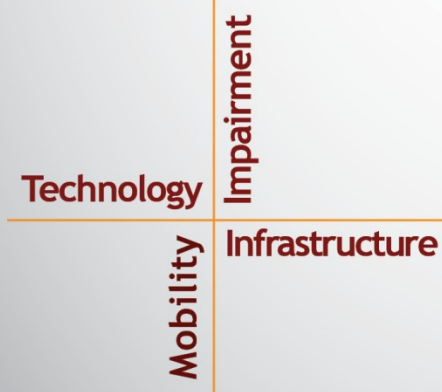# NSTSCE

National Surface Transportation Safety Center for Excellence

Integrating the Adaptive Lighting Database with SHRP 2 Naturalistic Driving Data

Yingfeng Li • Ronald B. Gibbons • Alejandra Medina

Submitted: September 16, 2015

Impairment

Technology

Infrastructure

Mobility

ACKNOWLEDGMENTS

EXECUTIVE SUMMARY

Recently, the Federal Highway Administration (FHWA) sponsored the project "Strategic Initiative for the Evaluation of Reduced Lighting on Roadways" to develop more sophisticated adaptive lighting systems that respond to real-time traffic and environmental factors. As part of this project, Virginia Tech Transportation Institute (VTTI) researchers developed an Adaptive Lighting Database (ALD) with detailed in situ lighting performance data and associated safety, traffic, and roadway data for seven states: Washington (WA), North Carolina (NC), California (CA), Delaware (DE), Minnesota (MN), Vermont (VT), and Virginia (VA). That study utilized a Roadway Lighting Mobile Measurement System (RLMMS) to collect detailed in situ measurements of in- and out- vehicle illuminance, and glare illuminance measured vertically from within the vehicle.

During approximately the same time frame when the lighting data were collected, the Second Strategic Highway Research Program (SHRP 2) carried out a large data collection effort as part of the Naturalistic Driving Study (NDS) project. The data collection focused on six sites around the nation, including Bloomington, Indiana; central Pennsylvania; Tampa Bay, Florida; Buffalo, New York; Durham, North Carolina; and Seattle, Washington. The collected naturalistic driving data contained detailed information about driver behavior, driver demographics, and vehicles. In addition, SHRP 2 was developing a Roadway Information Database (RID) that would include detailed traffic and roadway information for the six NDS sites.[8] The availability of NDS, RID, and detailed lighting data made it possible for researchers to investigate the detailed relationship between different lighting characteristics, roadway configurations, and roadway safety. However, before any meaningful data analyses could be conducted using these datasets, linkages needed to be established among them to enable seamless data integration.

This project was sponsored by the National Surface Transportation Safety Center for Excellence (NSTSCE) to address the aforementioned challenge and to make available an integrated dataset containing detailed naturalistic driving, roadway, and lighting information for researchers, the transportation community, and other interested organizations and individuals. During this project, the research team completed the following major activities:

- Developed an in-depth description of the NDS database structure, data elements, and database relationships;
- Documented in detail the data entities, format, and content of the VTTI ALD; and
- Developed and demonstrated two Geographic Information System (GIS)-based approaches for integrating NDS and ALD data.

During this project, the research team acquired access to the SHRP 2 NDS database and conducted a detailed review of the database structure, data entities, relationships, and contents. The research team also described the VTTI ALD in detail to provide the foundation for the development of systematic data integration approaches. The review of the two databases suggested that both databases contained internal spatial references that could be utilized to facilitate data integration. For the NDS database, the time series data contained Global Positioning System (GPS) coordinates that would allow users to pinpoint their locations. All other data tables within the database, such as the event, trip, driver, and vehicle data, are either

directly or indirectly linked to time series data. In addition, within the VTTI ALD, the detailed lighting measurement data also contained GPS coordinates defining the location where an individual measurement was taken. As such, the research team decided to explore methods that could spatially integrate the two databases.

Based on a thorough understanding of the database structures, the research team developed two spatial data integration approaches, each of which targeted different needs and requirements:

- Data integration directly between the ALD and NDS time series data. The researchers demonstrated how time series data could be extracted from the NDS database and directly matched to the ALD roadway network. By matching both lighting and time series data points onto the same roadway network, simple spatial joins or linear referencing mechanisms could be used to relate individual points from both datasets. The spatial data integration approach involved both existing and custom tools developed on the ArcGIS platform. Other data items in both databases could then be integrated together through spatial and relational joins. The researchers used data for the State of Washington to demonstrate the approach and associated advantages and challenges. During this activity, time series data representing approximately 2,800 nighttime NDS trips were matched to the ALD roadway network.

- Data integration of the ALD and NDS data based on RID. Within the NDS database, time series data points are matched to a digital map of the roadway network defined by links (uniquely identified by a LINKID). That matching effort provides researchers with the ability to directly isolate time series data collected on the links of interest, eliminating the need for additional spatial processing. The purpose of this approach was to utilize those results to enable more efficient data integration. Instead of using the NDS time series data, the research team matched the RID roadway network with the ALD roadway network. After the process, each ALD roadway segment was assigned with the LINKID of the corresponding RID roadway segment. The LINKID information within both databases can then be used to link data elements by performing relational database joins, which are many orders of magnitude faster than spatial joins. To demonstrate this approach, the research team used a draft version of the RID roadway data and lighting data for the State of North Carolina.

The first integration method illustrates a process to link NDS data directly to custom roadway data (e.g., the ALD roadway network and lighting data). This linking provides a flexible means to cater to unique data integration needs, especially when requiring roadway-related data elements that are not included in the SHRP 2 RID data. Similarly, the approach can be also used to link NDS data with roadways that are not included in the current RID database. However, the first approach requires the matching of a large number of time series data points spatially to the roadway network. This effort can be time-consuming, especially when using personal computers and desktop GIS software packages. In contrast, the latter integration method utilizes the roadway matching results completed by VTTI under contract to SHRP 2 and therefore requires much less demanding computing resources and processing time.

In summary, this research served as the bridge for integrating the VTTI ALD and SHRP 2 NDS data and provides insights in how other spatially related data can be integrated with the NDS data to cater custom safety study needs. By integrating detailed roadway lighting measurement data with the rich SHRP 2 NDS information, a large variety of safety analyses become possible.

Based on this project, the researchers recommend the following:

- SHRP 2 NDS-related improvements:

  o Develop comprehensive, user-friendly documentation particularly around the integration of datasets relevant to the SHRP 2 NDS database. As would be expected of an ongoing and dynamic project, documentation is needed, particularly in areas such as logical and physical data models depicting data entities, attributes, and relationships.

  o Consider incorporating documentation from this project into the SHRP 2 NDS documentation assuming the final data structure is similar to the current one. The data integration approaches developed in this project provide examples of how custom roadway or other spatially referenced data can be integrated with NDS data with minimal modification. Inclusion of documentation from this effort into training materials to facilitate users with similar needs could be beneficial to current and potential NDS data users.

  o Facilitate researchers in understanding how to fully use the SHRP 2 roadway matching results. In many research efforts, once time series data on roadway segments of interest are retrieved from the SHRP 2 Linking Table, users need to further locate data points within individual links. This project uses GIS techniques to match individual time series data points spatially onto roadway segments. The SHRP 2 matching results also enable users to use speed and timestamp data of the time series data points as a linear referencing mechanism to compute their exact locations within a roadway segment. Providing researchers with documentation (including training materials) of how to perform such tasks will permit users to integrate data with high efficiency and enable location-specific analysis as required.

- ALD-related improvements:

  o Develop more aggregated lighting measurements to reduce data processing demand. The VTTI lighting measurements were collected in situ at a frequency of 20 Hz. This practice resulted in a large number of lighting data points densely located next to each other. To facilitate potential users, it would be beneficial to aggregate consecutive lighting measurement points with similar values into single line features. This process significantly reduces the number of lighting data records without sacrificing much of the data accuracy.

  o Linearly reference lighting data points. The ALD research team should linearly reference the lighting measurement data points onto the official state roadway

networks as used in the database. Embedding the data points into the state official linear referencing system will enable seamless integration between the lighting measurements and other datasets maintained at state Departments of Transportation (DOTs), such as roadway information, traffic data, crash data, infrastructure data, and roadway project or work zone data. Based on the linear referencing information, such data integration can be performed accurately not only with a GIS but also using relational database management tools.

- Further studies and next steps. The integration of the VTTI in situ lighting measurements and the SHRP 2 NDS data has enabled opportunities for a number of meaningful studies towards a safer and more driver-friendly nighttime driving environment. Among those potential studies, the researchers recommend that the following high-priority studies be conducted in the near future:

  o Effects of detailed lighting characteristics on nighttime crashes and driver behavior. This analysis will require a comprehensive causation study of roadway lighting characteristics and nighttime crashes, near-crashes, speeds, and other surrogate measures of safety.

  o Nighttime traffic control visibility and effectiveness at different lighting levels. This study will involve an in-depth investigation of the visibility and effectiveness of certain traffic control measures as they relate to roadway safety. The detailed RID roadway and traffic control information, NDS sequence of events and video data, and VTTI in situ lighting performance measurements provide a unique opportunity to complete this study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

AADT	Average Annual Daily Traffic
ALD	Adaptive Lighting Database
ATR	Automatic Traffic Recorder
CA	California
Caltrans	California Department of Transportation
DE	Delaware
DOT	Department of Transportation
FHWA	Federal Highway Administration
GIS	Geographic Information System
GPS	Global Positioning System
HPC	High Performance Computing
HSIS	Highway Safety Information System
IRB	Institutional Review Board
MN	Minnesota
NC	North Carolina
NDS	Naturalistic Driving Study
NSTSCE	National Surface Transportation Safety Center for Excellence
QA/QC	Quality Assurance/Quality Control
RAM	Random Access Memory
RID	Roadway Information Database
RLMMS	Roadway Lighting Mobile Measurement System
SAFETEA-LU	Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users
SHRP 2	Second Strategic Highway Research Program
SQL	Structured Query Language
TMC	Traffic Message Channel
VA	Virginia
VT	Vermont
VTTI	Virginia Tech Transportation Institute

CHAPTER 1. INTRODUCTION

BACKGROUND

More than 50% of all fatal crashes occur at night, even though nighttime traffic volumes only constitute approximately 25% of all traffic. The resulting fatality rate for drivers is three times higher at night.[1,2,3] In general, adding roadway lighting has been considered a crash countermeasure, and there are currently more than 13 million streetlights nationwide.[4] However, alongside the desire at state Departments of Transportation (DOTs) to promote traffic safety is the desire to save energy consumption, especially when many of these agencies continue experiencing funding shortages.

Previous research on the safety impact of roadway lighting mostly focused on how the presence of lighting affected crash rates by comparing highways with and without lighting and the relationship between day and night crashes.[5] Such studies lacked the support of detailed lighting measurement data and therefore could not identify exactly what level of lighting is needed to maintain safety. One early study that was conducted to determine the relationship between crash rates and illumination levels[6] used data on 203 miles of sample roadways but could not determine a causal relationship between illumination levels and crash rates.

Many lighting factors can affect crash risk, such as the level of vertical and horizontal illuminance, roadway luminance, and lighting uniformity. The problem becomes more complex when variables such as human perception and behavior, roadway configuration, and traffic control measures are considered. The limited information with traditional safety and lighting data makes it extremely difficult to explore the complex interactions between characteristics of roadway lighting, crash risks, and driver behavior—yet this knowledge is urgently needed by transportation agencies that want to reduce energy consumption while maintaining safety.

Recently, the Federal Highway Administration (FHWA) sponsored the project "Strategic Initiative for the Evaluation of Reduced Lighting on Roadways" to develop more sophisticated adaptive lighting systems that respond to real-time traffic and environmental factors.[7] As part of this project, Virginia Tech Transportation Institute (VTTI) researchers developed an Adaptive Lighting Database (ALD) with detailed in situ lighting performance data and associated safety, traffic, and roadway data for seven states: Washington (WA), North Carolina (NC), California (CA), Delaware (DE), Minnesota (MN), Vermont (VT), and Virginia (VA). That study utilized a Roadway Lighting Mobile Measurement System (RLMMS) to collect detailed in situ measurements of illuminance inside and outside the vehicle, and glare illuminance measured vertically on the vehicle's windshield.

During approximately the same time frame as when the lighting data were collected, the Second Strategic Highway Research Program (SHRP 2) carried out a large mobile data collection effort as part of the Naturalistic Driving Study (NDS) project.[8] The data collection focused on six sites around the nation, including Bloomington, Indiana; central Pennsylvania; Tampa Bay, Florida; Buffalo, New York; Durham, North Carolina; and Seattle, Washington. The collected naturalistic driving data contain detailed information about driver behavior, driver demographics, and vehicles. In addition, the SHRP 2 Roadway Information Database (RID) provides detailed traffic and roadway information for the six NDS sites.[8] In parallel, SHRP 2 is developing

linkages between the driving and roadway data. The effort will permit researchers to very quickly isolate driving data on particular road segments of interest.[9] The availability of NDS, RID, and detailed lighting data makes it possible for researchers to investigate the complex relationship between different lighting characteristics, roadway configurations, and roadway safety. However, before any meaningful data analyses can be conducted using these datasets, linkages needed to be established among them to enable seamless data integration.

This project was sponsored by the National Surface Transportation Safety Center for Excellence (NSTSCE) to address the aforementioned challenge and to make available an integrated dataset containing detailed naturalistic driving, roadway, and lighting information for researchers, the transportation community, and other interested organizations and individuals.

PROJECT APPROACH

The primary objective of this project was to investigate the feasibility of integrating the VTTI ALD database with the SHRP 2 NDS data. During this project, the research team completed the following major activities:

- Developed an in-depth description of the NDS database structure, data elements, and database relationships;
- Documented in detail the data entities, format, and content of the VTTI ALD; and
- Developed and demonstrated two Geographic Information System (GIS)-based approaches for integrating NDS and ALD data.

The data integration approaches developed during this research made available an integrated dataset with rich driving, roadway, and in situ lighting performance information. This integrated dataset enables a large variety of analyses that may yield results of great significance to safety researchers, transportation agencies, and the traveling public. First, the results will help the transportation community understand in detail how lighting variables (including absence of lighting) contribute to crashes and driver behavior. Second, the results will help DOTs develop more effective lighting (not necessarily higher level of lighting) at different roadway settings to reduce nighttime crashes. Finally, the results will enable more efficient lighting design standards that minimize roadway energy consumption without sacrificing safety.

REPORT STRUCTURE

The rest of this report contains the following chapters:

- Chapter 2: Naturalistic Driving Study Data. This chapter describes in detail the SHRP 2 NDS study and the NDS database, including database structure, data entities, relationships, and contents as they relate to the VTTI lighting database.

- Chapter 3: Adaptive Lighting Database. This chapter provides a detailed background about the in situ lighting performance data VTTI collected as part of the FHWA project, including the data collection approach, data format, contents, and data elements.

- Chapter 4: NDS and Lighting Database Integration. In this chapter, the researchers describe in detail the two GIS-based approaches developed for integrating the lighting data spatially with the NDS data in two separate scenarios. The first approach directly relates the NDS time series data points to the lighting data points using GIS tools. The second approach utilizes the RID roadway network and provides a spatial linkage between the two databases by matching the ALD roadway network with the RID network. The first approach demonstrates a method for locating NDS data points on a custom roadway network using GIS techniques, while the second partially utilizes the SHRP 2 roadway data matching effort and therefore requires less time for data integration.

- Chapter 5: Conclusions and Recommendations. This chapter contains a brief summary of the project, including its major research activities and the conclusions and recommendations drawn based on the research findings.

CHAPTER 2. NATURALISTIC DRIVING STUDY DATA

BACKGROUND

In 2005, Congress authorized SHRP 2 through the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU).[10] SHRP 2 focuses on four research areas: safety, renewal, reliability, and capacity.[11] The program was originally authorized through fiscal year 2009 and continuing resolutions extended the research program to March 2015.

The primary goal of the SHRP 2 safety research is to address the role of driving performance and behaviors impacting traffic safety.[12] Since its establishment, SHRP 2 has funded a series of safety projects in the context of a large-scale naturalistic driving study. Figure 1 shows a general overview of the SHRP 2 safety projects according to the SHRP 2 safety research plan revised in February 2012.[13] Readers should note that the figure does not reflect the exact number, content, and timing of the awarded contracts. For example, as shown in Figure 1, the last S01 project was finished in late 2010 instead of early 2009.[14]
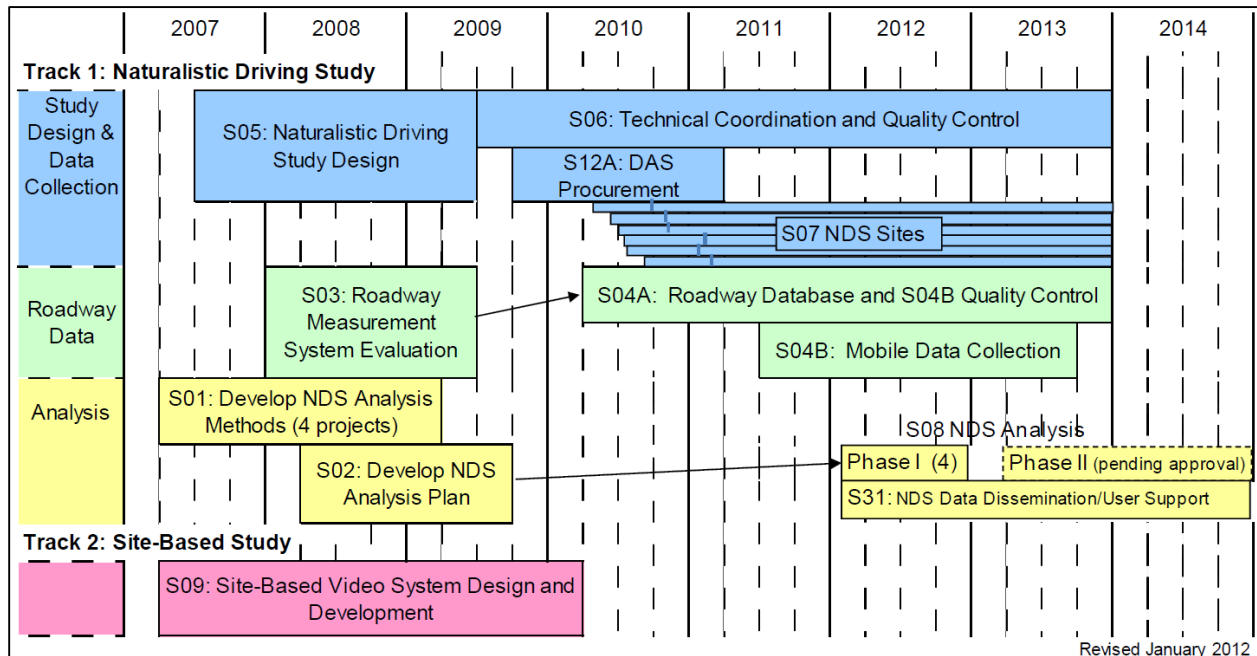
Figure 1. Diagram. SHRP 2 safety projects overview.[13]

SHRP 2 NATURALISTIC DRIVING STUDY DATA

The NDS field studies were conducted between 2010 and 2014. Data collection efforts were carried out at six sites:[15]

- Bloomington, Indiana;
- Central Pennsylvania;
- Tampa Bay, Florida;

- Buffalo, New York;
- Durham, North Carolina; and
- Seattle, Washington.

During data collection, each study vehicle was instrumented with a data acquisition system consisting of a forward radar, four video cameras, accelerometers, Global Positioning System (GPS) receivers, computer vision lane-tracking capability, and data storage equipment (Figure 2).
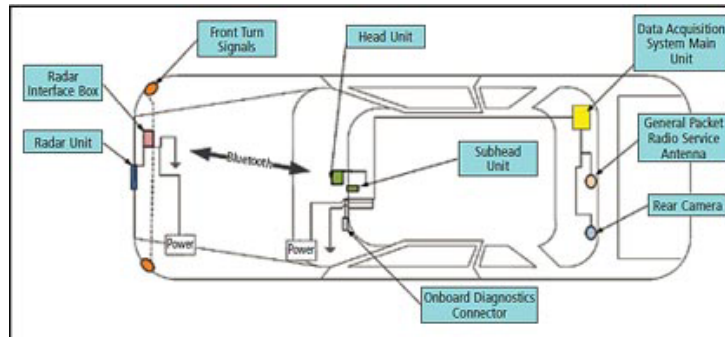
Figure 2. Diagram. NDS data acquisition system schematic view.[15]

The final NDS dataset contains several types of data files and was expected to approach 2 PB in size. As of May 2014, the SHRP 2 NDS data website[16] showed a total number of approximately 3,400 vehicles and 3,600 drivers that participated in the study.

In parallel with the naturalistic driving data collection efforts, two separate projects were carried out to collect necessary roadway information at the six NDS sites.[15] The resulting RID incorporated both roadway inventory data maintained by state DOTs and data collected by instrumented vehicles on selected routes.

The following is a summary of the major data files in addition to the recorded video and imagery data in the current NDS database:[16]

- Driver data, which include several types of data describing the study participant drivers, such as demographic information, driving history and skills, and interview/test results of physical and psychological characteristics that may affect their driving.

- Vehicle data, which include detailed variables describing each vehicle that was instrumented for the study.

- Time series data, which include a set of vehicle/event descriptors collected from the participating vehicles with the onboard data acquisition system at various frequencies, such as speeds, acceleration rates, coordinates, and vehicle conditions. All data points of the same time series file were assigned with the same file ID.

- Trip summaries, which contain summary information of continuous driving data files based on the time series data collected in the field. Each record describes the basic characteristics of an individual trip such as its duration, distance, and maximum speed.

- Event details, which list all identified events of interest, such as crashes, near-crashes, and baseline events (these are data selected for comparison purpose rather than due to the presence of conflicts) and information describing the events.

- Roadway data, which describe basic characteristics of selected roadways at the study sites, such as horizontal and vertical curvature, grade, cross-slope/superelevation, travel lanes, shoulder, and the presence of certain traffic control measures. More details about the NDS roadway data are provided in the following section.

