RESEARCH

UDOT Keeping Utah Moving

Utah Department of Transportation - Research Division
4501 South 2700 West - P.O. Box 148410 - SLC,UT 84114-8410

Report No. UT-24.12

EFFECT OF ROADWAY LIGHTING ON SAFETY

Prepared For:

Utah Department of Transportation
Research & Innovation Division

Final Report
September 2024

DISCLAIMER

ACKNOWLEDGMENTS

The authors acknowledge the Utah Department of Transportation (UDOT) for funding this research, and the following individuals from UDOT and other agencies on the Technical Advisory Committee for helping to guide the research:

- Jeff Lewis (UDOT Safety Programs Engineer)
- Ming Jiang (UDOT Lighting Engineer)
- Tam Southwick (UDOT Region 2 Traffic Engineer)
- L. Todd Wright (UDOT Electronics Supervisor)
- Travis Jensen (UDOT Consultant Project Manager)
- Tyler Laing (UDOT Region 2 Traffic Engineer)
- Suyanka Neupaney (UDOT Data & Safety Manager)
- Matthew Smith (UDOT Transportation Technology Project Manager)
- Kaehan Shour (Provo City Traffic Engineer)
- Dave Pearson (Salt Lake City Street Lighting Manager)

TECHNICAL REPORT ABSTRACT

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UT-24.12	N/A	N/A

4. Title and Subtitle EFFECT OF ROADWAY LIGHTING ON SAFETY	5. Report Date September 2024 6. Performing Organization Code N/A
7. Author(s) Samuel L. Runyan, EIT; David Bassett, P.E.; Grant G. Schultz, Ph.D., P.E., PTOE; Brad Brimley, Ph.D., P.E., PTOE, RSP; Gregory L. Snow, Ph.D.; Adam Simpson; Benjamin Dahl	8. Performing Organization Report No. N/A
9. Performing Organization Name and Address Avenue Consultants (Prime) 6605 S Redwood Rd Ste 200 Taylorsville, UT 84123 Brigham Young University (Sub) Department of Civil and Construction Engineering 430 Engineering Building Provo, UT 84602	10. Work Unit No. 5H092 61H 11. Contract or Grant No. 23-8591
12. Sponsoring Agency Name and Address Utah Department of Transportation 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410	13. Type of Report & Period Covered Final Jan 2023 to Sep 2024 14. Sponsoring Agency Code UT22.318

15. Supplementary Notes

Prepared in cooperation with the Utah Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration

16. Abstract

A significant portion of severe crashes and severe crashes involving vulnerable road users (VRU), 29 percent and 39 percent respectively, occur at night in Utah. Nighttime crash causes are less understood than daytime crashes as there are fewer nighttime crashes, but it is generally assumed that increasing street lighting can help mitigate these crashes. To test this assumption, the Utah Department of Transportation (UDOT) was interested in developing Crash Modification Factors (CMFs) specific to the state of Utah which correlate the quantity and quality of light to crash outcomes, to observe how specific light quantities and qualities might reduce nighttime crashes. Average illuminance and light uniformity metrics were measured for several arterial roads in Utah and statistical models were used to build models for estimating CMFs.

CMFs indicated that average illuminance was not correlated to a significant change in crashes except when increasing illuminance above 0.3 foot-candles (fc). This indicates that some light is better than none, but the benefit of increasing existing light levels may not be significant. However, CMFs for the uniformity of light indicated that less uniformity correlated to fewer crashes at lower average illuminance, and more crashes at higher illuminance. This may show that it is more important to light specific locations than to light entire road segments evenly, unless light levels are already high. Despite the possible benefit of increasing uniformity on roadways with high light levels, most uniformity CMFs suggested that future research should be focused on locations like crosswalks where lighting is most important.

17. Key Words	18. Distribution Statement	23. Registrant's Seal
Street Lighting, Road Safety, CMF, Crash Prediction, Illumination	Not restricted. Available through: UDOT Research Division 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410 www.udot.utah.gov/go/research	N/A

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	108	N/A

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AADT	Annual Average Daily Traffic
AASHTO	American Association of State Highway and Transportation Officials
CCD	Charge-Coupled Device
CCT	Correlated Color Temperature
CIE	*Commission Internationale de l'Éclairage* (International Commission on Illumination)
CMF	Crash Modification Factor
CMFunction	Crash Modification Function
DMI	Distance Measurement Instrument
DOT	Department of Transportation
FDOT	Florida Department of Transportation
FHWA	Federal Highway Administration
GIS	Geographic Information System
GLM	Generalized Linear Model
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HDOP	Horizontal Dilution of Precision
HSM	Highway Safety Manual
IES	Illuminating Engineering Society
LED	Light-Emitting Diode
NHTSA	National Highway Traffic Safety Administration
NOAA	National Oceanic and Atmospheric Administration
NYSDOT	New York State Department of Transportation
RHMVM	Rate Per Hundred Million Vehicle Miles
RLMMS	Roadway Lighting Mobile Measurement System
RTK	Real Time Kinematic
S/P	Scotopic to Photopic Ratio
STV	Small Target Visibility
STV-H	Weighted Average Target Visibility
TxDOT	Texas Department of Transportation

UDOT	Utah Department of Transportation
USNO	United States Naval Observatory
VLR	Veiling Luminance Ratio
VRU	Vulnerable Road User
WSDOT	Washington State Department of Transportation

EXECUTIVE SUMMARY

A significant portion of severe crashes and severe crashes involving vulnerable road users, 29 percent and 39 percent respectively, occur at night in Utah. Nighttime crashes are less understood than daytime crashes as there are fewer nighttime crashes, but it is generally assumed that increasing street lighting can help mitigate these crashes. To test this assumption, the Utah Department of Transportation was interested in developing Crash Modification Factors (CMFs) specific to the state of Utah which correlate the quantity and quality of light to crash outcomes, to see how specific light quantities and qualities might reduce nighttime crashes. Pavement illuminance data were collected for several urban arterials in Utah using a mobile light data collection system, and these data were used to calculate average illuminance and light uniformity metrics for each road segment. A negative binomial with log-link function structure was used to build models for estimating CMFs using two methodologies, a bivariate method with a single cutoff for each variable and a single variable method with two cutoffs.

The research estimated a CMF of 0.99 to increase the average illuminance from low (0 to 0.7 fc) to medium (0.7 to 1.4 fc) and 0.94 to increase average illuminance from medium to high (1.4 to 1.9 fc). This indicated that the average illuminance had minimal impact on crashes in Utah. However, a CMF of 0.73 to increase average illuminance above 0.3 fc given standard deviation above 0.3 fc showed that some lighting is generally better than none. CMFs for the uniformity of light (0.81 to increase standard deviation above 0.6 fc given average illuminance less than 0.9 fc; 0.85 to increase max/min above 2.5 given average illuminance less than 0.9 fc; and 0.82 to decrease lighting frequency below 30 fluctuations per mile given average illuminance below 0.6 fc) indicated that less uniformity resulted in fewer crashes at lower average illuminance. This may show that it is more important to light specific locations than to light entire road segments evenly. However, a CMF of 0.86 to increase lighting frequency above 30 fluctuations per mile given average illuminance above 0.6 fc showed that more uniformity may be desirable at some higher light levels. While the results show that segment-wide lighting and uniformity is important in some cases, future research should be focused on target locations where lighting is most important.

1.0 INTRODUCTION

1.1 Problem Statement

Nighttime conditions introduce unique problems to roadway safety because of reduced visibility. For example, a significant portion of severe crashes and severe crashes involving vulnerable road users (VRU), 29 percent and 39 percent respectively, occur at night in Utah. Nighttime crashes are less understood than daytime crashes as there are fewer nighttime crashes, but it is generally assumed that increasing street lighting can help mitigate these crashes. To test this assumption, research was conducted for the Utah Department of Transportation (UDOT) to identify the effects of street lighting on road safety in Utah. UDOT was interested in the quantity and quality of lighting whereas much of the previous research focused only on the presence of lighting. Quantity and quality of lighting needed to be defined and a data collection method developed for this research.

1.2 Objectives

The primary objective of this research was to develop Crash Modification Factors (CMFs) and/or crash modification functions (CMFunctions) describing the effects of street lighting on crashes. Since lighting is intended to improve safety during nighttime hours, the CMFs were developed for nighttime crashes specifically using the American Association of State Highway and Transportation Officials' (AASHTO) Highway Safety Manual (HSM) methodology (AASHTO, 2010).

Other objectives for this research include the following:

- Conduct a literature review of street lighting safety.
- Define quality and quantity of light as they relate to this research.
- Develop experimentation methodology for street lighting research.

1.3 Scope

The tasks for this project included conducting a literature review, researching the current lighting state of practice in Utah, developing experimentation, collecting data, evaluating data, and documenting results. The data were limited to arterials in Utah and Salt Lake Counties to maintain consistency in statistical analysis. Selected locations were also limited to speed limits of 45 mph or less for the safety of data collectors who were required to drive under the speed limit while collecting data.

1.4 Outline of Report

This report consists of nine chapters, including the introduction. The remaining chapters describe the literature review, state of practice in Utah, research methods, data collection, data evaluation, statistical analysis, conclusions, recommendations, and implementation. The report chapters are outlined below.

1. Introduction: The introduction outlines the objectives and scope of the research.

2. Literature Review: The literature review was conducted for existing published research, government reports, and other appropriate documentation related to lighting and safety.

3. State of Practice in Utah: Lighting professionals at UDOT, Provo City, and Salt Lake City were consulted, and Utah lighting standards were reviewed to identify the state of practice in Utah.

4. Research Methods: An experimentation method was developed for collecting roadway, crash, and lighting data. A method of light data collection using instruments attached to a moving vehicle was developed for this research.

5. Data Collection: The process of collecting and cleaning data is described in this chapter. This included the implementation of the experimentation developed, application of geographic information systems (GIS), and equations used to develop research metrics from raw data.

6. Data Evaluation: This chapter describes initial investigations into the data including histograms and correlation plots to help understand how variables would interact with each other during the statistical analysis.

7. Statistical Analysis: Statistical models and methods used to develop CMFs are described in this chapter. An explanation is provided for why the final statistical method was chosen.

8. Conclusions: The results of this research are summarized in the conclusions chapter, including CMFs. The limitations and challenges of the research are described.

9. Recommendations and Implementation: Recommendations for implementation of CMFs and future research are described in this chapter. An implementation plan is outlined.

2.0 LITERATURE REVIEW

2.1 Introduction

The HSM (AASHTO, 2010) contains CMFs for the presence of roadway lighting as a binary variable, but these do not account for the quantity and quality of light. The purpose of this literature review is to identify current practices for CMF development and experimentation and current roadway lighting standards used by government agencies. This information provides relevant context for developing Utah-specific CMFs based on the quantity and quality of roadway lighting. The sections of this literature review include a discussion on how quantity and quality of light are defined, a discussion of current lighting standards, a discussion of the importance of pedestrian visibility in lighting, a discussion of how CMFs are developed, a discussion of current CMFs for street lighting, a discussion of data collection methods, and a discussion of the limitations of current knowledge about the safety effects of lighting.

2.2 Defining Quantity and Quality of Light

The literature identifies four key areas which define the quantity and quality of street lighting. These are the level of lighting, lighting uniformity, lighting of surroundings, and the level of glare (Yoomak and Ngaopitakkul, 2018). The following subsections define these and other elements of lighting in more detail.

