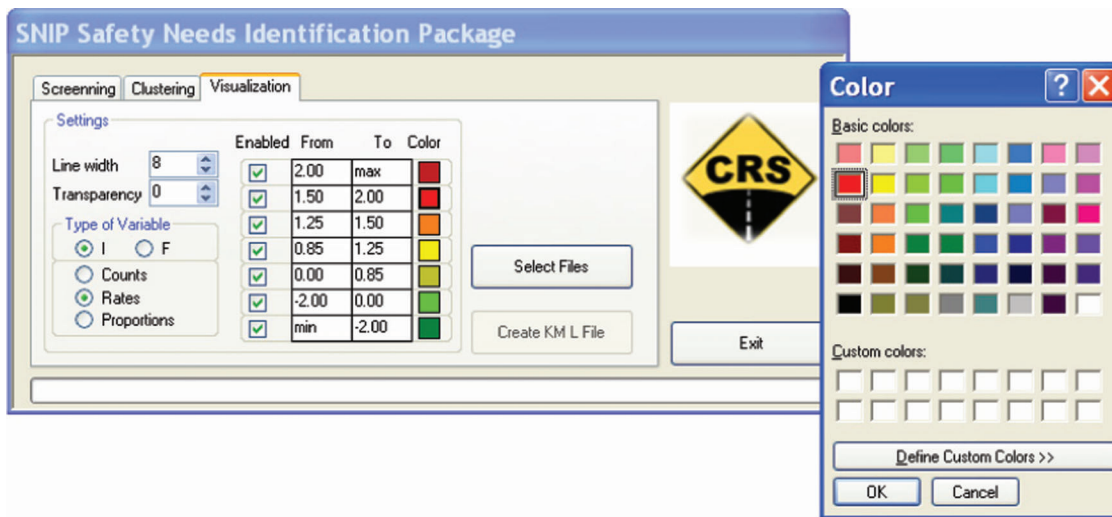Figure 20   SNIP Visualization Tool

Figure 21   Changing Color Code

The display settings include:

- Width of the displayed lane
- Level of transparency
- Criterion for adding elements to a cluster (I2 or threshold F)
- Safety performance measure (frequency, rate, proportion)
- Enabling/disabling ranges of values
- Minimum and maximum values of the ranges
- Color displayed for the range

The user should be aware that changing between I and F displays the default ranges of values, thus any user-entered values will be wiped out.

Select files

Clicking on the "Select Files" button shown in Figure 22 opens the Windows file selection window where the user may select the files with the results for display. Unlike in the case of clustering, where user has to open all the files at once, this time - the files for visualization can be converted to the KML format one by one (multiple KML files created) or all at once (one combined KML file created). Even in the case of multiple files, the user can still displayed all the created KML files in Google Earth simultaneously.

Creating KML File

Clicking on the "Create KML File" button opens the Windows standard creating file window where you should set the name of the new file to be created and also select the folder to save it (Figure 23).

KML is a file format used to display geographic data in a GIS browser such as Google Earth or Google Maps. To display the result Google Earth must be
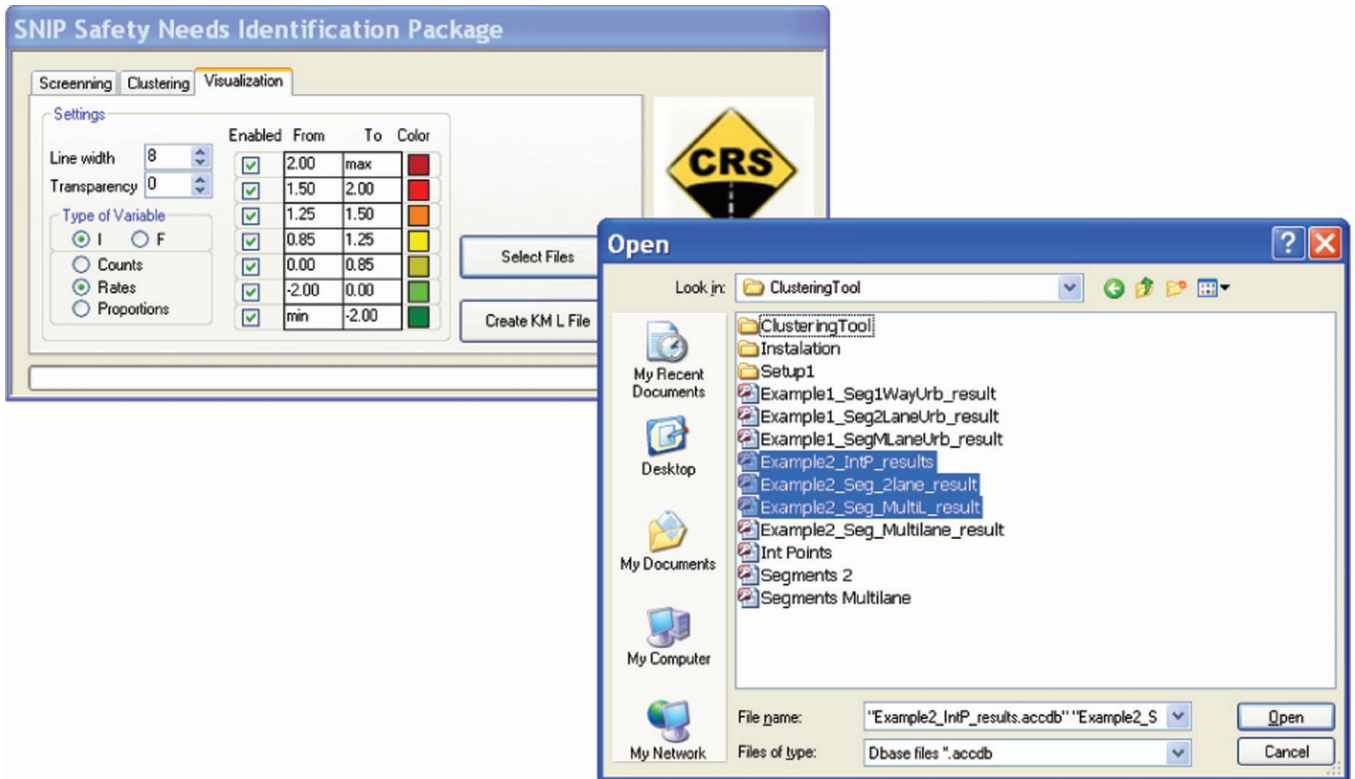
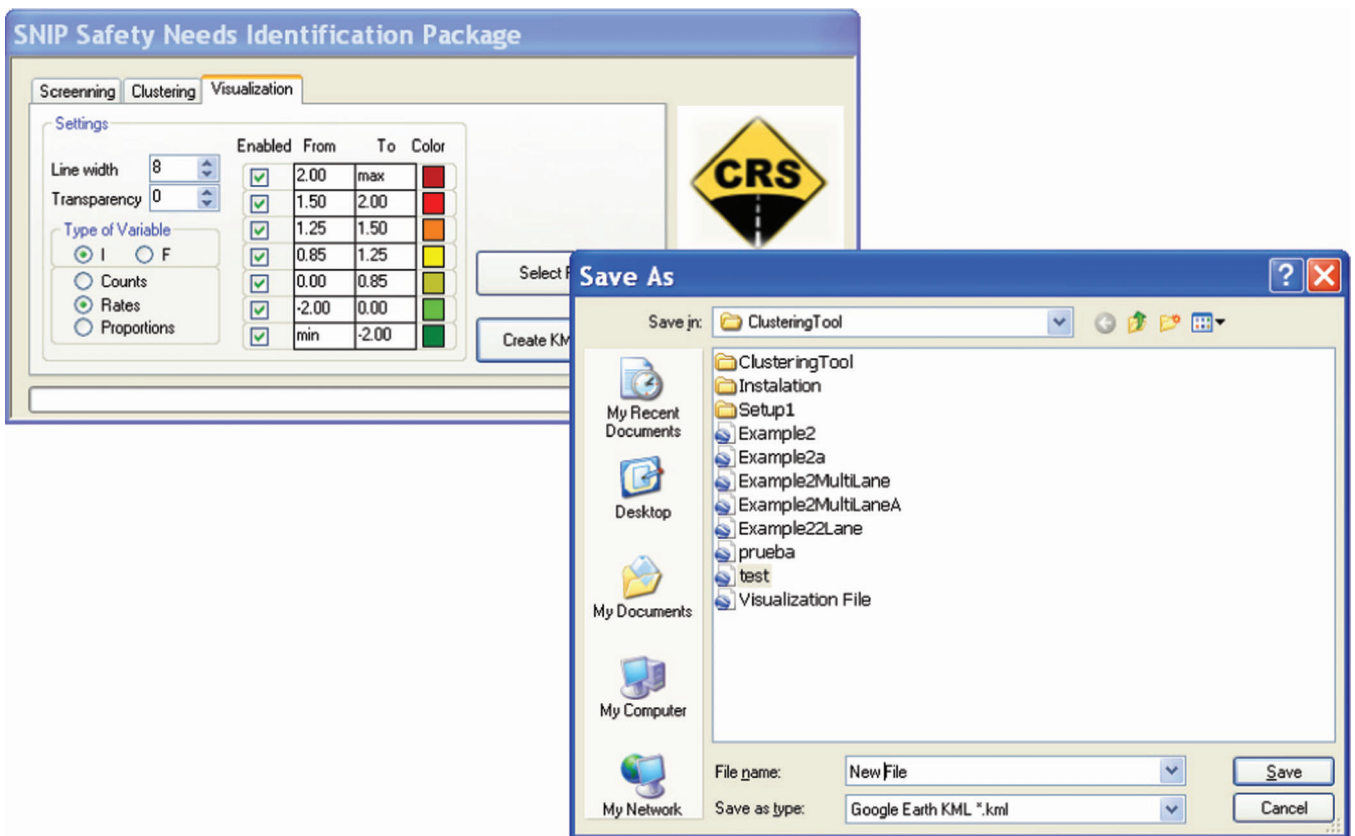Figure 22   Selecting Files for Visualization
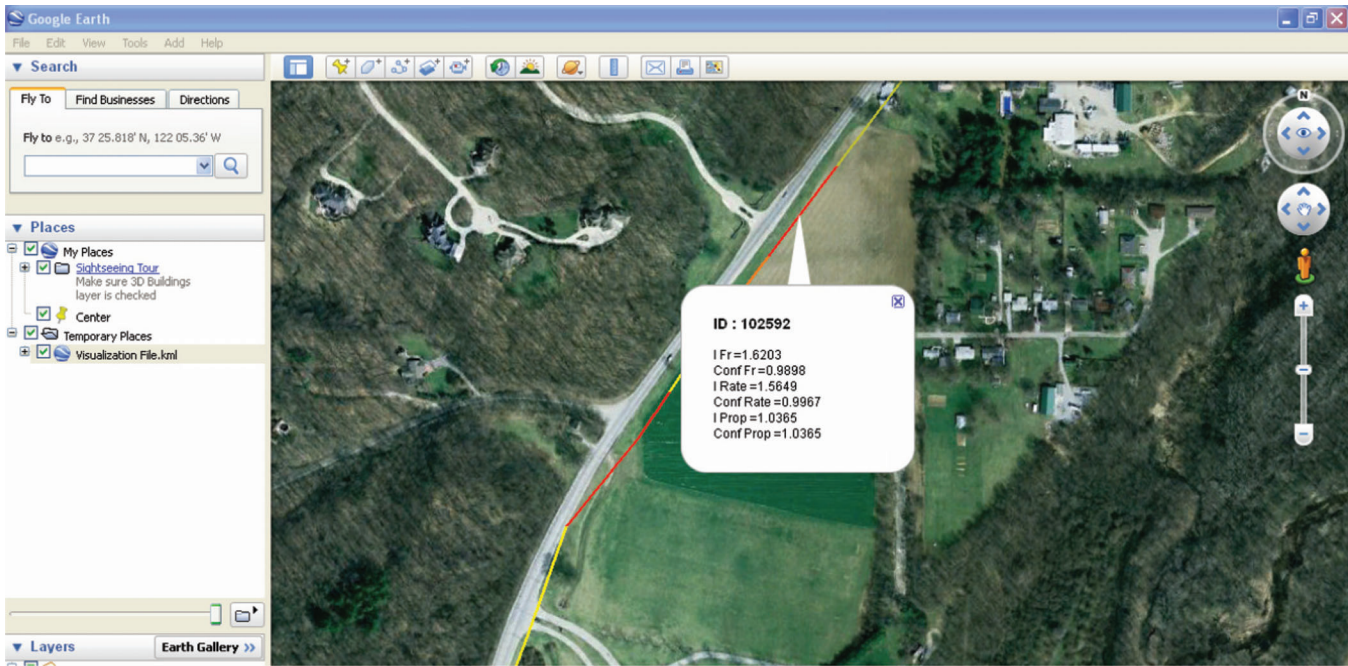

Figure 23   **Creating** a KML file

Figure 24   KML file displayed in Google Earth

installed. If Google Earth is not installed, it can be downloaded at no charge from the following web page: http://www.google.com/earth/index.html.

Upon opening (double clicking) the SNIP-created KML file (Google Earth is called automatically), Google Earth displays the full extent of the KML file with all the elements shown according to the user's settings. For zooming in, just double click on the zone you want to zoom. The zoom level can also be controlled with the mouse wheel.

By clicking on one of the displayed elements, a pop up window with the calculated indexes and confidence values and the element ID is shown (Figure 24).

For more information about how to use Google Earth go to the following website: http://www.google.com/earth/learn/.

4.2 ArcGIS Tools for Visualization

ArcGIS provides many visualization tools that can accomplish this job. Among the various features available in ArcGIS, the following three visualization tools are widely used:

- Symbology
- Labels
- Selection By Attributes

Symbology refers to visualization of feature (i.e., a single element), categories of elements, quantities, etc., by colors or symbols. Symbology has a special procedure to prepare charts like bar charts or pie charts to be shown as part of the individual element.

The step-by-step approach for "symbology" is presented as an example.

Step 1: Joining Output with Shape File

Open the output file and join with the existing shape file that corresponds to it. Here, the segment shape file needs to be joined with the output table that ranks the rural two-way lanes having narrow shoulders. For joining purposes, the CFID from the segment file should match with the CFID in the output file.

Step 2: Select Symbology Settings

After the output table is joined with the segment table, right click on the joined shape file and click on "Properties."

Click on the "Symbology" tab and then to "Quantities" and "graduated colors." Now the user needs to add the total crash field (or any other field to symbolize) as input for "Field." The user may choose a specific color ramp for symbology. For example "red to green" ramp can show the high crashes in red and low crashes in green.

ArcGIS provides options for user-defined and standard classification for Symbology. In order to change the default classification scheme and values user needs to 'Classify' button. The classification window will appear. Under the drop-down list in 'Classification' window there are a number of options for classification methods, for example; Equal Interval, Quantile, Natural Breaks etc. Also there is an option called
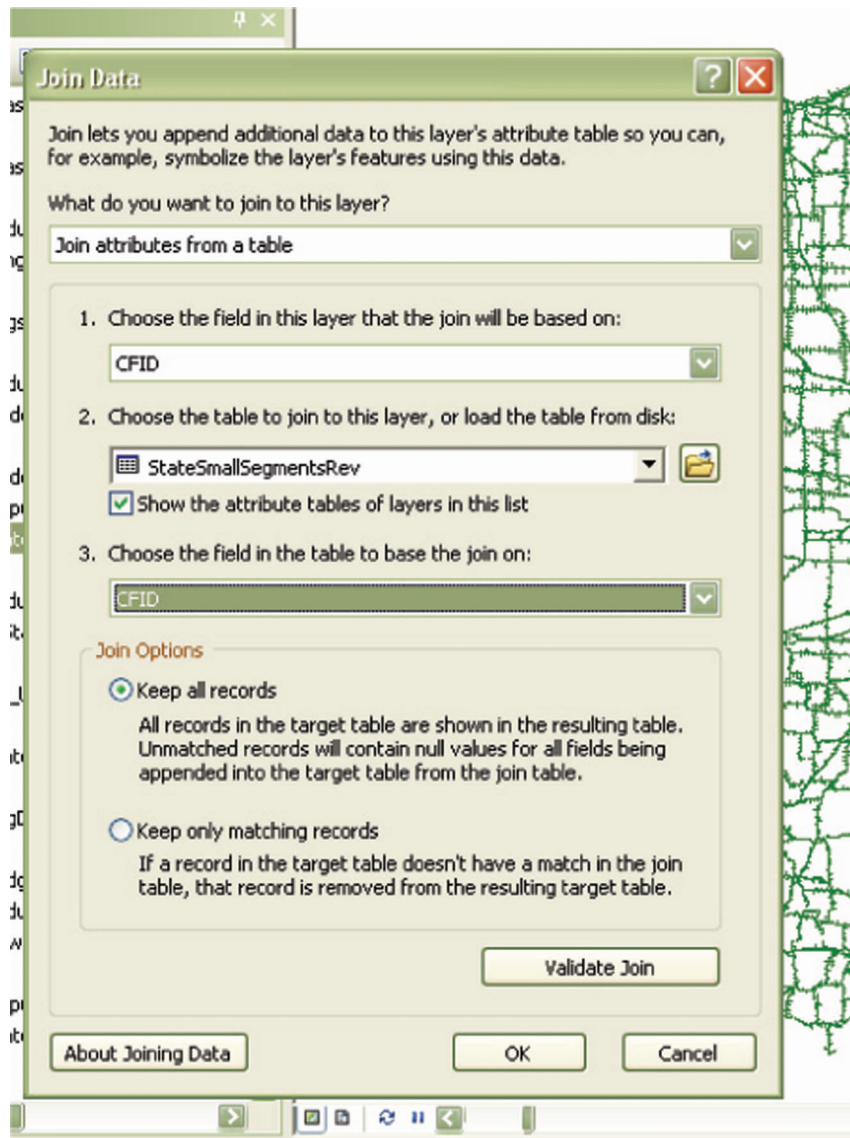
Figure 25 Joining Shape File with Output

'Manual' for user defined classification setting. User can see the distribution of the 'Index' in the window along with vertical lines showing the current classification (Figure 27).