The NDS data are currently hosted at the VTTI International Center for Naturalistic Driving Data Analysis (formerly Smart Data Center).[17] A peta-scale IBM® high performance computing (HPC) storage system facilitates the long-term storage of raw data and processes NDS data while maintaining data in an online status. To improve efficiency, the NDS data are stored in an IBM DB2-based relational database structure that supports data analysis and retrieval through Structured Query Language (SQL).

SHRP 2 ROADWAY INFORMATION DATABASE

As part of the SHRP 2 NDS, Project S04A (Roadway Information Database Developer and Technical Coordination and Quality Assurance of the Mobile Data Collection) and Project S04B (Mobile Data Collection) were launched to develop a roadway information database of the six NDS states to support NDS data-based safety analyses.[18] The resulting roadway database incorporates both roadway inventory data maintained by state DOTs and data collected by instrumented vehicles on selected routes.

The research team obtained draft RID data for North Carolina (although final data submitted to SHRP 2 may vary in format and content) for the purpose of this research. The dataset did not contain detailed documentation (such as standard metadata documents, technical memorandums, or project reports). It was organized in an Esri® file geodatabase and contained several feature classes and database tables listing the values for key attribute fields of the feature classes.

Among the RID feature classes, the Links Polyline feature class depicts all state-maintained roadways and local roadways for selected counties in the Durham-Raleigh region (Figure 3). The key attribute of this feature class is LINKID, which is a unique road segment identifier used across SHRP 2 projects. A separate SHRP 2 effort[9] matched NDS time series data to the same LINKIDs, establishing the linkage between driving data and roadway data. Examples of the other attributes of the Links feature class include from- and to-node IDs, route name, and functional class.
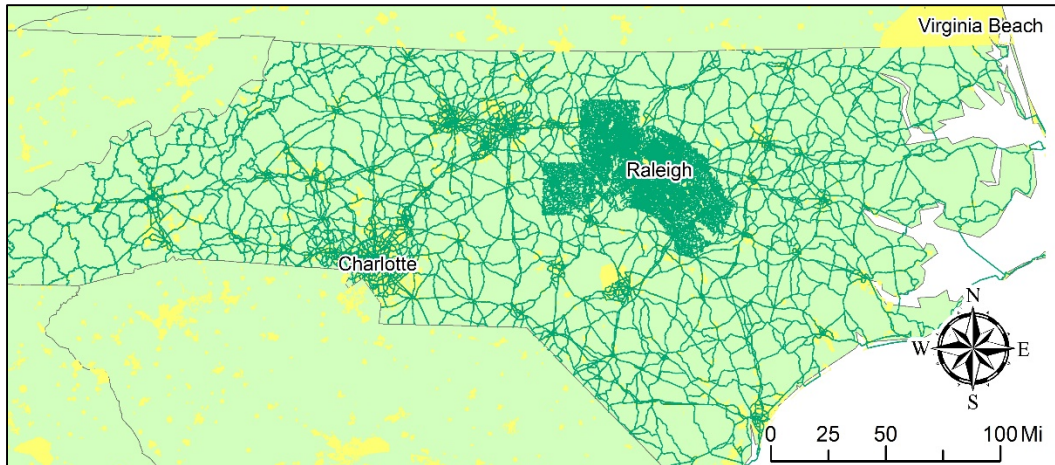
Figure 3. Map. RID Links feature class for North Carolina.

Other feature classes included in the RID database for North Carolina are:

- Routes, which is a Polyline M feature class with the same coverage as that of the Links feature class. The Routes feature class is intended to be used as the route layer for linearly referencing RID roadway data onto the corresponding roadways. The major attributes of the feature class include route name, and from and to measures.

- Sections, which is a Polyline M feature class containing roadway segments (many overlapping), each of which depicts one of the following variables: through lanes, ownership, maintenance operations, facility type, county code, functional system, urban code, access control, and average annual daily traffic (AADT). This feature class covers only state roadways (Figure 4).
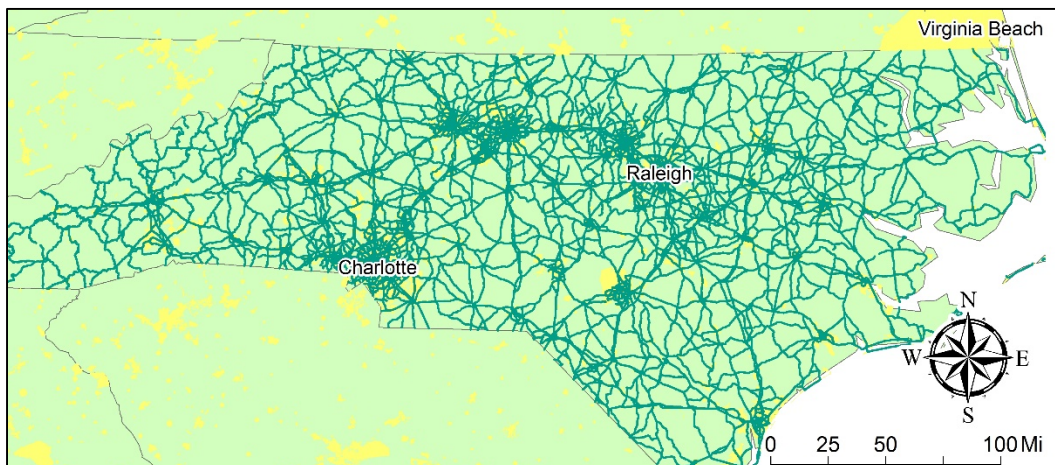
Figure 4. Map. RID Sections feature class for North Carolina.

- TOPS, which is a Polyline feature class depicting several types of information for a scattered sample of roadway segments (Figure 5), such as traffic information (e.g.,

AADT), roadway geometry (e.g., curves and lanes), pavement condition (e.g., rutting and pavement thickness), and other general roadway information (e.g., functional class and facility type).
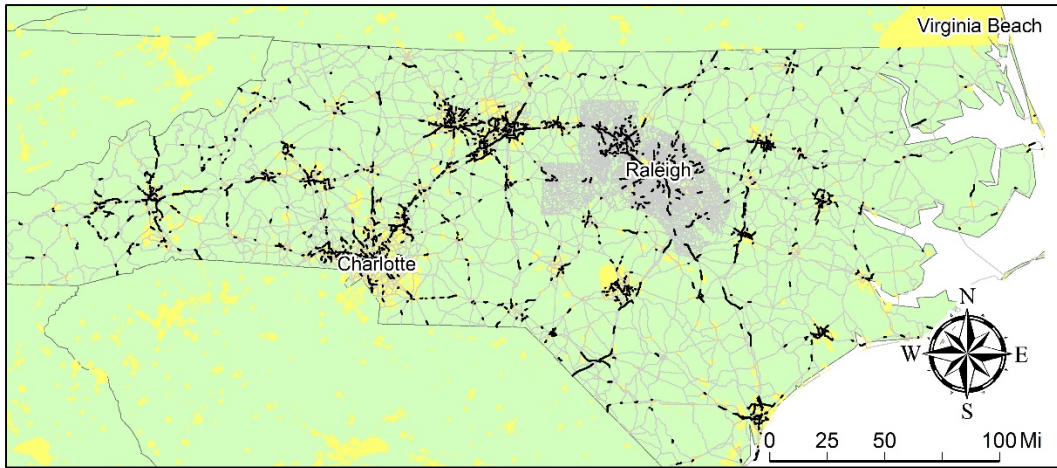
Figure 5. Map. RID TOPS feature class for North Carolina.

- Shoulder, which is a Polyline feature class indicating if a paved shoulder is present for selected roadway segments (Figure 6).
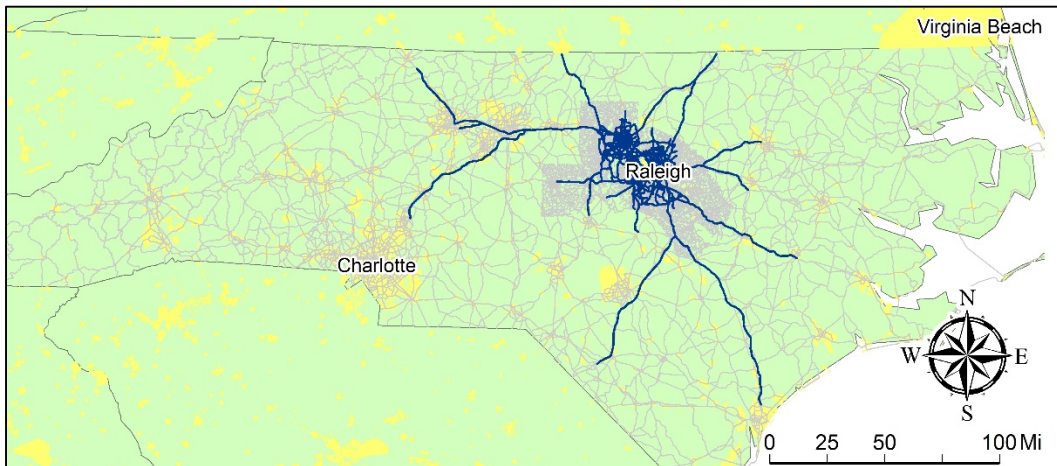
Figure 6. Map. RID Shoulder feature class for North Carolina.

- Lighting, which is a Polyline M feature class indicating the segments of the selected routes shown in Figure 6 that have roadway lights (Figure 7).
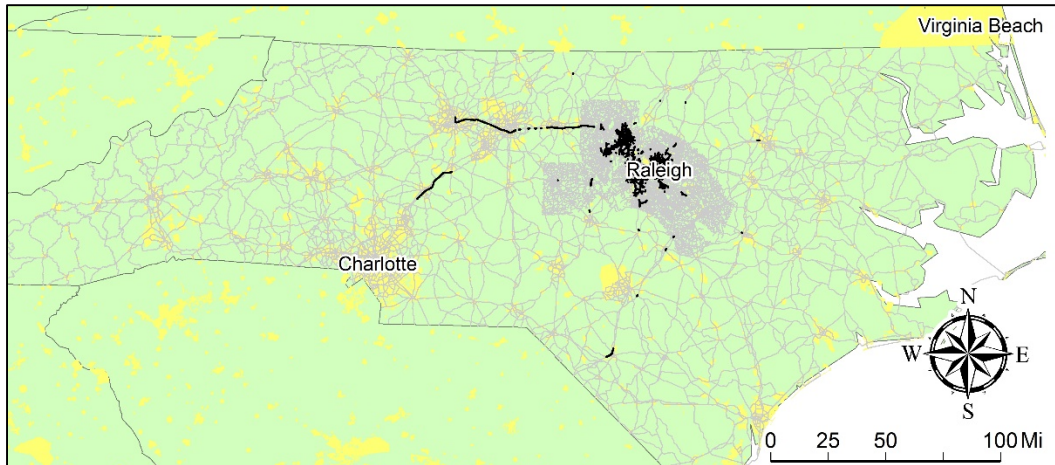
Figure 7. Map. RID Lighting feature class for North Carolina.

- Alignment, which is a Polyline feature class that depicts roadway alignment elements such as tangents, curves, and superelevation for a selected sample of roadways shown in Figure 6.

- Barrier, which is a Polyline feature class depicting roadside barrier information (e.g., barrier type and start/end treatment type) for selected roadways shown in Figure 6.

- Lane, which is a Polyline M feature class that depicts the lane configuration for selected routes shown in Figure 6, such as number of lanes, presence of turn and acceleration/deceleration lanes, and lane width.

- Location, which is a Polyline M feature class depicting the roadway grade and cross slope values along the selected routes shown in Figure 6.

- MedianStrip, which is a Polyline M feature class depicting the roadway median type along the selected routes shown in Figure 6.

- RumbleStrip, which is a Polyline M feature class depicting the location and type of rumble strips along the selected roadways shown in Figure 6.

- RouteIntersections, which is a point feature class depicting the locations and other basic information of the intersections along the selected roadways shown in Figure 6.

- Signs, which is a point feature class depicting the location and type of traffic signs along the selected roadways shown in Figure 6.

- Nodes, which is a point feature class depicting the locations of beginning and ending points based on the Link feature class.

As part of the SHRP 2 project,[9] the association of the driving and roadway datasets was done in a MATLAB® algorithm that interprets the connectivity of the road network, one node to the next. The algorithm snaps the GPS points to road segments identified with the unique LINKID (or LINK_ID in the linking results) and records the timestamp when the vehicle entered the link and when the vehicle departed the link. A detailed description of the data matching process was included in the final project report.[9]

Figure 8 illustrates the results of this matching process. The table identifies the trip data on each link by identifying the file (containing time series data points of the same trip) and the time series data timestamps at the start and end of the link. A sequence number is provided that indicates the order in which the links were traversed in the file. The straight-line distance between the closest time series data point and the digital map node at the start of the link is also provided to indicate how close the point was to the digital map node at the time the vehicle entered the link. Using this linking table, the roadway information for any NDS data point can be identified either directly or through a linear referencing process based on sequential speed and timestamp information.

FILE_ID	SEQUENCE_NO	LINK_ID	TIMESTAMP_START	TIMESTAMP_END	VEHICLE_TO_NODE_DIST
1000019	13	771927354	181461	191461	29.910603771603082
1000019	14	771927355	191461	192461	31.48441077439329
1000019	15	122666697	192461	202461	25.67103791161108
1000019	16	771927351	202461	213462	43.94671697815785
1000019	17	41839022	213462	246462	19.821665018948593
1000019	18	41835835	246462	248462	6.555630753366433

Figure 8. Screen capture. NDS time series point matching result table.[18]

The SHRP 2 linking table stores only the LINKID attribute to provide a more manageable granularity for researchers to access the data. Data containing the LINKID field can readily be joined with the RID roadway information according to the needs of the researcher.

SHRP 2 NDS DATABASE STRUCTURE

After going through a rigorous SHRP 2 data sharing protocol and the VTTI Institutional Review Board (IRB) process (see Appendix A for the IRB Protocol and Appendix B for the Data Sharing Agreement), the research team obtained limited data access to the NDS data. This section describes the database structure of the NDS data as well as how different NDS data items relate to each other. Readers should notice that this description was developed when the SHRP 2 data processing was ongoing. Final database structure, data entities, and relationships may change after the SHRP 2 NDS team finishes data processing.

The SHRP 2 NDS data reside on multiple databases. Most sensor data collected from NDS participating vehicles and their immediate linking tables are stored on the IBM DB2 database server. Intermediate and final processed data, including event and trip data, participant interview results, and participant vehicle data, reside on development and production SQL servers of relatively smaller sizes. During data processing, intermediate results are first stored on development servers to go through a quality assurance/quality control (QA/QC) process. Final

processed data (e.g., events and trips) are released quarterly to the production server linked to the SHRP 2 NDS InSight website.[16]

Figure 9 shows the SHRP 2 NDS database structure with an emphasis on the data items listed on the SHRP 2 NDS InSight website. The figure illustrates how different NDS data items can be linked together with the database. Note that this diagram is not intended to provide an exhaustive depiction of data elements and relationships. In reality, different data items can often be linked in multiple ways. In addition, many variables can be derived from multiple data tables or sources. For example, users may extract vehicle speed based on either GPS speed or vehicle network speed collected by different sensors. Certain trip-related data elements can be either directly available in the processed trip summary table or derived from time series data.

At the time when this document was written, the SHRP 2 NDS InSight website suggested that approximately 3 million trips and almost 10,000 events were processed. The authors were not able to develop an accurate assessment of the progress associated with the roadway–time series data matching effort.

Figure 9. Diagram. SHRP 2 NDS database structure.

CHAPTER 3. ADAPTIVE LIGHTING DATABASE

ADAPTIVE LIGHTING DATABASE

During the approximate time frame when NDS data were collected, VTTI collected detailed lighting performance data and associated information for more than 2,000 miles of highways in seven states: California, Delaware, Minnesota, North Carolina, Vermont, Virginia, and Washington (Figure 10).
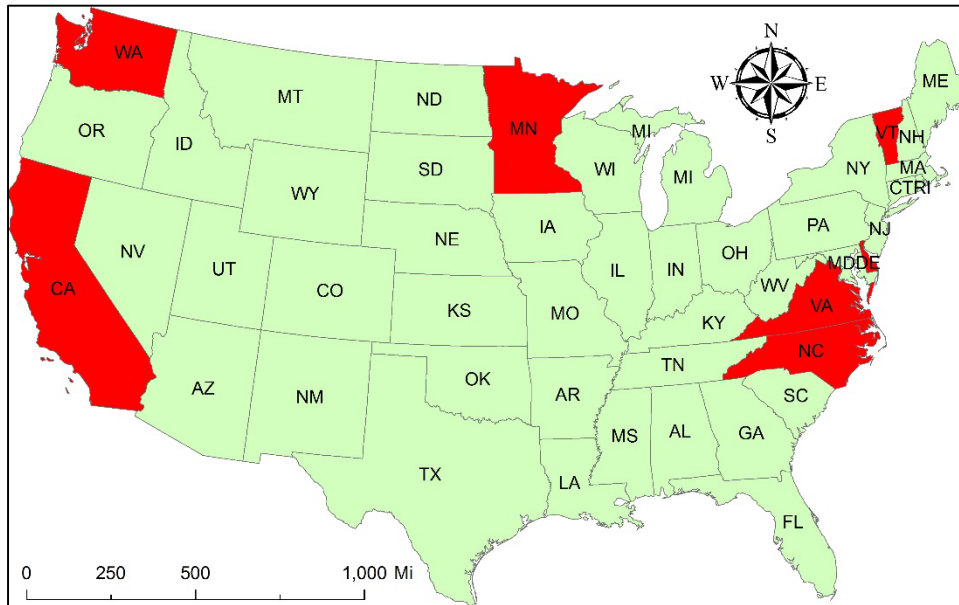
Figure 10. Map. States where light data were collected.

Lighting performance data were collected on the field using the RLMMS developed by the research team. The RLMMS has the ability to collect horizontal illuminance, roadway luminance, glare from oncoming traffic and other external light sources, and GPS position data. The system also has input buttons to flag special features in the data stream (Figure 11).
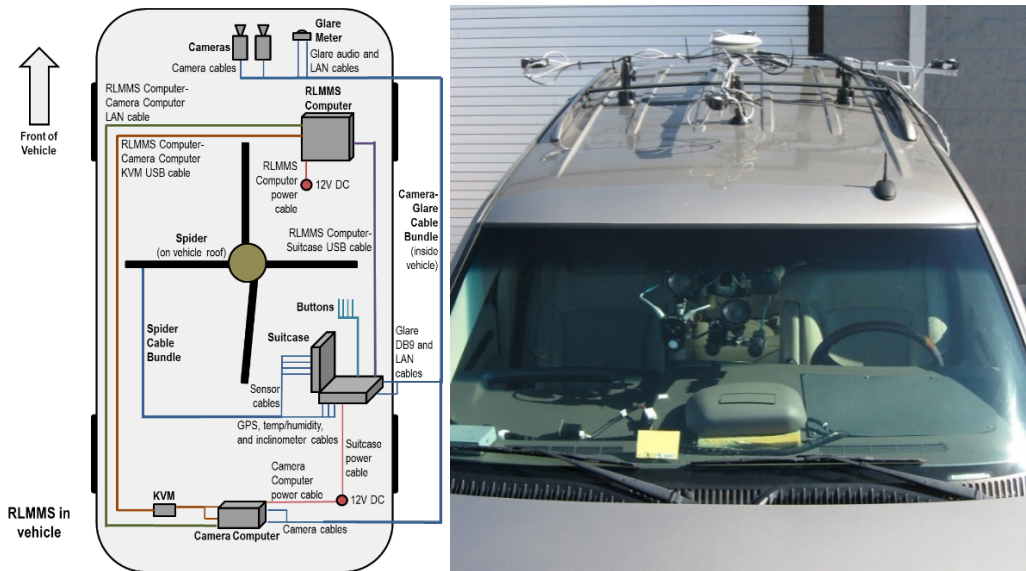
Figure 11. Illustration. Roadway Lighting Mobile Measurement System.[19]

The research team drove along the selected roadways in the seven states at the speed limits and collected lighting data at a frequency of 20 Hz and at a video capture rate of 3.75 frames per second (fps). For multi-lane highways, the research team collected lighting performance data for both the rightmost and leftmost lanes to ensure an accurate measurement. The entire data collection effort lasted from late 2012 to mid-2013. The resulting dataset comprised seven shapefiles, each of which contained point lighting performance measurements for an individual state. Table 1 and Table 2 provide further details about the time when the in situ lighting performance data were collected and the covered roadway miles covered.

Table 1. Roadway lighting performance data collection timeframe.

State	Data Collection Time	NDS State?
California	December 11–14, 2012	No
Delaware	November 27–30, 2012	No
Minnesota	November 15–16, 2012	No
North Carolina	April 16–19, 2013	Yes
Virginia	March 11–15, 2013; May 15–17, 2013	No
Vermont	October 31–November 1, 2012	No
Washington	January 23–25, 2013; June 5–7, 2013	Yes

Table 2. Lighting data collection highway miles by state.

Roadway Type	CA	DE	MN	NC	VA	VT	WA	
Interstate	160	41	68	167	372	22	151	981
Other Freeways and Expressways	77	27	27	59	61	7	101	359
Other Principal Arterial	31	155	7	51	102	47	98	490
Minor Arterial	--	47	51	30	52	25	28	232
Major Collector	--	--	--	9	26	--	16	52
Total	269	269	153	317	613	100	394	2,114

To enable meaningful analyses, VTTI also collected weather data, traffic data, crash data, roadway data, and lighting infrastructure data, and organized all datasets in a systematic manner in the ALD (Table 3 and Figure 12).

Table 3. Primary state-level data sources.