2.2.1 Level of Lighting and Uniformity

The level of lighting and its uniformity receive the most attention in the literature and are usually defined by horizontal and vertical illuminance, which are measures of light falling onto horizontal and vertical surfaces, respectively. Horizontal illuminance is measured from the pavement level to represent light landing on the pavement. Vertical illuminance is typically measured from about 5 feet above the pavement to represent light landing vertically on pedestrians. The level of lighting is generally measured as the average of one of these metrics in an area, and uniformity is a measure of how evenly distributed the light is (e.g., standard deviation or max/min ratios).

Luminance is also used to quantify lighting, and may even be preferred, as it is the light that reaches drivers' eyes, having reflected off the pavement and objects within the field of view. Luminance is usually measured by specially calibrated cameras which measure light reflected from the surrounding environment toward the driver's eye level. Alternatively, it is possible to estimate luminance when the horizontal illuminance, pavement reflectivity, and incident angle are known (Rice et al., 2020). However, illuminance measurements are often sufficient for CMF development because they are closely related to luminance.

Additional developments in photometry have identified the effects of correlated color temperature (CCT) on human light perception. CCT is an approximate measure of the chromaticity of a light source. Human vision perceives wavelengths of light predominantly in the photopic region (i.e., light visible to the cones of the retina) under high light levels and the scotopic region (i.e., light visible to the rods of the retina) under low light levels. The ratio of Scotopic to Photopic (S/P) vision under given light conditions is the Mesopic region (Arecchi et al., 2007). Because current lighting standards are largely designed for the photopic region only, there is room for further research on the practical applications of the S/P ratio in roadway lighting applications.

2.2.2 Lighting of Surroundings and Glare

The third lighting element listed above, the lighting of surroundings, is a broad category which includes visibility of objects outside the roadway. This includes, for example, illumination of roadway signs (accounted for by some CMFs). Conversely, there are no known CMFs which account for glare, the fourth common element of light quality. Disability glare is still considered an important factor in roadway safety and is often regulated by transportation agencies. It is usually measured by the veiling luminance ratio (VLR), which is the ratio of maximum to average luminance from the driver's perspective.

2.2.3 Other Elements of Lighting

Another parameter that can be used to quantify the lighting quality is small target visibility (STV). STV characterizes the ability to immediately see and identify an array of small objects at a distance under a given light condition (Adrian, 1989; Janoff, 1993). This measure involves human test subjects for field verification and the photometric data is difficult to collect. However, Keck (2001) developed an alternative to STV which accounts for the influence of vehicle headlights as well as fixed lighting. This method is called STV-H and it uses computer modeling to calculate the visibility of an array of targets, removing the need to measure photometric data in the field. There are currently no CMFs which account for STV or STV-H. The Illuminating Engineering Society (IES) of North America has changed STV from a recommended design criterion to a selection criterion for designs, since it is useful, but less practical, for roadway lighting design procedures (ANSI/IES, 2018).

2.3 Lighting Standards

The *Commission Internationale de l'Éclairage* (CIE) and the IES are responsible for defining lighting standards internationally and in North America, respectively. Both include specifications for roadway lighting. Specific policies in the United States are determined by state departments of transportation (DOTs) or local municipalities. State DOTs commonly adopt IES standards for roadway lighting based on illuminance, luminance, light uniformity, and VLR.

Florida Department of Transportation (FDOT) specifies a range of horizontal illuminance between 1.0 fc to 1.5 fc for retrofitted signalized intersections and 1.5 fc to 3.0 fc for new signalized intersections, and an average-to-minimum uniformity ratio of 4 or less, regardless of classification (FDOT, 2024). Washington State Department of Transportation (WSDOT) also uses minimum average illuminance and maximum uniformity but applies the standards over a broader range of intersection and roadway categories (WSDOT, 2023). UDOT, like many DOTs, uses IES RP-8-18 (ANSI/IES, 2018) and the AASHTO Roadway Lighting Design Guide (AASHTO, 2018). Other DOTs that directly adopt the IES and AASHTO guidelines include the Texas Department of Transportation (TxDOT), the New York State Department of Transportation (NYSDOT), and UDOT (NYSDOT, 1995; TxDOT, 2018; UDOT, 2021). Many states also specify vertical illuminance standards, especially at locations that experience pedestrian traffic, such as crosswalks. FDOT, for example, requires 1.0 fc vertical illuminance at intersections and midblock crossings (FDOT, 2024). More details on the practices used by UDOT and municipalities in Utah are discussed in Chapter 3.0.

2.4 Pedestrian Visibility

Pedestrian fatalities have been of particular concern in recent years, especially with nighttime crashes. According to the National Highway Traffic Safety Administration (NHTSA) Fatality Analysis Reporting System, pedestrian fatalities have been rising and accounted for 18 percent of all traffic fatalities in 2022, and 78 percent of pedestrian fatalities occurred in dark conditions (NHTSA, 2024). Furthermore, nighttime pedestrian deaths rose by 58 percent between 2010 and 2019 (Kirley et al., 2023). This has spurred much discussion on the effects of roadway lighting on pedestrian safety.

As discussed in Section 2.3, many standards recognize vertical illuminance as a strong indicator of pedestrian visibility because of its usefulness in facial recognition. Edwards and Gibbons (2008) determined that pedestrian visibility increases with vertical illuminance, up to 20 lux (1.9 fc). However, pedestrian visibility may not always be the best indicator of pedestrian nighttime safety, as shown by Niaki et al. (2016). Further research may be required to specifically determine the relationship between illuminance and pedestrian crashes, highlighted by the fact that there are currently no CMFs developed for this.

2.5 Developing CMFs

The literature on developing CMFs allows researchers some flexibility to choose their own methodology, but there are certain protocols to consider. Two sources, "Recommended Protocols for Developing Crash Modification Factors" (Carter et al., 2012) and "A Guide to Developing Quality Crash Modification Factors" (Gross et al., 2010) were written to help researchers understand protocols and aid them in developing high-quality CMFs. For example, Carter et al. (2012) explains that CMFs may not be determined using near-misses, speed reductions, or vehicle path deviations, and they must apply to infrastructure-related treatments, not changes in policy, enforcement, or behaviors. Additionally, researchers should consider various study methodologies carefully rather than relying solely on "protocols" to choose the best methodology.

The protocols include several variations on cross-sectional studies (studies which compare data at similar locations with and without the treatment) and before-after studies

(studies which compare data at the same location at different times). These studies are susceptible to their own issues and biases which the researcher should be aware of when choosing the best option. Therefore, Gross et al. (2010) created a flowchart, shown in Appendix A, to help researchers identify which study might be best for the specific CMF they are developing.

Researchers should be especially wary of confounding variables when developing CMF studies. For street lighting CMFs, these include roadway conditions which apply differently at night than during the day. An effective way to mitigate these may be to determine the night-to-day crash rate ratio (Bhagavathula et al., 2015; Box and Alroth, 1971; Gibbons et al., 2014b; Keck, 2001; Scott, 1980; Wang et al., 2017; Zhao et al., 2015). Since street lighting is believed to only have an effect during the night, this is one way to account for confounding variables. However, van Schalkwyk et al. (2016) claim that there is insufficient evidence in the literature of the effectiveness of the night-to-day crash ratio due to the drastic difference in nighttime and daytime conditions. An additional method proposed by Li et al. (2021) was to use a matched case-control methodology to account for confounding variables that still occur at night. This requires comparing nighttime crashes at locations with similar characteristics to rule out confounding variables.

Studies also need to use a sufficiently large sample size of study locations to ensure significance of the results. When only the presence of lighting was considered, some studies have successfully performed a before-after study, which doesn’t require as large of a sample size. However, researchers that considered specific light levels could not easily perform before-after analyses and instead opted for cross-sectional analyses. These require a large sample size to compare many sites with similar characteristics. Li et al. (2021) used a sample size of 2,444 roadway segments and Wei et al. (2016) used a sample size of 1,234 intersections. Other studies had similar sample sizes, ranging from 100 to 3,000 sites. Since several studies used mobile light measurement systems (see Section 2.7), researchers divided roadways into small segments to create a sufficiently large sample size.

2.6 Existing CMFs and Studies

Most CMFs based on fixed street lighting were developed only with the presence of street lighting in mind without accounting for photometric properties (Abdel-Aty et al., 2014; Bullough et al., 2013; Donnell et al., 2010; Elvik, 1995; Sacchi and Tayebikhorami, 2021; Wanvik, 2009; Ye et al., 2009). However, four studies conducted in urban Florida resulted in CMFs based on photometric properties. The CMFs produced by these studies are summarized in Table 2-1. They modify all severity level crashes, and some apply to night-to-day crash rate ratios rather than standard crash rates.

Table 2-1 Summary of Photometric CMFs

Study	Illuminance Property	Countermeasure	CMF	Crash Type	Location	Rating
Wei et al. (2016)	Average	(< 0.2 fc) to (≥ 0.2 fc and < 1.1 fc)	0.47	Night	3-leg	3
Wei et al. (2016)	Average	(< 0.2 fc) to (≥ 0.2 fc and < 1.1 fc)	0.48	Night	4-leg	3
Wei et al. (2016)	Average	(< 0.2 fc) to (≥ 0.2 fc and < 1.1 fc)	0.52	ND Ratio	3-leg and 4-leg	3
Wei et al. (2016)	Average	(≥ 0.2 fc to < 1.1 fc) to (≥ 1.1 fc)	1.16	ND Ratio	3-leg and 4-leg	3
Wang et al. (2017)	Max/Min	(≥ 6) to (< 6)	0.98	ND Ratio	Segments	4
Wang et al. (2017)	Average	(1 fc to X fc)	$X^{-0.0773}$	ND Ratio	Segments	4
Yang et al. (2019)	Average and Standard Deviation	(0.44 fc to M fc) and (0.01 to SD)	Equation 2-1	ND Ratio	Segments	3
Li et al. (2021)	Average	(≤ 0.5 fc) to (> 0.5 fc and ≤ 1.0 fc)	0.68	Night	Segments	4
Li et al. (2021)	Average	(≥ 0.5 fc and ≤ 1.0 fc) to (> 1.0 fc)	0.58	Night	Segments	4
Li et al. (2021)	Max/Min	(≤ 10) to (> 10)	1.39	Night	Segments	3

Note: ND = Night-to-day crash rate ratio

Equation 2-1 CMF for Average and Standard Deviation of Illuminance (Yang et al., 2019)

$$CMF_{E_h} = exp\left(-0.212 \cdot \Delta E_h^M + 0.022 \cdot {\Delta E_h^M}^2 + 0.022 \cdot \Delta E_h^{SD} + 0.065 \cdot {\Delta E_h^{SD}}^2\right),$$

$$M \in [0.07,\ 1.43],\ \ SD \in [0.01,\ 0.9],$$

Where:

$$\Delta E_h^M = difference\ in\ mean\ illuminance,$$

$$\Delta E_h^{SD} = difference\ in\ standard\ deviation\ of\ illuminance$$

$$M\ =\ final\ value\ for\ mean\ illuminance\ in\ fc$$

$$SD\ =\ final\ value\ for\ standard\ deviation\ of\ illuminance\ in\ fc$$

CMFs in Table 2-1 support the existence of a relationship between average illuminance and nighttime crash risk. While this relationship seems evident, many studies have also found that the relationship does not extend to high levels of illuminance (Bhagavathula et al., 2015; Gibbons et al., 2014b; Li et al., 2021; Wang et al., 2017; Yang et al., 2019; Zhao et al., 2015). For example, the CMFs developed by Wei et al. (2016) are valid only up to illuminance values of 1.1 fc.