Also when 'Manual' option is selected user can slide the vertical line by mouse drag and change the upper/lower limits for a particular class.

Click "apply" once and all of the parameters for symbology are selected. Now, the user will be able to see the high crash location in red (see Figure 28).

5. EXAMPLE STUDY

In this chapter, we demonstrate the SNIP applied to two studies: 1) identification of Indiana rural road segments with excessive numbers of nighttime (dark conditions) crashes and 2) identification of Indiana urban roads with excessive numbers of rear-end crashes. Complete step-by-step explanations of all the operations to screen the Indiana state network, to cluster the high-crash roads, and to display the obtained results are provided for the first study of nighttime crashes. Only the results of the second study are presented.

5.1 Nighttime Crashes Study

The specific conditions for this case study are listed below:

- Scope: Indiana
- Element: segments and intersection points
- Crash selection criteria: crashes in dark/not lighted conditions
- Road selection criteria: rural two-lane and multilane roads and intersection points on these roads
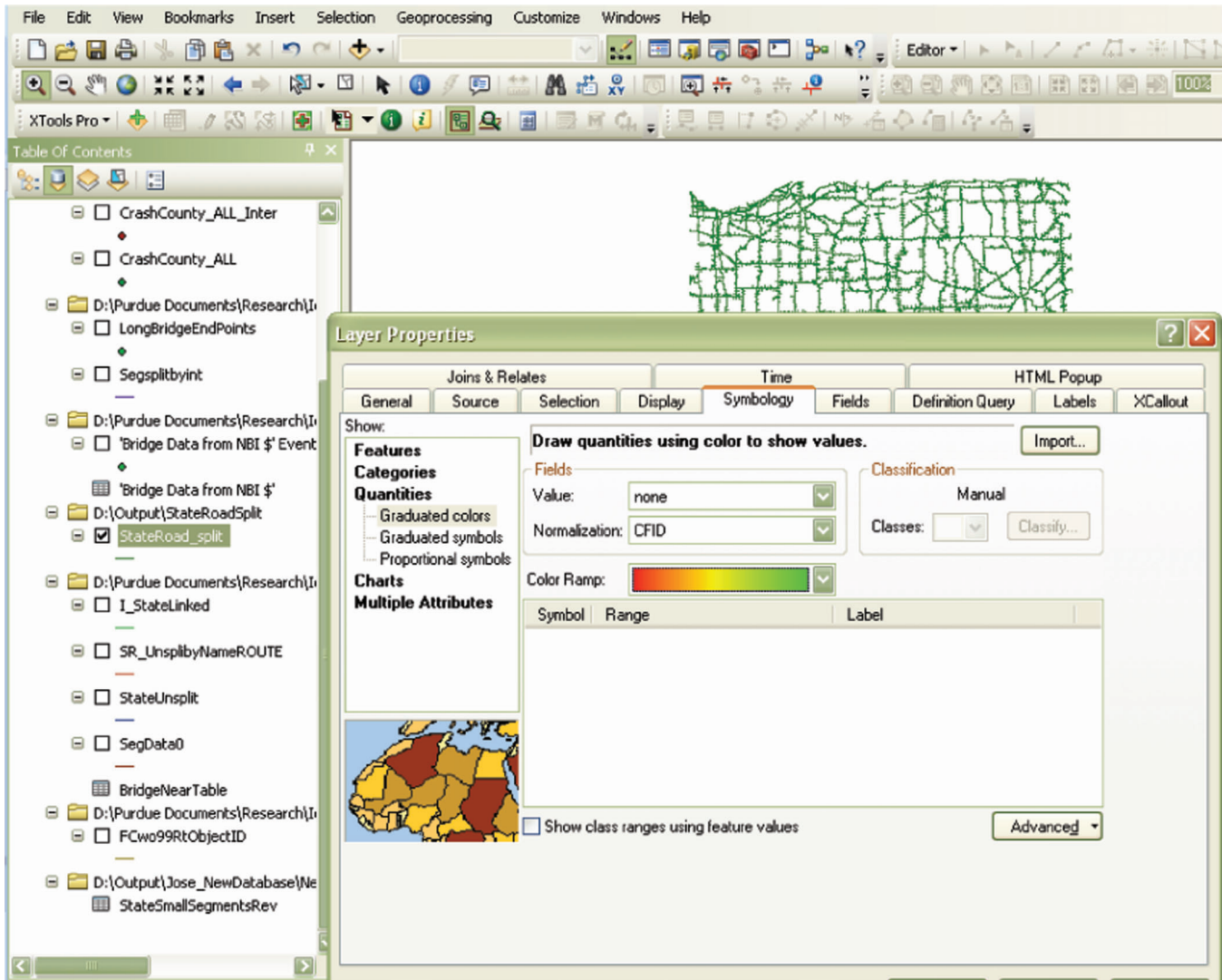
Figure 26   Choosing Setting for Symbology

- Safety performance measures: crash frequency, crash rate, proportion in all crashes
- Exposure for segments: VMT
- Exposure for intersections: volume of entering traffic

Screening network

Start the SNIP program and click on the "Start Screening" button as shown in Figure 29.

Before running the analysis, the database was compacted to reduce the tables to their original conditions and to eliminate the space taken by temporary tables used in intermediate calculations. To complete this operation, go to the main control menu of Microsoft Access and click the option ""Manage/ Compact and Repair Database" (Figure 30, version 2007 shown). If at any point during the analysis Access fails to display the current results (the table remains empty or it displays old results), the user should stop

the analysis and recompress the database. After the process of recompression is completed, the user should reset the analysis by pressing the Reset button in the Scope/Element tab and should redo the last analysis from the beginning.

The screening analysis starts by going to the tab "Scope/Element" and selecting STATE for the Scope as and the class of elements as being SEGMENTS (Figure 31 and Figure 32).

Then move to the "Crash Selection Criteria" tab and select the Light Conditions as being "Dark/ not Lighted". Leave all other conditions preselected as "All" (Figure 33). Pressing the "Extract Crash Data" button (Figure 34) will extract the crash data of interest.

The user should then wait until a small pop-up window displays the message that the data extraction is finished (Figure 35). The button "View Extracted Crashes" will become visible and let the user review
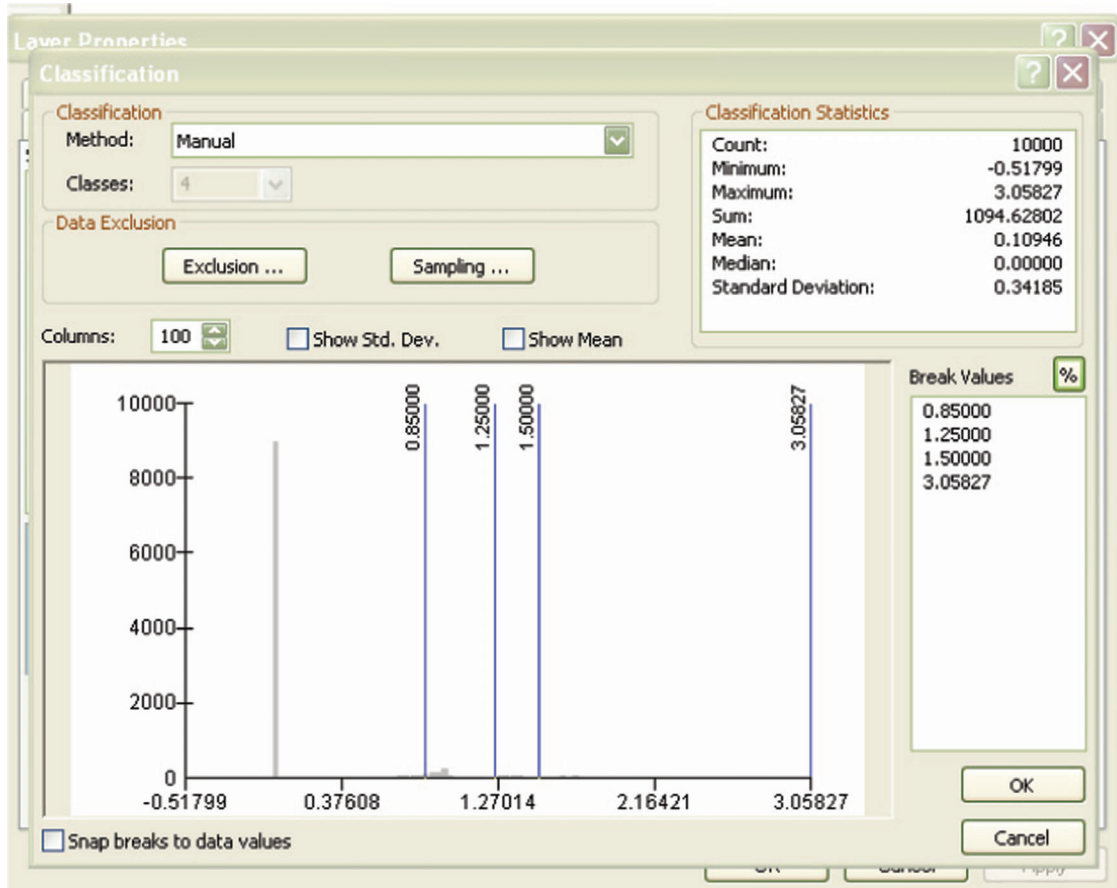
Figure 27   Classification Settings in Symbology

the crashes that happened during dark hours, in not lighted conditions.

At this point, he/she should proceed to the "Element Selection Criteria tab".

In the "Element Selection Criteria" Tab, the user will be interested in the Rural Two Lane and Rural Multilane" Segment Types. In this first version of the tool, only one type of segment can be analyzed at a time. So the user will perform the analysis first for Rural Two Lane and then repeat the process for Rural Multilane.

At this point, the user selects the "Rural Two Lane" (Figure 36) and leaves all other options as "All". The user will then press the "Extract Element Data" button (Figure 37) and once more wait for the pop-up window that indicates that the data extraction is complete.

The user now should have the crash and element data available for analysis and needs to go to the "Metrics" Tab to select the measure of exposure that will be adopted in the rate calculations. Since the class of elements chosen was Segments, the available measures of exposure are "Length" and "VMT." For this analysis, we choose 'VMT" (Figure 38).

The last step in the metrics selection process is the choice of the crash population that will be considered for those proportion calculations. These crashes will be

the standard against which the crashes extracted using the Crash Selection Criteria will be displayed as a percentage of. In our analysis, we wish to look at all crashes. We notice that the field for Light Conditions will have all options except for "Dark/Lights off" (our choice in the "Crash Selection Criteria") tab and "All" grayed out. We choose "All" (Figure 39).

After pressing the "Extract Reference Crashes" button, the user should wait until the "Finished Data Extraction" popup shows (Figure 35).

The user is now ready to start the calculation of frequencies, rates, and proportions for crashes in dark (no lighting) conditions that occurred on the two-lane rural segments in the state.

For the analysis to start, the user should go to the 'Results" Tab and press the "Start Screening Calculations" button. It will take a few minutes for the creation of several intermediate tables. The user may occasionally see a few messages like the one displayed in Figure 40, warning that existing tables (from previous analyses) will be deleted. The user should always press the "Yes" button to proceed with the analysis.

Once the creation process is completed, a message that the calculations are finished will pop up. After pressing OK, the user will see the final results table.
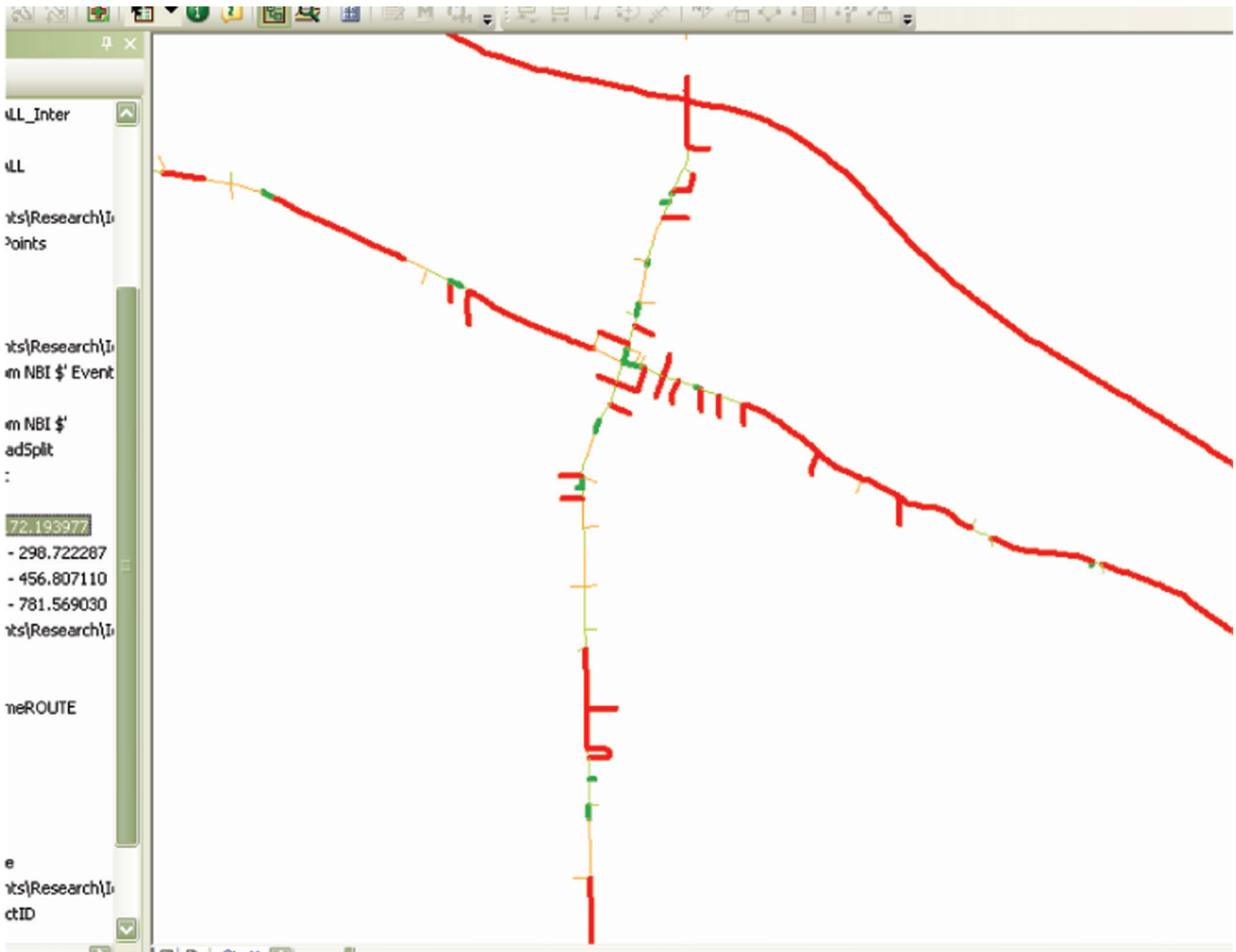
Figure 28   Final Results of Symbology in Map
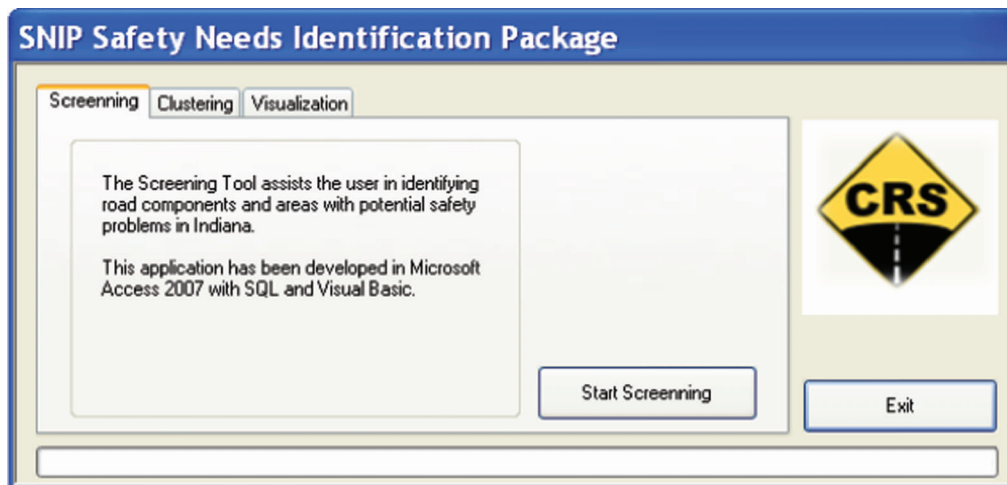