State	CA	DE	NC	MN	VA	VT	WA
GIS Road Network	CALTRANS	DE DOT	NC DOT	MN DOT	VA DOT	VT DOT	WSDOT/HSIS
Lighting Performance Data	VTTI	VTTI	VTTI	VTTI	VTTI	VTTI	VTTI
Lighting Design Data	CALTRANS	DE DOT	NC DOT	MN DOT	VA DOT	VT DOT	WA DOT; City of Seattle
Crash Data	HSIS	DE DOT	HSIS	HSIS	VA DOT	VT DOT	WA DOT/HSIS
Roadway Data	HSIS	DE DOT	HSIS	HSIS	VA DOT	VT DOT	WA DOT/HSIS
Traffic Data	CALTRANS	DE DOT	NC DOT	MN DOT	VA DOT	VT DOT	WA DOT

Caltrans: California Department of Transportation.
HSIS: Highway Safety Information System.
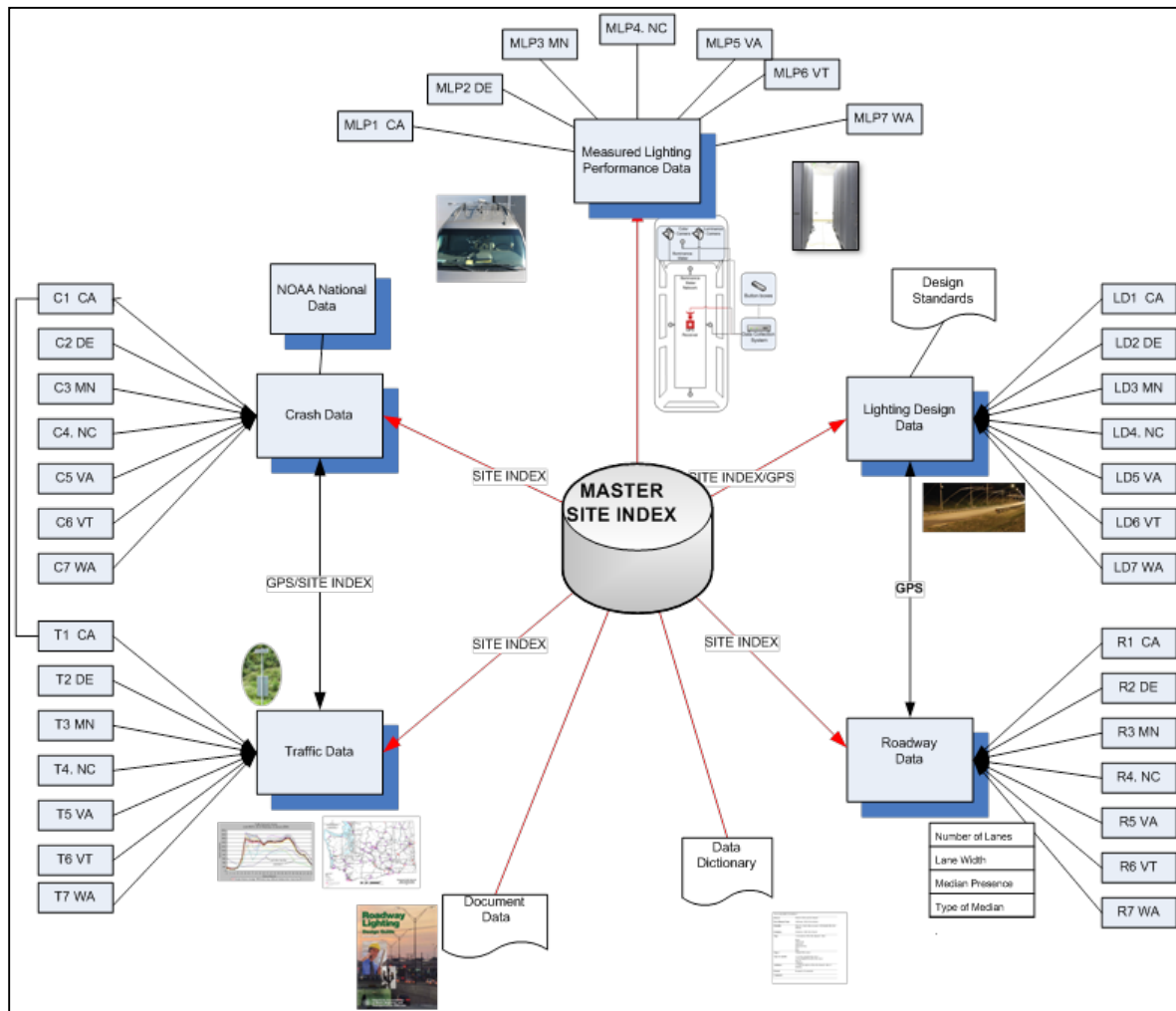
Figure 12. Diagram. ALD structure.

As Table 3 shows, crash data were obtained primarily from the Highway Safety Information System (HSIS) and complemented with data from certain state DOTs. For most states, VTTI obtained crash data for 2004–2008 due to data availability. For Delaware and Vermont, the research team obtained crash data for 2006–2010. The research team also obtained traffic data from state DOTs, including automatic traffic recorder (ATR) counts at permanent stations in tabular format and roadway segment AADT data in shapefile format for corridors where lighting performance data were collected. The traffic data were for the years between 2004 and 2008. Lighting infrastructure data were obtained for Delaware, Minnesota, Virginia, Vermont, and Washington from the state DOTs. The data depicted the locations of roadway light poles along major state highways in each state.

The research team obtained roadway network data for each state from either the corresponding state DOT or HSIS. The state roadway shapefiles contained Polyline M features depicting the locations of the public roadways. For most states, these shapefiles were based on the official state route networks developed for linear referencing purposes at state DOTs and only included very basic attributes. The route networks were roadbed-based with individual roadbeds represented by separate lines (i.e., double lines for divided highways).

PRELIMINARY PROCESSING OF LIGHTING PERFORMANCE DATA

The two states where both lighting performance data and NDS data were collected (i.e., North Carolina and Washington) were of particular interest to this project. The Washington lighting performance data were primarily collected on major highway corridors in and around the Seattle-Tacoma-Bellevue metropolitan area (Figure 13). The data collection effort resulted in approximately 1.6 million lighting performance data points along 646 miles of roadways (by roadbeds), the majority of which were freeways. The North Carolina dataset included 743,695 lighting data points on approximately 538 miles of roadways (by roadbed), the majority of which were urban freeways around the cities of Winston-Salem, Greensboro, Durham, and Raleigh (Figure 14).
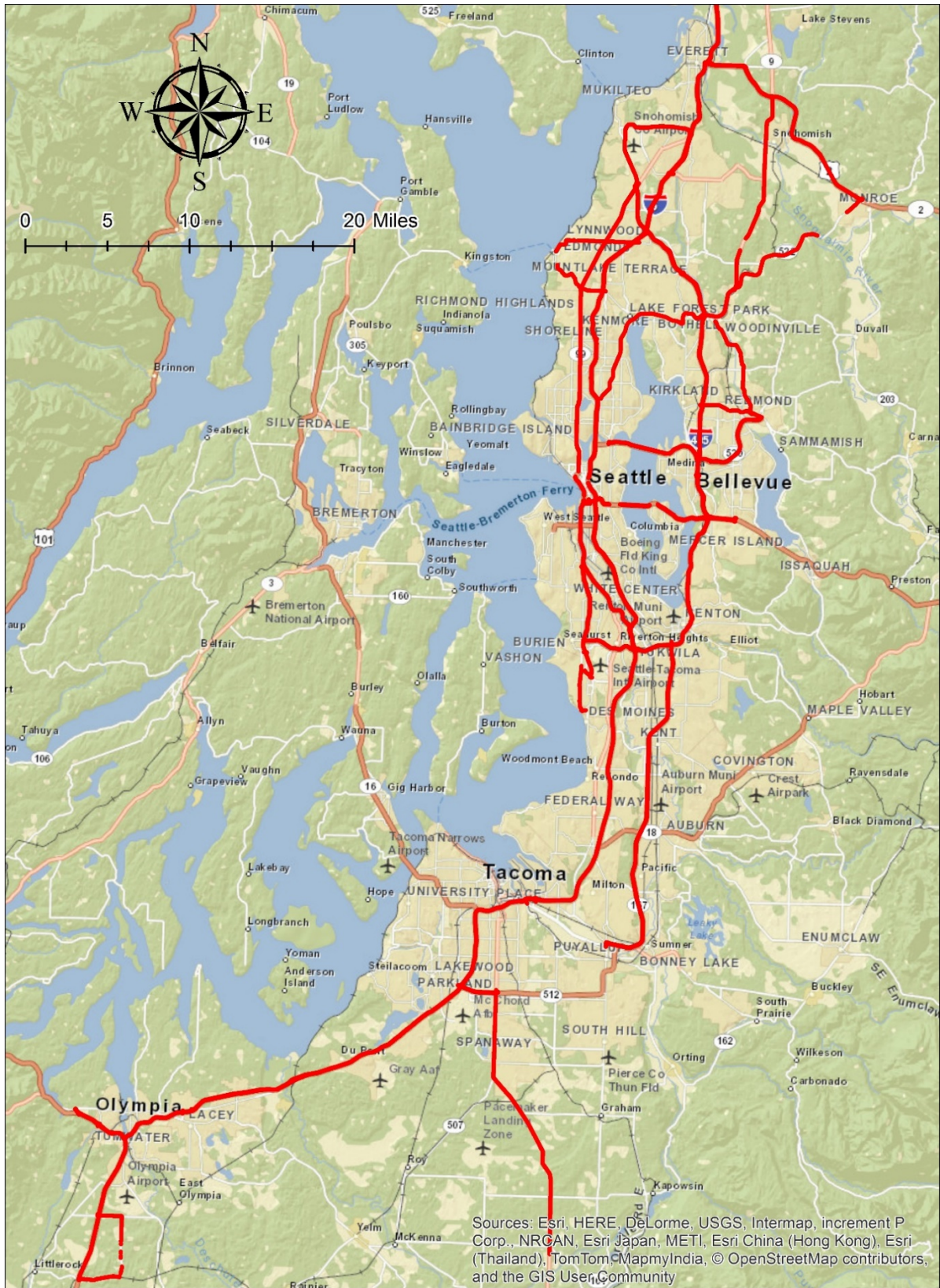
Figure 13. Map. Washington roadways with lighting performance data collected.
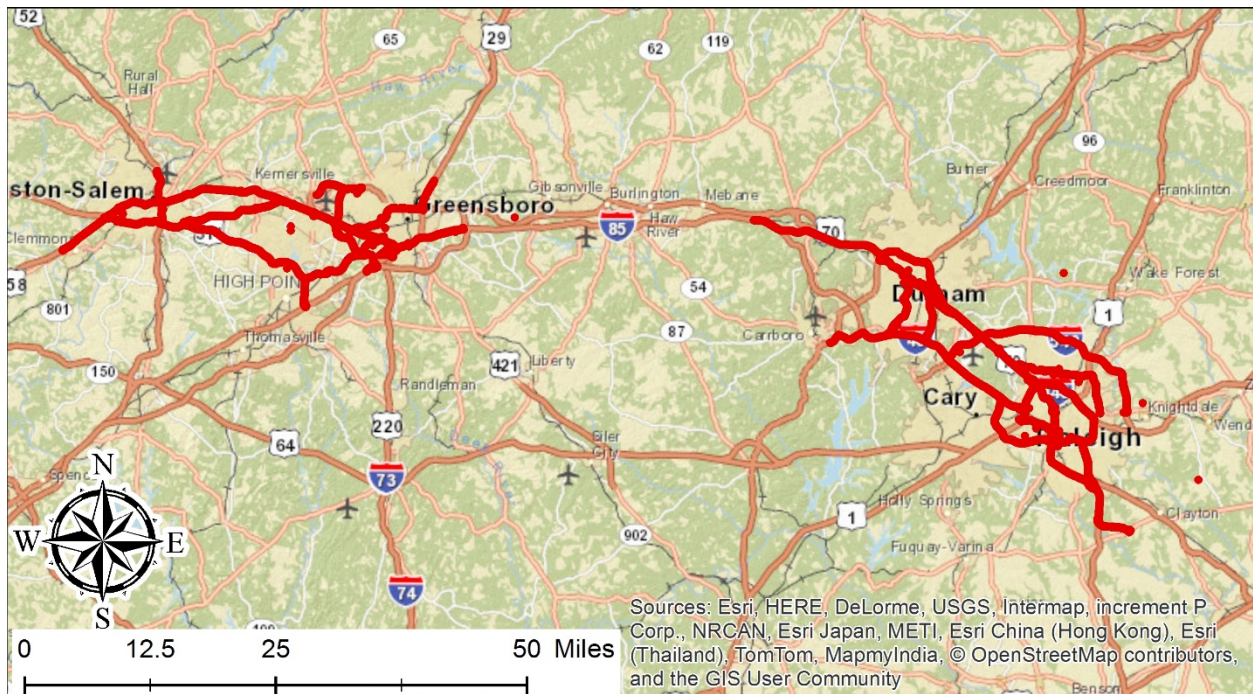
Figure 14. Map. North Carolina roadways with lighting performance data collected.

Due to the lack of relational linkages between the lighting performance data and NDS data, and because both ALD and NDS data contained spatial references, it became evident that the data integration could be most effectively achieved using spatial data conflation. Such procedures require the use of a common roadway network (or different yet directly matched roadway networks) as the spatial linkage. When both datasets are spatially linked to the roadway network, they are considered matched since straightforward spatial join procedures can be performed to extract attributes from both datasets into a common table. To support rigorous spatial data conflation between the two datasets, however, the research team needed to ensure relatively accurate spatial data integrity within the ALD.

Both the Washington and North Carolina lighting data used the official state DOT's route networks that form the basis for various state and federal data programs. These route networks, therefore, were used for linking the lighting performance data with the NDS data during this project. An examination of the lighting performance data, however, unveiled spatial discrepancies between the lighting performance data points and the underlying roadway networks that needed to be corrected. Among the various issues identified, due to the varying levels of spatial data accuracy between lighting data and the route networks, and because of the fact that the route networks use centerlines to represent roadbeds while the lighting data points were based on individual lanes, the lighting performance data points scattered around the roadways when overlying them. Figure 15 and Figure 16 illustrate this phenomenon.
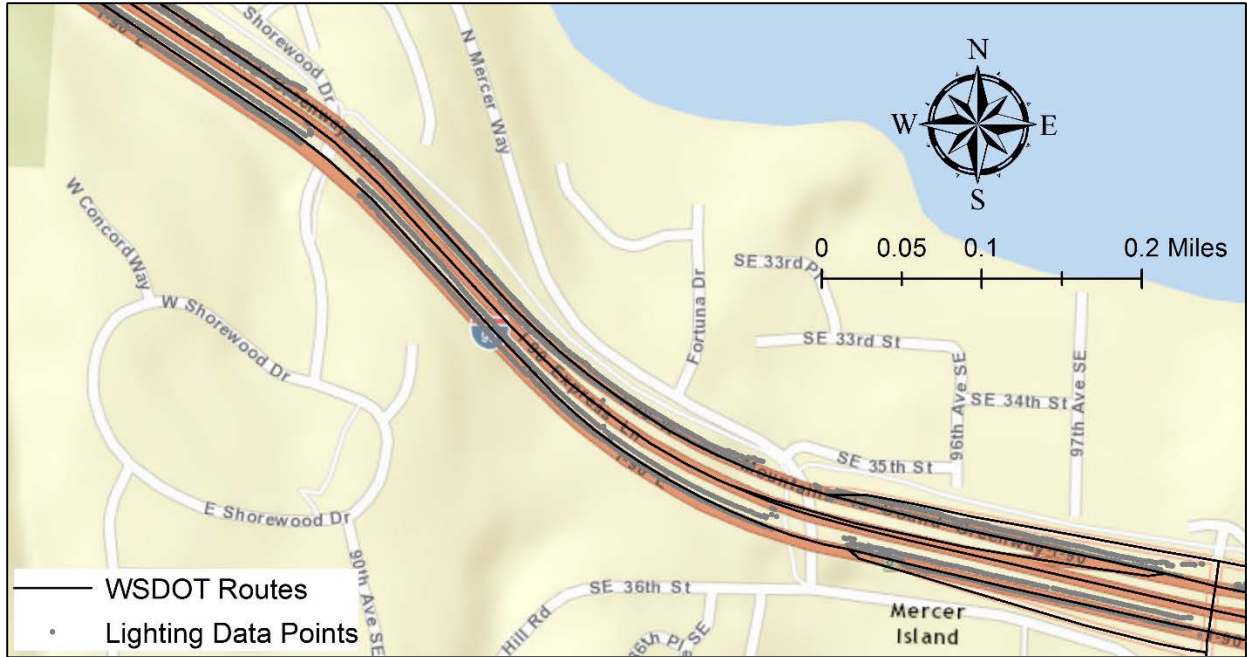
Figure 15. Map. Spatial discrepancy between lighting data points and roadway network.
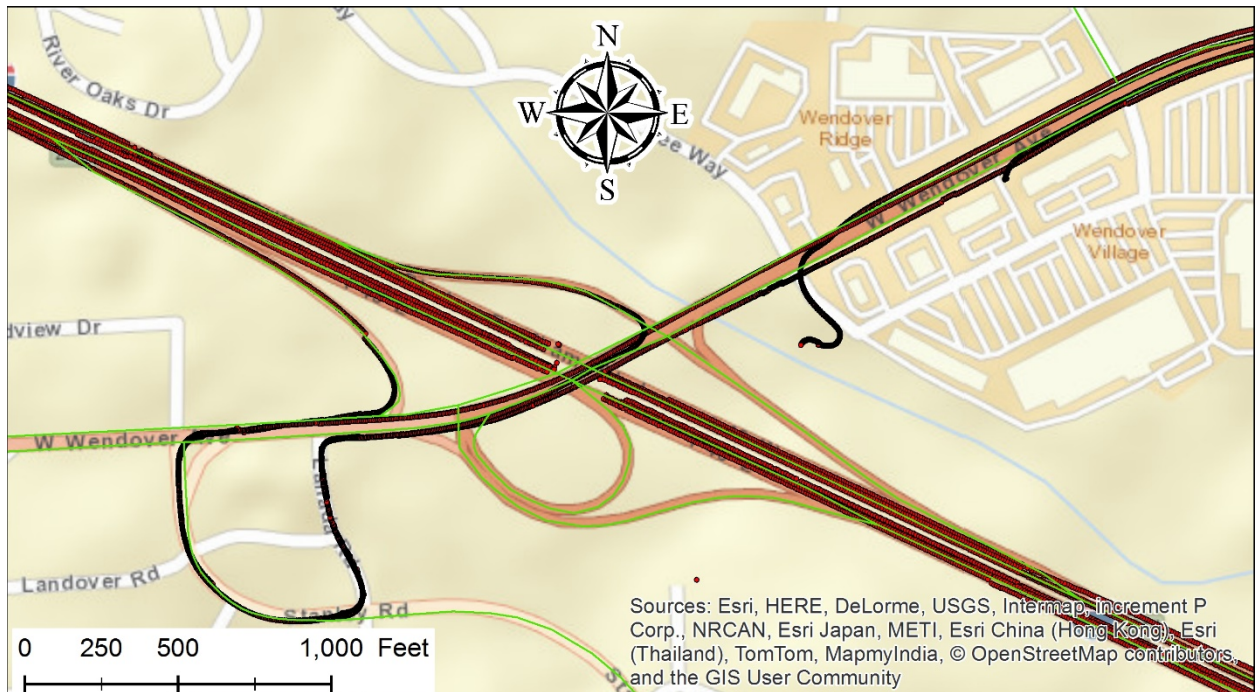
Figure 16. Map. Lighting data points before processing.

To enable spatial data integration, the research team needed to first match the lighting performance data points to the state route network for both Washington and North Carolina. However, the lighting performance data did not include linear referencing information or

attributes, such as roadway name and travel direction that can be used to facilitate the data matching. Therefore, the spatial relationship between the lighting performance data points and roadway features became the only usable criterion. The research team developed the following geoprocessing procedure based on the ArcGIS® platform to perform the data matching:

1) Clean roadway network for data matching. During this step, the research team deleted the features interacting the roadways for which lighting data were collected from the roadway network to create a clean version of the roadway network. This step helped reduce the possibility of matching lighting data points incorrectly to nearby or intersecting roadway segments. The features deleted during this step included, for example, ramps, express lanes on freeways with a separate roadbed, minor intersecting roads, and frontage roads where lighting data were not collected. Figure 17 illustrates the differences between the original and cleaned roadway networks.
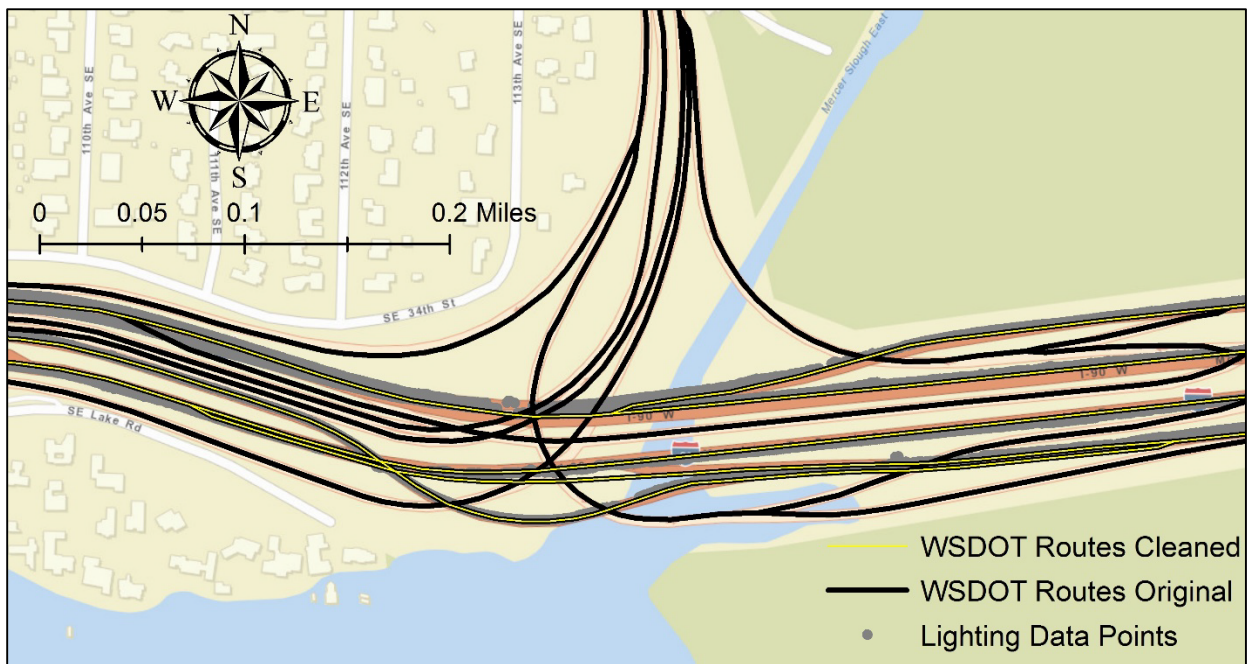
Figure 17. Map. Roadway network cleaned to improve matching accuracy.