Uniformity is a valuable measure of lighting quality, but there is no consistent measurement for light uniformity, which is sometimes measured as the standard deviation of illuminance measurements and sometimes as a ratio involving the minimum, maximum, and average illuminance. Gibbons et al. (2014b), Li et al. (2021), and Wang et al. (2017) used the max-to-min ratio to measure light uniformity but reached different conclusions about its effectiveness. Gibbons et al. (2014b) used an algorithm that detects whenever illuminance fluctuates above or below average. Li et al. (2021) used smaller segments to determine localized uniformity by comparing max-to-min ratios at locations where measurements drop below average. Both of these methods produce a more descriptive measure of uniformity than segment-wide max-to-min ratios. Wang et al. (2017) produced a higher-ranking CMF using segment-wide max-to-min ratios. Yang et al. (2019) and Zhao et al. (2015) identified a significant correlation between the standard deviation and nighttime crash reduction, but the measure could not be used on its own due to collinearity to the mean.

For intersections, Bhagavathula et al. (2015) and Wei et al. (2016) studied the effectiveness of horizontal illuminance. Both studies determined that increasing illuminance was not effective for reducing total crashes at higher lighting levels, but Wei et al. (2016) determined that higher illuminance (≥ 0.9 fc) correlated to reduced fatal and severe injury crashes.

Other metrics which may be considered by future CMFs are vertical illuminance, glare, and luminance. Bhagavathula et al. (2021) determined that vertical illuminance up to 10 lux (0.9 fc) at midblock crossings increased pedestrian safety. They also recommended horizontal illuminance of at least 14 lux (1.3 fc) at intersections to offset disability glare from oncoming vehicles. Although Gibbons et al. (2014b) did not specifically evaluate pedestrian crashes, the crash rate reduction from vertical illuminance on segments was unsurprisingly very similar to horizontal illuminance. They used the vertical-to-horizontal illuminance ratio to measure glare but lacked sufficient data to determine a significant relationship with the crash rate. They also measured luminance, but this produced unexpected results likely due to the confounding factors of the headlights and the wide range of the luminance camera.

2.7 Data Collection Methods

Development of a CMF requires accurate data on both the crashes that have occurred over a defined period of time and details that characterize the road environment. Often the road environment is described with values of traffic volumes, geometric features (i.e., lane width, lane configurations, grade, radius), and traffic control. The use of lighting and photometric values such as illuminance and light uniformity also characterize the environment. To examine the impacts of different qualities of light, photometric data must be collected from multiple locations with different lighting conditions.

Original documentation from IES recommends taking horizontal illuminance measurements along a roadway at locations that form a grid of quarter-lane spacing with intervals less than 5 meters (16.4 feet) apart (IES, 1989). With early technologies, this required data collectors to stand in the middle of the roadway to gather photometric readings. With advances in technology that improve data collection speeds, instruments can now be mounted to a vehicle to create a mobile system that collects photometric data along with coordinates from Global Positioning System (GPS) or Global Navigation Satellite System (GNSS) satellites. Such systems have been used in recent studies (Johnson et al., 2014; Suk and Walter, 2019; Tomczuk et al., 2021; Zatari et al., 2005; Zhou et al., 2009; Zimmer, 1988). Figure 2-1 shows an example of a mobile light data collection system called the Roadway Lighting Mobile Measurement System (RLMMS) developed by Gibbons et al. (2018).

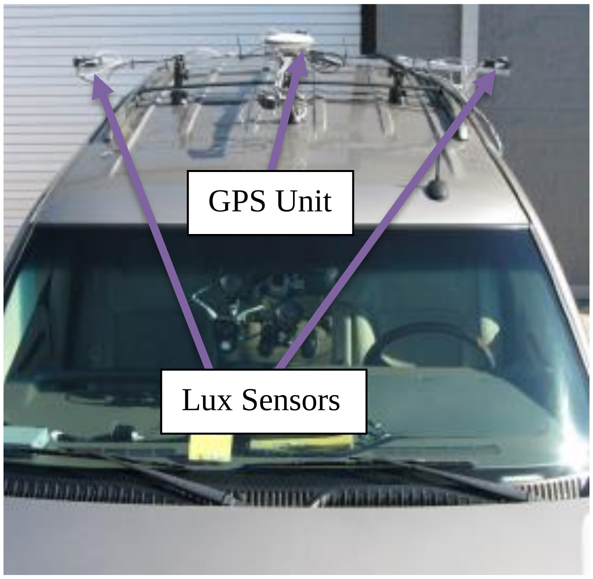


Figure 2-1 Image of the RLMMS (Gibbons et al., 2018)

One drawback of mobile systems is that measurements from the top of the vehicle need to be extrapolated to the ground level to comply with IES recommendations for photometric data collection. Zhou et al. (2009) accomplished this using the locations and heights of the two nearest luminaires to each individual measurement and extrapolating with the inverse square law. However, they also discovered that the median difference between measurements at the top of the vehicle and at the pavement level were not statistically different from zero at the 95 percent confidence level, indicating that extrapolation may not be strictly necessary.

Location data can be obtained in multiple ways. One method is to use GPS or GNSS receivers to provide the approximate location. A second method is to use distance measurement instruments (DMIs) to provide scalar distance along a roadway, a method supported by several researchers who determined that DMIs provide more precise location measurements than standard GPS/GNSS signals (Johnson et al., 2014; Zhou et al., 2009). Another alternative, proposed by Tomczuk et al. (2021), is to use GPS/GNSS receivers with Real Time Kinematic (RTK) corrections.

Usually, horizontal illuminance is measured using lux meters, now typically attached externally to the top of a vehicle. However, Zatari et al. (2005) devised a way to measure illuminance, luminance, and glare using Charge-Coupled Device (CCD) cameras instead of lux meters. This method was advantageous because cameras could measure multiple luminaires at once and they could provide directionality without needing to be reoriented.

For many mobile photometric data collection systems, special attention should be paid to collecting data under consistent nighttime conditions. Attaching the sensors to the top of the vehicle helps to minimize the influence of vehicle headlights from surrounding vehicles (Tomczuk et al., 2021). Additionally, Suk and Walter (2019) decided to take photometric measurements only during overcast sky conditions to minimize the effect of moonlight. They accomplished this using the Clear Dark Sky Chart website (http://www.cleardarksky.com/csk/).

2.8 Limitations

A recent WSDOT study of continuous mainline roadway lighting included an extensive literature review which identified several recurring issues in safety analysis for roadway lighting (van Schalkwyk et al., 2016). One of these issues was that there was no consistent method for classifying nighttime crashes. The authors recommended determining nighttime conditions as the time outside of civil twilight (when the sun is 6 degrees below the horizon), rather than relying on the reported light conditions in crash report forms. The authors also questioned the assumption that night-to-day crash rate ratios sufficiently overcome confounding factors, claiming the assumption is not justified since nighttime conditions are so different than daytime conditions. Replacing or supplementing the night-to-day crash rate ratios was also discussed in Section 2.5.

The lighting CMFs developed in the HSM (AASHTO, 2010) come from a study by Elvik and Vaa (2004) but which van Schalkwyk et al. (2016) discovered was conducted originally by Elvik (1995). This study involved a meta-analysis of 37 previous studies, many of which did not use very rigorous statistical methods. Van Schalkwyk et al. (2016) believed that this meta-analysis was less reliable because it combined results across varying environments and was likely also impacted by publication bias (Elvik, 1995). Since this study was published in the HSM it may have set the expectation for future studies to contain similar bias.

As with any research, it is critical to understand and consider the limitations of previous studies. Every study will have limitations, but some of these can be mitigated with effective experiment design. For example, while developing a lighting CMF it is important to remember that roadway lighting often serves an aesthetic purpose as well as a safety purpose. Excessive light levels cause light pollution that negatively affects the environment and human wellbeing on top of the energy cost to maintain those levels (Gibbons et al., 2014a). Therefore, agencies are faced with balancing the need to provide effective lighting with costs of light pollution and energy.

3.0 STATE OF THE PRACTICE IN UTAH

3.1 Overview

The research team interviewed road lighting professionals at UDOT and Provo City to ascertain the current state of practice of road lighting design and standards in Utah. The purpose of this was to identify any characteristics of roadway lighting that might be most important to consider in the analysis, and to understand how standards are created and maintained in Utah. The following sections outline the findings from UDOT and Provo City lighting practices.

3.2 UDOT Standards

UDOT adopts its photometric standards from the AASHTO Roadway Lighting Design Guide (AASHTO, 2018) and IES RP-8-18 (ANSI/IES, 2018). These include threshold values for illuminance, luminance, light uniformity, and VLR. However, UDOT allows designers some flexibility to deviate from these standards as necessary. In practice, designers use AGi32 software (Revalize, 2021) to conform to these photometric standards and builders install luminaires which output the luminous flux specified by the AGi32 design. There are also standards for luminaire design such as pole height, distance between poles, and slip-base requirements. Many of these are specified by UDOT standard drawings or federal standards (UDOT, 2021).

Pedestrian safety regarding roadway lighting, especially on local roads, has been of particular interest to UDOT in recent years and policy changes are being considered to emphasize this. Currently, UDOT recommends designing for vertical illuminance at pedestrian crossings, but there are no specific requirements for this. UDOT standards state the following about vertical illuminance (UDOT, 2021):

> "Vertical illuminance has been identified as an important element in lighting crosswalks associated with intersections. Higher levels of vertical illuminance produce a better positive contrast. If all the light from the luminaire is directed downward, the vertical profile of the pedestrian will not be adequately illuminated."

With language lifted from Federal Highway Administration (FHWA) research conducted by Edwards and Gibbons (2008), UDOT standards also include the statement that (UDOT, 2021):

> "A vertical illuminance level of 20 lux [1.9 fc] measured at 1.5 m (5 ft) from the road surface allowed drivers to detect pedestrians in midblock crosswalks at adequate stopping distances under rural conditions."

The illumination requirements used by other state DOTs were presented previously in Section 2.3. With specifications for several different functional classifications, UDOT appears to have more specific illumination requirements than some other DOTs. In comparison, FDOT specifies a range of horizontal illuminance between 1.0 fc to 3.0 fc for intersections with an average-to-minimum uniformity ratio of 4 or less, regardless of classification (FDOT, 2024). For local-to-local intersections with low pedestrian activity, UDOT permits horizontal illuminance values below 1.0 fc and an average-to-minimum uniformity ratio of 6. Major-to-major intersections with high pedestrian activity are to have horizontal illuminance values above 3.0 fc (UDOT, 2021). This is also a departure from other DOTs which set lighting standards based on a range of lighting values. While UDOT's standards are more specific than others, they also come directly from IES standards. For example, UDOT recommends that lighting designers balance between the Luminance Method and the Illuminance Method from IES RP-8, but there is some flexibility for the designer to decide how to apply these standards. Similarly, TxDOT, NYSDOT, and other DOTs also recommend following IES standards as well as the AASHTO Roadway Lighting Design Guide (NYSDOT, 1995; TxDOT, 2018). These uniform applications of lighting standards support developing specific standards for Utah roads.

3.3 Provo City Standards

Lighting specialists for Provo, Utah were also interviewed to discuss the implementation of lighting standards in their city. Provo City sets standards primarily based on IES RP-8 and the AASHTO Roadway Lighting Design Guide (AASHTO, 2018). The city is currently in the process of updating their standards to match the most recent version of these guidelines. Developers are required to build lighting that conforms to Provo City standards and Provo City maintains the lighting. Luminaires are purchased from trusted suppliers and designers use AGi32 software to ensure photometric standards are met. Provo City adjusts light levels as necessary to address public concerns or design inadequacies. Fortunately, luminaires are provided with a dial to adjust light levels in these cases.