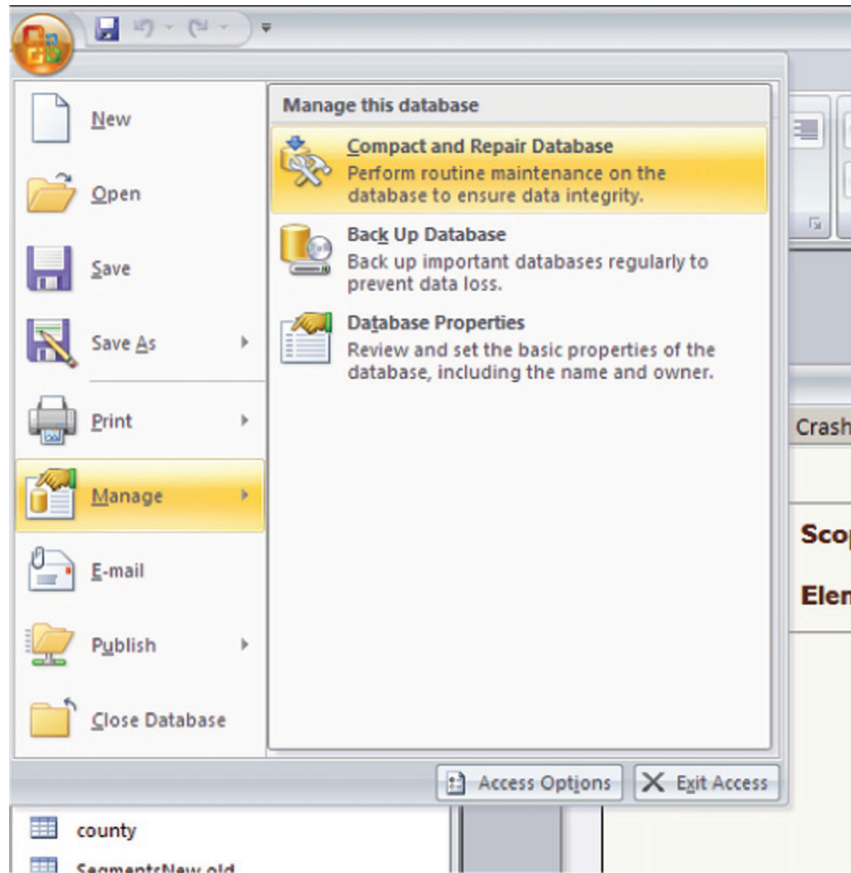Figure 29   SNIP – Screening tool calling

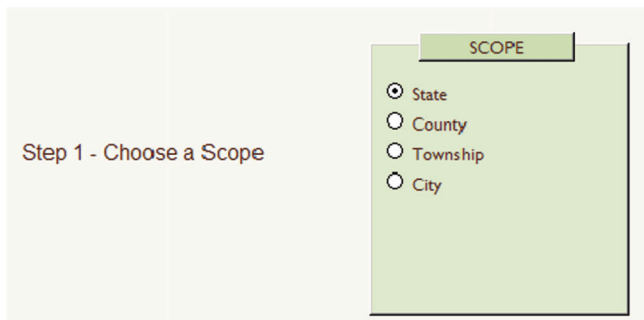Figure 30 Compact Database Using MS. Access Menu

Figure 31 SCOPE = State

This table, based on the selections of this example, will contain all of the rural two-lane segments in Indiana, the basic segment identifiers, and the number of crashes meeting the dark/no lighting conditions assigned to them. The significance levels and indexes for crash frequency, crash rates (by VMT in this example), and crash proportions (as a percentage of all crashes in those segments in this example), are also displayed (Figure 41).

The Results file can then be exported as a MS Access File by pressing the button "Export Results to Access
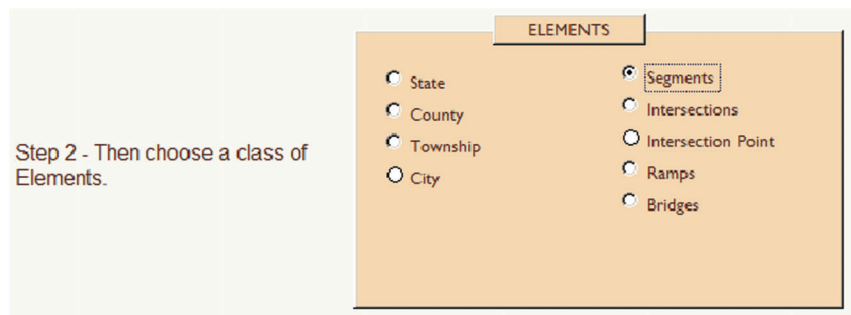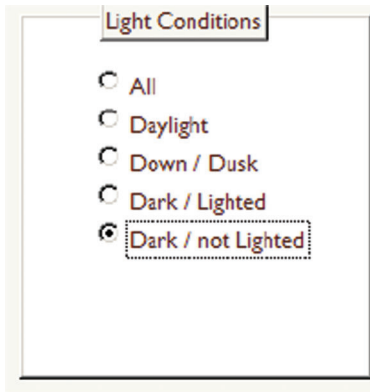

Figure 32 Element = Segments

Figure 33   Light Condition chosen as Dark/not Lighted
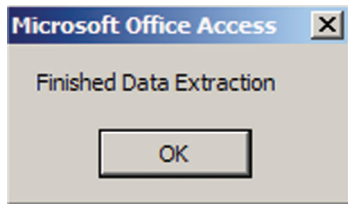

Figure 34   Extract Crashes of Interest

Figure 35   Data Extraction is Complete

File" that will become visible right next to the "Start Screening Calculations" button (Figure 42).The default name for the exported file will be Seg_result.accdb for segments and IntP_result.accdb for intersection points, but the user can save the results table with any name of his/her choice.
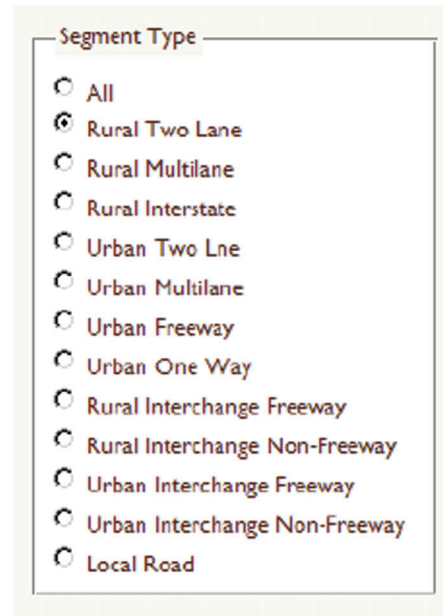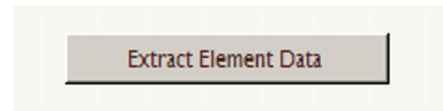

Figure 36   Select Segment Type as Rural Two Lane

Figure 37   Extract Element Data Button

The user will then repeat all the former steps selecting Segment Type as "Rural Multilane" to obtain the second segment type. At the end of the process, another Results Table will be created for the new segment type of choice. This file should also be exported to a MS Access file using the above described procedure. The file should have a different name from the one containing the rural two-lane segments (Figure 43).

The steps for the analysis of rural intersection points is identical, except that the choices of element is "Intersection Point" (Figure 44), the choice of "Intersection Rural or Urban" in the "Element
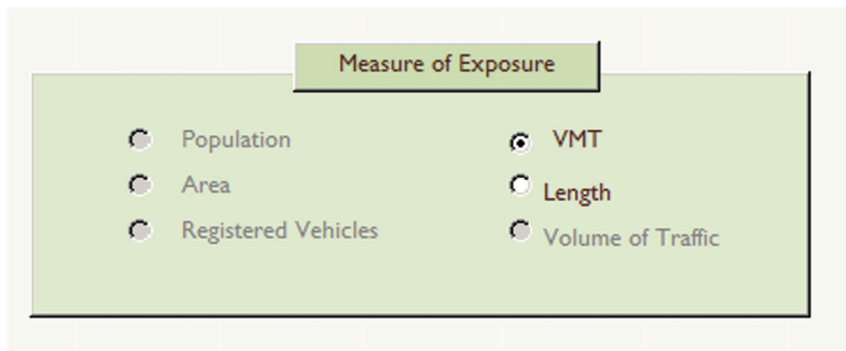

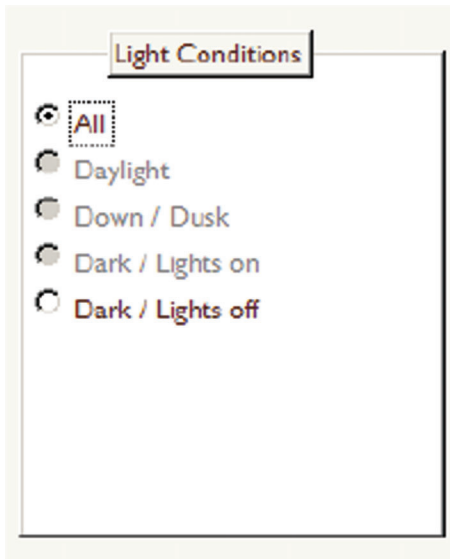Figure 38   Measure of Exposure chosen as VMT

Figure 39 Choice of All Light Conditions for Proportion Calculations

Selection Criteria" tab should be "Rural" (Figure 45). The only measure of exposure available for Intersection Points will be "Volume of Traffic," which the user should choose (Figure 46).

After the new choices are made, running the case will produce a third results table, which can also be exported to a MS Access file using the export button described in Figure 42 and Figure 43. A third name, reflecting the choice of interesection points should be used. These three files will later be used by the clustering and visualization tool.

Clustering high-crash locations by using the Clustering Tool

The initial values needed by the clustering process were set up as follows: Option F for controlling adding elements; I_1 = 1.25; F=0.5; and Allow Cluster to Switch Roads, as shown in Figure 47. Then the files of the screening tool results were opened.

Once the settings and the data import were completed, the clustering process was run. At the end of the clustering process, a Summary window appears with a summary report showing the total number of clusters found, the total segments included, and the total intersection points included, as shown in Figure 48.

Visualization using KML files

The initial values needed by the embedded visualization process were set up as the default values for variable I, as shown in Figure 49.

Figure 50 shows the high-crash rural non-freeway locations in Indiana according to the color code shown in Figure 49.

Figure 51 shows the high-crash rural non-freeway locations in Porter County and Figure 52 shows the high-crash locations during dark conditions on US-6 near Valparaiso. Figure 53 presents clusters obtained for this area.

Visualization of Safety Screening Output by ArcGIS

Convert Output to .dbf Format

In order to visualize the standard screening and clustering results on an ArcGIS map by using the standard visualization technique (symbology), the results must be imported to ArcGIS and then joined with the existing network shape file. ArcGIS supports many different non-GIS formats but the most commonly used are dbf, xls, and csv. Of these three types, a file in the .dbf format can be seamlessly joined with an existing shape attribute table; thus, the screening and clustering results for display should be in this format.

In this example, the screening results were saved in the MS Excel format. Standard MS Excel tools do not have the capability to export into the .dbf format. Therefore, the Excel file must be opened in MS Access and then saved as a .dbf file. Figure 54 shows the screening results in the Excel format for rural two- and multilane roads.

Open the MS Access and create a blank database. You may also import the file into an existing Access database. Click on the tab "External Data" and then click on the Excel logo. A window "Get External Data-Excel Spreadsheet" will appear. Click on "browse" button and locate the Excel file (screening output) to be imported into MS Access (Figure 55).
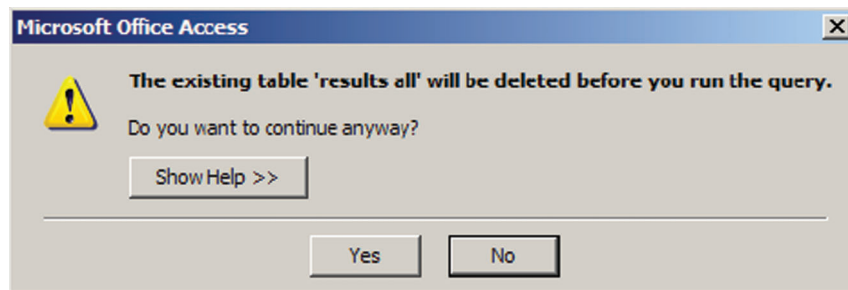

Figure 40 Writing Overwrite Message

Segment_Type	Length_Ft	VMT	X_Start	Y_Start	X_End	Y_End	Crashes	Confidence_Freq	Index_Freq	Confiden
Rural Two Lane	205.2	62566	-86.406088783	41.3033819756	-86.405499	41.303701	18	1.000	4.20	
Rural Two Lane	137.7	157716	-86.3994950001	39.763019	-86.39951	39.762641	10	1.000	3.10	
Rural Two Lane	521.9		-84.877353	41.629415	-84.8792616045	41.6293899133	10	1.000	3.10	
Rural Two Lane	523.7	538563	-85.9601124681	38.3631462801	-85.961345	38.364208	10	1.000	3.10	
Rural Two Lane	537.0	365566	-87.370618	39.593337	-87.3706179995	39.5948118328	9	1.000	2.94	
Rural Two Lane	594.0	702748	-87.159754	41.187619	-87.1619125026	41.1876473707	9	1.000	2.94	
Rural Two Lane	505.6	312692	-86.771254	41.662825	-86.7721303728	41.6640477203	9	1.000	2.94	
Rural Two Lane	572.8	279483	-85.528216	40.16247	-85.5264544306	40.1632730357	9	1.000	2.94	
Rural Two Lane	534.4	329623	-86.086464	39.196742	-86.0883275133	39.1965145155	8	1.000	2.76	
Rural Two Lane	570.2	216756	-87.49256	38.333059	-87.4945115564	38.3327989012	8	1.000	2.76	
Rural Two Lane	545.2	93707	-86.7391564071	41.4927633429	-86.739064	41.491268	8	1.000	2.76	
Rural Two Lane	227.6	17438	-86.582751	41.218056	-86.583574	41.21799	8	1.000	2.76	
Rural Two Lane	183.0	31605	-85.136107	38.705617	-85.1367060946	38.7054375807	8	1.000	2.76	
Rural Two Lane	658.8	285121	-85.885683	41.217132	-85.8869392287	41.2156184213	8	1.000	2.76	
Rural Two Lane	544.9	275808	-86.318903	39.161808	-86.3170109927	39.1615430153	8	1.000	2.76	
Rural Two Lane	512.9	312362	-86.2711641447	40.0115035289	-86.271812	40.012821	8	1.000	2.76	
Rural Two Lane	82.5	28861	-86.439667	40.758269	-86.43995	40.75834	8	1.000	2.76	
Rural Two Lane	529.7	213120	-85.8368934125	39.0866290012	-85.837028	39.085178	7	1.000	2.57	
Rural Two Lane	232.4	79390	-86.3949690002	40.279796	-86.394985	40.279158	7	1.000	2.57	
Rural Two Lane	477.2	368086	-85.198332367	41.4449038518	-85.196592	41.444885	7	1.000	2.57	
Rural Two Lane	246.5	34858	-86.8273515226	40.0230241836	-86.828202	40.023199	7	1.000	2.57	
Rural Two Lane	562.2	32129	-86.177285	41.237596	-86.1793283735	41.2376550548	7	1.000	2.57	
Rural Two Lane	539.3	163862	-85.4142346765	40.8409413909	-85.412555	40.84019	7	1.000	2.57	
Rural Two Lane	496.0	136807	-85.544861	38.912076	-85.5457907915	38.9132280658	7	1.000	2.57	
Rural Two Lane	528.6	135001	-86.0005715998	41.3628032007	-86.000555	41.361352	7	1.000	2.57	
Rural Two Lane	532.2	483221	-85.797177	41.667564	-85.7991240049	41.6675106661	7	1.000	2.57	
Rural Two Lane	499.3	137357	-85.0011468764	39.6324496287	-84.999602	39.633122	7	1.000	2.57	

Record: I◄ ◄ 1 of 88647 ► ►I ►✱   ✖ No Filter   Search

Figure 41   Final Results Table

Once the Excel file is converted to an MS Access table, export the table as a .dbf file. Figure 56 shows the snapshot of the exporting procedure. Here the table "Intersection Points" is the imported screening results from the Excel file. Before exporting as .dbf file user needs to make sure that all field names are unique in the table (to be exported) and also have 8 (eight) or less number of characters. This is the basic requirement to successfully export into a .dbf file. The newly created files are SEG.dbf with segments and INTERSEC.dbf with intersection points.