2) Match lighting data points to nearest roadway segments. The research team used the Near tool available on ArcGIS Desktop 10.2. The ArcGIS Near operation calculates distance and additional proximity information between the input features and the closest feature in another feature class. Depending on user specifications, it outputs information such as the nearest XY coordinates on the nearest roadway segment for each lighting data point, the ID of the nearest roadway segment, and the distance. During this operation, the research team used a search tolerance of 50 ft.; a lighting performance point was considered unmatchable if a roadway segment was not found within this distance.

3) Perform QA/QC. The research team conducted QA/QC to visually identify, to the extent possible, the lighting data points that were incorrectly mapped. During this process, the researchers used techniques such as color coding the processed lighting data points by the

roadway name/ID attributes produced during the previous step, and/or focusing on lighting performance data points with a relatively larger matching distance (e.g., 25 ft. or higher). The mismatched data points were exported into a new feature class for further processing.

4) Repeat previous steps for mismatched data points. During this iteration, the research team further cleaned the route network to retain only those roadway segments needed for matching. The researchers then ran the Near tool for another iteration to rematch any previously mismatched lighting performance data points.

5) Merge all results into one feature class. After the previous steps, it was necessary to merge the resulting lighting performance data point feature classes into one feature class with all lighting performance data points matched.

6) Generate a new lighting performance data feature class based on matching results. The resulting lighting performance data feature class now contained XY coordinates of the nearest point on the corresponding roadway segment for each lighting data point. Based on this information, the research team used the ArcGIS Display XY Data tool to generate a new feature class on which the lighting performance data points were spatially aligned with the roadway network.

Figure 18 compares the locations of the lighting performance data points before and after the data matching process. As seen from the figure, the lighting data points were snapped to the corresponding roadway segments, enabling a spatial linkage between the lighting data and the roadway network. To ensure accuracy, the research team randomly checked 333,909 lighting performance data points, among which 325,456 points were matched correctly, representing an accuracy rate of 97.5%. Among those that were mismatched, most were located at interchange areas and intersections where small numbers of points were mistakenly matched to intersecting roadways or ramps that were physically closer. A more thorough QA/QC process could correct this issue. However, the researchers' assessment was that the small percentage of isolated data points would not result in significant impact on data analyses yet the process to correct them would be relatively time consuming.
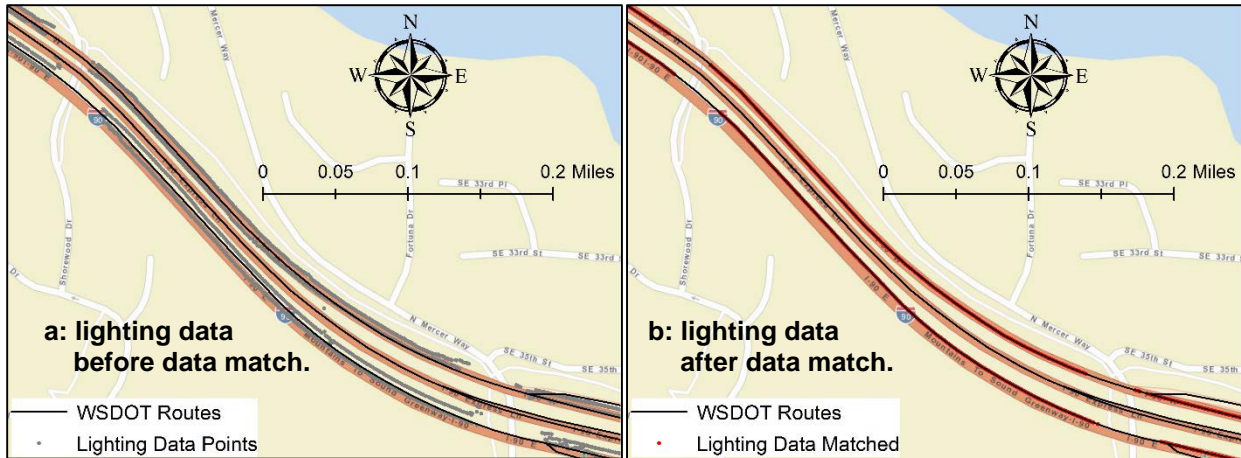
Figure 18. Map. Spatial comparison of lighting data before and after data match.

Using a similar method, the researchers later spatially matched the North Carolina lighting data points onto the roadway network. Figure 19 shows the lighting data points after having been spatially matched onto the state roadway network.
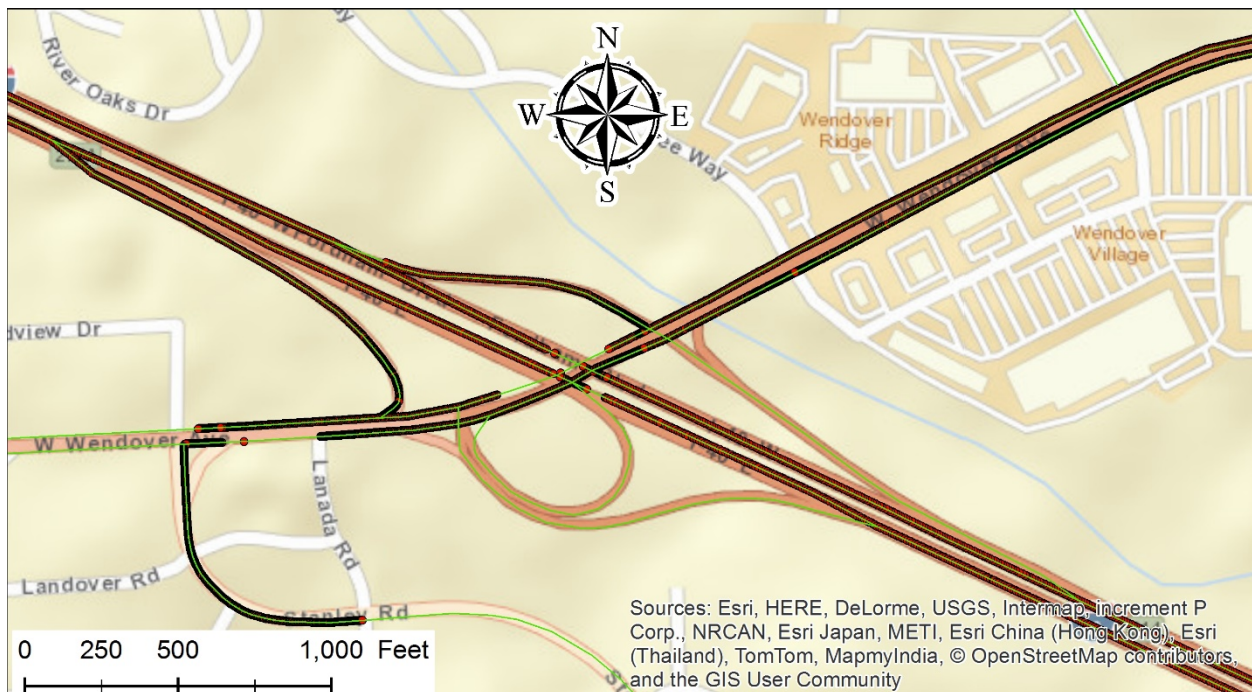
Figure 19. Map. Lighting performance data points matched to NC roadway network.

CHAPTER 4. NDS AND ALD DATA INTEGRATION

INTRODUCTION

During this project, the research team developed separate spatial data integration procedures for two scenarios:

- Data integration based on SHRP 2 time series data. The research team developed a spatial data matching approach that directly linked NDS time series data with the ALD roadway network. The approach performed data matching at the individual data point level, without decreasing the spatial resolution of the original NDS data. In addition, the approach directly linked NDS data to the custom roadway network, therefore eliminating additional sources of error. To demonstrate the data matching approach, the research team used sample NDS and ALD data from the State of Washington.

- Data integration via the RID digital map. The RID includes the unique LINKIDs used as the basic linkages within the SHRP 2 database. The research team decided to match the ALD to the RID to permit referencing the ALD using the common field.

The research team performed most of the spatial data matching tasks on the Esri ArcGIS platform. ArcGIS is currently one of the most commonly used spatial analysis software packages. It provides relatively user-friendly interfaces and a large number of readily accessible tools. The following sections describe these two scenarios and the associated spatial matching procedures in detail.

DATA INTEGRATION BASED ON NDS TIME SERIES DATA

As illustrated in Figure 9, all commonly used NDS data items are linked with the time series data through relational tables within the SHRP 2 NDS database. The NDS time series data include a set of vehicle/event descriptors collected from the participant vehicles with the onboard data acquisition system at varying frequencies, such as timestamp, speed, acceleration rates, coordinates, and vehicle condition. The GPS coordinate information within time series data provides a basis for spatially integrating the time series data with the ALD. Figure 20 shows this linkage graphically.
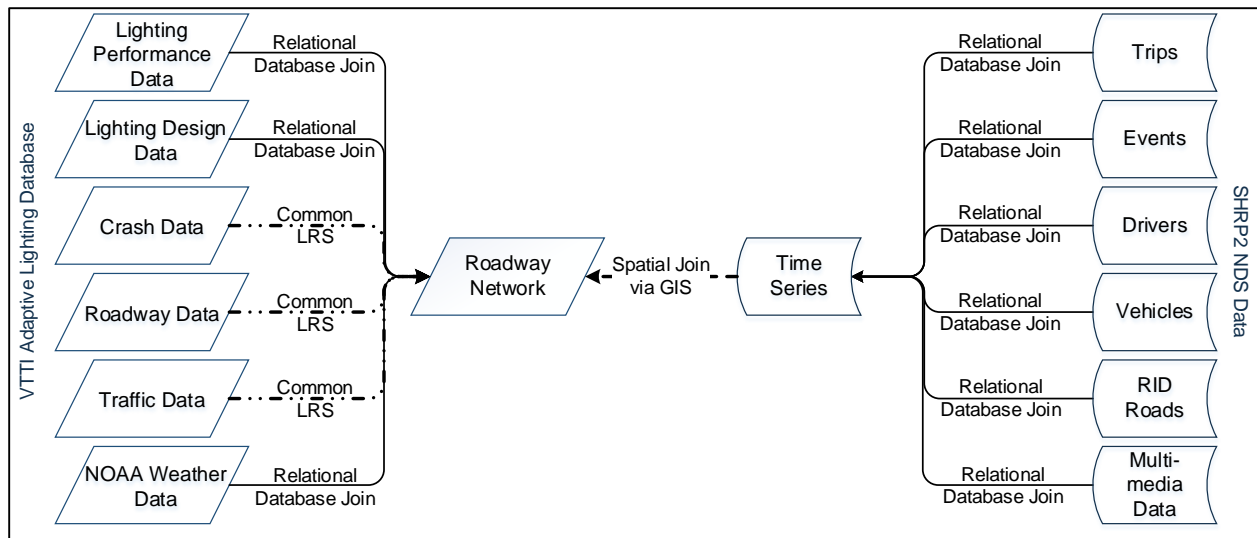
Figure 20. Diagram. Linking NDS data with ALD data based on time series data.

Time Series Data Extraction

As shown in Figure 9, NDS time series data include a range of sensor data elements collected at different intervals. To store the large quantity of data in the SHRP 2 NDS, which are collected at different sample rates, the database architecture uses a columnar format. In this architecture, a data record includes variable ID, a timestamp, and one sensor reading. Readings from different sensors are stored in the same column and the variable names are used to isolate them. Additionally, data are parsed into tables according to their type (e.g., FLOAT vs. INT) and how frequently researchers need to query them (HOT vs. COLD). Most time series tables contain a large number of records. The HOT_FLOAT_SENSOR_DATA table, for example, contains 2,186,924,755,924 rows of data depicting 476 variables.

External data extraction tasks are generally performed by the SHRP 2 NDS data team upon request due to their direct access to NDS data, powerful database servers, and their familiarity with the NDS data structure. For the purpose of this study, the research team requested direct access to the NDS databases in an effort to develop a complete understanding of the database structure and data elements. This knowledge was critical for this research in order to thoroughly assess the potential linkages between the NDS and ALD data. The knowledge was also critical to identify and recommend meaningful analyses that can be performed with the integrated data.

After carefully evaluating the data extraction effort and the amount of data needed for this feasibility study, the research team extracted 11 sample time series variables for 5,000 nighttime (e.g., 9:30 p.m. to 5:00 a.m.) NDS trips in the Seattle-Tacoma-Bellevue metropolitan area. The resulting dataset included 3,929,713 records and was imported into a Microsoft® Access database. The extracted variables included:

- File ID: unique identifier for NDS data files. Each file is a collection of time series data points collected by onboard sensors for a vehicle trip. This is one of the primary keys used for linking time series data to other data such as trip information.

26

- Vehicle ID: unique identifier for each vehicle. This field is used to link file IDs to detailed vehicle information.

- Driver ID: unique identifier for each driver. This field is used to link file IDs to detailed driver information.

- Location: a two-digit code that identifies the state where the trip occurred. The location code for Washington is 58.

- Local Date and Time: local date and time when the trip took place.

- Timestamp: timestamp for each time series data point.

- GPS Speed: measured GPS speed.

- Latitude: latitude of each time series data point.

- Longitude: longitude of each time series data point.

Figure 21 shows the SQL statement the research team used to extract the sample time series data.

```
@export on;
@export set filename = "C:\ELI Local Data\Projects\451114 NDS and Lighting
   Database\Naturalistic Driving Database Documentation\WASHINGTONTIMESERIESDATA.csv";
SELECT T4.*, T1.FILE_ID, T1.STATUS, T1.TIMESTAMP, T1.DATA AS GPS_SPEED, T2.DATA AS
   LAT, T3.DATA AS LON
FROM (SELECT FILE_ID, VTTI_VEHICLE_ID, VTTI_PARTICIPANT_ID, VTTI_DRIVER_ID,
   VTTI_LOCATION_CODE, LOCAL_DATETIME FROM SHRP2NDS_MAIN.FILE_SUMMARY_INFO WHERE
   VTTI_LOCATION_CODE = 58 and (TIME(LOCAL_DATETIME) not between '05:00:00' and
   '21:30:00') FETCH FIRST 5000 ROWS ONLY) T4
JOIN (SELECT * FROM SHRP2NDS_STAGING.HOT_FLOAT_SENSOR_DATA WHERE VARIABLEID = -330) T1
   ON T4.FILE_ID = T1.FILE_ID
JOIN (SELECT * FROM SHRP2NDS_STAGING.HOT_FLOAT_SENSOR_DATA WHERE VARIABLEID = -317) T2
   ON T1.FILE_ID = T2.FILE_ID AND T1.TIMESTAMP = T2.TIMESTAMP
JOIN (SELECT * FROM SHRP2NDS_STAGING.HOT_FLOAT_SENSOR_DATA WHERE VARIABLEID = -318) T3
   ON T1.FILE_ID = T3.FILE_ID AND T1.TIMESTAMP = T3.TIMESTAMP
WHERE (T2.DATA >46.814 AND T2. DATA < 48.195) AND (T3.DATA >-123.039 AND T3.DATA < -
   121.936);
@export off;
```

Figure 21. Illustration. SQL statement for extracting time series data.

During the data extraction, the research team accessed the NDS data through a client computer remotely linked to the database servers. According to the data sharing agreement, the research team did not have write privileges on the servers and all queries had to be executed within random access memory (RAM). As such, it was extremely time-consuming to query and retrieve large amounts of data.

Event Data Extraction

At the time of the data extraction, the SHRP 2 NDS data team had preliminarily processed 2,716 events, including 2,042 baseline events, 229 crash events, 6 crash-relevant events, 433 near-crash events, and 6 non-subject conflict events. This data table was imported to the Access database for further analysis. The event types are defined as follows by the NDS data team:

- Crash: Any contact that the subject vehicle has with an object, either moving or fixed, at any speed in which kinetic energy is measurably transferred or dissipated.

- Near-crash: Any circumstance that requires a rapid evasive maneuver by the subject vehicle, or any other vehicle, pedestrian, cyclist, or animal to avoid a crash. A rapid, evasive maneuver is defined as steering, braking, accelerating, or any combination of control inputs that approaches the limits of the vehicle capabilities.

- Crash-relevant: Any circumstance that requires an evasive maneuver on the part of the subject vehicle, or any other vehicle, pedestrian, cyclist, or animal that is less severe than a rapid evasive maneuver, but greater in severity than a normal maneuver to avoid a crash. A crash avoidance response can include braking, steering, accelerating, or any combination of control inputs.

- Non-subject conflict: Any incident that gets captured on video, crash-relevant, near-crash, or crash, that does not involve the subject driver.

- Baseline: Any epoch of data selected for comparison to any of the conflict types listed above rather than due to the presence of conflict.

The research team joined the 2,716 events to the extracted sample time series data and identified six associated events including two near-crash and four baseline events. Note that these 2,716 events are a small subset of the final number of events in the SHRP 2 dataset.

Linking NDS Time Series Data to the ALD

Figure 22 shows the extracted time series data points mapped based on GPS coordinates. A close examination of the time series data points showed several challenges that needed to be addressed in the GIS data linking process:

- Data points on minor roads. NDS trips involved both minor and arterial roadways. In contrast, the ALD roadway network contains mostly state-maintained roadways. The time series data points on minor roads that were not included in the ALD roadway network could cause decreased data matching accuracy and processing speed.

- Data points with problematic GPS coordinates. A number of time series data points had problematic GPS coordinates. After mapping, such points appeared at locations without any roadways. These points could not be linked to any roadways with spatially based procedures.

- Points at intersections and interchange areas. Intersections and interchange areas involve roadways crossing each other and ramps close to main lanes (see Figure 23). At such locations, it becomes extremely difficult to match time series data points merely based on their spatial relationships with the closely located and/or crossing roadways.

- Closely located roadway features. In addition to interchange areas, some roadways include express lanes or frontage roads on separate roadbeds. Such roadway features are closely located from each other and some time series data points along such roads can be difficult to accurately link to the correct roadway segments, as illustrated in Figure 24.

- Inaccurate GPS coordinates. GPS accuracy is affected by multiple factors such as atmospheric effects, sky blockage, receiver quality, and the use of different satellites. Thus, some NDS trips closely followed the roadway network when overlaid on a map while others were tens of feet away from the corresponding roadway centerlines.

- Roadway network accuracy. The locations of some roadway segments, particularly those with continuous curves or at interchange areas, are not accurately depicted on the GIS roadway network.

- The sheer number of data points that need to be processed. The NDS data included a massive number of time series data points. The large amount of data resulted in challenges for data matching and required that any successful data matching algorithm be extremely efficient.
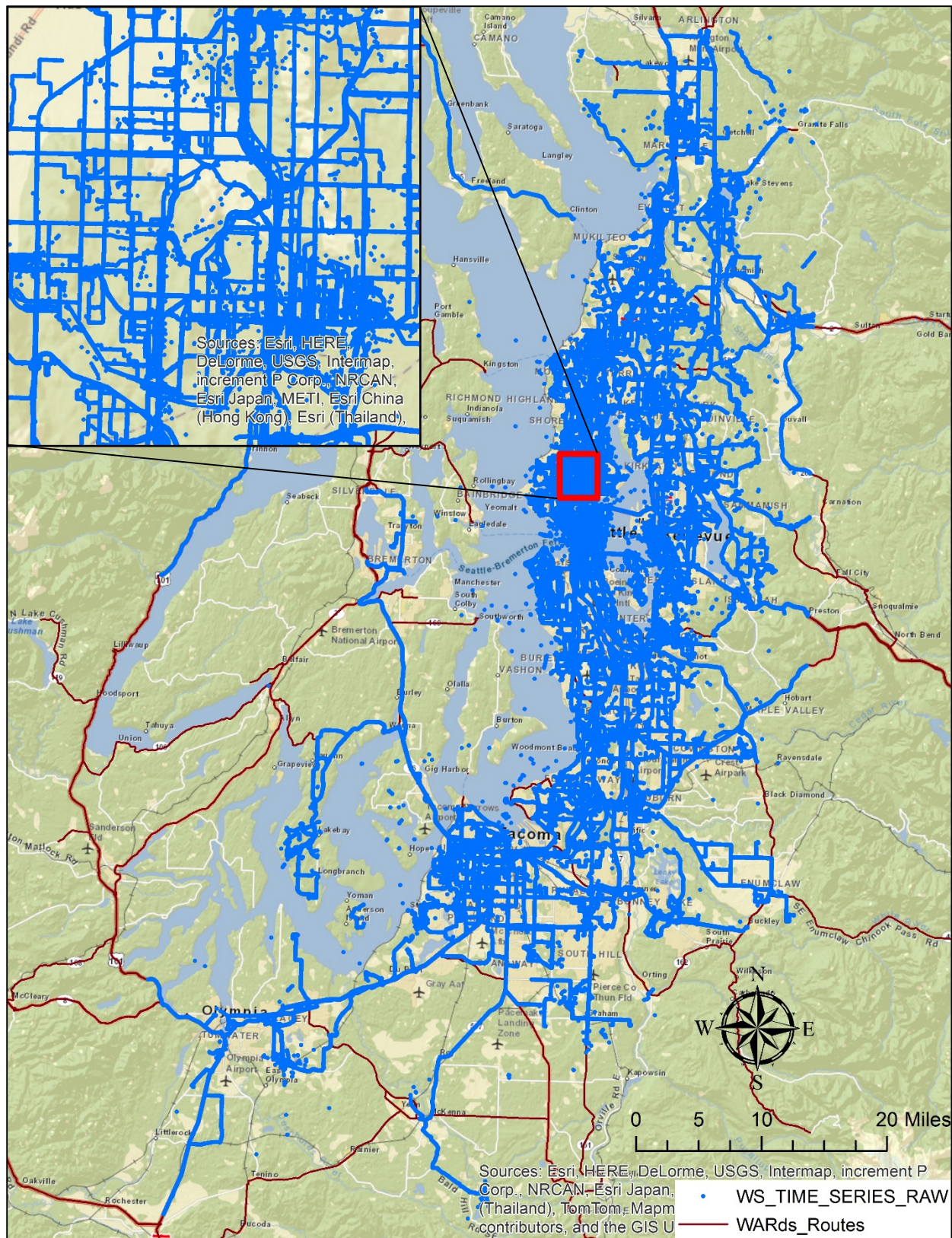
Figure 22. Map. Time series data points mapped based on GPS coordinates.