A concern raised by Provo City is that drivers may notice a contrast between older and newer lighting designs, and they are working to update lighting designs wherever feasible. Differences in lighting designs mean that drivers experience changes in light color or light uniformity. While light uniformity is considered a safety concern, Provo City also values consistency in color and works to make the lighting on continuous roadway segments as consistent as possible (K. Shour, personal communication, May 16, 2023).

3.4 Summary

There seems to be a considerable difference between the focus of DOTs such as UDOT and cities such as Provo City when it comes to roadway lighting. While UDOT is interested in pedestrian safety and luminaire design, Provo City's main concern is satisfying the public and building a uniform lighting system. However, both rely on IES RP-8 and the AASHTO Roadway Lighting Design Guide when designing roadway lighting. These standards are used consistently among other state DOTs such as TxDOT and NYSDOT, whereas other states such as FDOT have endeavored to create state-specific standards.

The next chapter describes how the research methods were developed prior to data collection and evaluation. Justification for these methods is also provided.

4.0 RESEARCH METHODS

4.1 Overview

The focus of this research is the impacts of fixed roadway lighting on crashes on urban arterials at night. Roadway lighting is a typical feature of many urban arterials, but there can be inconsistencies in lighting applications from one arterial to another due to local policies or practices and adoptions of different technologies and materials. For example, some agencies require higher light levels at intersections and midblock crossings. Other agencies or locations may have advanced lighting technology such as adaptive lighting or Light-Emitting Diode (LED) fixtures. Stray lighting from adjacent developments also impacts the quality and consistency of illumination along a corridor and is considered in this research as it is part of the driving environment. Other light-related variables, such as the height and quality of vehicle headlights, are ignored because they do not contribute to the consistent roadway environment.

Crash model development requires two types of data: crash data and facility data. Crash data includes information such as crash frequency (total observed/reported crashes), severity, collision type, time of day, and contributing factors. Crash data are often collected over a period of multiple years. In this study, the crash data come from UDOT through the AASHTOWare Safety software, powered by Numetric. Since roadway lighting is believed to impact nighttime crashes exclusively, the analysis separates nighttime and daytime crashes. Although the crash data contain an indicator for the light conditions, as recorded by the responding law enforcement agents, this alone may be insufficient to describe actual nighttime conditions at the time of crash. Data from the National Oceanic and Atmospheric Administration (NOAA), with historical solar calculations based on the date, time, and location, are used to accurately determine nighttime conditions (NOAA, 2015).

Facility data are the characteristics of the road to connect with crashes. Geometric features typically included are the number of lanes and their widths, horizontal or vertical curvature, the presence or type of median, and the length of the road segment, if applicable (models for crashes at intersections do not include segment length). Roadside features can also be considered, including shoulder presence, shoulder width, and sidewalk presence. Traffic volumes, typically measured as annual average daily traffic (AADT), are included to account for exposure. The final relevant facility information is the lighting qualities that vary from one study site to another. These may include the types and layouts of luminaires used, the type of lamp used, and measured photometric properties such as the illumination on the pavement. The facility data and crash data are discussed in the following sections.

4.2 Facility Data

Facility data include the roadway data, luminaire data, and photometric data, as discussed in the following subsections.

4.2.1 Roadway Data

UDOT defines roadway segments based on characteristics such as lane configuration and functional class. Roadway data for each segment were obtained from UDOT and include the following characteristics:

- Segment length
- Number of lanes in each direction
- Lane width
- Speed limit
- Shoulder width
- Sidewalk presence
- AADT
- Functional class

This study included 17 distinct urban state highways in Utah. The highways selected by the research team were in metropolitan areas of Salt Lake County and Utah County. The intent was for the sites to have roadway features that are relatively consistent, while having diversity in lighting properties. To accomplish this, segments were first selected from those that had uniform roadway characteristics, particularly in functional class, number of lanes, and speed limit. Segments were considered suitable for data collection if they were arterials with a speed limit between 35 mph and 45 mph with no more than 6 lanes. Then, the light levels were visually estimated as “high,” “medium,” or “low” based on the density of luminaires at each segment. After a long list of all suitable segments was determined, a random sample of each roadway type/light level grouping was selected using a random sampling function.

Figure 4-1 depicts all suitable locations identified by the research team while Figure 4-2 shows the randomly sampled locations. Figure 4-2 also shows segments where additional data were collected out of convenience from driving between the selected segments. Light data collected at these additional locations were limited and not useful for this research except for arterial locations where there was less lighting, thus requiring less data. Additional locations excluding this exception were not included in this research but may be useful for future research. The full list of suitable locations is shown in Appendix B to show which areas this research is most applicable to.

4.2.2 Luminaire Data

In addition to the measured photometric data, luminaire data were also used for characterizing the quality or quantity of light on a facility. The details and locations of roadway luminaires are contained in GIS data obtained from Salt Lake City, Provo City, and UDOT. However, luminaire data from other municipalities were not obtained, so luminaire data was investigated but not ultimately used in the analysis. Luminaire data include the following characteristics:

- Luminaire height
- Lamp type (High Pressure Sodium, LED, Induction, Metal Halide, Halogen, Mercury Vapor, Incandescent, Compact Fluorescent Light).
- Luminous flux (lumens)
- Angle of lamp
- Distance between luminaires

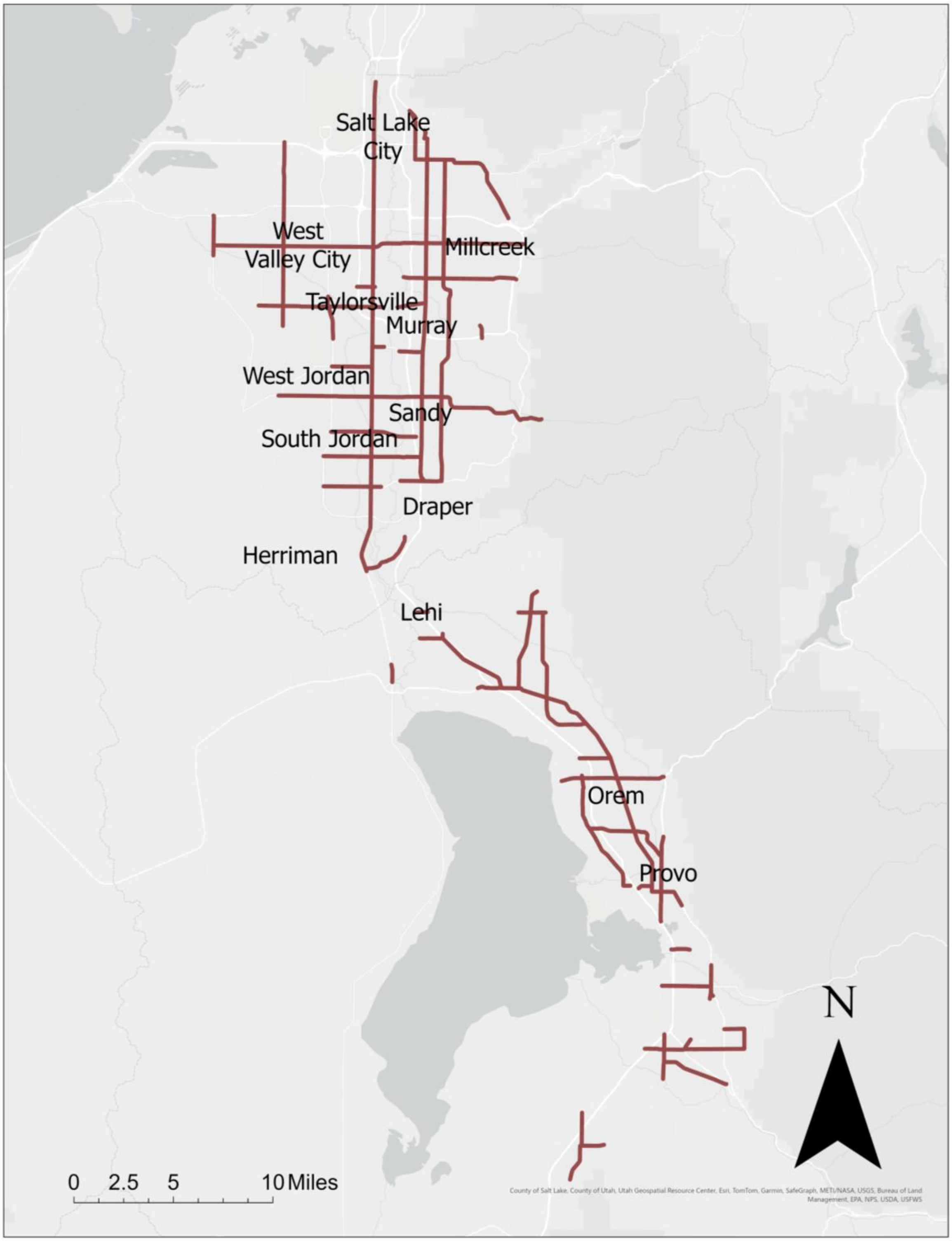


Figure 4-1 All Suitable Study Locations

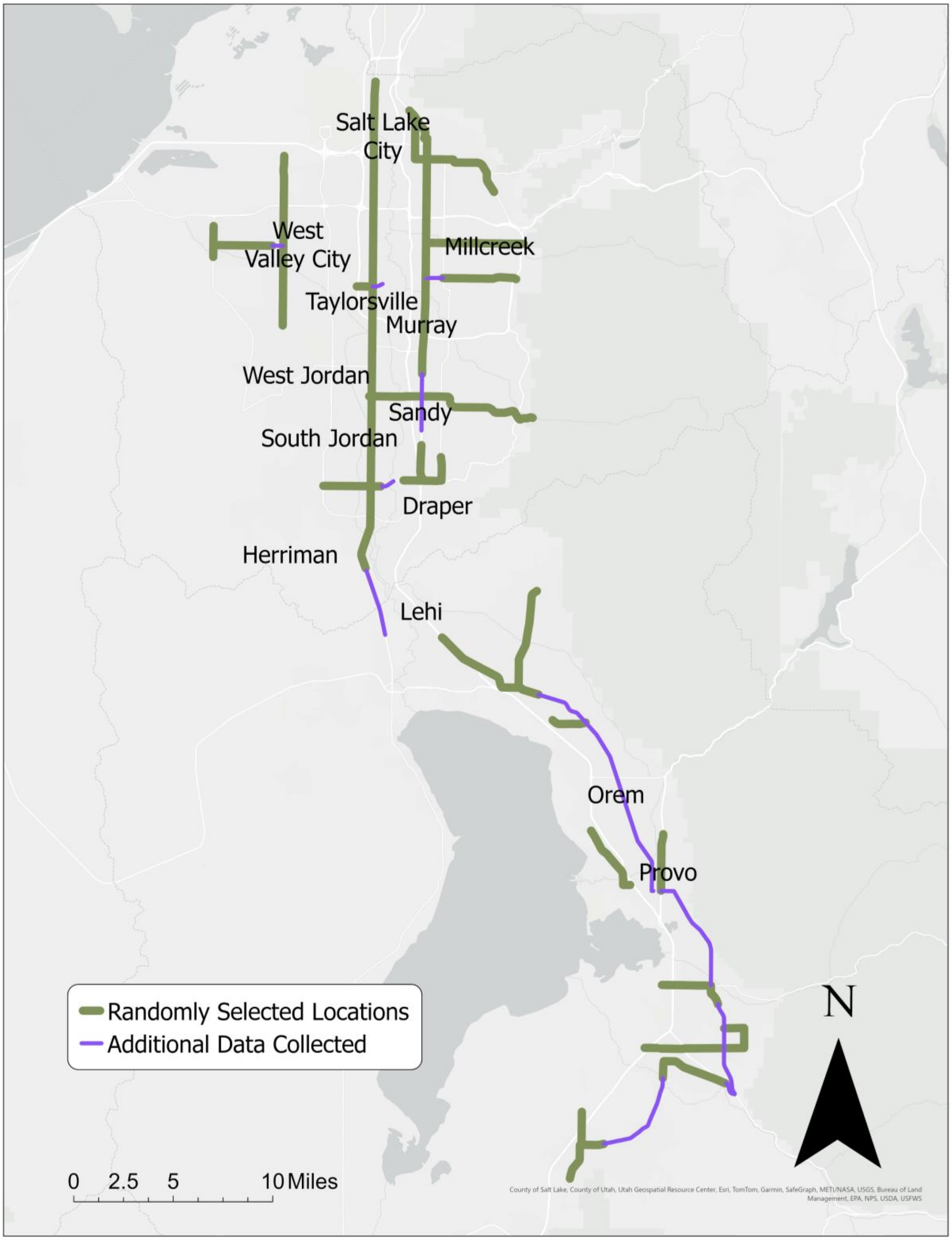


Figure 4-2 Randomly Sampled Locations

4.2.3 Photometric Data

It was preferable for photometric data to be collected during low-light moon phases to minimize the effects of moonlight. The U.S. Naval Observatory (USNO) Astronomical Applications Department moon phase tool was used to accomplish this (USNO, 2024). However, this did not account for all potential sources of excess light such as spillover from vehicle headlights or reflections in clouds or fog. If conditions were not suitable for data collection on a given night, the data were not collected at that time.