It is important to note that the standard screening results were originally created as a MS Access table, and then this file was saved as an Excel file (recall Figure 41 and Figure 43). Since for visualization purposes a .dbf file is critical, it is worthwhile to export the standard screening output file directly to the .dbf format.

Adding Required Data in ArcMap
Once the .dbf files are created, add both the SEG.dbf and INTERSEC.dbf files in ArcMap (see Figure 57). Also add the network and intersection shape files. There now should be two GIS layers (i.e. shape files) and two non-GIS data tables (i.e., output file in .dbf format) in the ArcMap.

Joining Non-GIS Data with GIS Attribute Table
In the following few sections, the procedure to visualize the rural two-lane and multilane roads with nighttime crashes is shown. The same procedure needs to be followed for the intersection points on these roads to visualize all the results.

The two .dbf output tables for segments and intersection points contain six safety performance measures calculated for each road element included in these tables: confidence F calculated for the crash

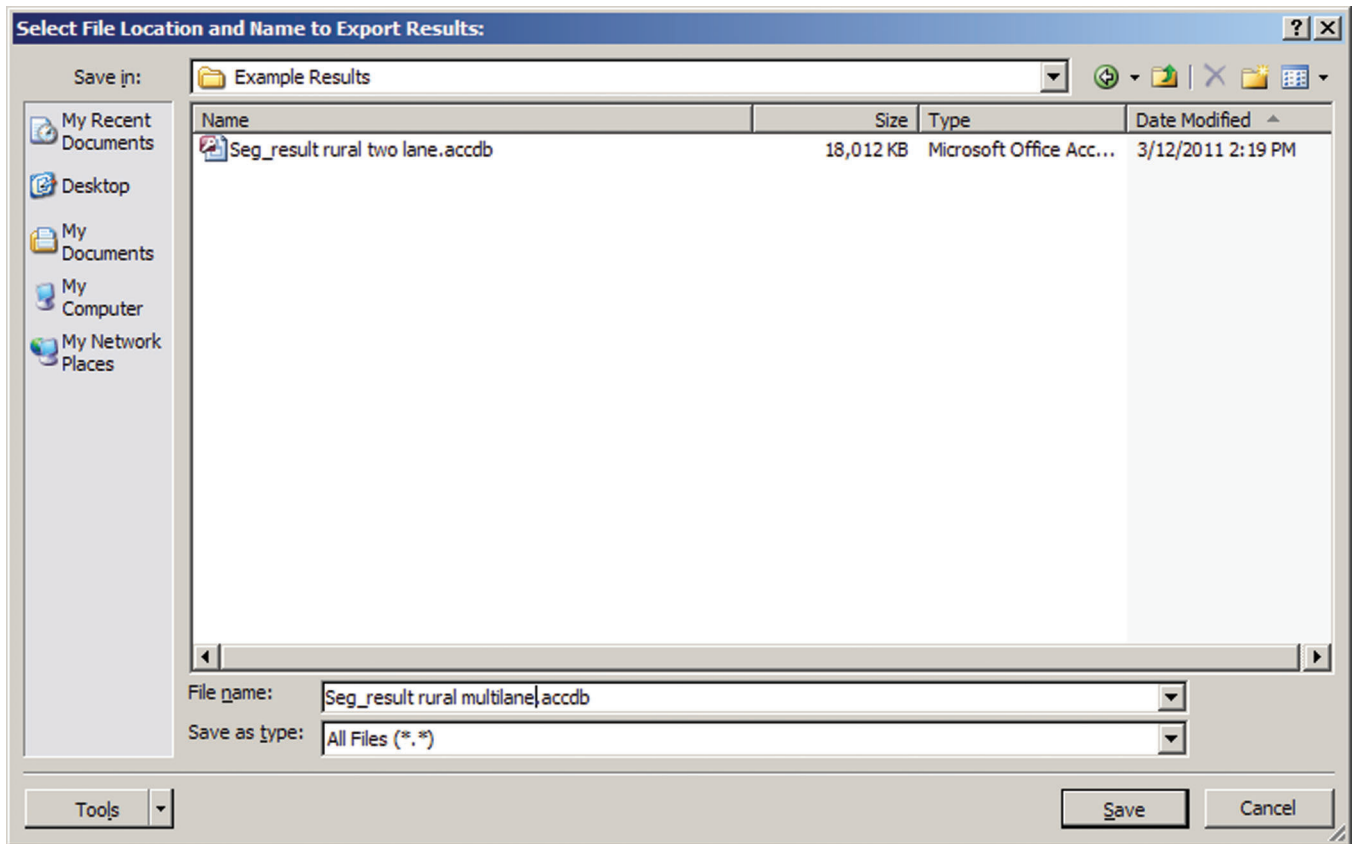Figure 42   Export to MS Access button

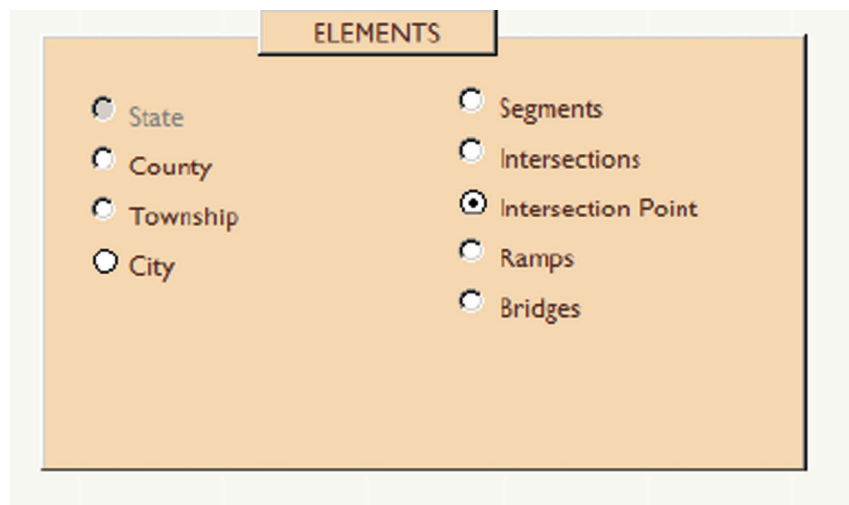Figure 43   Exporting Results to MS Access

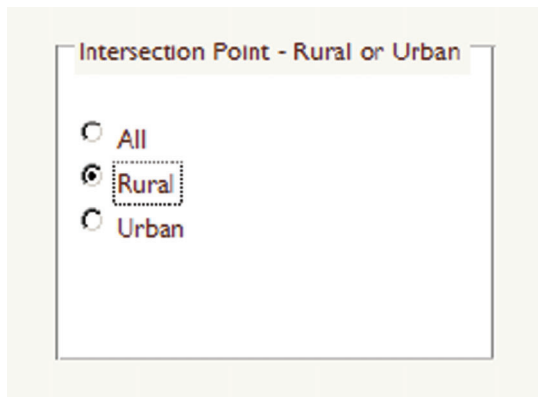Figure 44   Choosing Intersection Points

Figure 45   Choosing Rural Intersection Points

frequency, rate, and proportion, and index I also calculated for the crash frequency, rate and proportion.

For visualization purposes, two files are needed: (1) the .dbf output table (which is specific to a particular screening operation) and (2) the GIS attribute table of the network shape file. ArcGIS has the capability to create a virtual table made of these two tables by joining them via a common variable. Right-click on the network layer and click "Join" and then select the common variable (column) in both the tables. Here, the join variable is "CFID" (Figure 58 and Figure 59).

Any visualization operation can be made easily on such a joint table. After finishing the visualization operation, the user can easily separate the two datasets.

Once the tables are joined together, the symbology operation can proceed. Figure 60 shows a joined table view for segments.

Selecting Map Settings for Visualization

Right click on the joined table and click on "Properties." Click on "Symbology" tab and then click on "Quantities" and "Graduated Colors."

The user must select the name of the variable which he wants to visualize. Click on the drop-down list for Field values and select the variable "Index of Crash Rate."

The overall range of the values for this variable will be automatically determined and broken into several ranges with corresponding initial symbols, labels, and colors (see Figure 61) arranged in ascending order. The user can change these initial settings.

In the presented example, red was selected as the color for high index values (1.5 and more), orange for intermediate values (1.25-1.5), light green for low values (0.85- 1.25), or dark green for very low values (<0.85). These colors and the corresponding symbol/color legend are on subsequent maps with the displayed results.

Visualization in Maps Standard Screening and Clustering Output

Once all of the settings for Symbology are selected, click "Apply" to start visualization. Figure 62 shows the high-crash rural non-freeway locations in Indiana. Locations with higher crash rates are shown in red (>1.5) or orange (1.25-1.5).

For better identification, we may now label the road based on segment name. Right click "Properties" and then "Label." Select Road name (SEGNAME) as "Label Field" and click the check box "Label Features in This Layer" (Figure 63).

Once labeling is completed, zoom into the Porter County area for a closer look. Figure 64 shows the spatial distribution of crashes on rural two-lane and multilane roads under the dark condition. We can easily find locations (US-6 and State Road 130) having high indices of crash rate (1.25 and higher).

We can also symbolize the intersection points and segments together in ArcMap. We need to follow the same procedure for intersections and bring these two on the map.

Finally, let's zoom to a section of US-6 road near Valparaiso (Figure 65).