Sources: Esri, HERE, DeLorme, USGS, Intermap, increment P Corp., NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, Mapm... contributors, and the GIS U...

WS_TIME_SERIES_RAW
WARds_Routes
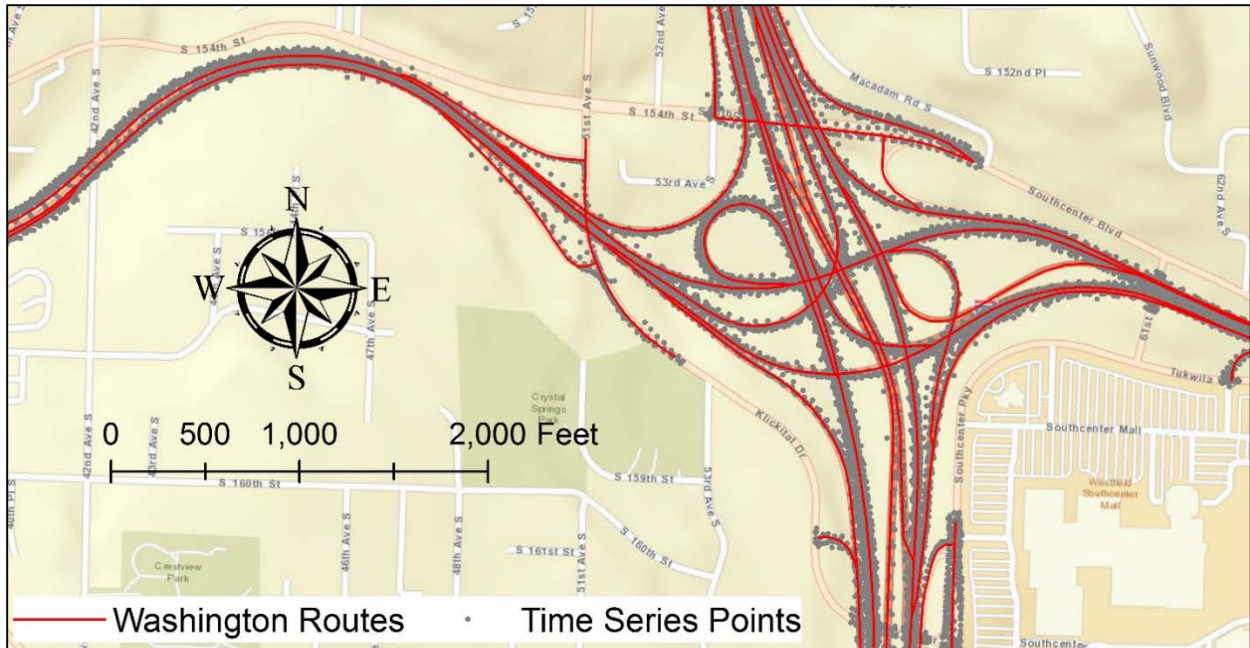
Figure 23. Map. Time series data points at freeway interchange area.
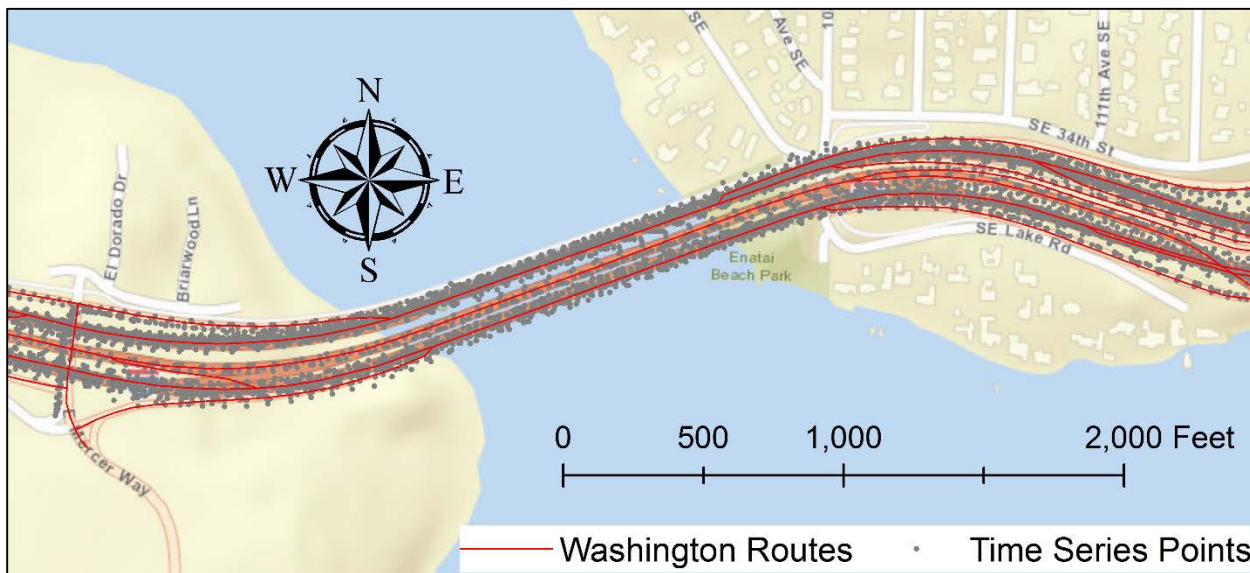
Figure 24. Map. IH 90 section with multiple main-lane roadbeds.

To address the various challenges, the research team developed a GIS-based data matching approach that took into account both spatial and attribute information. This approach incorporated two major considerations to improve matching accuracy:

- Matching based on trip segments instead of individual time series data points. It was extremely challenging to correctly match individual time series data points onto roadway segments purely based on spatial relationship at interchange areas and roadways with closely located express lanes or frontage roads. This challenge could be overcome to a

large extent by combining the continuous time series data points into much longer trip segments and then matching them to the nearest parallel roadway segments.

- Matching based on both spatial proximity and route name. The aforementioned procedure would result in route names being matched to the time series data points. The purpose of the NDS data matching task was to snap the time series data points to the ALD roadway network to enable precise linkages between the ALD data points and time series data points. The researchers' approach utilized a modified spatial join approach that snapped time series data points to the nearest roadway segment with the same route name.

The following are the major steps of the data matching approach (also see Figure 25), most of which were performed using the ArcGIS for Desktop package:

1. Clean time series data by deleting data points that were located more than 40 ft. away from an existing roadway. The researchers generated a 40-ft. buffer around each roadway segment and then deleted all time-series data points that did not fall in the buffers.

2. Identify correct routes to which time series data points should be matched. The purpose of this step was to assign the route name of each corresponding roadway segment to the time series data points. The route name information added to the time series data points would be later used to improve the accuracy of the spatial join.

- Generate time series trip lines. Using the ArcGIS Points to Line tool, such trip lines were generated by sequentially linking time series data points of each NDS trip based on timestamp information. Trip lines will be matched with roadway segments instead of individual time series data points to improve matching accuracy. Figure 26 and Figure 27 illustrate the generated trip lines at different zoom levels.
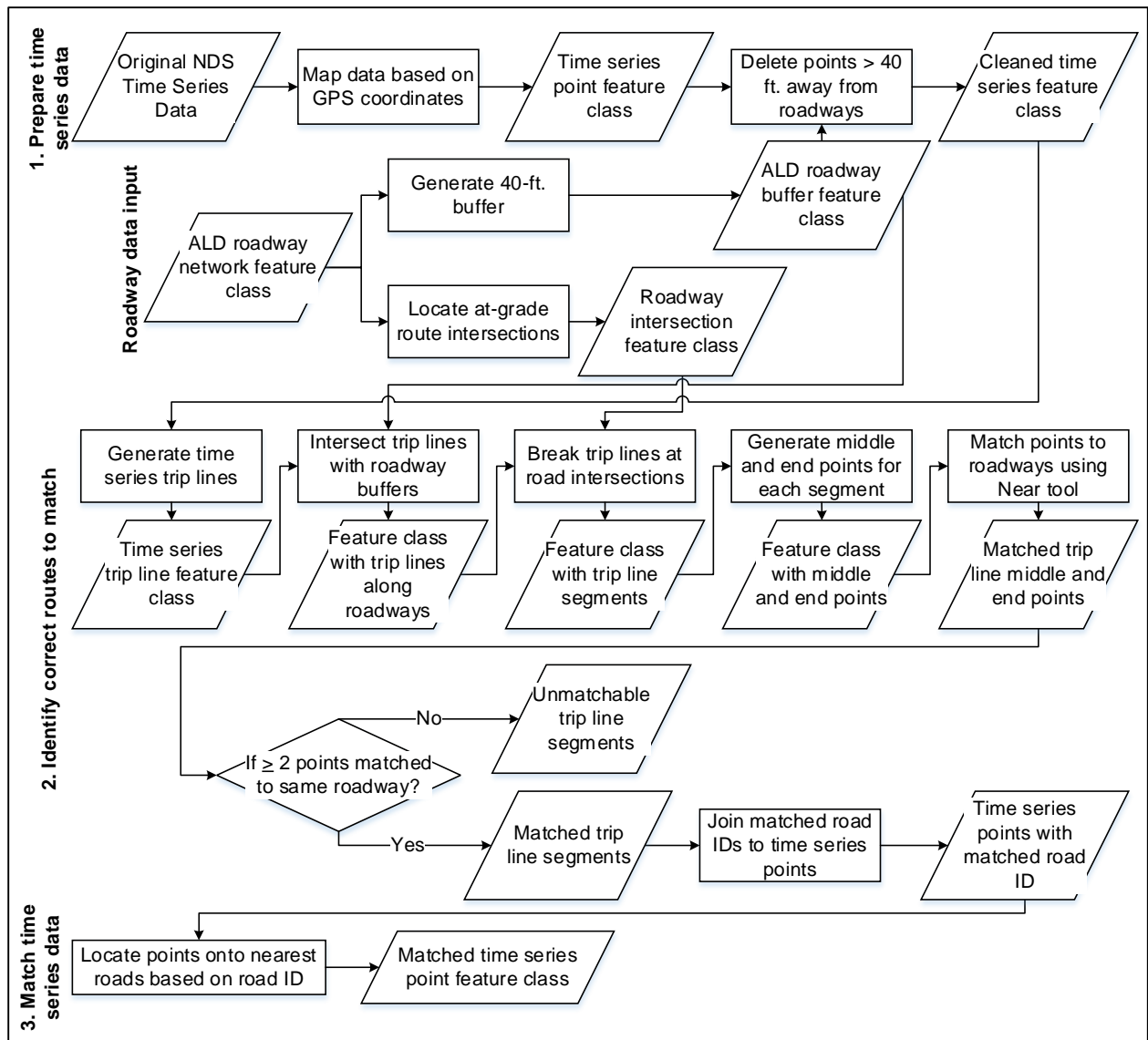
Figure 25. Diagram. Steps for linking NDS time series data points to roadway network.
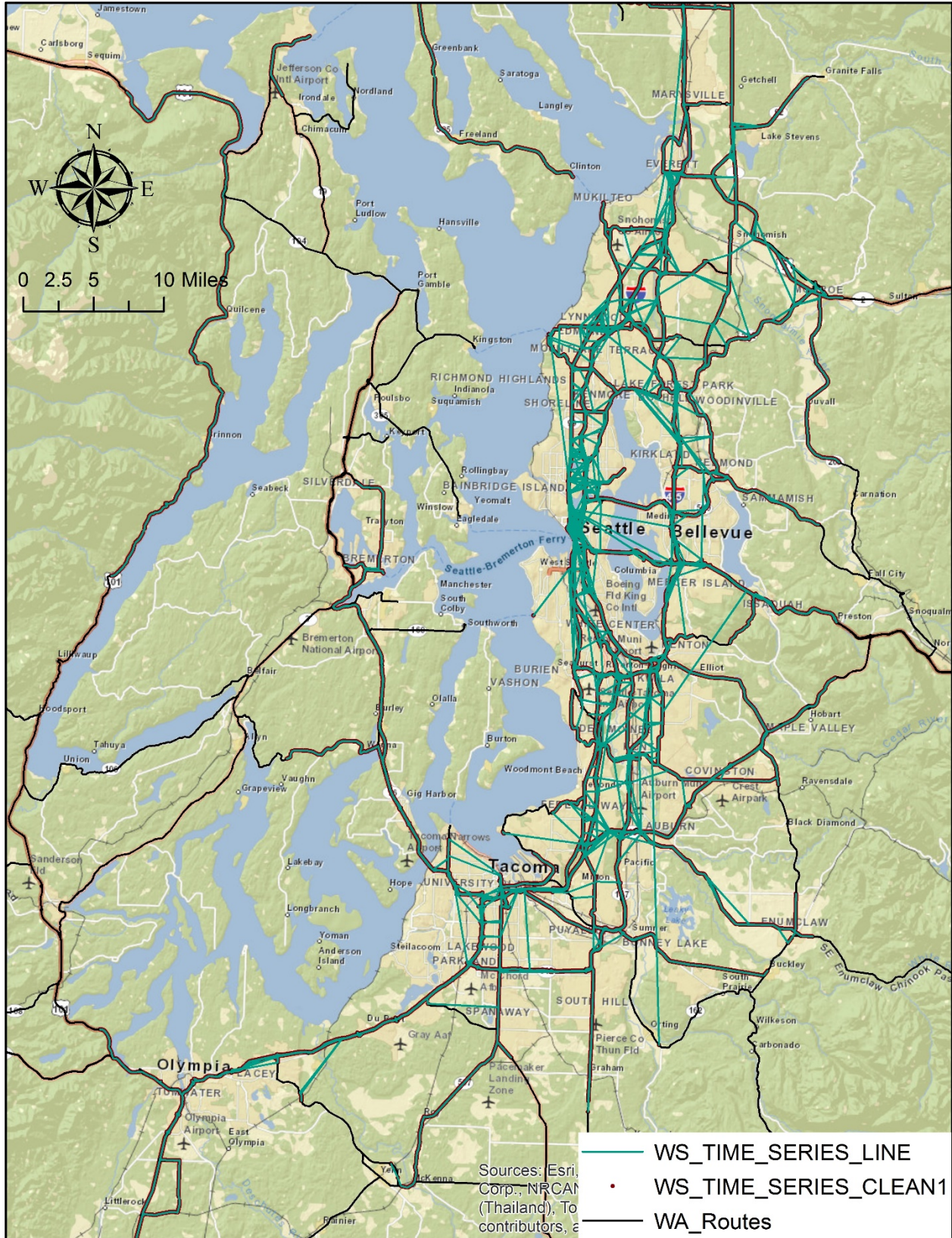
Figure 26. Map. Lines generated based on time series points.
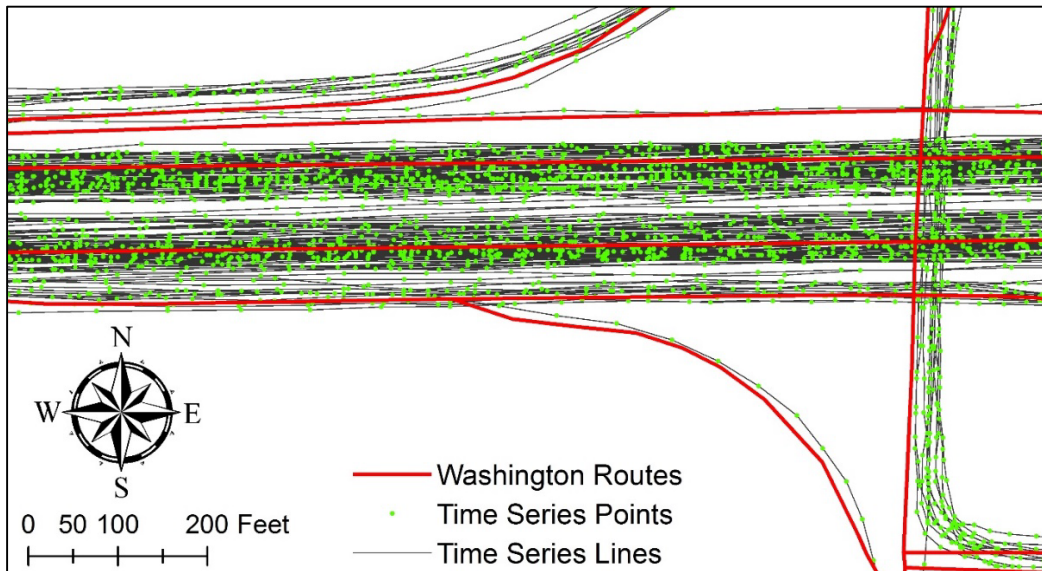
Figure 27. Map. Lines linking sequential time series data points.

- Filter out line segments that do not follow existing roads. As shown in Figure 26, the previous step resulted in straight lines that did not follow existing roadways. These lines were caused by intermediate points that were deleted due to problematic GPS coordinates, or because they were located on minor roads that were not included in the roadway network. This task was done with the following steps:

  o Create 40-ft. buffers for roadway segments. During the buffer operation, all overlapping buffers were dissolved into single polygons for efficiency.

  o Intersect buffers with time series trip lines. Using the ArcGIS Intersect tool, this step resulted in only trip line portions within the buffers. Notice that during this step, the intersecting trip lines were unintentionally broken at intersecting points. The ArcGIS Dissolve tool was used later to combine the broken trip lines back together.

- Match trip line segments to routes. During this step, the researchers first split the trip lines at all at-grade intersections. The trip line segments were then matched to the nearest roadway sections. The nearest roadway section was identified as the roadway that was nearest to at least two of the three key vertices along the trip line segment: two end points and the middle point.

  o Break time series trip lines into segments at intersections. Using the ArcGIS Split Line at Point tool, the research team broke the trips at all at-grade intersections. An intersection layer was generated for this purpose using the ArcGIS Intersect tool. The research team further processed the intersection layer and deleted the points at grade-separated intersections.

o Convert lines to points. Using the ArcGIS Feature Vertices To Points tool, the researchers converted each trip line segment into three points including the two end points and the middle point.

o Match trip line points to roadways. Using the Near tool, the researchers identified the nearest roadways to all trip line points. The resulting table was then analyzed to identify the roadways that were the nearest to at least two of the three points for each trip line. The route names of the matched roadways were kept in the resulting table.

o Join route names back to time series trip line segments. The route name information from the previous step was later joined back to the original trip line segments. The resulting time series trip line segments contained the route names of the matched roadway segments.

3. Snap time series data points to roadways. This step involved the following activities:

• Join route name information to original time series data points. Using the ArcGIS Near tool, the researchers joined the route name information of the trip line segments to the time series data points (the latter were exactly intersected with the former). The Spatial Join tool could be used for this purpose as well but it would be much more time consuming compared with the Near tool.

• Locate time series data points onto the nearest roadways with the same route name. For this purpose, the researchers developed a custom tool within the ArcGIS ModelBuilder environment that enabled Near operations based on attributes (Figure 28 and Figure 29). The model included a sub-model that performed the Near operation iteratively for each group of time series data points and roadways with the same route name. To reduce processing time, the researchers deleted the small number of time series data points that were matched to ramps.
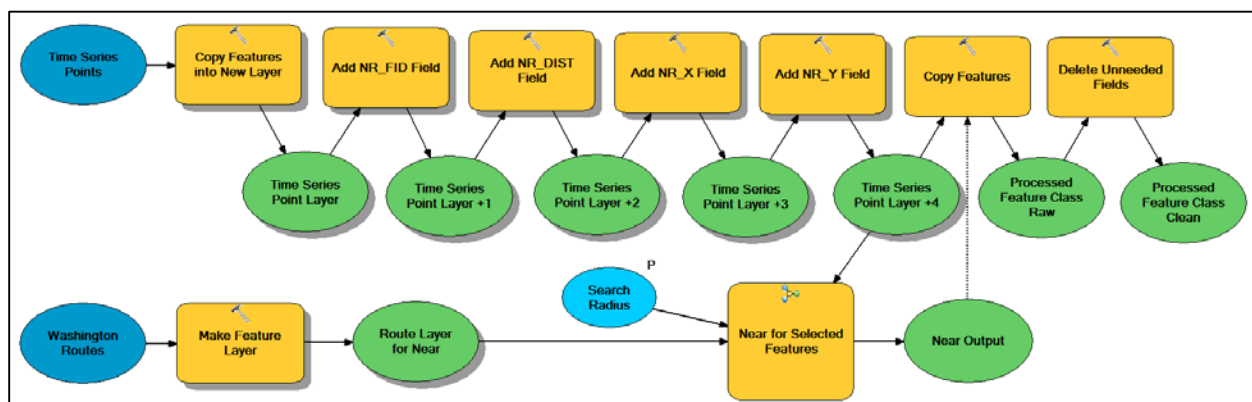
Figure 28. Diagram. ArcGIS model enabling Near based on route name.
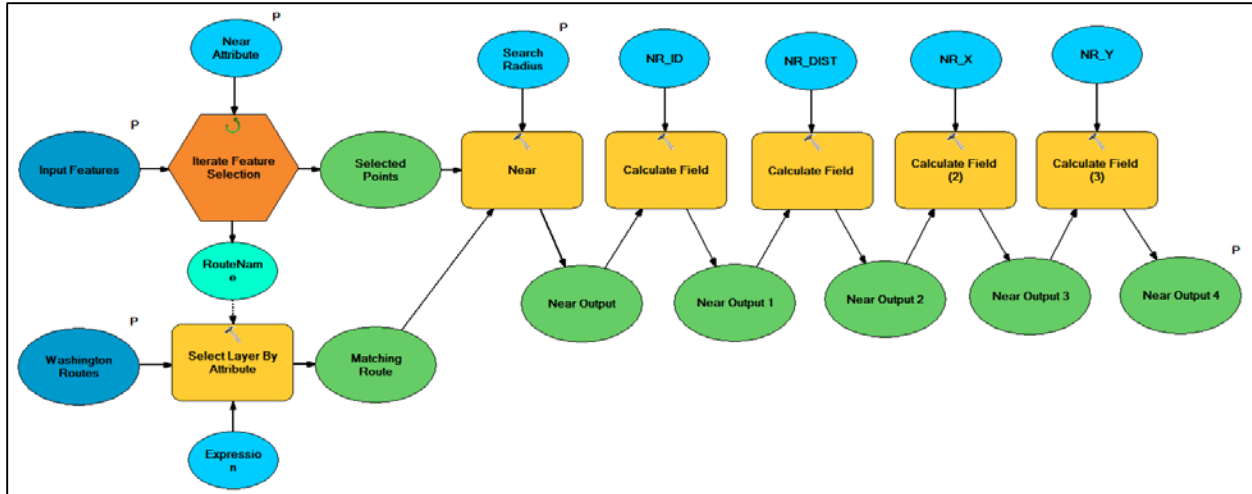
Figure 29. Diagram. Sub-model that performs Near for each route.