The photometric data were collected by two LI-210R photometric sensors. The instruments are manufactured by LI-COR (LI-COR 2024a). The sensors were fastened to the top of the vehicle as shown in Figure 4-3 to continuously collect illuminance readings at regular intervals while moving. The sensors were placed 6 feet apart to approximate quarter-lane spacing between measurements. This is recommended by IES documentation and allows for heat maps to be generated showing the distribution of illuminance on the roadway. These measurements were paired with GPS readings collected simultaneously by an LI-1500 data logger (LI-COR 2024b), allowing continuous readings along a single lane of a segment. By driving at a constant speed, the photometric readings represent light levels collected at a regular spacing. If the vehicle is stopped or moving slowly, readings accumulate in a small area or at the same location. To account for these inconsistencies, readings were averaged at 50-foot intervals.

Although high-quality equipment was used for light data collection, the readings fluctuated around the actual light value based on the amount of voltage and the quality of power supplied to the device as shown in Figure 4-4. Since electricity supplied from a car is inconsistent and variable, an external battery was used to power the device instead.

Multiple readings along a segment in different lanes and with regular spacing were combined to determine variables such as average illuminance and uniformity. Uniformity was determined using an algorithm that calculates max-to-min and avg-to-min ratios whenever there is a fluctuation above or below the mean illuminance. The maximum, minimum, and average of these ratios along a roadway segment were used to represent the uniformity of the entire segment. This method was also compared to using the standard deviation of illuminance to determine uniformity for the entire segment.

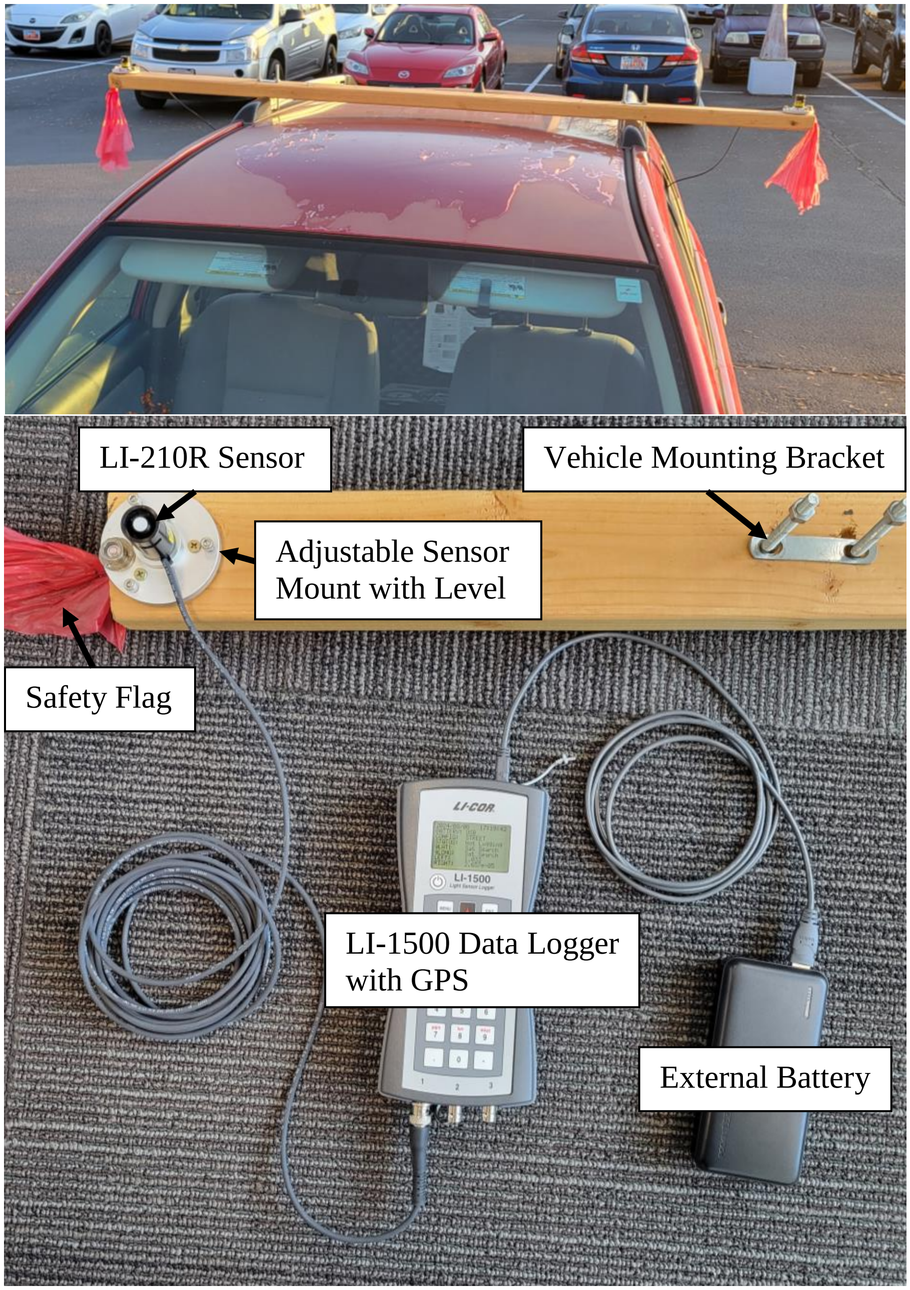


Figure 4-3 Mobile Light Data Collection System

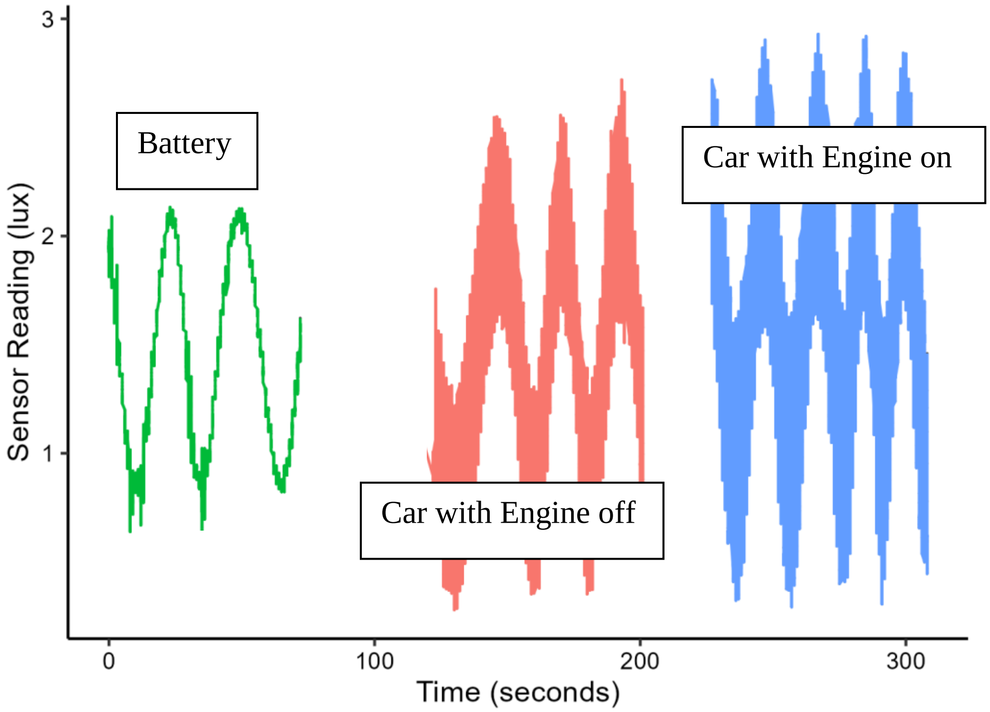


Figure 4-4 Light Sensor Performance Based on Power Source

4.3 Crash Data

The crash data used in the study include the total crashes and severity levels 2 (potential injury) through 5 (fatal) crashes occurring on a segment, and the specific time of the crash (to determine the lighting conditions when the crash occurred). The crash data were collected over 5 years (2018-2022). Therefore, precautions were taken by identifying any roadway lighting projects on the study corridors so that roadway lighting on the selected roadways did not change significantly during the 5-year period. As discussed previously, the solar calculation tool developed by NOAA was used to determine whether nighttime conditions existed at the exact day, time, and location of each crash. This tool used Jean Meeus's *Astronomical Algorithms* (Meeus, 1991) to determine sunrise and sunset. However, the period between sunset and sunrise includes civil twilight, which needed to be ignored for the purposes of this study. Civil twilight occurs between sunrise/sunset and the time when the sun is 6 degrees below the horizon. Therefore, an approximation of 30 minutes before/after sunset/sunrise was used to exclude civil twilight.

It was originally planned to separate crash data by whether they occurred in the "summer" months between April through September or the "winter" months between October through March. This would allow for variables associated with time of year such as driver behavior to be accounted for. These crashes would also be associated with photometric data collected separately during the summer and winter months to account for any changes in nighttime lighting. However, tests on a short segment of 900 East in Provo, UT showed that light levels were not significantly different in the summer and winter, so this step was not included in the research.

4.4 Summary

Roadway facility and crash data were combined to identify relationships between lighting quality and crashes. Roadway lighting was measured by a photometric sensor while continuously driving along state-owned highways in urban areas. Except for lighting quality, the individual highways were intended to have similar features to help isolate the effects of lighting on crashes. Reported crashes from a 5-year period are evaluated in this study. The next chapter outlines the data collected and the methods used to prepare the data for analysis.

5.0 DATA COLLECTION

5.1 Overview

Data collected for this study can be separated into three main categories: crash data, general roadway data, and light data. Crash data and roadway data allow for a calculation of crash outcome based on roadway characteristics. Light data are important specifically because the goal of this study is to develop a CMF based on measured light attributes. The following sections outline how data from each of these categories were collected along with a discussion on the overall compilation of the data.