The output from clustering has been symbolized similar to the standard screening. The results of clustering are shown in Figure 66. The cluster IDs are labeled instead of segment name in this case. The indices for crash rates for individual clusters are used to symbolize the segments. The clustering has been done such that none of the clusters has the Index I of crash rate lower than 1.25. Un-clustered segments are not shown. We have chosen only two ranges (1.25-1.5) and (I >1.5) for cluster visualization.
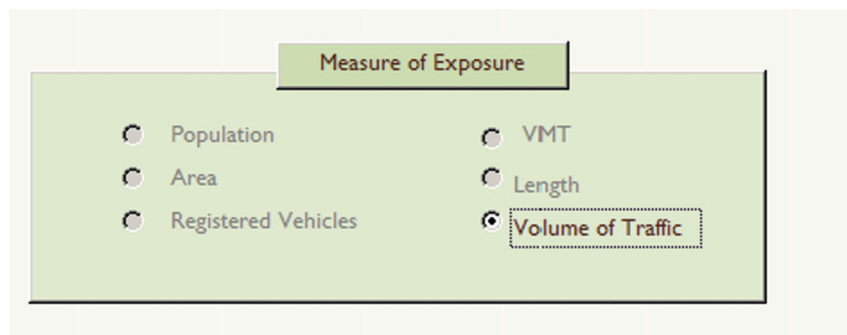
Figure 46   Volume of Traffic is the Only Exposure for Intersection Points and Intersections
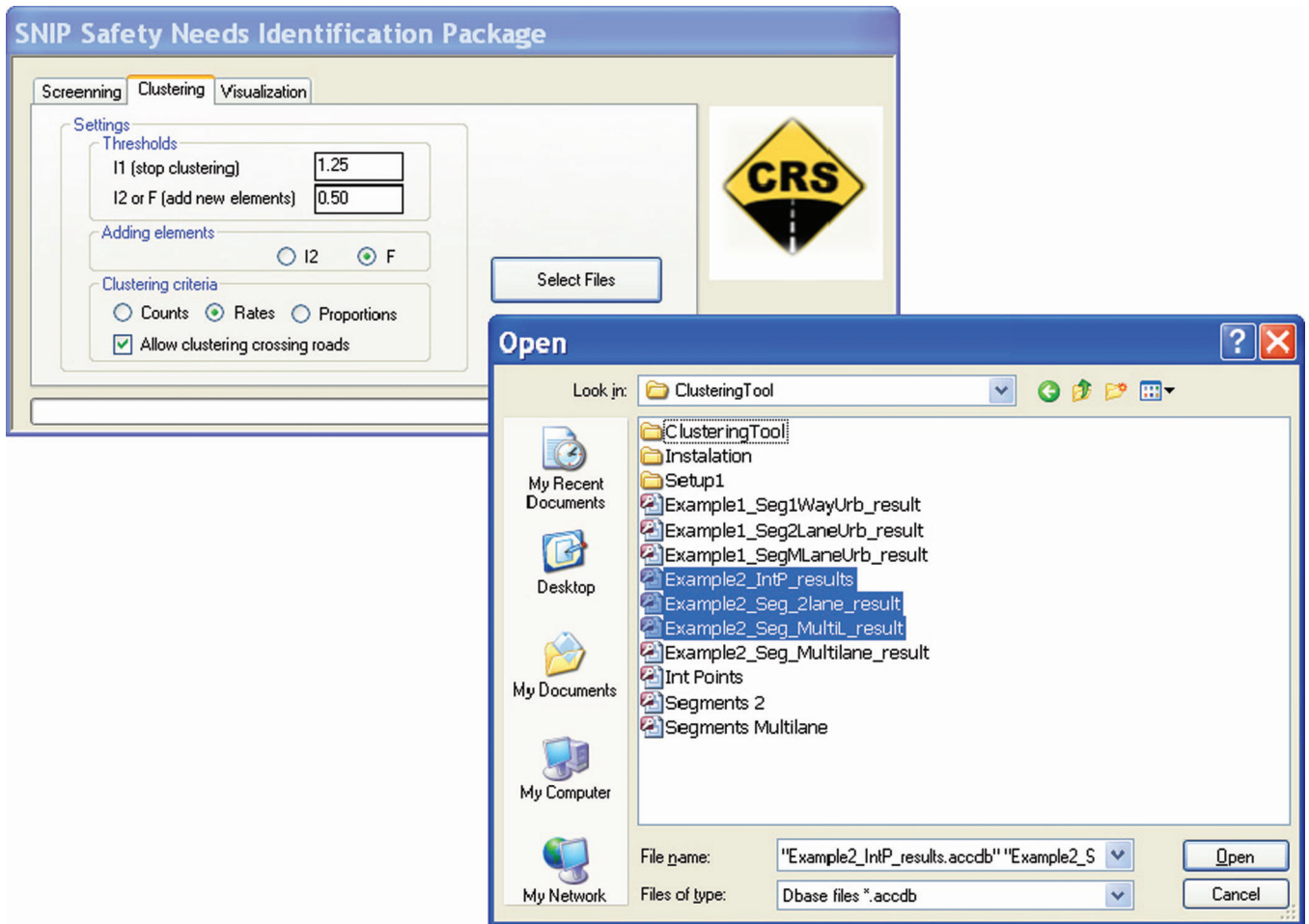
Figure 47   Selecting Files for Clustering
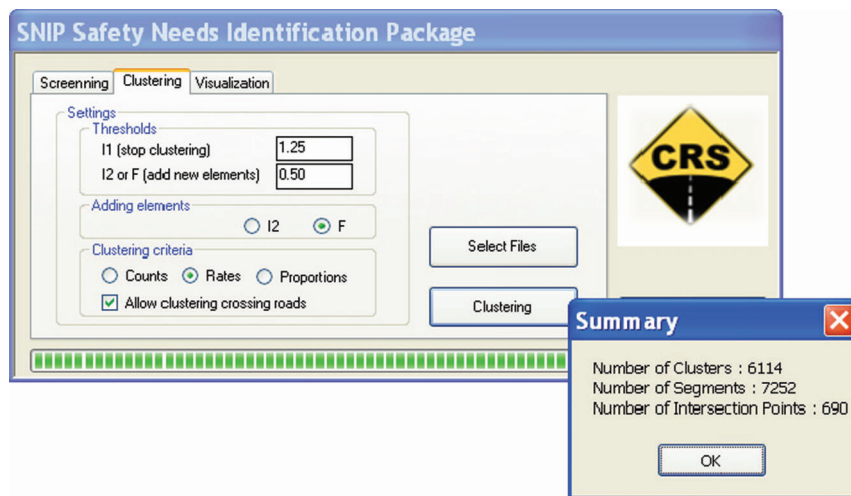
Figure 48   Clustering Summary
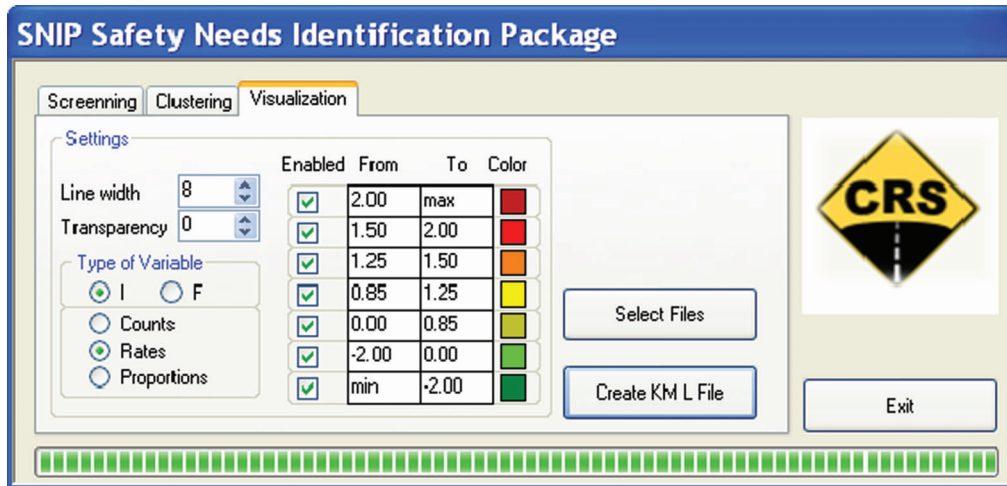
Figure 49   Embedded visualization tool settings
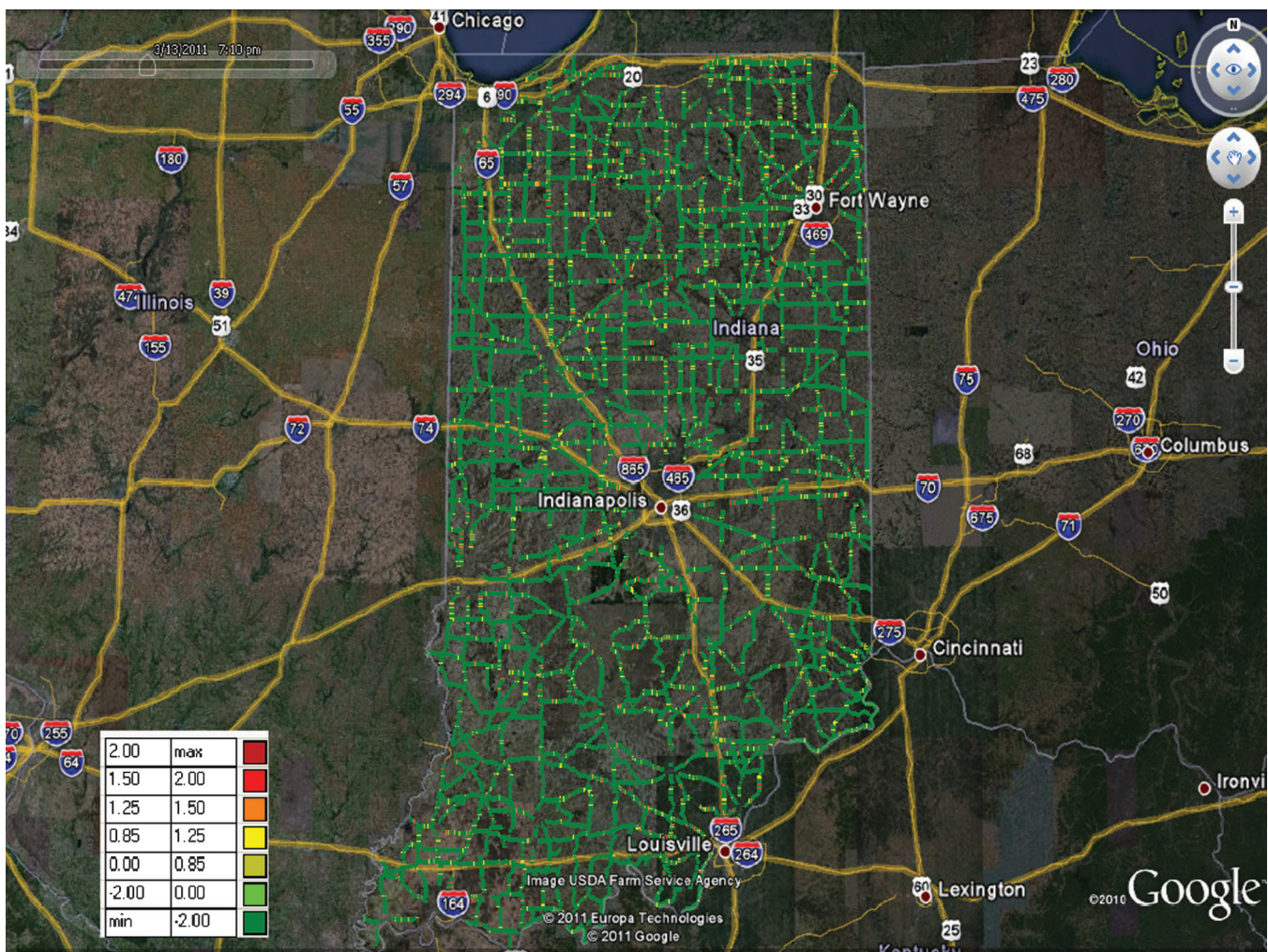
Figure 50   High-crash rural non-freeway locations in Indiana
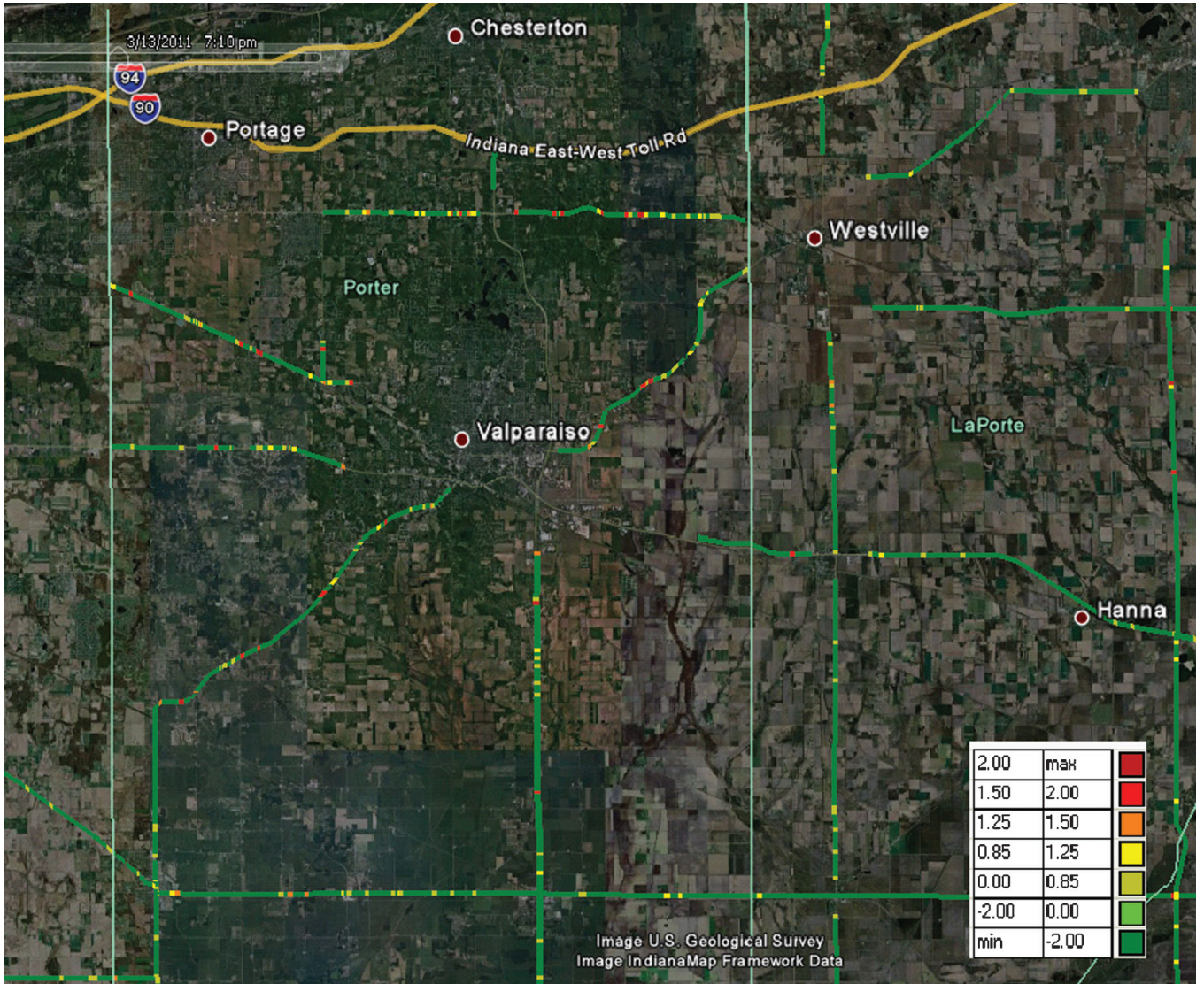
Figure 51   Porter County: High-crash Rural Locations in Dark Conditions

Figure 52   High-crash locations during dark conditions on US-6 near Valparaiso, IN
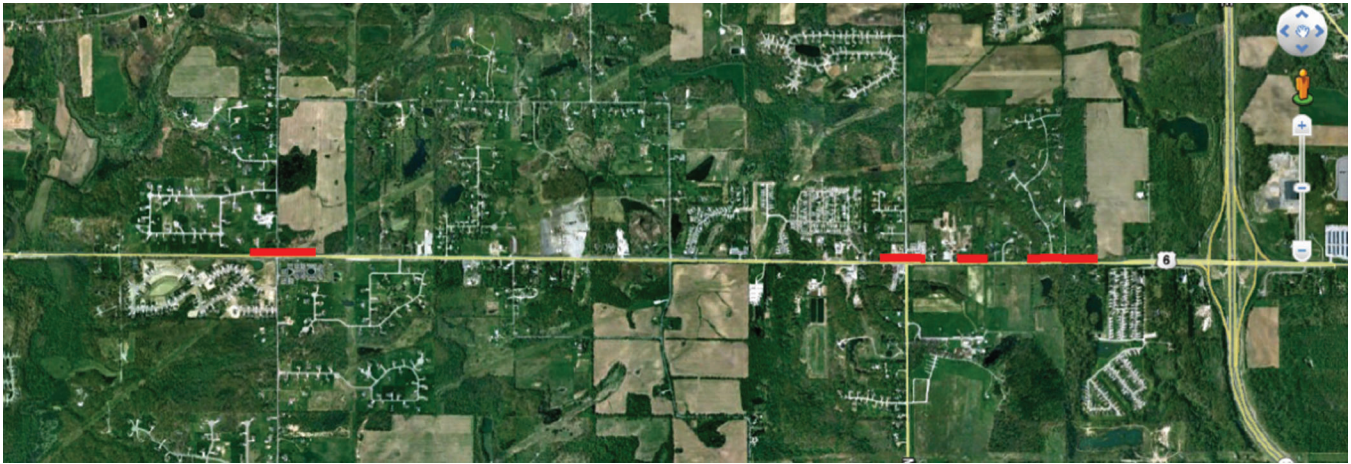
Figure 53 Clusters during dark conditions on US-6 near Valparaiso, IN

It can be seen from Figure 66 that the individual segments shown in Figure 65 have been grouped together into clusters, each having a unique cluster ID to differentiate between them. Also, it is obvious from the figure that some elements are excluded from the cluster due to an overall low index of crash rate (i.e., lower than the minimum threshold value). The clustering process emphasizes the spatial pattern of high-crash locations that helps scoping the potential safety improvement projects.

5.2 Screening Results for Rear End Collisions on Urban Roads

The following example shows the standard screening and clustering results for rear-end crashes on urban

roads in Indiana. The demonstration focuses on Tippecanoe County. The specific conditions for this case study are listed below:

- Scope: Indiana
- Element: segments
- Crash selection criteria: rear-end manner of collision
- Road selection criteria: urban roads
- Safety performance measures: crash frequency, crash rate (only displayed), proportion in all crashes
- Exposure for segments: VMT

The screening, clustering, and visualization procedures are similar to the ones presented in Section 5.1. The figures shown below compare the high-crash segments before and after clustering at the county and corridor levels. The upper images in each figure have been created

	A	B	C	D	E	F	G	H	M	N	O	P	Q	R	S	T
	COUNTY	Expr1	Expr2	Expr3	CFID	SEGTYPE_OL	Length	VMT	crashes	m	n	vi	ratem	ratea0	ratea	s
	ALLEN	80342	25000	02-S-930-0-01	105383	5	490.1034871	1243154.137	125	1.019034736	20699	0.742645654	3.506416064	4.74091E-05	4.74091E-05	187
	LAKE	66096	48528	45-S-053-0-01	18524	5	345.6993979	696431.0298	121	1.019034736	20699	0.695498697	1.964339641	4.74091E-05	4.74091E-05	168
	LAKE	54414	33466	45-U-041-0-01	123770	5	595.174011	1651989.821	115	1.019034736	20699	0.627677092	4.65956994	4.74091E-05	4.74091E-05	156
	LAKE	66096	48528	45-S-055-0-01	96960	4	202.5479388	267286.9752	86	1.019034736	20699	0.348911221	0.753904376	4.74091E-05	4.74091E-05	122
	LAKE	66870	68220	45-U-030-0-01	123684	5	466.5389549	1082408.566	86	1.019034736	20699	0.348911221	3.053020273	4.74091E-05	4.74091E-05	125
	ALLEN	80342	25000	02-S-930-0-01	90882	5	382.9353183	1185215.114	80	1.019034736	20699	0.301381432	3.342994396	4.74091E-05	4.74091E-05	97
	ALLEN	80342	25000	02-S-930-0-01	837	5	466.9543587	1445260.692	77	1.019034736	20699	0.278921011	4.076473832	4.74091E-05	4.74091E-05	120
	ALLEN	80342	25000	02-U-027-0-01	928	5	174.7434198	390422.6404	77	1.019034736	20699	0.278921011	1.101218407	4.74091E-05	4.74091E-05	122
	MARION	59058	36003	49-U-031-0-01	125856	5	464.7807087	1334750.756	75	1.019034736	20699	0.264430535	3.764771684	4.74091E-05	4.74091E-05	115
	TIPPECANOE	22342	40788	79-U-052-0-01	138542	5	531.9691834	1071482.993	68	1.019034736	20699	0.216757644	3.022203818	4.74091E-05	4.74091E-05	93
	VIGO	34546		84-U-041-0-01	139980	5	537.1724609	972020.453	67	1.019034736	20699	0.210333741	2.741661737	4.74091E-05	4.74091E-05	89
	MADISON	1486	01468	48-S-009-0-01	125280	5	419.7679061	535327.8484	65	1.019034736	20699	0.19777582	1.509935181	4.74091E-05	4.74091E-05	156
	HAMILTON	13060	10342	29-U-031-0-01	12113	5	497.2965255	2191722.801	60	1.019034736	20699	0.168072	6.181930147	4.74091E-05	4.74091E-05	72
	LAKE	66870	19270	45-U-030-0-01	97163	5	238.6172239	553611.4923	60	1.019034736	20699	0.168072	1.561505667	4.74091E-05	4.74091E-05	116
	TIPPECANOE	22342	40788	79-U-052-0-01	138539	5	555.5979123	1059018.551	60	1.019034736	20699	0.168072	2.987046859	4.74091E-05	4.74091E-05	90
	TIPPECANOE	22342	40788	79-U-052-0-01	34976	5	531.8035187	1005466.208	58	1.019034736	20699	0.156866866	2.835998174	4.74091E-05	4.74091E-05	90
	HAMILTON	54198	54180	29-S-037-0-01	11963	5	527.7393015	1219512.519	57	1.019034736	20699	0.151409241	3.439732982	4.74091E-05	4.74091E-05	69
	HOWARD	11422	40392	34-U-031-0-01	14210	5	582.787344	1728099.109	57	1.019034736	20699	0.151409241	4.874242298	4.74091E-05	4.74091E-05	109
	LAKE	54414	33466	45-U-041-0-01	123767	5	550.9911983	768434.7083	56	1.019034736	20699	0.146048243	2.167431797	4.74091E-05	4.74091E-05	80
	PORTER	82484	12412	64-S-049-0-01	131666	5	510.0412793	1061247.14	56	1.019034736	20699	0.146048243	2.993332773	4.74091E-05	4.74091E-05	70
	HAMILTON	54198	54180	29-S-032-0-01	94721	5	205.0668273	255299.0684	55	1.019034736	20699	0.140783873	0.720091522	4.74091E-05	4.74091E-05	72
	DEARBORN	42480	42462	15-U-050-0-01	6556	5	234.2044459	605026.0921	53	1.019034736	20699	0.130545016	1.706524674	4.74091E-05	4.74091E-05	89
	HOWARD	11422	40392	34-U-031-0-01	118692	5	544.7231125	1447688.776	52	1.019034736	20699	0.125570529	4.083322438	4.74091E-05	4.74091E-05	64
	ST JOSEPH	13222	65988	71-S-933-0-01	135009	5	467.6852357	815874.1984	52	1.019034736	20699	0.125570529	2.301238688	4.74091E-05	4.74091E-05	94
	TIPPECANOE	79352	82862	79-U-052-0-01	138520	5	589.4818379	1464832.595	52	1.019034736	20699	0.125570529	4.131677956	4.74091E-05	4.74091E-05	57
	SCOTT	79100	68526	72-S-056-0-01	101724	5	199.4898777	291554.0154	50	1.019034736	20699	0.115911439	0.822351511	4.74091E-05	4.74091E-05	83
	VANDERBURGH	59670	22000	82-S-062-0-01	35521	5	153.3249847	443769.3508	48	1.019034736	20699	0.106638859	1.251687087	4.74091E-05	4.74091E-05	53
	TIPPECANOE	22342	40788	79-U-052-0-01	138555	5	468.3659639	613501.0147	46	1.019034736	20699	0.097752789	1.730428874	4.74091E-05	4.74091E-05	70
	LAKE	66870	68220	45-U-041-0-01	19053	5	255.61126	513200.1632	45	1.019034736	20699	0.093454696	1.447522268	4.74091E-05	4.74091E-05	83
	VANDERBURGH	59670	22000	82-U-041-0-01	139336	5	591.7790153	1490013.668	45	1.019034736	20699	0.093454696	4.202703193	4.74091E-05	4.74091E-05	55