- Snap time series data points to roadways. The resulting time series layer of the previous step contained the XY coordinates of the nearest location on the matched roadway for each time series data point. Using the Add XY Data tool in ArcGIS, the time series data points were remapped onto the corresponding roadways based on the XY coordinates. Figure 30 illustrates the time series data points after they were snapped to the matching roadways.
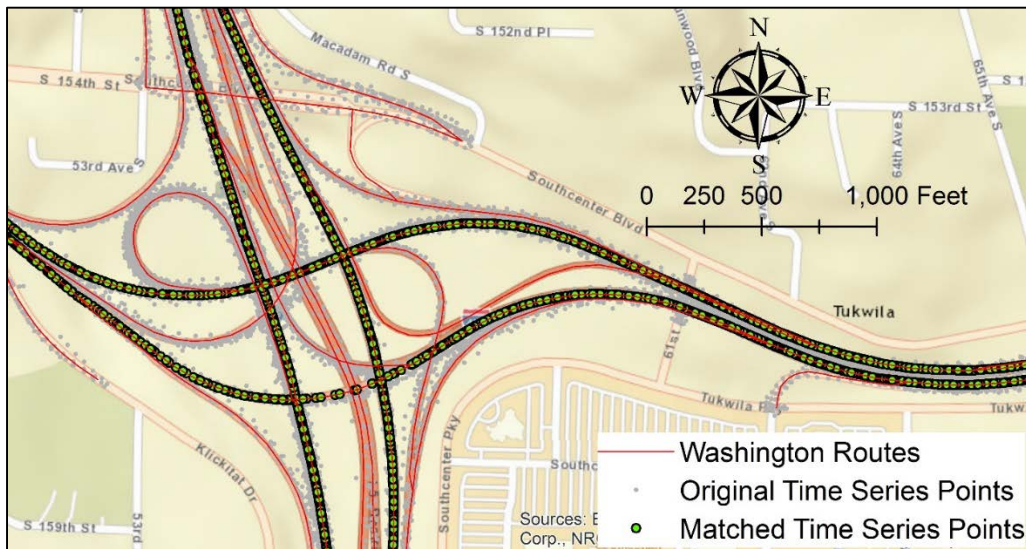
Figure 30. Map. Time series points snapped to roadways.

The result of this process was a significant proportion of the extracted time series data points being linked to the ALD roadway network. Through the time series data, users can then link the lighting performance data with all other NDS data items, including event, driver, and vehicle data. Alternatively, the NDS events can be matched to the roadway network spatially using a similar approach.

Table 4. Summary statistics of time series data matching results.

Number of trips in total	5,000
Number of original time series data points	3,929,713
Number of time series data points within 40-ft. of WA routes	1,313,511
Total trips linked	2,815
Total points linked	883,347

DATA INTEGRATION BASED ON THE SHRP 2 RID

Figure 31 shows how the NDS and ALD data can be linked through the RID roadway network. As the figure shows, for this scenario, the key for the database integration is the matching of the RID roadway network and the ALD roadway network. The research team used a draft copy of the RID data for North Carolina to illustrate the data matching approach.
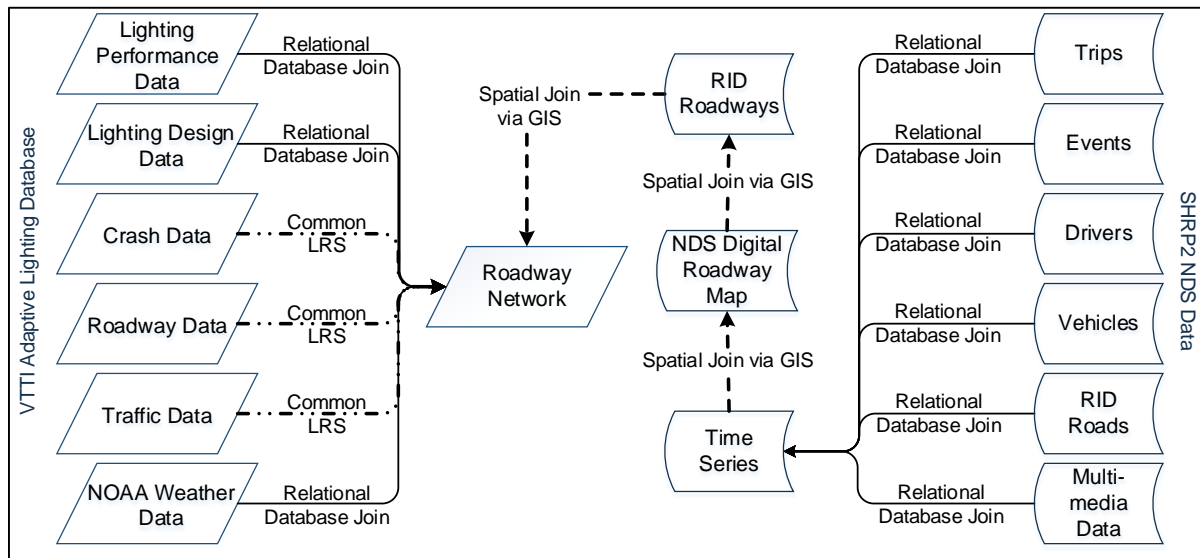
Figure 31. Diagram. Linking NDS data with ALD based on RID.

Matching the RID roadways with the ALD roadway network was less challenging due to the relatively small number of records involved in both datasets. The data matching task involved a spatial linking process between two line layers that do not share a common linear referencing system. The two layers involved in this linking process were the ALD route layer for North Carolina and the RID Links layer that contained LINKID information.

Figure 32 and Figure 33 spatially compare the North Carolina route network used in the ALD with the RID Links feature class. The comparison suggested that:

- The two roadway feature classes align relatively well. In most cases, the offsets between the same roadway features in the two feature classes varied from a few feet to more than 10 ft. In some cases, the offsets approached 25 ft.

- The two feature classes contained roadway segments of different lengths. In general, because the ALD route layer was developed for linear referencing purposes, it contained much longer roadway features, which frequently represented entire roadways determined based on very basic characteristics such as roadway name and type. In contrast, the Links layer contained much shorter segments varying from hundreds of feet to multiple miles.

- The two roadway feature classes contained inconsistent roadbed representations. For example, some roadways are represented in one layer by single lines as undivided roadways while in the other layer by double lines as divided roadways.

- In most areas, the ALD route layer contained more detailed roadway features, including local roads. However, in the Durham-Raleigh area, the RID Links layer contained detailed local roadways, many of which were not included in the ALD route layer.
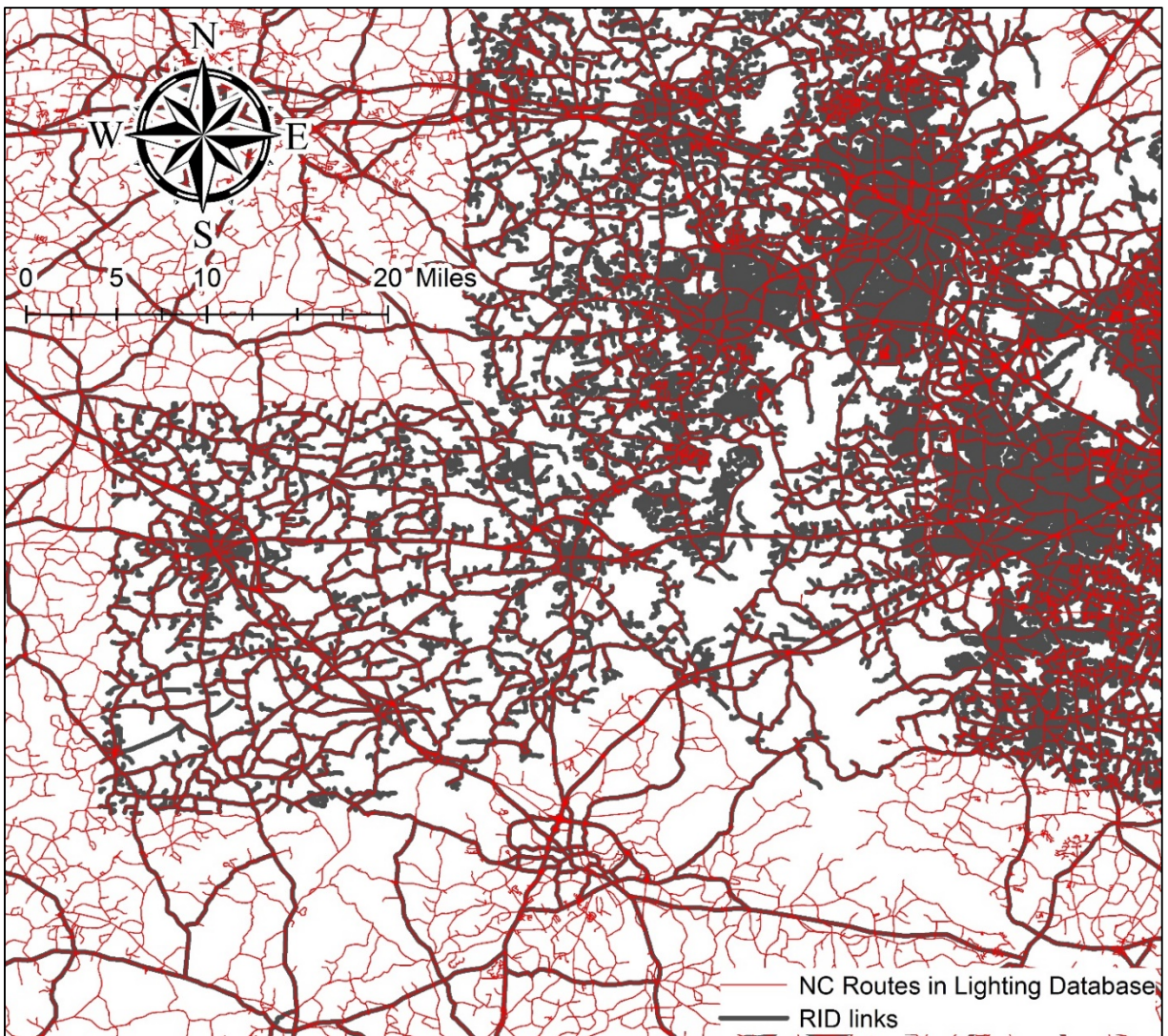
Figure 32. Map. ALD route network and RID Links.
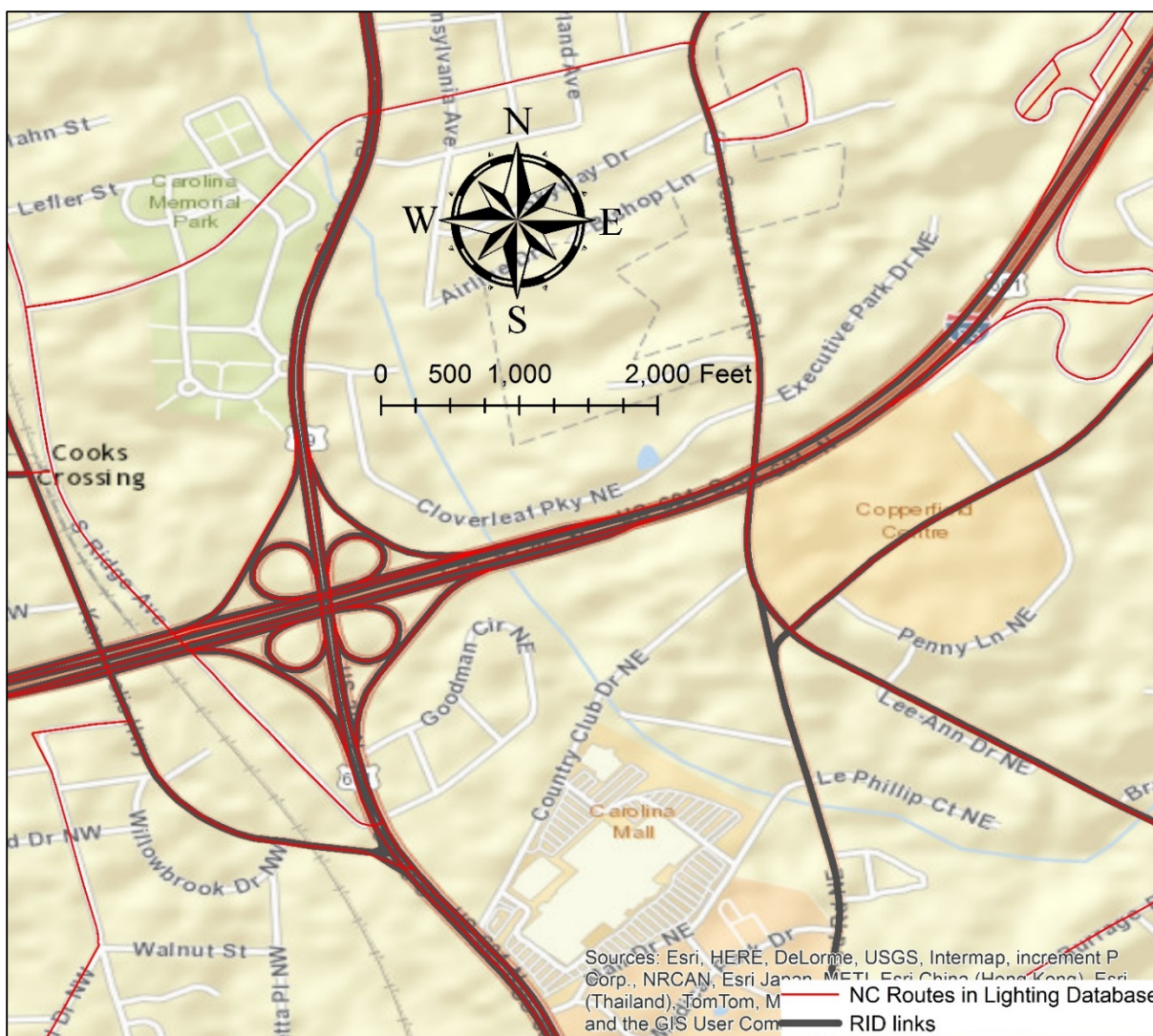
Figure 33. Map. Detailed ALD route network and RID Links.

The research team developed a spatial data conflation procedure to match the RID roadway network with the network used for the North Carolina lighting performance data. The procedure included the following key steps (Figure 34):

1. Split ALD roadway network into shorter segments based on the RID Links feature class. The different routing segmentation criteria used by the ALD roadway network and the RID network prevented researchers from performing accurate data matching. The research team needed to first divide the ALD roadways into equal or shorter segments so that none of the ALD roadway segments overlapped with more than one RID roadway segment. This procedure involved the following sub-steps:

   - Generate end points of RID roadway segments. Using the ArcGIS Feature Vertex to Points tool, the research team generated a point feature class that contained all end points for the RID roadway segments.

- Dissolve RID end points. The point feature class generated during the previous step contained all end points for RID roadway segments. Because both end points of each feature were generated, there were two overlapping points at each location where two adjacent RID road segments connected. To improve data processing efficiency and accuracy, the research team combined overlapping points into single points using the Dissolve tool. The result of this sub-step was a point feature class that contained a single point at each location where a RID roadway segment ended.

- Split ALD roadway network according to the RID network. The research team used the ArcGIS Split Lines at Points tool to divide all lighting database roadway segments at locations where RID roadway segments ended (i.e., locations represented by the point features). The result of this sub-step was a feature class containing the ALD roadways re-segmented according to the RID roadway segments.
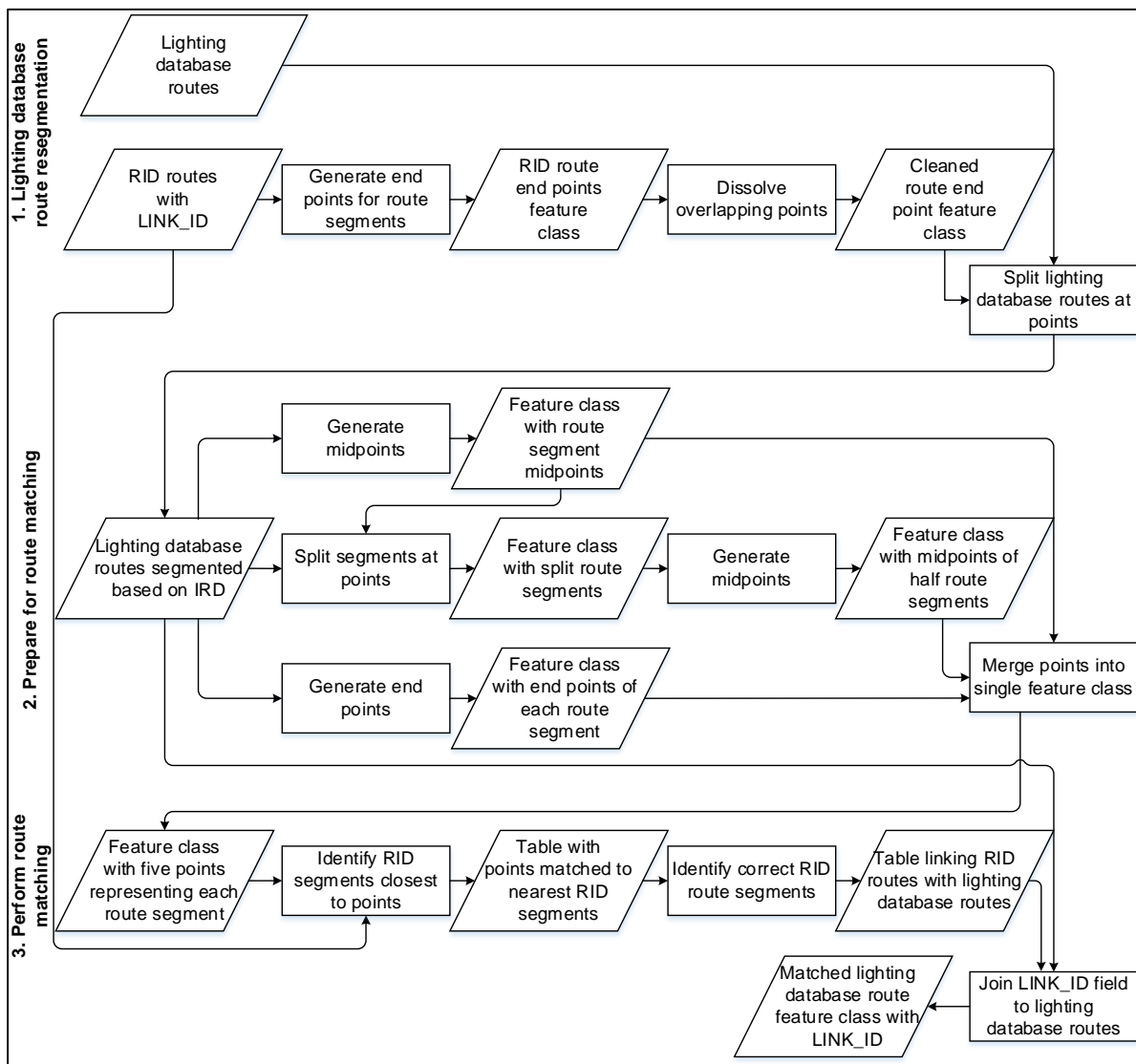
Figure 34. Diagram. Procedure for matching RID routes with ALD routes.

2. Prepare feature class for linking. To address the aforementioned spatial discrepancies between the two roadway networks to be matched, the research team decided to use a point feature class, within which each ALD roadway segment was represented by five key vertices including the two end points and the three points located at quarter length. During the matching process, the nearest RID roadway segment from each point within a 50-foot search radius was identified. The RID segment that was the closest to at least three key points of each ALD roadway segment was considered as the matching segment. This step included the following sub-steps:

- Generate points for the lighting database roadway segments. The current ArcGIS version (i.e., ArcGIS 10.2.2) only has options to extract midpoint, start, end, both ends, or all vertices of a line feature. As such, the five vertices along each segment had to be generated in multiple steps.

  o Generate end points. The researchers first generated a point feature class that contained both end points of each ALD roadway segment with the ArcGIS Feature Vertices To Points tool.

  o Generate midpoints. Using the same tool, the researchers generated a separate feature class containing the midpoints of the ALD roadway segments. In addition to being used as one of the five key vertices, the points were used to split the lighting database segments into two halves so that the midpoint of each half segment could be generated.

  o Split line at midpoints. The lighting database roadway segments were split into two segments of equal length based on the midpoint feature class using the Split Lines at Points tool. The new feature class was then used as the basis to generate the additional two points at each quarter length of the original lighting database roadway segments.

  o Generate midpoints for new segments. Using the same tool, the researchers generated another point layer containing the two midpoints for the newly split roadway segments.

  o Merge point feature classes. Finally, the researchers combined the previously generated point feature classes into a single feature class using the ArcGIS Merge tool. This feature class contained the five key vertices representing each lighting database roadway segment, including its two end points, the midpoint, and the two points at quarter length of each segment.

3. Match LINKID information from the RID roadway network to the lighting roadway network. During this step, the researchers first located the nearest RID roadway segment to each of the five vertices representing a lighting database roadway segment and then identified the RID segment that was the nearest to at least three of the five vertices as the matching segment.

- Identify the segment nearest to each of the five vertices. The researchers used the ArcGIS Near tool for this purpose. This step added the Object IDs of the nearest RID roadway segments to the corresponding point records.