5.2 Crash Data

Crash data were obtained from UDOT through the AASHTOWare Safety software, powered by Numetric (https://udot.aashtowaresafety.com/). These data include characteristics about the crash, the time of the crash, and the location of the crash. Among the crash characteristics is an indicator of whether the crash occurred during daytime or nighttime conditions, as documented in the official crash record by law enforcement responding to the crash. To remove human error and judgment, daytime or nighttime conditions were determined from astronomical calculations with the recorded time of the crash (see Section 4.3). Additionally, initial crash data also included an indicator of whether the crash location had roadway lighting or not, which as a binary indicator is insufficient to describe lighting conditions. Since this study is focused on quantity and quality of light, detailed light data were also collected as explained in Section 5.4. Crash data were collected over a 5-year period from 2018 to 2022. This required the assumption that roadway characteristics did not change during this period (see Section 5.3).

Besides crash time and location, the following crash characteristics were included in the study: severity, weather conditions, involvement of intoxication, drowsy driving, and presence of VRU. There is also a data indicator assigned by the reporting police officer of whether the crash was intersection-related or not. A crash was considered intersection-related for this study if it had the indicator and the crash location was within the functional area of an intersection. This functional area was determined by UDOT based on intersection type using a radius between 100 ft to 500 ft from the center of the intersection as shown in Table 5-1. In the rare cases where crashes fell within the functional area for multiple intersections, the crash was assigned to whichever intersection was closer.

Table 5-1 UDOT Intersection Functional Area Classification

Intersection Type	Area of Influence (ft.)
Signal Control	300
Minor Leg Stop Control	150
All-Way Stop Control	100
Yield Control	100
Uncontrolled	100
Roundabout	300
Offset Left-Turn	400
Median Thru-U Turn	400
Restricted Crossing U-Turn	400
Single-Point Urban Interchange	500
Diverging Diamond Interchange	400
Active Transportation Only	100
Railroad Crossing	100

5.3 Roadway Data

Basic roadway data were included in a roadway segments file provided by UDOT, which included route name, beginning and ending milepoints for each segment, AADT, median type, and lane configuration. Crashes were assigned to segments in the file. The segments were divided at locations where a roadway attribute changes, ensuring each segment has uniform characteristics. With route and milepoint information, segments were geolocated to summarize light data as explained in Section 5.4.2.1.

5.4 Light Data

Light data include quantitative and qualitative attributes, which describe the explanatory variables for CMFs produced by this study. While most of the light data were measured as part of the experimentation developed for the study, some municipalities also provided information on the use of luminaires within the study areas. The following subsections describe how luminaire data were obtained from municipalities, how photometric data were measured, and how qualitative observations about lighting were made.

5.4.1 Luminaire Data

Luminaires used for street lighting come in various configurations and sizes with a variety of lamp types. These details can be specific to a jurisdiction and roadway project, making it difficult to obtain the desired data from a municipality. Data on the height, angle, and type of lamp used as well as the luminaire configuration were all desired; however, only Salt Lake City and Provo City provided this information. Evaluation of these characteristics will require a separate analysis.

Since accurate data on luminaires were difficult to obtain, the researchers estimated luminaire height and lamp type while collecting photometric data. It was observed that luminaires tended to be a consistent height across roadway segments, so this was a reasonable way to approximate the missing data. Luminaire height was later used to extrapolate the measured photometric data as explained in Section 5.4.2.2. Luminaire data were also summarized by roadway segment to assist in analysis.

5.4.2 Photometric Data

Photometric data collection constituted the bulk of the data collection efforts for this study. It was determined that horizontal illuminance would be measured to represent the photometric properties of the roadway lighting. The process for collecting this data was outlined previously in Section 4.2.3. To ensure accurate data collection, the researchers used the checklist shown in Figure 5-1 every time data were collected.

The following subsections explain the processes used to clean the raw photometric data, how the data were extrapolated from the vehicle roof level to the pavement level, and how the data were summarized by road segment.

Light Data Collection, Before You Go!

Did you remember to....

- Mount the rack correctly (arrow pointing forward)?
- Mount the left and right sensors to the corresponding side and tightly secure with the mounting screws?
- Plug the left sensor into input 1 and the right sensor into input 2?
- Mount the safety flasher to the **back** of the car and plug into car power?
- Plug the data logger into the external battery?
- Place the GPS approximately in line with the light sensors? (GPS is behind the screen of the data logger.)
- Log information in the binder? (i.e. date, name, weather, moon phase, car mileage start and end, notes, routes and lanes to collect data on, traffic, luminaire heights, and other information)
- Ensure the logging configuration is set to "STREET" and the GPS is enabled and working?
- Press "START STOP" to start logging and input the necessary information?
- Drive no faster than 25 mph while logging data on the study site?
- Turn the flasher on while driving below the speed limit?
- Obey all traffic laws?
- Record ending mileage and time when finished?
- Only collect data during complete night? (Approximately 11pm to 4am MDT during the summer and 8pm to 5am MST during the winter)

Figure 5-1 Light Data Collection Checklist

5.4.2.1 Data Cleaning

The raw output of the data is a table including the date and time of recording; GPS coordinates (labeled LAT and LONG); horizontal dilution of precision (HDOP), a measure of GPS precision; the measured illuminance value (in klux, later converted to fc) for each sensor (labeled LEFT and RIGHT); and the user input for the route name (labeled ROUTE). Table 5-2 shows a sample of the raw data.

Table 5-2 Sample of Raw Light Data

DATE	TIME	LEFT (klux)	RIGHT (klux)	LAT	LONG	HDOP	ROUTE
8/4/2023	39:52.0	0.002782	0	40.52695	-111.872	0.87	71
8/4/2023	39:53.0	0.000571	0.000585	40.52695	-111.872	0.87	71
8/4/2023	39:54.0	0.007452	0	40.52696	-111.872	0.87	71

The coordinates were used to calculate vector distance and direction between successive points. If points were closer than 50 ft apart, their average was calculated, and the new point was placed between the original points. Because illuminance was collected in continuous time intervals, this process solved the issue of points accumulating when the vehicle was stopped.

Even with the method of averaging data clumps, points were still spaced irregularly. Kriging, a method for spatial interpolation, was used within ArcGIS software to estimate values onto an evenly-spaced raster grid. This process smoothed data irregularities and measurement errors. Following the Kriging process, the rasterized light data were summarized by segment. To summarize light values by roadway segment, a GIS dataset was first created from the route and milepoint information of the segment's dataset. The segments were then spatially buffered using road lane information to form polygons that accurately represent the area of the roadway. This buffered roadway segment data and the rasterized photometric data obtained from Kriging were used as inputs to calculate summary statistics by segment. A tool within ArcGIS software calculated statistics such as the average, maximum, minimum, and standard deviation of illumination within each roadway segment.

5.4.2.2 Data Extrapolation

While the cleaned photometric data could be used for statistical analysis, there was a possibility that the resulting CMF would not be as useful because the data were collected at the height of the vehicle roof where the sensors were mounted. One option to address this was considered, which applies the inverse square law of light to extrapolate illuminance values to the pavement level. A value at the pavement level is more standard in roadway lighting design software such as AGi32 and is therefore more useful than a value from the vehicle's height.

However, since there was insufficient data about luminaire heights, this method could not be practically applied for the whole dataset. Data measured at the vehicle height were used in place of extrapolated data.

5.4.2.3 Uniformity Calculation

The average illuminance is a straightforward calculation, but there is not a standard way to calculate uniformity of light from illuminance. Some simple uniformity calculations include the standard deviation of illuminance and the maximum-over-minimum ratio of illuminance. However, these metrics do not account for how often light varies along a road segment.

To account for the frequency of light variation and other more detailed uniformity metrics, calculations could be made from the “illuminance profile” for each segment. These profiles were created by plotting the average, maximum, and minimum illuminance for 50-foot sections along each segment. Figure 5-2 shows an example of one of these illuminance profiles. The chart in this example intuitively shows that the segment has non-uniform lighting with frequent fluctuations, which means there should be a way to quantify this non-uniformity.

One way to measure uniformity from lighting profiles is to calculate the maximum-over-minimum ratios between adjacent fluctuations. This helps to “smooth” out the effect of outliers seen in traditional maximum-over-minimum ratios. However, this still does not address the frequency of fluctuations in the lighting. To calculate this, lighting profiles can be used to measure how often lighting fluctuates, which will be called “lighting frequency” in this research, measured as the number of fluctuations per mile. Lighting frequency can also be used to estimate how closely light poles or other light sources are spaced using Equation 5-1.

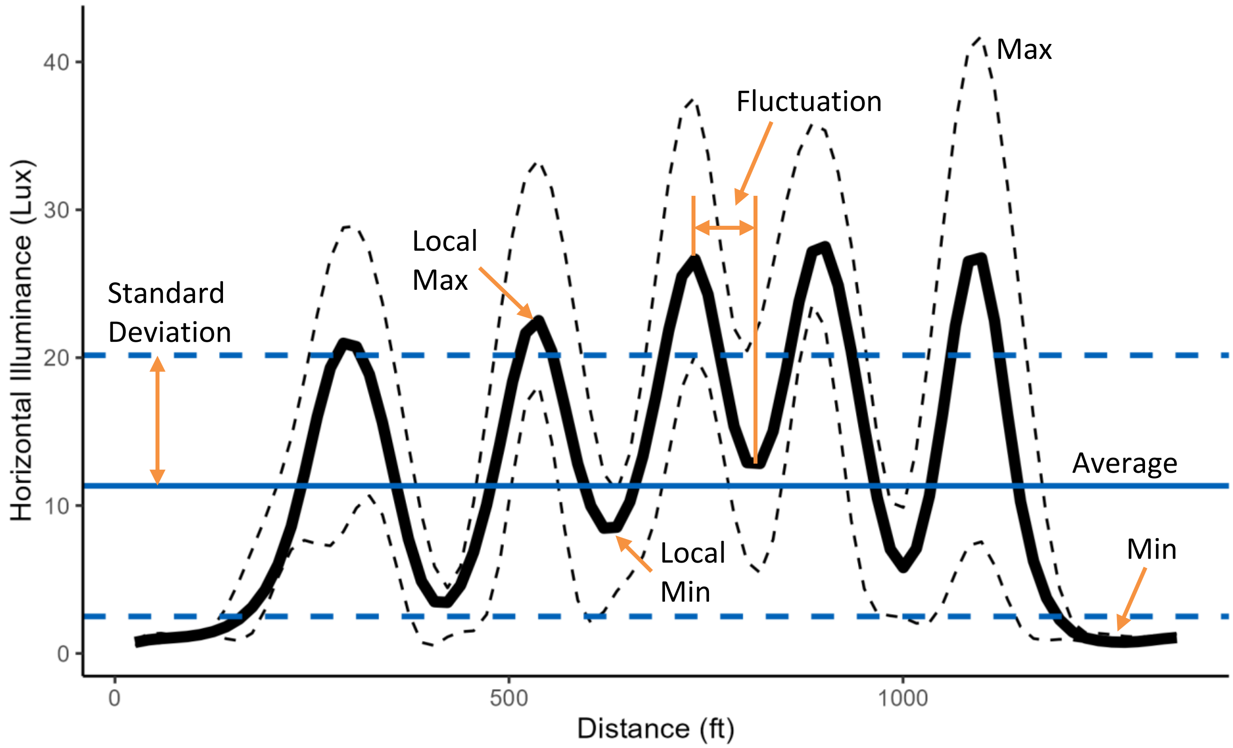


Figure 5-2 Illuminance Profile for US-89 from MP 330.20 to 330.46

Equation 5-1 Lighting Frequency to Spacing Conversion

$$Spacing\ (feet) = \frac{5280\ feet\ per\ mile}{Frequency\ (fluctuations\ per\ mile)} \times 2\ fluctuations$$

5.4.3 Qualitative Data

In addition to measuring illuminance, the researchers recorded qualitative observations related to street lighting. This included the reflectivity of the pavement and the quality of lighting on pedestrian walkways. Observations were simple to avoid assigning too much precision to human judgment (e.g., the pedestrian lighting could be classified as “good” or “bad”). While these qualitative observations were not empirically based, they were considered as potentially useful variables to include in the statistical analysis and were summarized by roadway segment for this purpose. Note, the checklist in Figure 5-1 listed some of the qualitative observations made. Although qualitative observations were evaluated, it was decided they were not useful for the analysis.