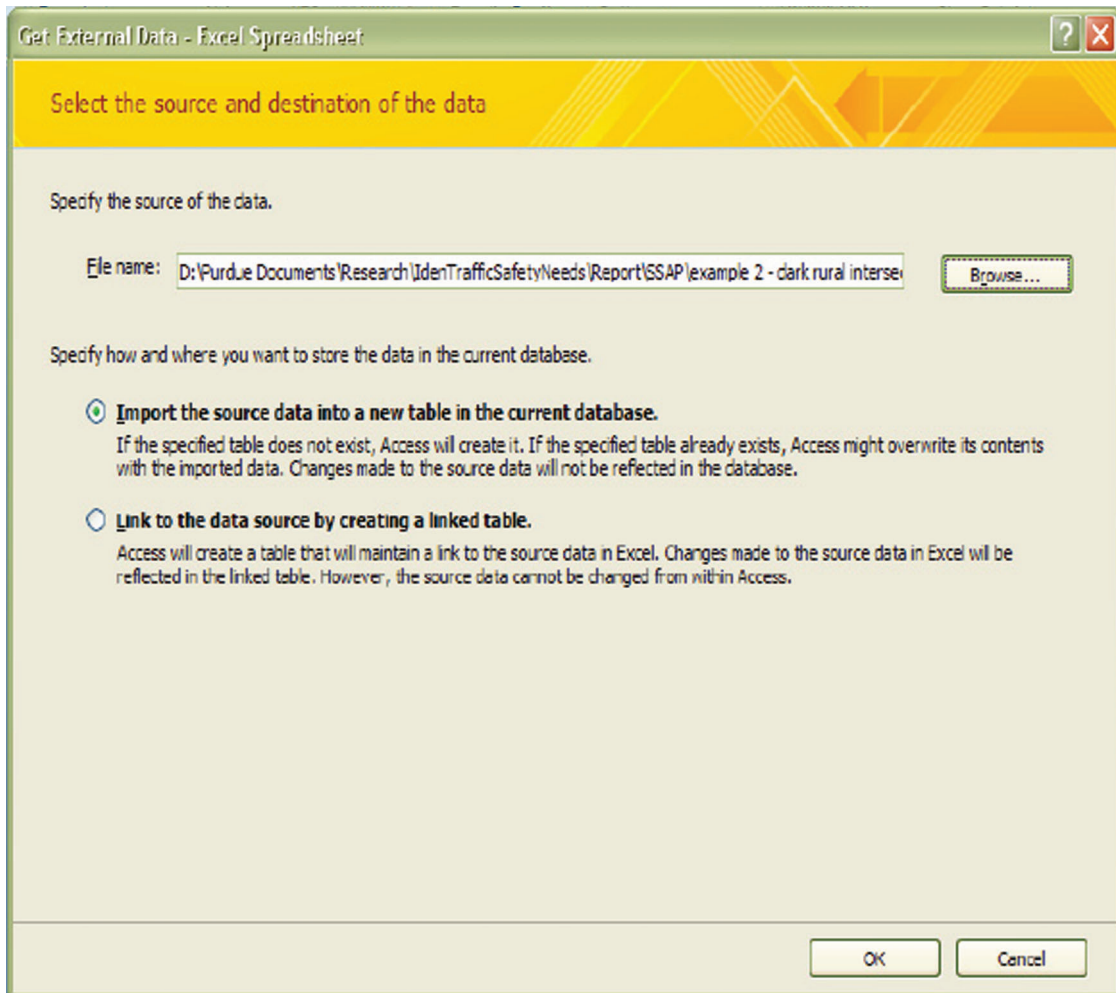Figure 54 Screening output in the Excel format

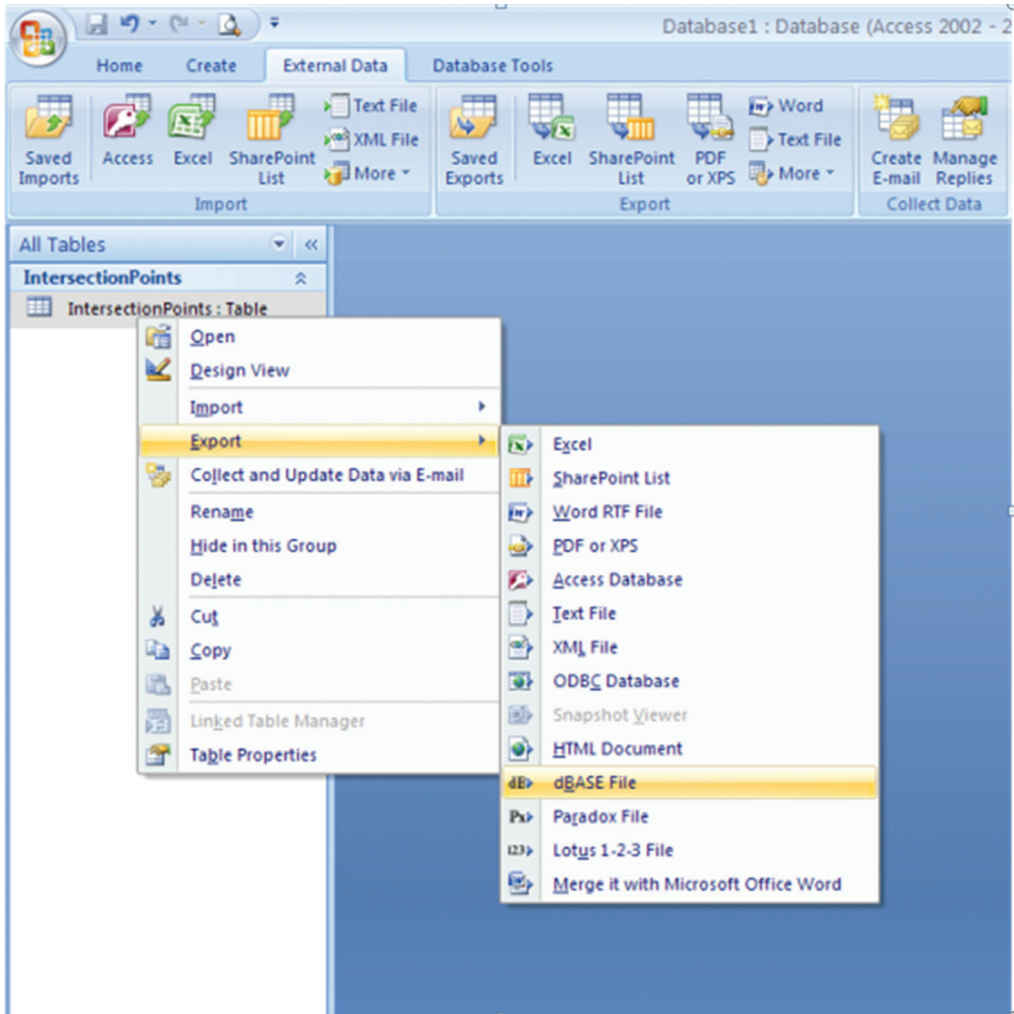Figure 55   Importing Output file in Access
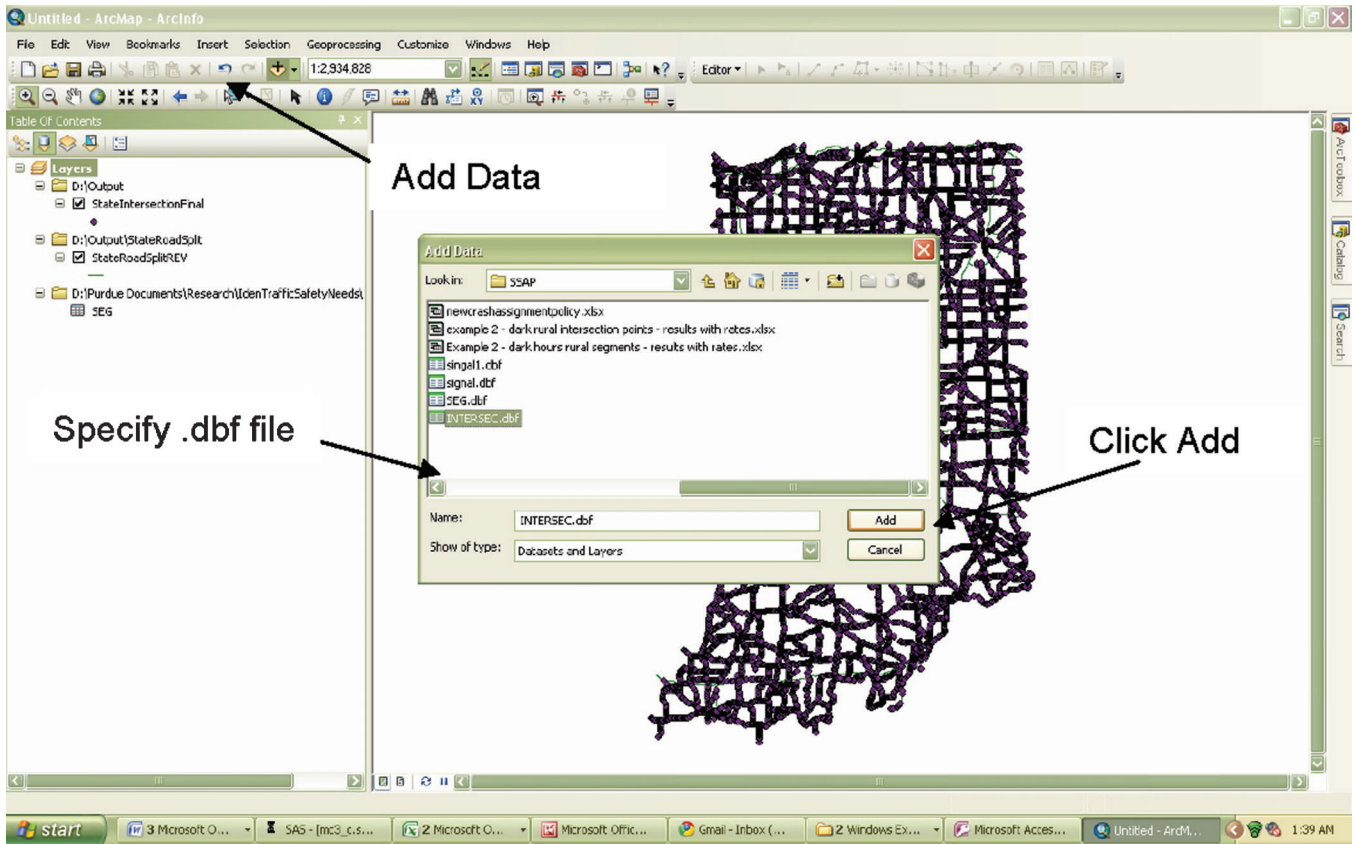
Figure 56   Export Access Table to DBF File
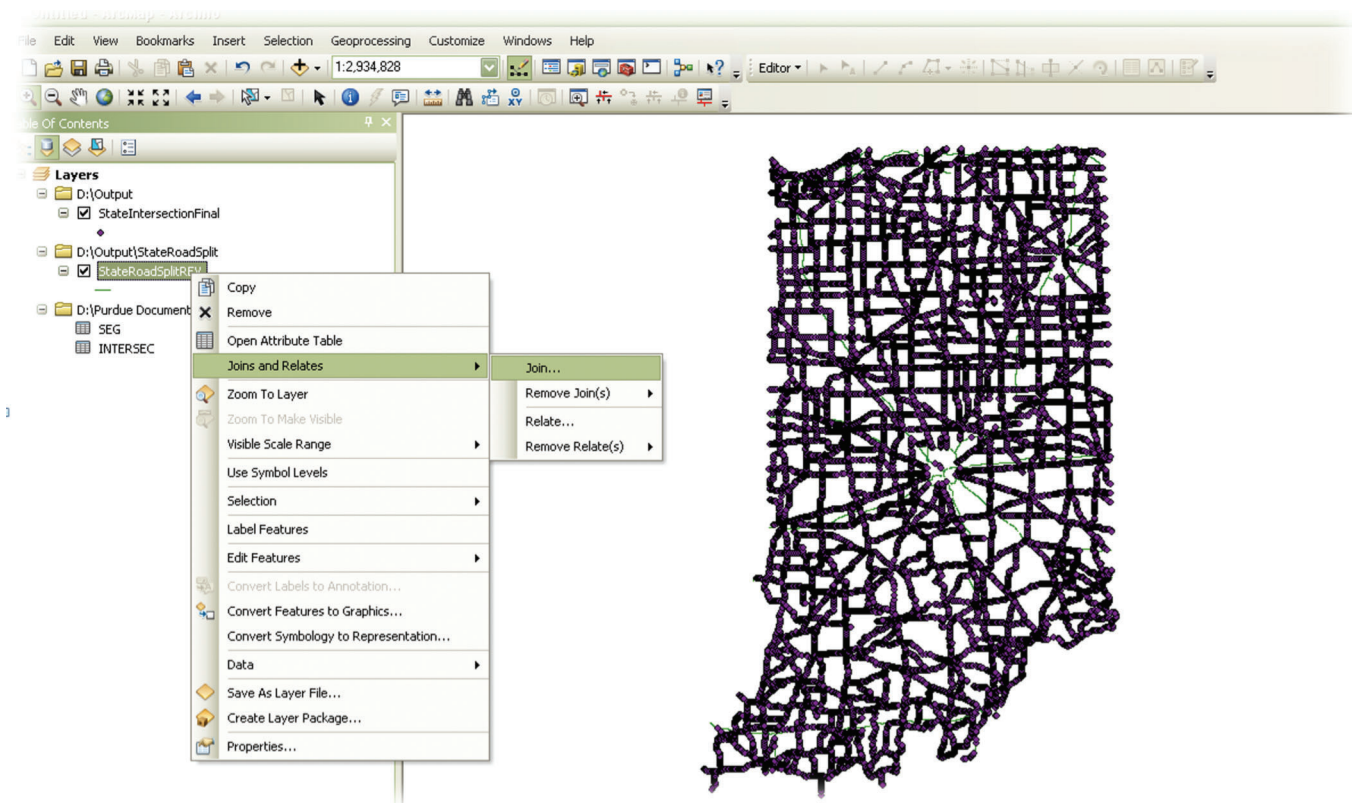
Figure 57   Adding Output Files in ArcMap

Figure 58   Joining Attributes

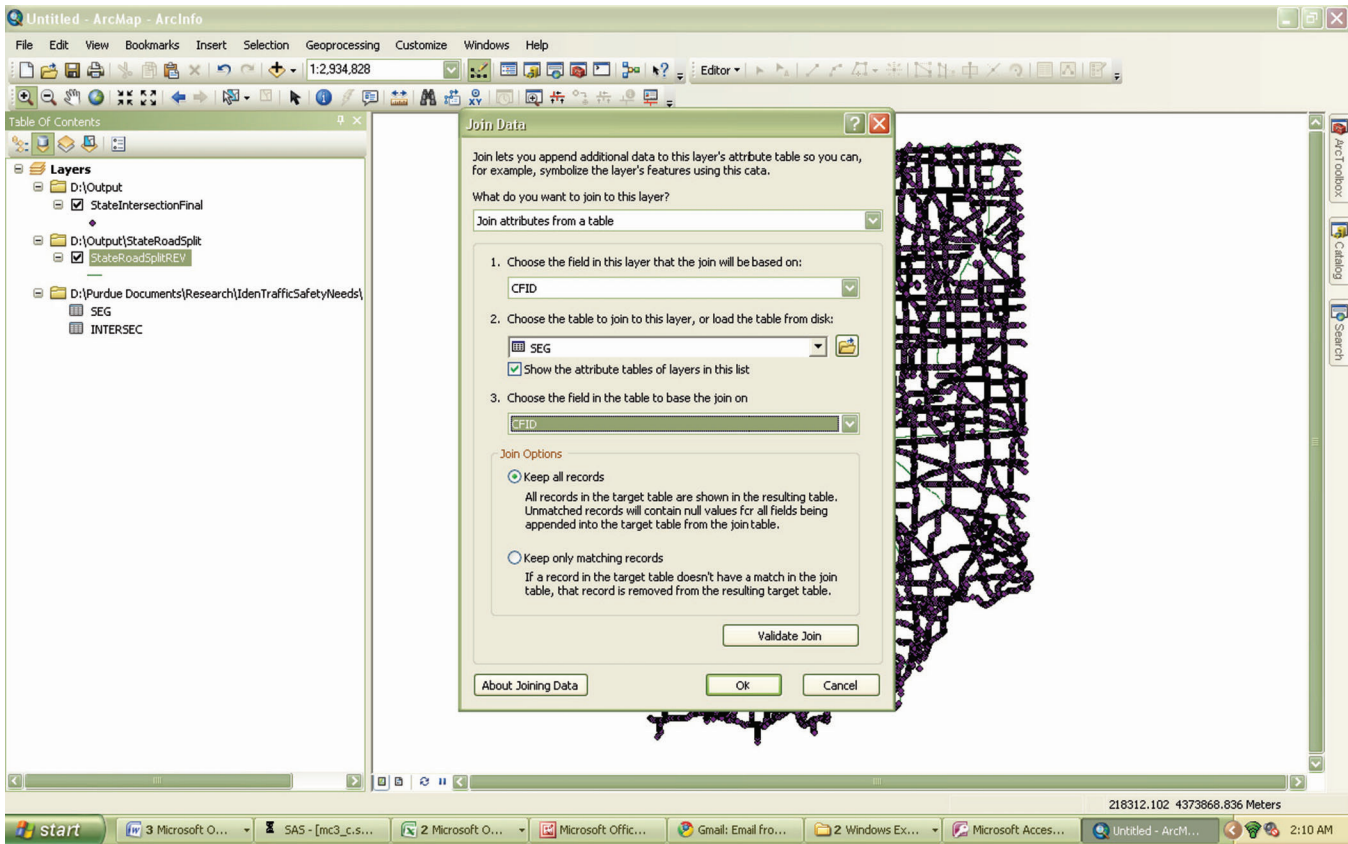Figure 59   Specifying Inputs for Joining