- Determine correct match. The attribute table of the point feature class resulting from the above step was first exported into a Microsoft Access table. The researchers developed a straightforward algorithm in Access and identified the RID roadway segment that was the nearest to at least three of the five vertices of each lighting database roadway segment. The result was a linking table containing Object IDs of ALD roadway segments and the matched RID roadway segments.

- Add LINKID information to lighting database roadway segments. During this step, the researchers joined the RID roadway feature class to the ALD roadway feature class using the linking table produced during the previous step. The LINKID information of the matching RID roadway segments was then copied into a new field of the lighting database roadway feature class.

Through the above process, the research team matched the roadway network used in the North Carolina lighting database with the RID roadway network. The resulting ALD roadway feature class contained roadway segments that were assigned with the correct LINKID information. Because the SHRP 2 NDS time series data were matched to the same LINKID information, and the lighting data points were matched spatially to the ALD roadway feature class, users will be able to link the lighting performance data points to NDS time series data points through SQL queries in conjunction with a basic spatial join process.
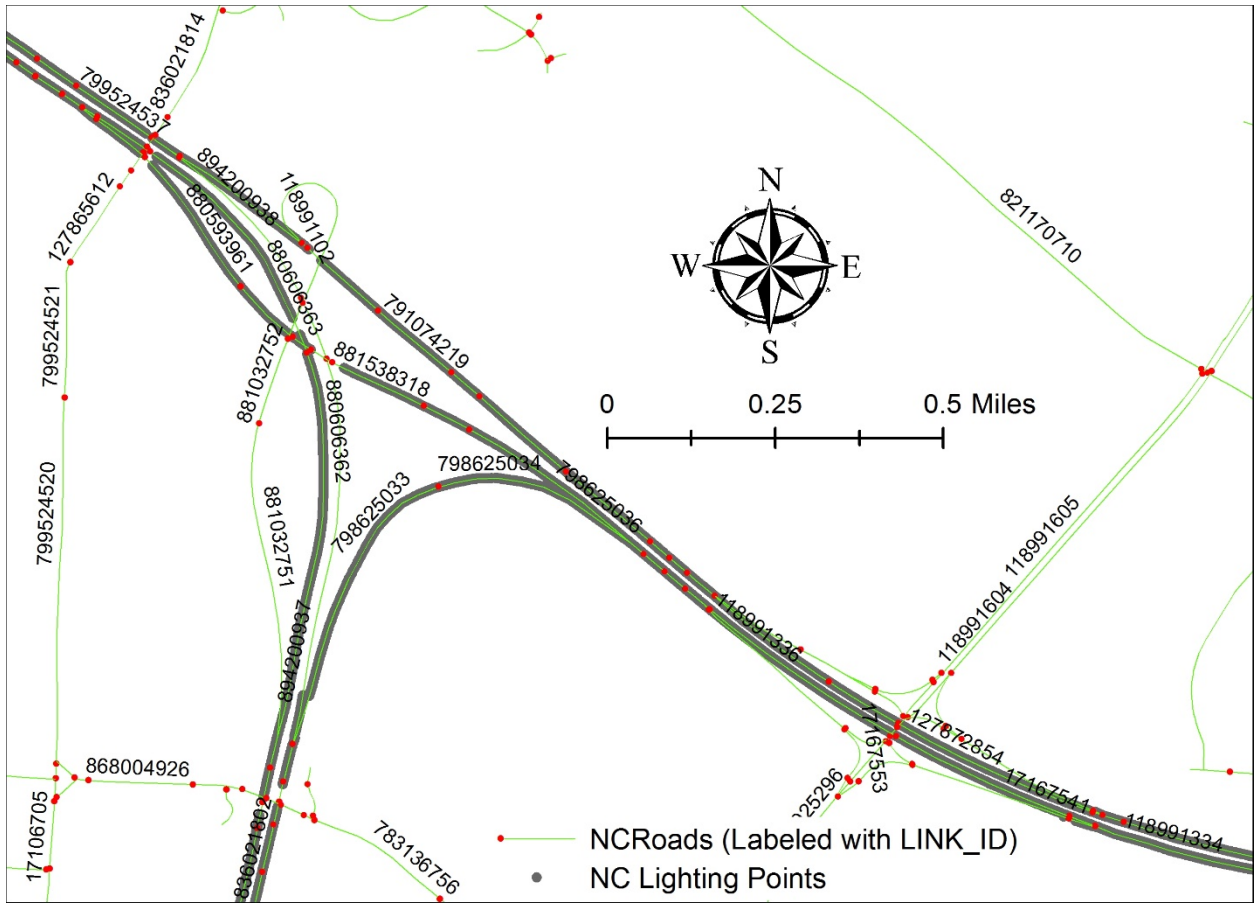
Figure 35: Map. Lighting data points and roadway segments with LINKID information.

CHAPTER 5. CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

The availability of SHRP 2 NDS data and the detailed VTTI lighting performance measurements on more than 2,000 miles of major roads across the nation have made it possible to evaluate and improve the use of roadway lighting as a safety countermeasure to a level that could not be achieved otherwise. However, before any meaningful analysis can be conducted, it was necessary to identify systematic approaches that could effectively integrate the data elements from these two very different data sources. The major purpose of this NSTSCE project was to address this challenge in an effort to make available an integrated dataset containing detailed naturalistic driving, roadway, and lighting information for researchers, the transportation community, and other interested organizations and individuals.

During this project, the research team received access to the SHRP 2 NDS database and conducted a detailed review of the database structure, data entities, relationships, and contents. The research team also described the VTTI ALD in detail to provide the foundation for the development of systematic data integration approaches. The review of the two databases suggested that both databases contained spatial references internally that could be utilized for data integration. In terms of the NDS database, the time series data contained GPS coordinates that would allow users to pinpoint their locations. All other data tables within the database, such as the event, trip, driver, and vehicle data, are either directly or indirectly linked to time series data. In addition, within the VTTI ALD, the detailed lighting measurement data also contained GPS coordinates defining the location where an individual measurement was taken. As such, the research team decided to explore methods that could spatially integrate both databases.

Based on the thorough understanding of the database structures, the research team developed two spatial data integration approaches, each of which targeted different needs and requirements:

- Data integration directly between the ALD and NDS time series data. The researchers demonstrated how time series data could be extracted from the NDS database and directly matched to the ALD roadway network. By matching both lighting and time series data to the same roadway network, individual points from both datasets could be related through simple spatial joins or linear referencing mechanisms. The spatial data integration approach involved both existing and custom tools developed on the ArcGIS platform. Other data items in both databases could then be integrated together through spatial and relational joins. The researchers used data for the state of Washington to demonstrate the approach and associated advantages and challenges. During this activity, time series data representing approximately 2,800 nighttime NDS trips were matched to the ALD roadway network.

- Data integration of the ALD and NDS data based on RID. Within the NDS database, time series data points are matched to a digital map of the roadway network defined by links (uniquely identified by a LINKID). That matching effort will provide researchers with the ability to directly isolate time series data collected on the links of interest, eliminating the need for additional spatial processing. The purpose of this approach was to utilize those results to enable more efficient data integration. Instead of using the NDS time

series data, the research team matched the RID roadway network with the ALD roadway network. After the process, each ALD roadway segment was assigned with the LINKID of the corresponding RID roadway segment. Through the LINKID information within both databases, users can link data elements by performing relational database joins, which are many orders of magnitude faster than spatial joins. To demonstrate this approach, the research team used a draft version of the RID roadway data and lighting data for the state of North Carolina.

The first integration method illustrates a process to link NDS data directly to custom roadway data (e.g., the ALD roadway network and lighting data). This linking provides users a flexible means to cater to unique data integration needs, especially when requiring roadway-related data elements that are not included in the SHRP 2 RID data. Similarly, the approach can be also used to link NDS data with roadways that are not included in the current RID database. However, the first approach requires matching a large number of time series data points spatially to the roadway network. This effort can be time-consuming, especially when using personal computers and desktop GIS software packages. In contrast, the latter integration method utilizes the roadway matching results completed by SHRP 2 and therefore requires much less demanding computing resources and processing time.

In summary, this research serves as the bridge for integrating the VTTI ALD and SHRP 2 NDS data and provides insights in how other spatially related data can be integrated with the NDS data to cater to custom safety study needs. By integrating detailed roadway lighting performance measurement data with the rich SHRP 2 NDS information, a large variety of safety analyses become possible. Examples of such analyses include the following:

- Lighting effects on traffic crashes. Researchers can now analyze how exactly lighting characteristics, such as glare, luminance level, and lighting uniformity affect nighttime safety (e.g., crash rate and severity) based on identified NDS events, including crashes, near-crashes, and baseline events. It is also possible to investigate how exactly lighting plays a role in the sequence of events of nighttime crashes and near-crashes. Findings of such studies may result in improved roadway lighting design requirements or specifications.

- Lighting effects on driver behavior. By analyzing detailed performance measures in the NDS database associated with driver behavior in conjunction with the detailed in situ lighting measurements, researchers may develop a better understanding of how lighting performance affects nighttime drivers in terms of behaviors related to speeding, distraction, driving workload, and drowsiness. Such studies will provide insights in how countermeasures can be developed to improve nighttime highway safety and reduce nighttime driving workload.

- Nighttime traffic control and roadway design improvements. By integrating the NDS data, including its RID data and the detailed lighting performance data, it becomes possible to conduct correlation analyses between safety surrogate measures and nighttime environmental factors such as roadway features, traffic control visibility, and lighting performance. The understanding of the complex relationship between these elements is a

critical step towards nighttime safety improvement through more effective traffic control measures and roadway design.

- Human-vehicle interactions during nighttime. With the detailed vehicle, driver, and naturalistic driving and event videos in the NDS database and the VTTI lighting performance data, researchers can carry out studies of how certain vehicle characteristics (e.g., headlamp brightness and operation angle, sight distances, mirror reflectiveness, and certain interior features including sight obstructions within vehicles) affect roadway safety and driving workload in specific roadway settings. Such results can be particularly beneficial to automobile manufacturers for improving vehicle interior and performance features affecting nighttime driving.

RECOMMENDATIONS

Based on this project, the researchers recommend the following:

- NDS-related improvements:

  o Develop comprehensive, user-friendly documentation, particularly for the integration of datasets relevant to the SHRP 2 NDS database. As would be expected of an ongoing and dynamic project, documentation is needed, particularly in areas such as the logical and physical data models depicting data entities, attributes, and relationships.

  o Consider incorporating documentation from this project into the SHRP 2 NDS documentation, assuming the final data structure is similar to the current one. The data integration approaches developed in this project provide examples for users to readily use with minimal modification to integrate custom roadway or other spatially referenced data with NDS data. Inclusion of documentation from this effort into training materials to facilitate users with similar needs could be beneficial to current and potential NDS data users.

  o Facilitate researchers in understanding how to fully use the SHRP 2 roadway matching results. In many research efforts, once time series data on roadway segments of interest are retrieved from the SHRP 2 Linking Table, users need to further locate data points within individual links. This project uses GIS techniques to match individual time series data points spatially to roadway segments. The SHRP 2 matching results also enable users to use speed and timestamp data of the time series data points as a linear referencing mechanism to compute their exact locations within a roadway segment. Providing researchers with documentation (including training materials) of how to perform such tasks will permit users to integrate data with high efficiency and enable location specific analysis as required.

- ALD-related improvements:

  o Develop more aggregated lighting performance measurements to reduce data processing demand. The VTTI lighting performance measurements were collected in

situ at a frequency of 20 Hz. This practice resulted in a large number of lighting data points densely located next to each other. To assist potential users, it would be beneficial to aggregate consecutive lighting measurement points with similar values into single line features. This process will result in a significantly reduced number of lighting data records without sacrificing a great deal of data accuracy.

- o Linearly reference lighting data points. The ALD research team should linearly reference the lighting measurement data points to the official state roadway networks as used in the database. Embedding the data points into the state official linear referencing system will enable seamless integration between the lighting performance measurements and other datasets maintained at state DOTs, such as roadway information, traffic data, crash data, infrastructure data, and roadway project or work zone data. Based on the linear referencing information, such data integration can be performed accurately not only with a GIS but also using relational database management tools.

- Further studies and next step. The integration of the VTTI in situ lighting measurements and the SHRP 2 NDS data has created opportunities for a number of meaningful studies towards safer and more driver-friendly nighttime driving environments. Among the aforementioned potential studies, the researchers recommend that the following high-priority studies be conducted in the near future:

  - o Effects of detailed lighting performance characteristics on nighttime crashes and driver behavior. This study will require a comprehensive causation study of roadway lighting performance characteristics and nighttime crashes, near-crashes, speeds, and other surrogate measures of safety.

  - o Nighttime traffic control visibility and effectiveness at different lighting levels. This study will involve an in-depth investigation of the visibility and effectiveness of certain traffic control measures as they relate to roadway safety. The detailed RID roadway and traffic control information, NDS sequence of events and video data, and VTTI in situ lighting performance measurements provide a unique opportunity for this study.

Currently, the research team is in the process of conducting a follow-up phase to investigate the feasibility of using SHRP 2 NDS data in conjunction with the data matching approaches and results to study the effectiveness of selected traffic control measures.

REFERENCES

1. Griffith, M. S. (1994). Comparison of the safety of lighting options on urban freeways. *Public Roads*, *58*(2). Retrieved from https://www.fhwa.dot.gov/publications/publicroads/94fall/p94au8.cfm
2. Zhu, Y., Kuhn, T. , Majo, P., & Hinds, W. C. (2006). Comparison of daytime and nighttime concentration profiles and size distributions of ultrafine particles near a major highway. *Environmental Science Technology, 40,* 2531–2536.
3. Illuminating Engineering Society of North America. (2000, reaffirmed 2005). *American national standard practice for roadway lighting* (ANSI/IESNA RP-8-00). New York: Author.
4. Lighting Research Center, Rensselaer Polytechnic Institute. (2008). Mesopic street lighting system conserves energy. Troy, New York: Author. Retrieved from http://www.lrc.rpi.edu/resources/newsroom/pdf/2008/MesopicGroton8511.pdf
5. Box, P. C. (1989). Major road accident reduction by illumination. *Transportation Research Record, 1247*, 32–38.
6 . Box, P. C. (1971). *Relationship between illumination and freeway accidents*. Illuminating Engineering Research Institute, 1971.
7. Federal Highway Administration. (2014). *FHWA research and technology agenda*. Retrieved from http://www.fhwa.dot.gov/research/fhwaresearch/agenda/printresearcharea.cfm?urlanchor=safety
8. Transportation Research Board of the National Academies. (2014). The SHRP 2 Naturalistic Driving Study (NDS). Retrieved from http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Pages/The-SHRP-2-Naturalistic-Driving-Study-472.aspx
9. McLaughlin, S., & Hankey, J. (2015). *Naturalistic Driving Study: Linking the Study Data to the Roadway Information Database*. SHRP 2 Report S2-S31-RW-3, Virginia Tech Transportation Institute, Blacksburg, Virginia.
10. Federal Highway Administration. (2014). *SAFETEA-LU – Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users*. Retrieved from https://www.fhwa.dot.gov/safetealu/
11. Transportation Research Board of the National Academies. (2014). Strategic Highway Research Program 2 (SHRP 2). Retrieved from http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/General.aspx
12. Transportation Research Board of the National Academies. (2010, December). *Safety program brief: Updating safety research in SHRP 2*. Retrieved from http://onlinepubs.trb.org/onlinepubs/shrp2/SafetyBrief.pdf
13. Transportation Research Board of the National Academies. (2012). *Strategic Highway Research Program (SHRP) 2 – Revised safety research plan: Making a significant improvement in highway safety*. Retrieved from http://onlinepubs.trb.org/onlinepubs/shrp2/RevisedSafetyResearchPlanMarch2012.pdf
14. Transportation Research Board of the National Academies. (2014). (2014). *SHRP 2 Safety Projects*. Retrieved from http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Pages/Safety_Projects_304.aspx

15. Campbell, K. (2012, September–October.). *The SHRP 2 Naturalistic Driving Study: Addressing driver performance and behavior in traffic safety*. *TR News*, *282*. Retrieved from https://insight-forums.shrp2nds.us/uploads/FileUpload/80/9cb98204a5cf10632df695a674b4c9.pdf
16. Virginia Tech Transportation Institute. (2014). *SHRP 2 NDS InSight – Data*. Retrieved from https://insight.shrp2nds.us/home/index
17. Virginia Tech Transportation Institute. (2014). *International Center for Naturalistic Driving Data Analysis at Virginia Tech*. Retrieved from https://www.vtti.vt.edu/facilities/data-services.html
18. SHRP 2 Request for Proposals for Roadway Information Database Development and Technical Coordination and Quality Assurance of the Mobile Data Collection Project (S04B). (2009, September 4.)

APPENDIX A. IRB PROTOCOL FOR NDS DATA ACCESS

VirginiaTech

Institutional Review Board
Existing Data Research Protocol

Note: complete this application only if this research project *only* involves the collection or study of **existing data**. Once complete, upload this form as a Word document to the IRB Protocol Management System: https://secure.research.vt.edu/irb

1. DO ANY OF THE INVESTIGATORS OF THIS PROJECT HAVE A REPORTABLE CONFLICT OF INTEREST? (http://www.irb.vt.edu/pages/researchers.htm#conflict)

☒ **No**
☐ **Yes,** explain:

2. WILL THIS RESEARCH INVOLVE COLLABORATION WITH ANOTHER INSTITUTION?

☒ **No,** go to question 3
☐ **Yes,** answer questions within table

IF YES
Provide the name of the institution [for institutions located overseas, please also provide name of country]:
Indicate the status of this research project with the other institution's IRB: ☐ Pending approval   ☐ Approved   ☐ Other institution does not have a human subject protections review board   ☐ Other, explain:
Will the collaborating institution(s) be engaged in the research?   (http://www.hhs.gov/ohrp/policy/engage08.html)   ☐ No   ☐ Yes
Will Virginia Tech's IRB review all human subject research activities involved with this project?   ☐ No, provide the name of the primary institution: ☐ Yes   *Note: primary institution = primary recipient of the grant or main coordinating center*

3. IS THIS RESEARCH FUNDED?

☐ **No,** go to question 4
☒ **Yes,** answer questions within table

IF YES
Provide the name of the sponsor [if NIH, specify department]:
Is this project receiving federal funds?   ☐ No   ☒ Yes

1

> **If yes,**
>
>> **Does the grant application, OSP proposal, or "statement of work" related to this project include activities involving human subjects that are <u>not</u> covered within this IRB application?**
>> - ☒ No, all human subject activities are covered in this IRB application
>> - ☐ Yes, however these activities will be covered in future VT IRB applications, these activities include:
>> - ☐ Yes, however these activities have been covered in past VT IRB applications, the IRB number(s) are as follows:
>> - ☐ Yes, however these activities have been or will be reviewed by another institution's IRB, the name of this institution is as follows:
>> - ☐ Other, explain:
>>
>> **Is Virginia Tech the primary awardee or the coordinating center of this grant?**
>> - ☐ No, provide the name of the primary institution:
>> - ☒ Yes

4. DOES THIS STUDY INVOLVE CONFIDENTIAL OR PROPRIETARY INFORMATION (OTHER THAN HUMAN SUBJECT CONFIDENTIAL INFORMATION), OR INFORMATION RESTRICTED FOR NATIONAL SECURITY OR OTHER REASONS BY A U.S. GOVERNMENT AGENCY?

For example – government / industry proprietary or confidential trade secret information

☒ No
☐ Yes, describe:

5. DOES THIS STUDY INVOLVE SHIPPING ANY TANGIBLE ITEM, BIOLOGICAL OR SELECT AGENT OUTSIDE THE U.S?

☒ No
☐ Yes

6. DESCRIBE THE BACKGROUND, PURPOSE, AND ANTICIPATED FINDINGS OF THIS STUDY:

> The FHWA project "Strategic Initiative for Evaluation of Reduced Lighting on Roadways" will evaluate all of the issues associated with the application of adaptive lighting to the roadway environment and develop recommended practices for adaptive lighting systems. VTTI is conducting an intensive data collection and processing efforts in eight states, for six primary types of data. Including: crash data, traffic data, lighting design/maintenance, and in-situ-data lighting data collection and roadway in NC and WA. As part of the SHRP 2 Naturalistic Driving Study Database, VTTI is also collecting Naturalistic driving data. The purpose of this project is to explore the mechanisms to link the VTTI lighting database to the NDS.

7. EXPLAIN WHAT THE RESEARCH TEAM PLANS TO DO WITH THE STUDY RESULTS:

For example - publish or use for dissertation

> The results will be compiled into a report for the sponsor, presentations and published in scientific journals.

8. WILL PERSONALLY IDENTIFYING STUDY RESULTS OR DATA BE RELEASED TO ANYONE OUTSIDE OF THE RESEARCH TEAM?

For example – to the funding agency or outside data analyst, or participants identified in publications with individual consent

2

☒ No
☐ **Yes,** to whom will identifying data be released?

9. WILL ANY STUDY FILES CONTAIN PARTICIPANT IDENTIFYING INFORMATION (E.G., NAME, CONTACT INFORMATION, VIDEO/AUDIO RECORDINGS)?

☒ **No,** go to question 10
☐ **Yes,** answer questions within table ⟶

IF YES
Describe if/how the study will utilize study codes:
If applicable, where will the key [i.e., linked code and identifying information document (for instance, John Doe = study ID 001)] **be stored and who will have access?** *Note: the key should be stored separately from subjects' completed data documents and accessibility should be limited.*

10. WHERE WILL DATA BE STORED?

Reduced data obtained from this current effort will be stored in secure digital project folders on a secure server at VTTI that are only accessible by personnel directly involved in the project, with the principal investigator's approval. The original data will remain stored as described in the IRB-approved protocol for the original project.

11. WHO WILL HAVE ACCESS TO STUDY DATA?

Only the personnel directly involved in the project will have access to the data for the purposes of this specific project. Any other projects requiring use of the linked dataset would undergo additional IRB review and would require an additional data sharing agreement.

12. DESCRIBE THE PLANS FOR RETAINING OR DESTROYING THE STUDY DATA:

The data from the original study will be retained as described in the IRB-approved protocols for the SHRP2 NDS (IRB# 09-953). The reduced data obtained for this current research effort, will be stored in a separate project folder on a secure server at VTTI, with summary and reduced de-identified data resulting from this project retained indefinitely.