5.5 Data Compilation

After the roadway-segment characteristic data, segment photometric data, and segment crash data were built, they needed to be combined for statistical analysis. For the crash data, this involved locating the segment where each crash occurred using route and milepoint information. This created a dataset which summarized each crash within the study area and period along with roadway characteristics for the segment where the crash occurred. Combining crash data with roadway segment data and segment photometric data required summarizing each relevant crash characteristic separately since there could be any number of crashes on a segment. The resulting dataset is more useful than the raw crash dataset because it includes crash counts summarized by roadway segment and separated by characteristics such as nighttime conditions, crash severity, and VRU involvement.

5.6 Summary

Crash data, roadway data, and light data were all successfully collected, cleaned, and compiled using the processes described in this chapter. The remainder of this report will explain how the data were used to build statistical models and develop CMFs for the quantity and quality of lighting.

6.0 DATA EVALUATION

6.1 Overview

For the purpose of statistical analysis, the data collected were classified as explanatory variables, response variables, and covariates. Since a CMF was developed to predict a crash response based on quantity and quality of light, potential explanatory variables included any light-related variables, and potential response variables included any crash metrics. Other data related to roadway characteristics were considered for covariates because these could account for any external crash response unrelated to the relationship between the explanatory and response variables. The sections in this chapter evaluate the validity and usefulness of potential explanatory variables, response variables, and covariates.

6.2 Explanatory Variable Evaluation

The explanatory variable in a statistical analysis refers to the variable which is assumed to cause a change in the response variable. In this study, the explanatory variable was originally defined as the "quantity and quality of street lighting." Through the literature review and experimentation development early in the study, it was decided that horizontal illuminance, uniformity, and qualitative observations would be reasonable measures of lighting to use as explanatory variables. These variables can be used together to create a CMFunction, or separately to create multiple CMFs or CMFunctions.

For horizontal illuminance, the average illuminance for each road segment was used. For uniformity, several options were calculated as discussed previously in Section 5.4.2.3. These are the standard deviation, maximum over minimum (max/min), maximum over average (max/avg), local max/min, and lighting frequency. Table 6-1 shows the benefits and drawbacks of each of these explanatory variables.

Table 6-1 Benefits and Drawbacks of Explanatory Variables

Explanatory Variable	Benefits	Drawbacks
Average Illuminance	Clearest metric representing "quantity" of light. Standard metric used in street lighting measurement. Easier to measure than luminance.	Does not account for uniformity. Possibly not as representative of street lighting as luminance.
Uniformity (Standard Deviation)	Easy to calculate. Not influenced as much by outliers.	Hasn't been proven effective in the literature. Doesn't account for adjacent differences in light.
Uniformity (Max/Min)	Easy to calculate. Most common uniformity metric used in the literature and lighting standards.	Calculation becomes problematic when zero light measured. Doesn't account for adjacent differences in light.
Uniformity (Max/Avg)	Easy to calculate. Metric used for the illuminance method in the AASHTO Lighting Design Guide.	May still be overinflated by outliers. Doesn't account for adjacent differences in light.
Uniformity (Local Max/Min)	Accounts for adjacent differences in light. Justification for a similar method given by Gibbons et al. (2014b).	This method is somewhat experimental.
Uniformity (Lighting Frequency)	Accounts for adjacent differences in light. Simple and intuitive way of understanding lighting.	This method is very experimental. Ignores traditional max/min ratio and treats this as a binary operator.

Average illuminance is a particularly important metric for this research because it is the primary variable used to quantify light. Figure 6-1 shows the distribution of horizontal illuminance by road segment for the study areas to evaluate the reasonableness of the data.

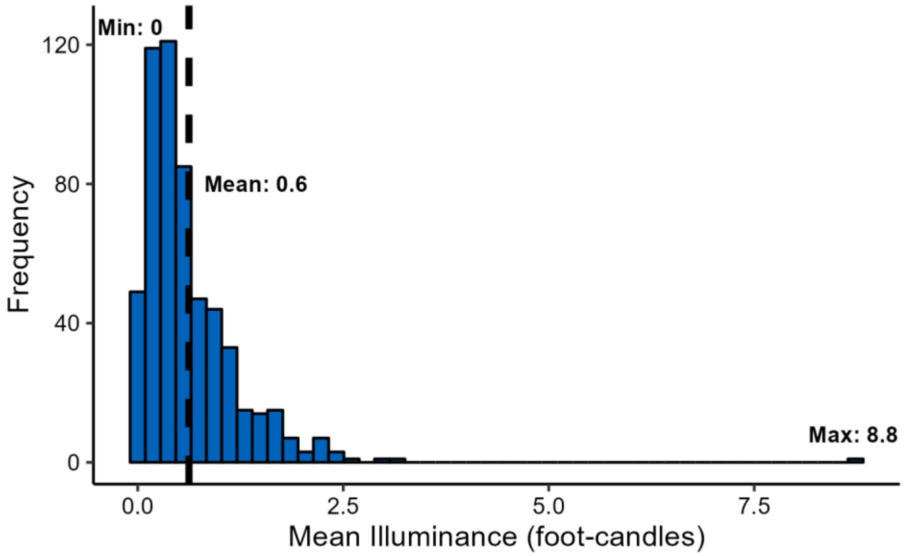


Figure 6-1 Histogram of Average Illuminance by Segment

The average illuminance distribution is rather right skewed, but this behavior is not entirely surprising because light dissipates quickly over distance. Unless a road segment is continuously illuminated with high light levels, the dark spots in the road will reduce the average drastically. The data also contains a very large outlier of about 8.8 fc which was removed during analysis.

Unlike average illuminance, there may be multiple ways to understand uniformity of light. The traditional metrics used for uniformity are max/min and max/avg illuminance ratios. However, there are several road segments with portions where the light measured zero making it impossible to calculate max/min. Also, many of the low-light measurements are less precise, so it was decided not to use max/min. Figure 6-2 shows the distribution of max/avg ratios. Standard deviation, a similar metric sometimes used, is shown in Figure 6-3. Both distributions are right-skewed just like the average illuminance distribution. Therefore, it is useful to note that there is more data in the lower ranges of these metrics.

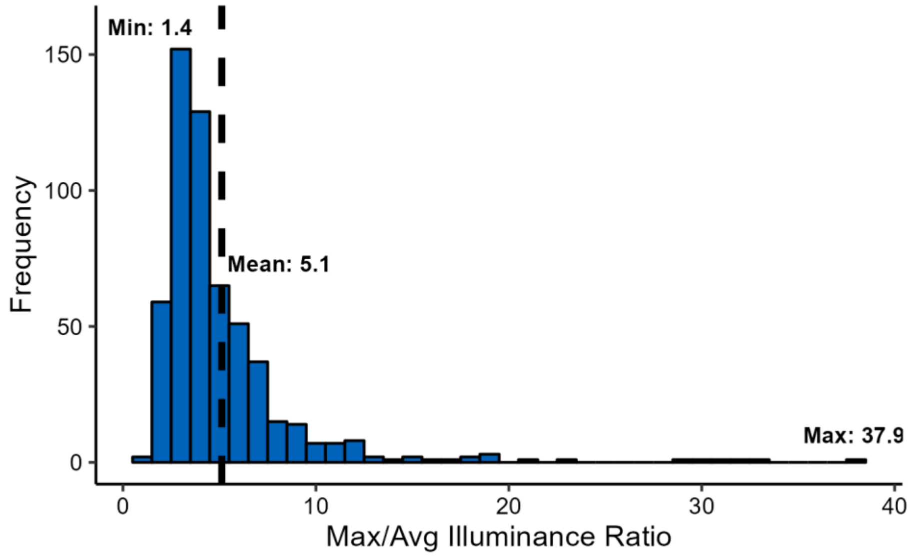


Figure 6-2 Histogram of Max/Avg Ratios by Segment

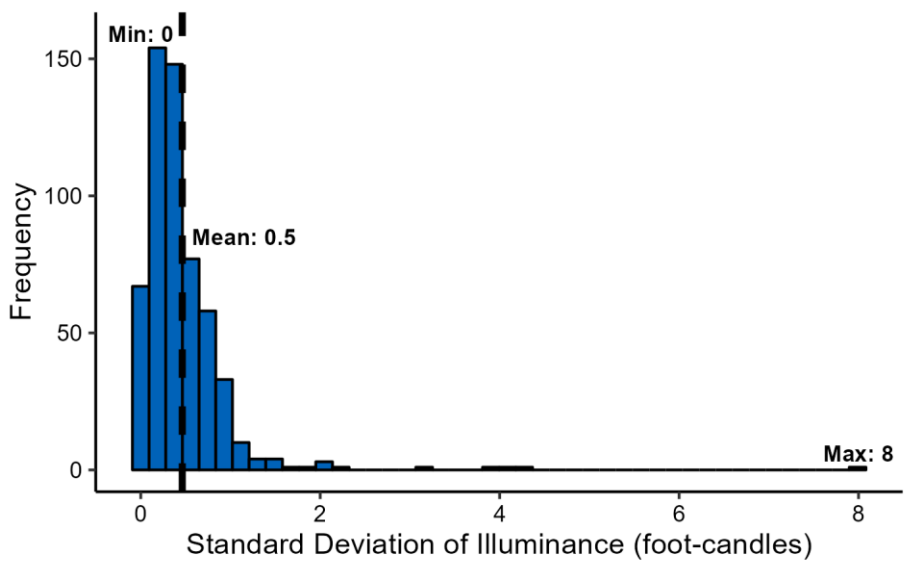


Figure 6-3 Histogram of Standard Deviation Illuminance by Segment

As mentioned previously, max/min was not considered as an explanatory variable because there is too much uncertainty in low-light measurements. However, local max/min addresses this issue by taking illuminance averages every 50 feet and by averaging all max/min ratios across a segment. The local max/min ratio distribution plotted on a log scale is shown in Figure 6-4. This metric is right-skewed even with a log transformation, which could be an indication that this metric is better at identifying segments with higher-than-normal light non-uniformity (contrast). However, this may also indicate that low-light measurement error still influences local max/min just as it does with standard max/min ratios. Taking the log transformation of local max/min may help to address this error.