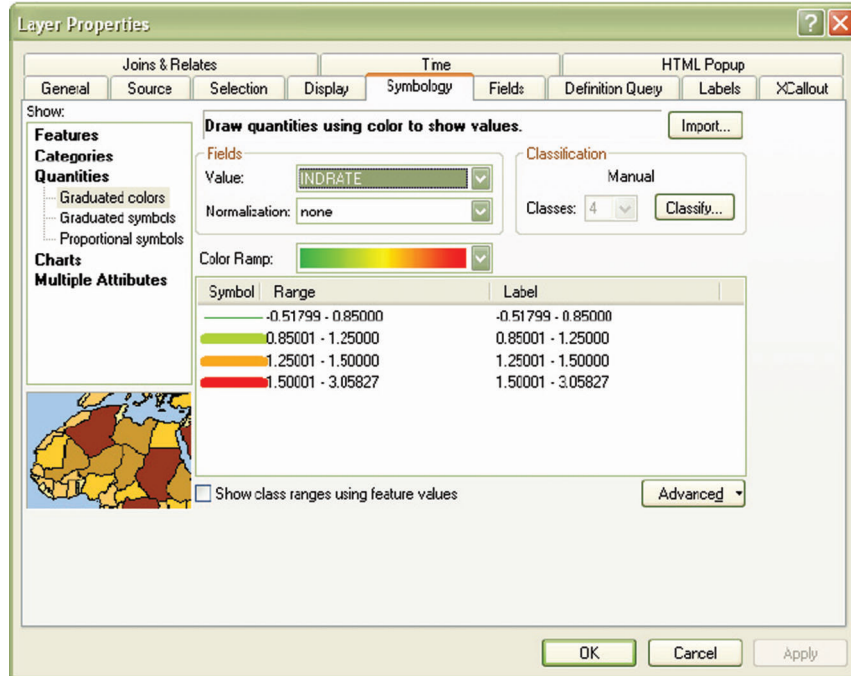Figure 60   Joined Table View (Segment Table + Segment Screening Output)

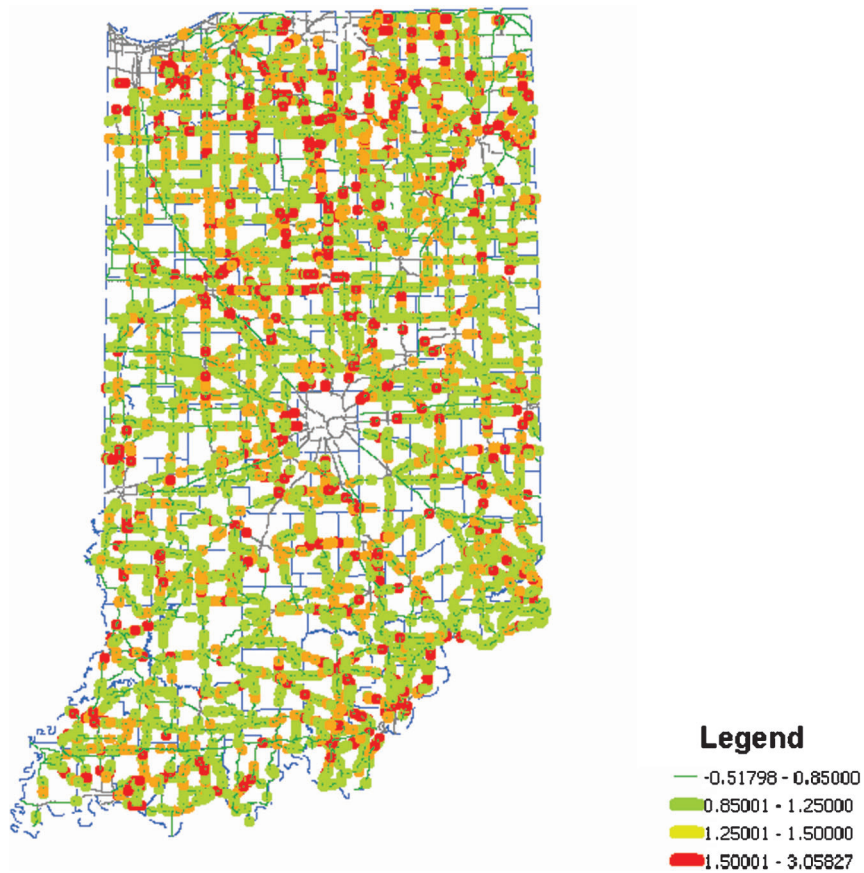Figure 61   Selecting Settings for Symbology (Rural Road Segments Crash Rate under Dark Not-lighted Conditions)
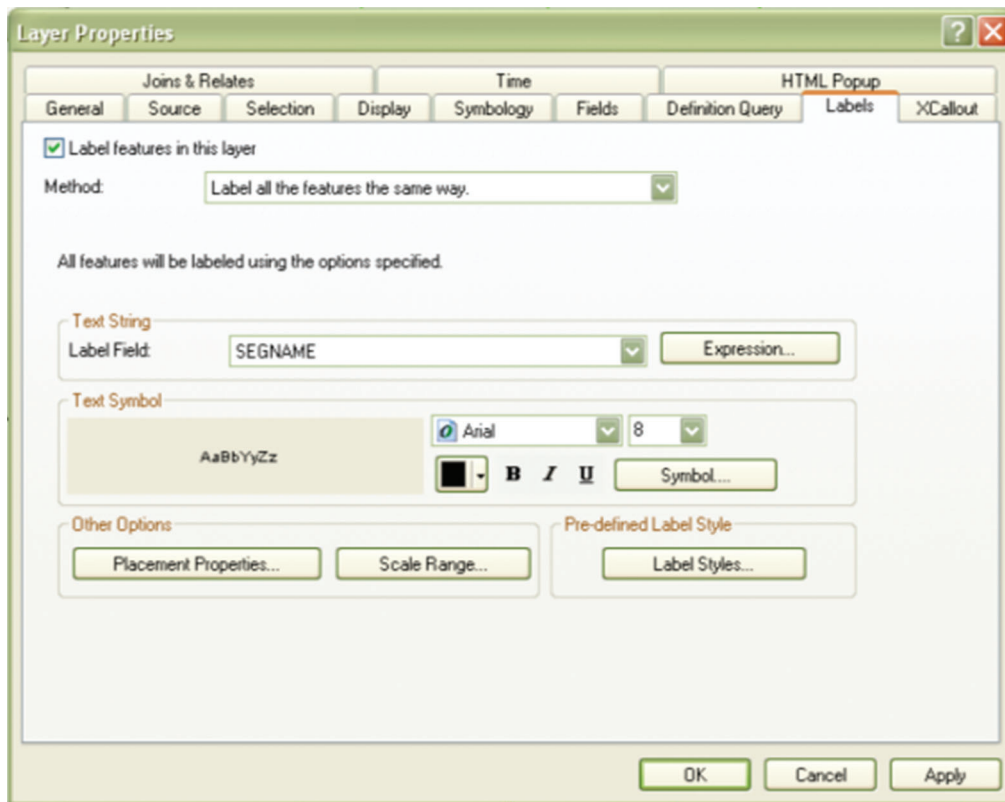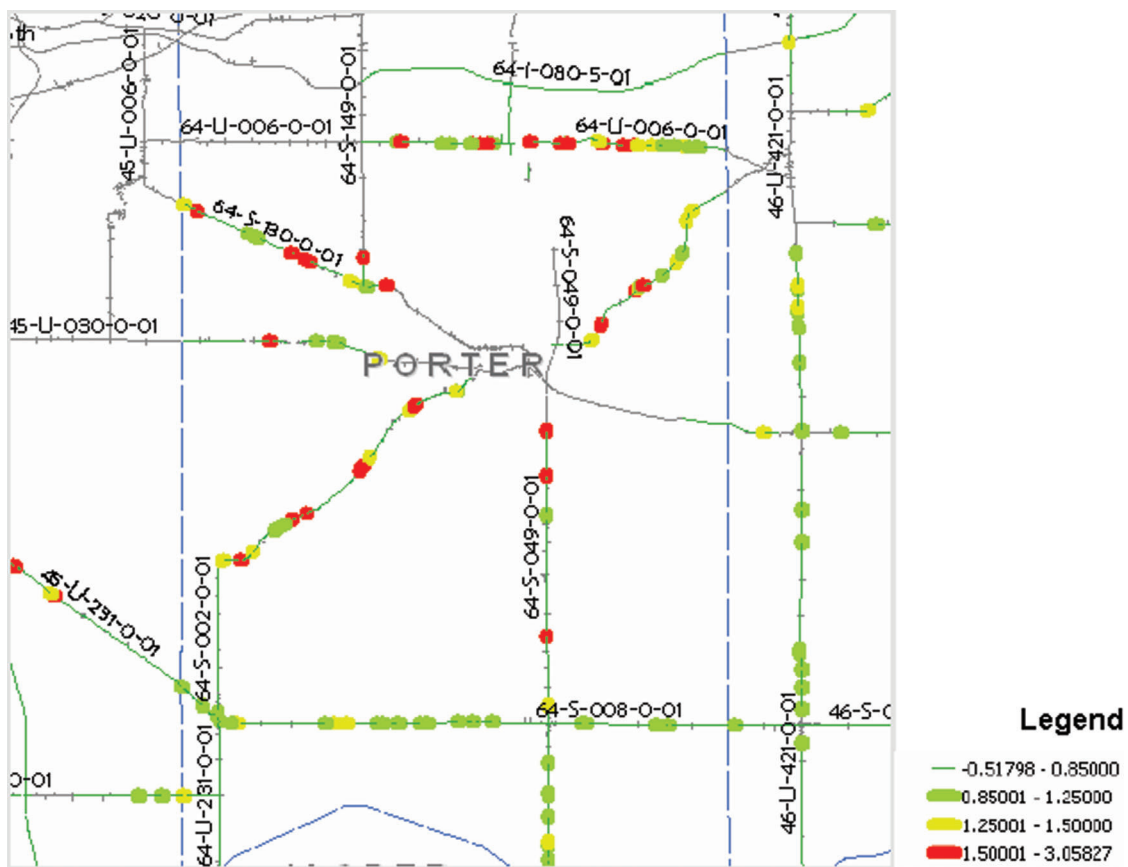