13. FROM WHERE DOES THE EXISTING DATA ORIGINATE?

The existing data were collected by the SHRP2 NDS (IRB# 09-953). Participants enrolled in this study agreed to the future use of their data by qualified researchers. The VTTI lighting database originated from the Strategic Initative for Evaluation of Reduced Lighting on Roadways, the first modern large-scale study to use field-collected lighting data to evaluate the relationship between actual lighting levels and traffic safety and represents the first effort to collect lighting performance data at a national level. To fulfill the project's special requirements, six primary data types are included in this database: crash data, traffic data, roadway design, lighting design, and in-situ-data lighting data collection. After an extensive initial data request and evaluation, seven states were selected to be included in this study: California, Delaware, Minnesota, North Carolina, Vermont, Virginia, and Washington. The data sources are shown below:

3

State	CA	DE	NC	MN	VA	VT	WA
GIS Road Network	CALTRANS	DE DOT	NC DOT	MN DOT	VA DOT	VT DOT	WA DOT/HSIS
Lighting Performance Data	VTTI	VTTI	VTTI	VTTI	VTTI	VTTI	VTTI
Lighting Design Data	CALTRANS	DEDOT	NCDOT	MN DOT	VDOT	VTDOT	WADOT/City of Seattle
Crash Data	HSIS	DE DOT	HSIS	HSIS	VA DOT	VT DOT	WA DOT/HSIS
Roadway Data	HSIS	DE DOT	HSIS	HSIS	VA DOT	VT DOT	WA DOT/HSIS
Traffic Data	CALTRANS	DE DOT	NC DOT	MN DOT	VA DOT	VT DOT	WA DOT

In addition, the research team also acquired sunrise/sunset data from the National Oceanic and Atmospheric Administration (NOAA). The GIS base road network was the key for the data structure. The research team has acquired directly from the DOT or the DOT website the GIS data for all seven states included in the study. The quality of the data was examined by cross-reference with multiple data sources, such as field collected GPS coordinates. Two primary geo-reference systems were used in this database: a linear referencing system (LRS) and a coordinate (latitude/longitude) system.

The lighting performance data conveys detailed information on the current performance of the lighting system data were measured in situ using a mobile data collection system Roadway Lighting Mobile Measurement System (RLMMS). The RRLMMS has the ability to collect illuminance, luminance camera, in-vehicle responder data, and GPS data. Members of the project team drove on the routes selected and collected lighting data at the frequency of 20 Hz and the rate of video capture is 3.75 frames per second (fps). The data collection vehicle adhered to the posted speed limit and followed the direction of roadway signage. Due to the significant number of miles where the lighting performance data were collected, multiples files were generated.

The acquired data were integrated with ARC-GISTM software with the capability of converting the roadway lighting data gathered and forming a visual database. Using GPS coordinates and the Minolta illuminance values collected via the RLMMS, the resulting dataset was uploaded to special software that can be used to generate maps highlighting all of the data points and detailing their illuminance.

14. PROVIDE A DETAILED DESCRIPTION OF THE EXISTING DATA:

The second Strategic Highway Research Program (SHRP2) Naturalistic Driving Study (S07 NDS) collected more than 35 million vehicle miles of naturalistic data, inclduing GPS coordinates that can be linked to roadway and other non-human subjects infrastructure databases. We feel that this proejct will enrich the usability of the SHRP 2 database for future users interested not only in roadway lighting questions but also roadway, crash and traffic data. The lighting and sunrise/sunset data are not human subjects data. The research team also acquired sunrise/sunset data from the National Oceanic and Atmospheric Administration (NOAA). The GIS base road network was the key for the data structure. The research team has acquired directly from the DOT or the DOT website the GIS data for all seven states included in the study. The quality of the data was examined by cross-reference with multiple data sources, such as field collected GPS coordinates. Two primary geo-reference systems were used in this database: a linear referencing system (LRS) and a coordinate (latitude/longitude) system.

The lighting performance data conveys detailed information on the current performance of the lighting system data were measured in situ using a mobile data collection system Roadway Lighting Mobile Measurement System (RLMMS). The RRLMMS has the ability to collect illuminance, luminance camera, in-vehicle responder data, and GPS data. Members of the project team drove on the routes selected and collected lighting data at the frequency of 20 Hz and the rate of video capture is 3.75 frames per second (fps). The data collection vehicle adhered to the posted speed limit and followed the direction of roadway signage. Due to the significant number of miles where the lighting performance data were collected, multiples files were generated.

The acquired data were integrated with ARC-GISTM software with the capability of converting the roadway

4

lighting data gathered and forming a visual database. Using GPS coordinates and the Minolta illuminance values collected via the RLMMS, the resulting dataset was uploaded to special software that can be used to generate maps highlighting all of the data points and detailing their illuminance.

15. IS THE SOURCE OF THE DATA PUBLIC?

☒ **No,** go to question 16
☐ **Yes,** you are finished with this application

16. WILL ANY INDIVIDUAL ASSOCIATED WITH THIS PROJECT (INTERNAL OR EXTERNAL) HAVE ACCESS TO OR BE PROVIDED WITH EXISTING DATA CONTAINING INFORMATION WHICH WOULD ENABLE THE IDENTIFICATION OF SUBJECTS:

- **Directly** (e.g., by name, phone number, address, email address, social security number, student ID number), or
- **Indirectly through study codes** even if the researcher or research team does not have access to the master list linking study codes to identifiable information such as name, student ID number, etc

or

- **Indirectly through the use of information that could reasonably be used in combination to identify an individual** (e.g., demographics)

☒ **No,** collected/analyzed data will be completely de-identified
☐ **Yes,**

If yes,

Research will not qualify for exempt review; therefore, if feasible, written consent must be obtained from individuals whose data will be collected / analyzed, unless this requirement is waived by the IRB.

Will written/signed or verbal consent be obtained from participants prior to the analysis of collected data?
No, requesting waiver from IRB

This research protocol represents a contract between all research personnel associated with the project, the University, and federal government; therefore, must be followed accordingly and kept current.

Proposed modifications must be approved by the IRB prior to implementation except where necessary to eliminate apparent immediate hazards to the human subjects.

Do not begin human subjects activities until you receive an IRB approval letter via email.

It is the Principal Investigator's responsibility to ensure all members of the research team who collect or handle human subjects data have completed human subjects protection training prior to handling or collecting the data.

----------END----------

5

APPENDIX B. DATA SHARING AGREEMENT FOR NDS DATA ACCESS

Data Sharing Agreement

Onsite Use of Identifying Video and Driving Data
Offsite Use of Non-Identifying Driving, Vehicle, Participant, and Crash Data

<u>Disclaimer:</u> This data sharing agreement template has been developed in accordance with the terms specified in the consent forms that participants signed and thus represents a required minimal set of safeguards for participant data. Additional safeguards going above and beyond what the consent document requires may be specified by the individual Institutional Review Boards as they review requests for analyses of the data. Thus, any future modifications or additions to this template will provide additional protections for the use of participant data and will never reduce the protections accorded in the consent document.

Use of Identifiable Video and Driving Data in Secure Data Enclaves

When a researcher, research team, or research institution (hereafter referred to as the receiving agency) requests access to an existing SHRP 2 dataset containing identifying data, the data analysis shall be conducted in a designated secure data enclave within VTTI's facilities. Identifying data for the purposes of this agreement include face video, GPS coordinates, and any other data by which the identity of the participant may be revealed. In this situation, the client comes to the data warehouse site to run analyses in coordination/cooperation with VTTI researchers. The data enclave will be physically and securely separated from other data reduction and analysis efforts at VTTI. All work will be monitored and supported by VTTI staff and completed within the confines of the enclave. There will be an hourly all-inclusive fee for use of the enclave/data which will include the cost of VTTI support.

Release of Non-Identifying Driving, Vehicle, Participant, and Crash Data

Release of streamed data describing driving epochs requires thorough de-identification of data prior to release. De-identification activities (performed by VTTI personnel) and shipping costs will be paid for by the receiving agency. De-identification includes removing at a minimum:

- Dates and times (for example, March 15, 2010 06:45am could be changed to March, Monday, 6am-12pm)
- Voiceprints
- Full face photos, videos, & comparable images
- De-identifying GPS coordinates
- Full trip files with starting and ending locations shown via forward video
- Files with identifiable highway signs and footage of a high-profile incident (such that research participant identity could be uncovered via news reports)
- Any other types of data that could be used to identify a research participant

Data Sharing Agreement

Use of the data enclave and offsite use of non-identifying driving epochs requires a data sharing agreement signed by the receiving agency. This document indicates agreement with the following:

- The receiving agency must provide a detailed proposal with researcher qualifications prior to beginning work with the dataset. Qualifications should indicate familiarity with and previous use of confidential or proprietary data using human research participants.
- The receiving agency must first obtain IRB permission to conduct the data analysis, and all parties who will be working with identifying data must undergo IRB training. The original research participant consent form will be shared with the receiving agency as part of this process (attached, with data sharing clauses highlighted).
- The receiving agency may not copy or remove files containing identifying data from the data enclave. Reduced, non-identifying files will be provided to the receiving agency by VTTI staff. To ensure data have been de-identified, it may be necessary for VTTI staff to further review the content of files before delivery.
- All personnel working with the data must agree to the working conditions such as leaving cell phones and cameras at the entrance of the data reduction laboratory.
- The receiving agency must agree not to attempt to learn the identity of research participants (e.g., using GPS and video data to locate the research participant's home or work address).
- If the receiving agency discovers identifying information or data in a dataset that was intended to be non-identifying, they must agree to provide that information to VTTI so that it can be properly de-identified for future use (for example, a pedestrian's face is visible and identifying in the forward view).

- The receiving agency must agree not to use data for purposes other than specified in the analysis plan; an additional data sharing agreement will be required for each new set of analyses.
- The receiving agency must agree not to show any identifying data at research conferences.
- The receiving agency agrees to properly acknowledge the source of the data in any reports or articles resulting from the analyses.
- Optional: The receiving agency agrees to return the reduced dataset to VTTI for to be made available to future researchers. In some cases the reduced dataset will have a proprietary nature and can be placed on hold for up to five years before it is provided to other researchers (for example, an OEM develops a crash avoidance algorithm that they hope to incorporate in their future fleet).
- All personnel who will be working with the data must agree not to release or share information leading to the identification of participants or to release or share non-identifying raw data.

Instructions – please fill out the form and send it back to the Data Sharing Manager at VTTI (datasharing@vtti.vt.edu). The Data Sharing Manager will review the information and send it back to you. You may then sign it and either send a scanned copy of the signed form to datasharing@vtti.vt.edu or fax it to 540-231-1555, attn. Suzie Lee. The Data Sharing Manager will then sign the form and send it back to you, at which point the data sharing process will begin, assuming that the appropriate fiscal contracts are in place at that point.

Data Sharing Agreement between <u>Receiving Agency</u> and Virginia Tech Transportation Institute.

1. **Please describe the scope of the proposed analysis (1 paragraph). Please include the full project title and the research sponsor.**

Project Title: Exploration of the integration of the Naturalistic Driving data with the VTTI Lighting Database

Project Sponsor: The National Surface Transportation Safety Center for Excellence (NSTSCE) at VTTI. NSTSCE was established by the Federal Public Transportation Act of 2005 to develop and disseminate advanced transportation safety techniques and innovations in rural and urban communities.

Project Scope: As part of an FHWA project, VTTI has developed an Adaptive Lighting Database for eight states with six primary types of data including: crash data, traffic data, lighting design/maintenance, and in-situ-data lighting data collection and roadway. In addition, the research team also acquired sunrise/sunset data from the National Oceanic and Atmospheric Administration (NOAA). The purpose of this project is to explore the possibility to incorporate the NDS data as an addition dataset in the City of Seattle state Database. One important consideration is that the lighting data collection was performed at the same time that NDS data was collected. The possibility to have a database that includes field lighting, traffic, safety, roadway, weather, and naturalistic driving data almost in the same timeframe or a very close time envelope represents a unique opportunity for the transportation community to perform critical safety and operations studies.

2. **Please describe the dataset you expect to receive (1 paragraph).**

Data Specification:
Time Series data: The major purpose of this project is to explore linkages between the Adaptive Lighting Database and the NDS database. To meet this purpose, the research team will attempt to link Time Series data points on a number of major highway corridors in Washington and North Carolina where the research team previously collected lighting data. The research team will not use origin and destination data points of the time series files as they are located on minor roads that are not covered by the research team's lighting database.
Specified data fields (Speed GPS Details, Speed Vehicle Network Details, Acceleration x-axis Details, Acceleration y-axis Details, Yaw Rate z-axis Details, Distance Details, Headlight Setting Details, Acceleration x-axis fast Details, Acceleration y-axis fast details, Acceleration z-axis Details, Acceleration z-axis

- fast Details, Day Details, Head Confidence Details, Heading GPS Details, Illuminance Ambient Details, Lane Marking Distance Left, Lane Marking Distance Right, Lane Marking Type Right, Lane Marking Type Left, Lane Width Details, Latitude Details, Longitude Details, Location Details, Month Details, Subject ID Details, Time Details, Vehicle ID Details, Video Frame Details, and Year Details) for:
  - data points within the lat/long boundary defined by (clockwise): (-123.039, 48.195), (-121.936, 48.195), (-121.936, 46.814), and (-123.039, 46.814) in Washington
  - data points within the lat/long boundary defined by (clockwise): (-80.443, 36.176), (-78.47, 36.176), (-78.47, 35.589), and (-80.443, 35.589) in North Carolina

Event Detail table data: Knowing that events have not been fully processed, the research team would like to identify how sample events can be linked to the Time Series data points that have been linked to the lighting database. The research team will use the sample data to understand the Event Detail table structure and identify potential types of the analyses that can be performed in conjunction with the research team's lighting data.

- Specified data fields (Event ID Details, Participant ID Details, Event Severity 1 Details, Event Severity 2 Details, Event Start Details, Subject Reaction Start Details, Impact of Proximity Time Details, Event End Details, Pre-Incident Maneuver Details, Maneuver Judgment Details, Precipitating Event Details, Event Nature 1 Details, Incident Type 1 Details, Crash Severity 1 Details, V1 Evasive Maneuver 1 Details, V1 Post-Maneuver Control 1 Details, Event Nature 2 Details, Incident Type 2 Details, Crash Severity 2 Details, V1 Evasive Maneuver 2 Details, V1 Post-Maneuver Control 2 Details, Vehicle Rollover Details, Driver Behavior 1 Details, Driver Behavior 2 Details, Driver Behavior 3 Details, Driver Impairments Details, Secondary Task 1 Details, Secondary Task 1 Start Time Details, Secondary Task 1 End Time Details, Secondary Task 1 Outcome Details, Secondary Task 2 Details, Secondary Task 2 Start Time Details, Secondary Task 2 End Time Details, Secondary Task 2 Outcome Details, Driver Seatbelt Use Details, Vehicle Contributing Factors Details, Infrastructure Details, Visual Obstruction Details, Lighting Details, Weather Details, Surface Condition Details, Traffic Flow Details, Contiguous Travel Lanes Details, Through Travel Lanes Details, V1 Lane Occupied Details, Traffic Density Details, Traffic Control Details, Relation to Junction Details, Alignment Details, Grade Details, Locality Details, Construction Zone Details, Fault Details, Final Narrative Details, and Version History) for:
  - Currently identified events associated with the above time series files

Route Linking Table data: This data table contains matching information between selected Time Series data points and NAVTEQ roadway segments. The table itself does not contain personally identifying information.
- All data fields for:
  - data points within the lat/long boundary defined by (clockwise): (-80.443, 36.176), (-78.47, 36.176), (-78.47, 35.589), and (-80.443, 35.589)

The research team will also view the following tables for the purpose of understanding their structures and key fields that can be used to link back with the Time Series data. With this information, the research team will identify the types of potential analyses that can be performed in conjunction with the research team's lighting data.

Vehicle Detail Table

Trip Summary Table

Driver Demographic Questionnaire

Driving History Questionnaire

Driver Behavior Questionnaire

1. Please describe the researcher qualifications (1 paragraph per researcher).

Ron Gibbons (PI, IRB certified). Ron Gibbons is director of the Center for Infrastructure Based Safety Systems at VTTI. He is also the Institute's lead lighting, visibility and weather research scientist. Dr. Gibbons is responsible for lighting- and visibility-associated research projects. He is currently the PI on projects that consider the lighting of overhead signage, modeling the human visual system, the development of lighting criteria for rural intersections and the assessment of Roadway Lighting Safety. He is the author of several published papers on roadway lighting, photometry, and target visibility.

Alejandra Medina Flintsch (IRB certified). Alejandra has more than 20 years of professional experience as a transportation engineer, researcher, and trainer with a strong commitment to excellence. Ms. Medina's main areas of expertise in the safety area includes Safety and Traffic Modeling, Accident Analysis, Accident Databases, Infrastructure/ Asset Management, Roadway Design and Traffic Control Benefit Costs Analysis, Road Safety Inspection, Road Safety Audits, She is the PI /Co-PI and key researcher for several national safety evaluation projects for FHWA, VDOT, NHRDCP and FMCSA including the NHRCP Evaluation of Sag Vertical Curves the FHWA Exploratory Research Idea Driving Behavior in Traffic, and FMCSA On Board Safety Measures Effectiveness.

Yingfeng (Eric) Li (IRB certified). Eric Li is currently a Research Associate at VTTI. He has tremendous experience in cutting edge research, technical support, and technology transfer pertaining to the nation's transportation system. Dr. Li has led, co-led, or played a critical role in numerous projects from national and local sponsors.

Feng Guo (IRB certified). Feng is an expert on traffic safety analysis and Bayesian modeling. Dr. Guo is an assistant professor at the department of statistics and has a joint appointment at the Virginia Tech Transportation Institute (VTTI). He holds dual Ph.D.s in statistics and in civil engineering. He was the leading statisticians for dozens of research projects sponsored by NHTSA, FMCSA, and NIH. He is the Principle Investigator (PI) or Co-PI of several projects related to development of statistical analyses method for naturalistic driving study and traffic safety research.

Dr. Guo's areas of expertise include traffic safety modeling, epidemiology study design, categorical data analysis, Bayesian method, and analyzing naturalistic driving study data.

Rajaram Bhagavathula (IRB certified). Rajaram Bhagavathula is a Research Associate at VTTI. He's research interests are primarily in the areas of roadway lighting and visibility, traffic safety, and traffic control. He has worked on several research projects sponsored by various national and local transportation agencies.

Travis Terry (IRB certified). Travis Terry is currently a Research Associate at the Center for Infrastructure-Based Safety Systems (CIBSS) of VTTI. He has played a critical role in many research projects pertaining to

Jianhe Du (IRB certified). Jianhe Du is a research associate in the Center of Sustainable Mobility. She has experiences in the areas of traffic engineering, transportation planning, GIS modeling, and transportation safety modeling. She has worked on several research projects including: SHRP II L10, Sag Curve Design Criteria for Headlight Sight Distance, Advanced Crash Avoidance Technologies, Backing Crash Countermeasures, and other human factor and GIS related research.

1. **Please describe what you plan to do with the data when your analyses are complete.**

The major purpose of this study is to explore the possibility of integrating NDS data with the research team's Adaptive Lighting Database. The data will be used to verify if the integration is feasible and how such an integration should be performed. The process will be documented in a research report, with any descriptions pertaining to NDS data in a highly aggregated manner to ensure the protection of potentially personal identifying information. In addition to the research report, the research may result in scientific journal papers based on the contents of the report. Based on the project timeline, the research team will need access to the requested data for 12 months immediately after the access is granted.

2. **Please provide proof of IRB permission to conduct the data analysis OR proof of an official exemption from IRB approval. As part of this, all researchers/analysts should provide proof of IRB training. These may be included as attachments.**
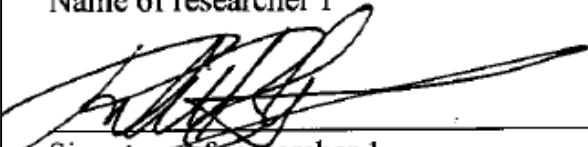
3. **In signing this data sharing agreement, the receiving agency agrees not to attempt to learn the identity of research participants.**

4. **In signing this data sharing agreement, the receiving agency agrees to not distribute the data to other entities or use it for purposes other than those specified in the scope of the proposed analysis. The receiving agency agrees to hold the data in reserve only to answer questions relating to the project described in this data sharing agreement, and to seek an additional data sharing agreement prior to using the data for any other purpose. An additional IRB approval will also be required for additional uses of the data.**

The receiving agency should not sign until all requested information has been received and the agreement has been approved and signed by the VTTI Data Sharing Manager.

Ron Gibbons
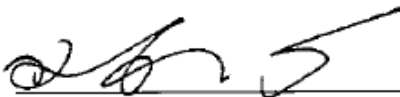Name of researcher 1

05/21/14
Date

Signature of researcher 1

Alejandra Medina Flintsch
Name of researcher 2

05/20/2014.
Date

Signature of researcher 2

Eric Li
Name of researcher 3

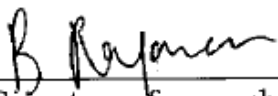05/20/2014
Date

Signature of researcher 3

Feng Guo
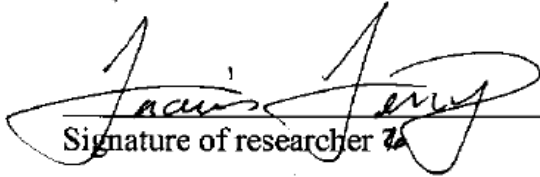Name of researcher 4

5/20/2014
Date

Signature of researcher 4

Rajaram Bhagavathula
Name of researcher 5

5/20/2014
Date

Signature of researcher 5

Travis Terry
_____
Name of researcher 6
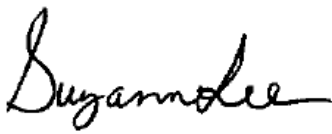
5/29/2014
_____
Date

_____
Signature of researcher 6

Jianhe Du
_____
Name of researcher 7

5/20/2014
_____
Date

_____
Signature of researcher 7

Suzanne Lee
_____
Name of VTTI Data Sharing Manager

May 20, 2014
_____
Date

_____
Signature of VTTI Data Sharing Manager

NOTE: Language from the Primary Driver Information Consent Form related to data sharing is attached to this data sharing agreement.