Figure 62   Dark-unlighted Crash Rates on Rural Two-lane and Multilane Roads

Figure 63   Labeling Segment Names

Figure 64   Porter County: High-crash Rural Locations in Dark Conditions
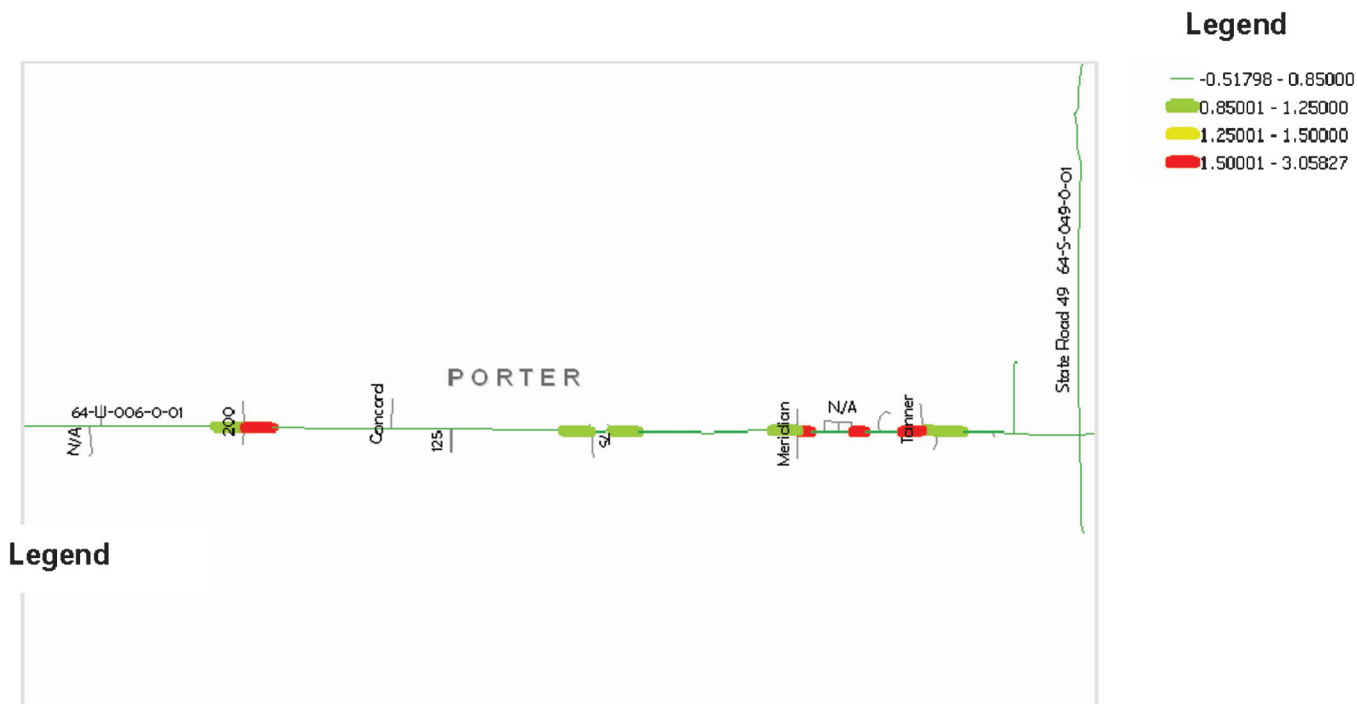
with the SNIP embedded visualization tool based on the Google Earth satellite images, while the bottom images have been created using the ArcGIS tools and the shape files taken from the CRS database. The minor differences between the upper and bottom results have been caused by using different ranges of I values to display the color codes. Figure 67 shows the screening results before clustering, and Figure 68 shows the results after clustering for Tippecanoe County. Figure 69 shows the screening results before clustering and Figure 70 after clustering for two segments crossing each other sections on routes US-52 and IN-26 in Lafayette.
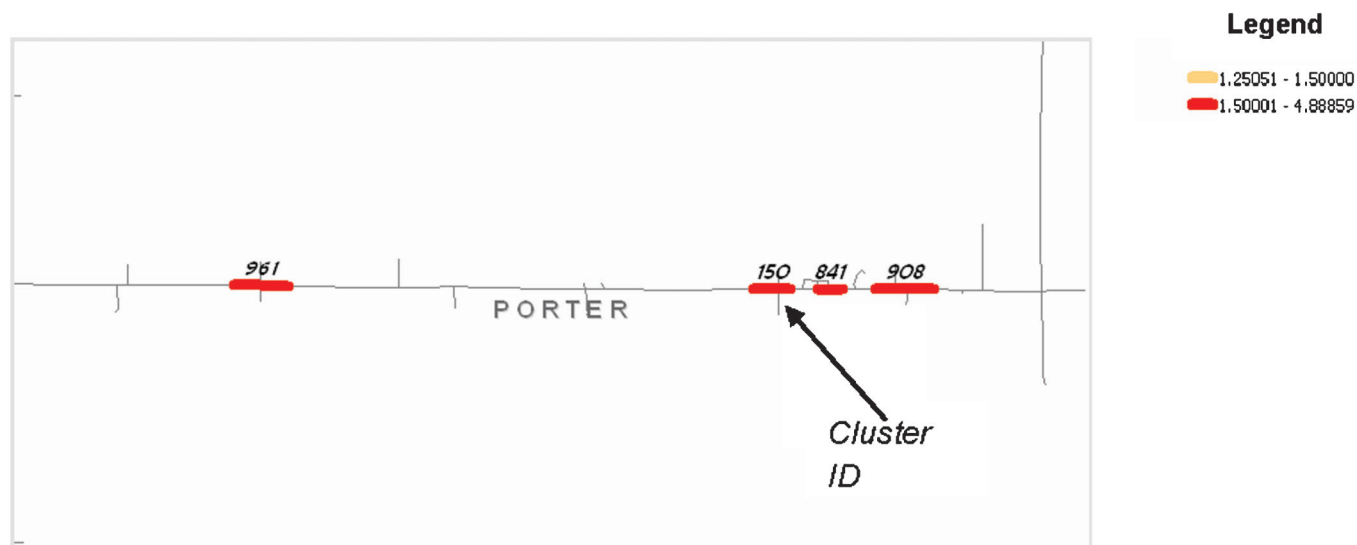
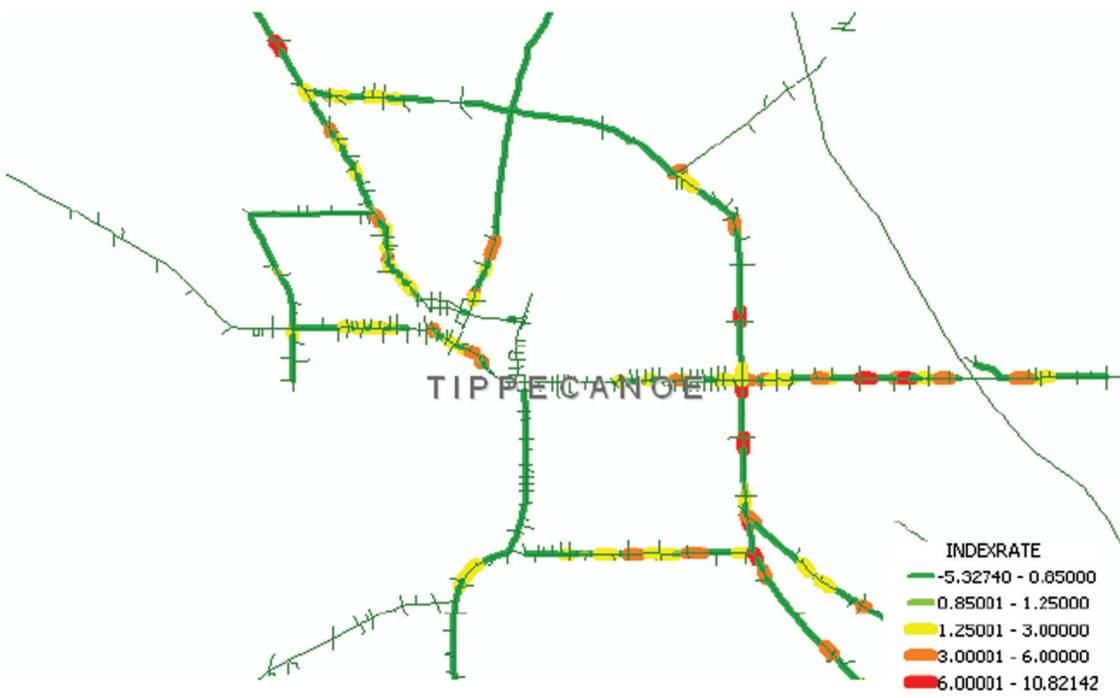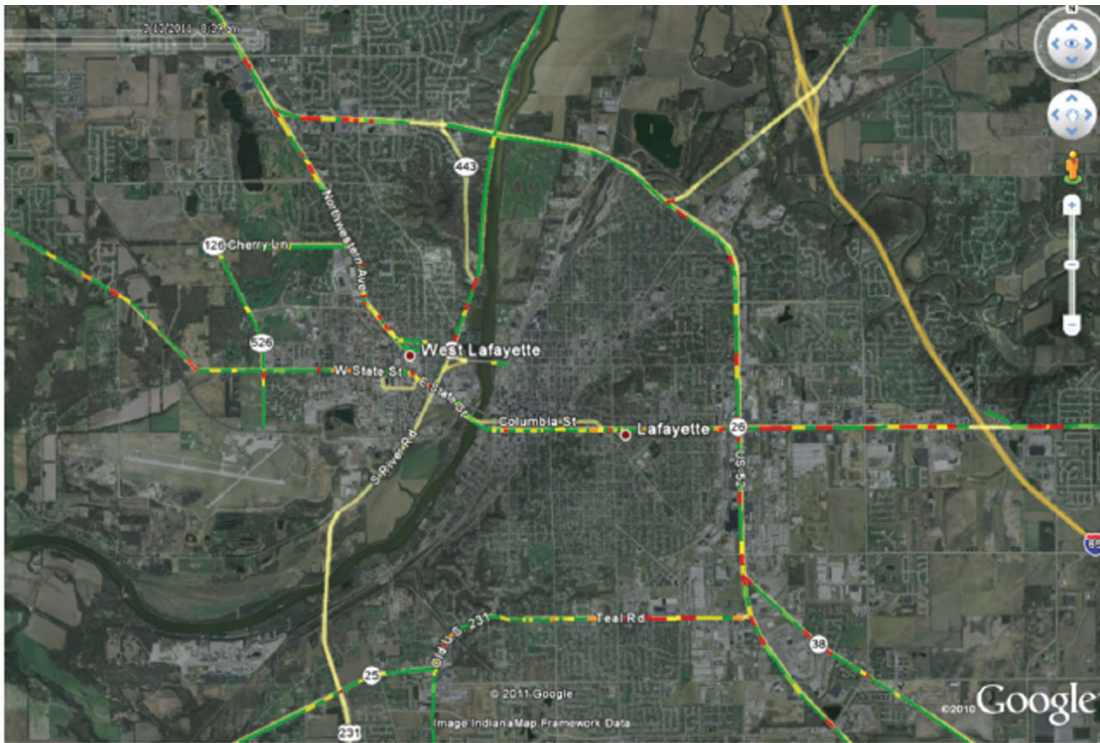Figure 66   Clustered Segments Having High Index of Crash Rate, Near Valparaiso, IN

Figure 67   Segments before Clustering; Rear-End Crash Rates, Tippecanoe County

INDEXRATE
— -5.32740 - 0.85000
— 0.85001 - 1.25000
— 1.25001 - 3.00000
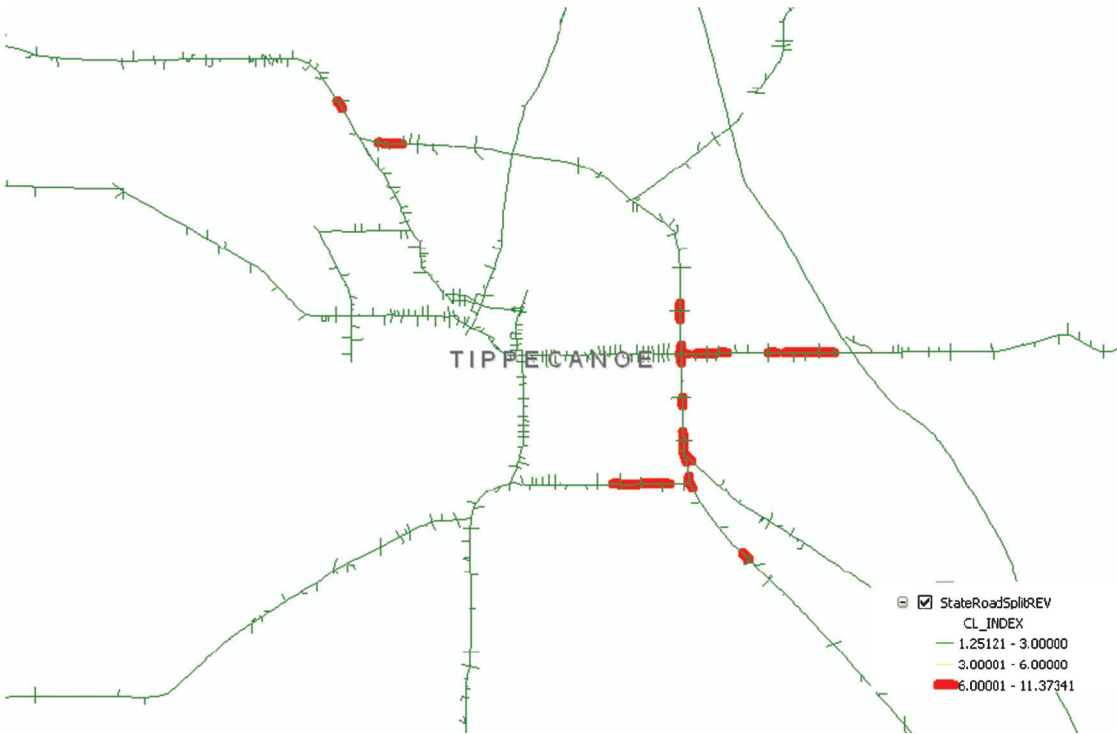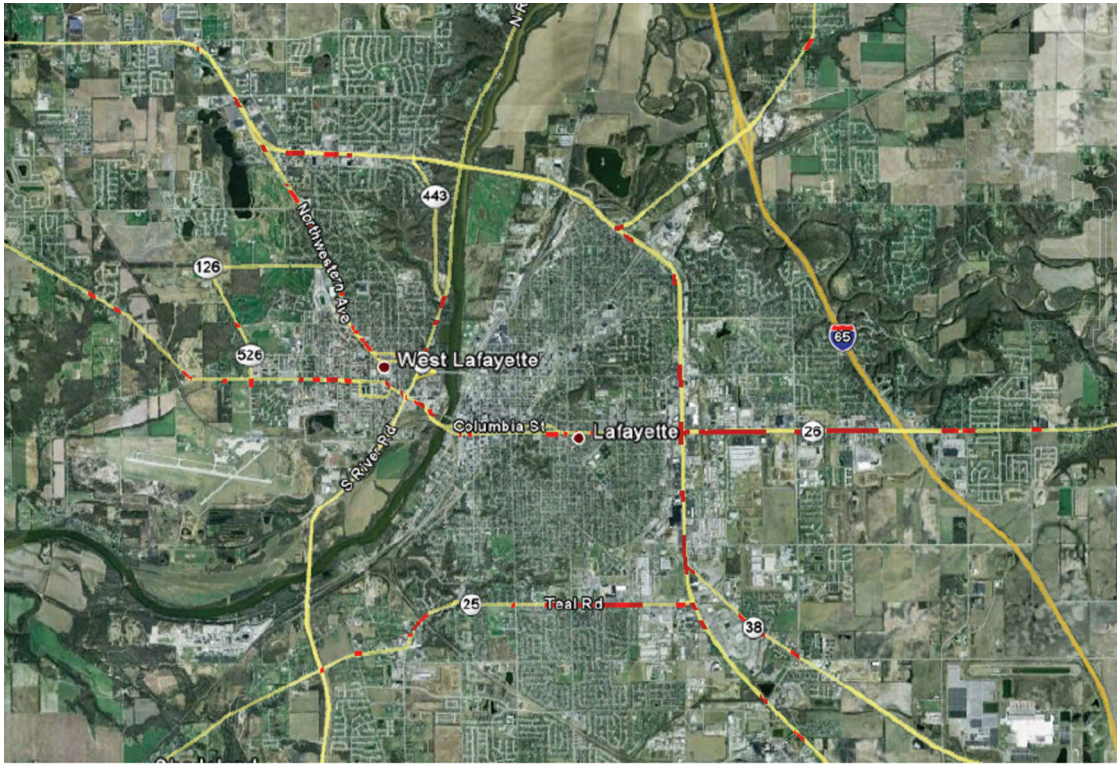— 3.00001 - 6.00000
— 6.00001 - 10.82142

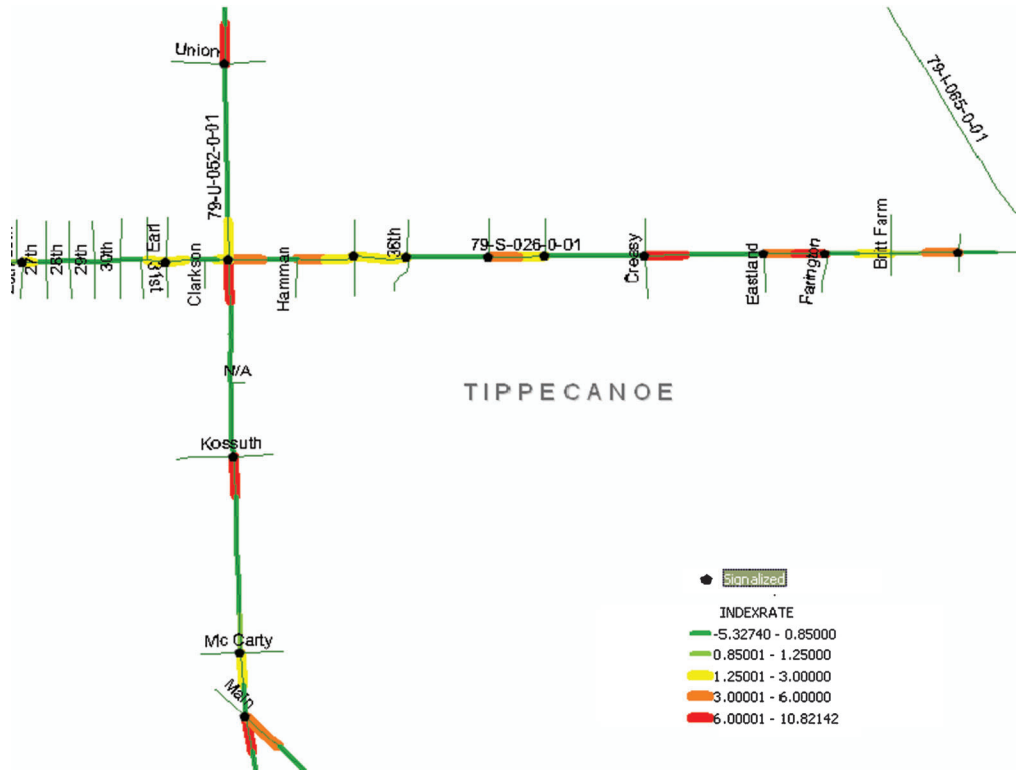Figure 68   Segments after Clustering; Rear-End Crash Rates, Tippecanoe County
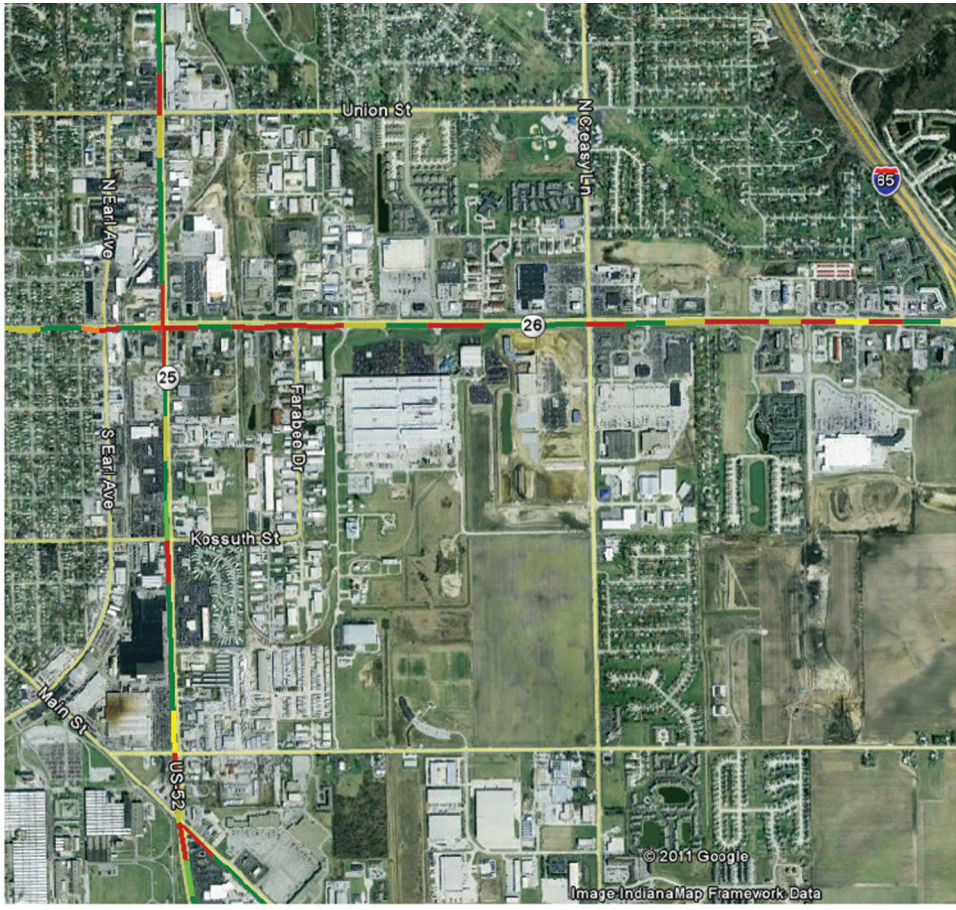
Figure 69   Segments before Clustering; Rear-End Crash Rates, US-52 and SR 26 in Lafayette

Figure 70   Segments after Clustering; Rear-End Crash Rates, US-52 and SR 26 in Lafayette