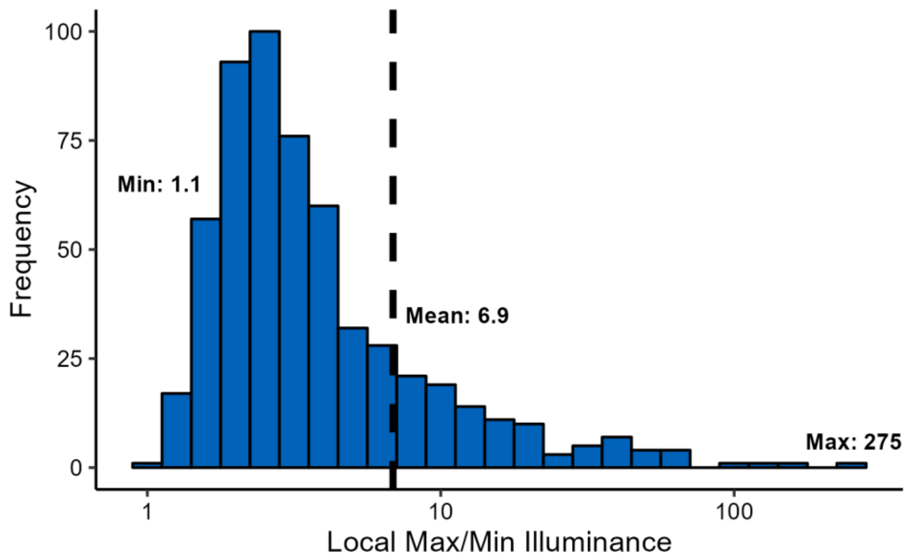


Figure 6-4 Histogram of Local Max/Min Ratio on a Log Scale

Another experimental lighting metric was lighting frequency. Figure 6-5 shows the distribution for this when a significant contrast of 0.1 fc is used (meaning only fluctuations where the difference between measurements was greater than 0.1 fc were counted). This significance is only enough to adjust for measurement error, so it theoretically includes any true lighting contrast. Using a higher significance makes the distribution more right-skewed with a high volume of zeros. Therefore, the 0.1 fc significance is most straightforward while still accounting for measurement error. Lighting frequency ignores the magnitude of contrast but is the only lighting metric which truly accounts for the frequency of contrast.

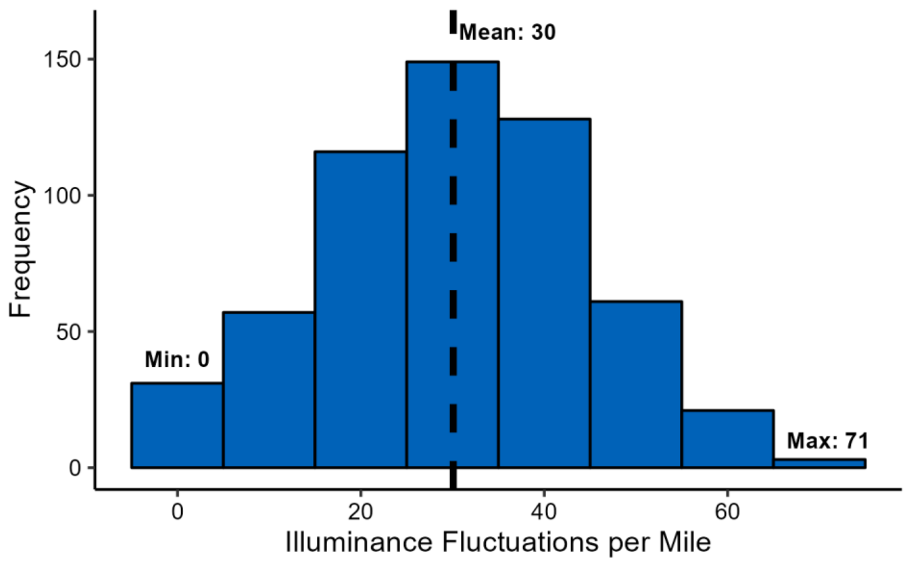


Figure 6-5 Histogram of Lighting Frequency by Segment (0.1 fc Significance)

Figure 6-6 shows correlation plots with Pearson's correlation coefficients (R) for standard deviation and lighting frequency. Both have a strong positive correlation with average illuminance. This means that the statistical models should account for the interaction between average illuminance and these metrics.

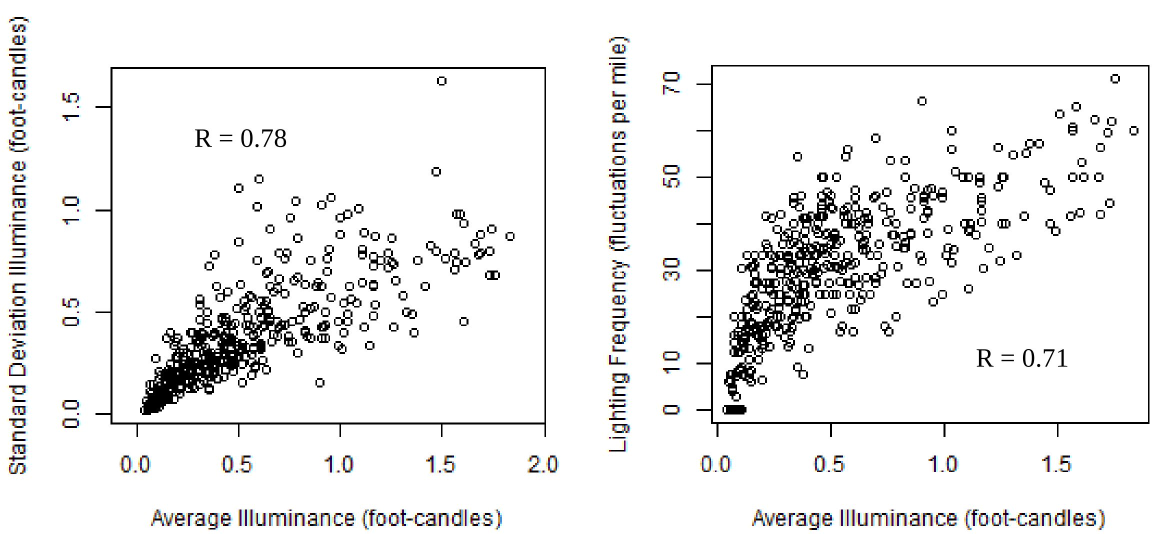


Figure 6-6 Correlation Plots for Standard Deviation and Lighting Frequency

Finally, the qualitative observations about shoulder and sidewalk lighting were taken because there was no way to measure lighting in these areas with the vehicle-mounted light data collection system. While this metric is not empirically based, the observations were kept simple so it could still be useful. This metric will not be used to calculate a CMF, but it may be useful for drawing relationships with pedestrian crashes. Any relationships identified by this may provide justification for further research on shoulder and sidewalk lighting.

6.3 Response Variable Evaluation

Vehicle crashes are the response variable for any study involving CMFs. Some consideration may be given to surrogate safety measures such as post-encroachment time or hard-braking data, but these were not considered in this analysis. Since street lighting should only affect nighttime crashes, apart from crashes specifically with light poles which may occur during the day, nighttime crashes were filtered using the process described previously in Section 4.3. Among nighttime crashes, crashes with VRU and higher severity crashes were of particular interest. Nighttime crashes with severity levels 2 (potential injury) through 5 (fatal) were considered for some CMFs. VRU crashes could be explained by external variables which data were not obtained for, including population density, so these were not considered as heavily. However, since severe VRU crashes are far more common at night, higher severity nighttime crashes served as a useful stand-in for these.

Another response variable considered for this analysis was the night-to-day crash ratio. By including daytime crashes in the response variable, it is theoretically possible to eliminate covariates, if the same covariates which affect daytime crashes would also affect nighttime crashes. However, since the nighttime environment is significantly different from the daytime environment, this assumption should be given some scrutiny. Figure 6-7 and Figure 6-8 compare nighttime crash counts to night-to-day crash ratios showing the night-to-day crash ratios were slightly more normally distributed. However, the skewness shown in Figure 6-8 implies that there are likely still other covariates not addressed by the night-to-day ratio.

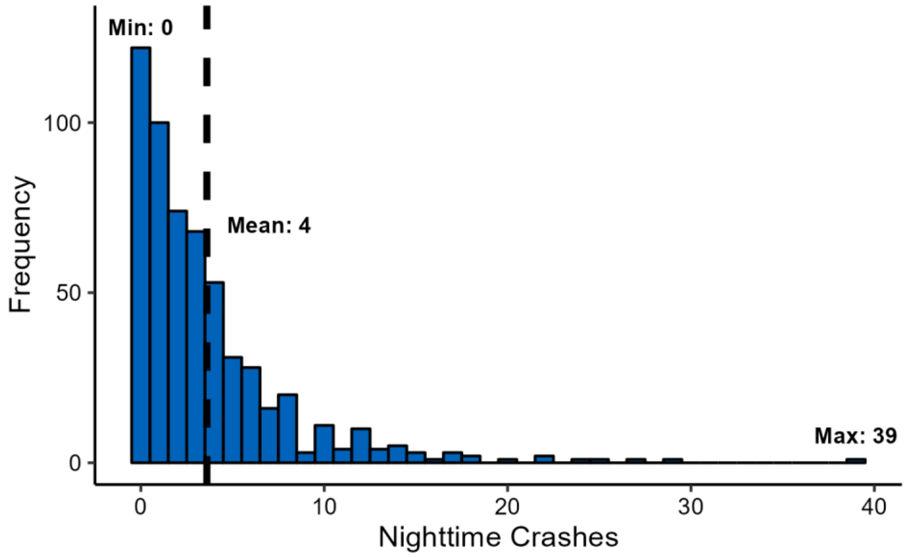


Figure 6-7 Histogram of Nighttime Crashes by Segment

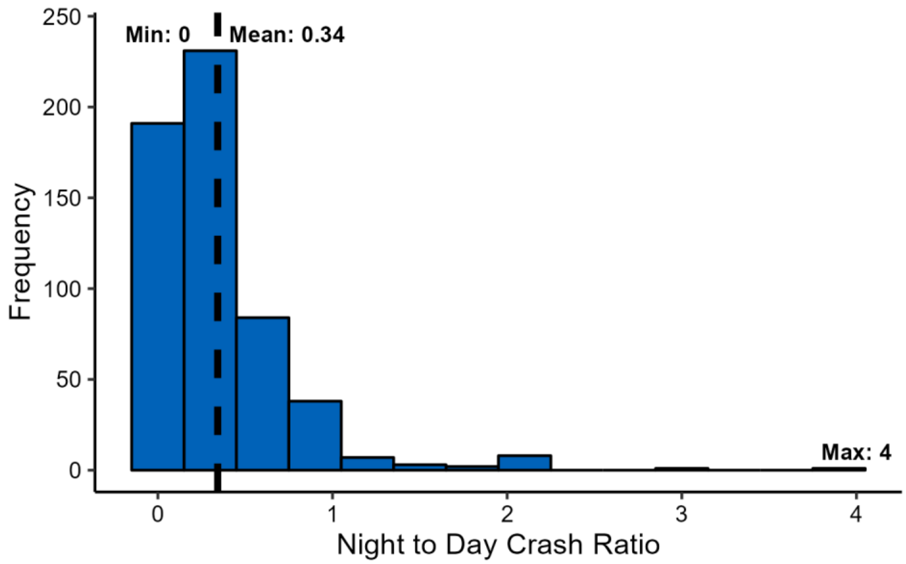


Figure 6-8 Histogram of Night-to-Day Crash Ratio by Segment

Another way to normalize the crash data could be to calculate a crash rate with road segment length and AADT. Crash rates per hundred million vehicle miles (RHMVM) are common to use for road-segment-related crashes. Figure 6-9 shows the distribution of these crash rates. This is more normally distributed than crash counts or night-to-day crash ratios, except there are many excess zeros. A statistical method called zero-inflation could be used to address this issue by assuming that road segments with zero crashes are inherently different from other road segments. There has been some debate on whether this method is appropriate for road safety analysis, but Pew et al. (2020) provides justification for zero-inflation.

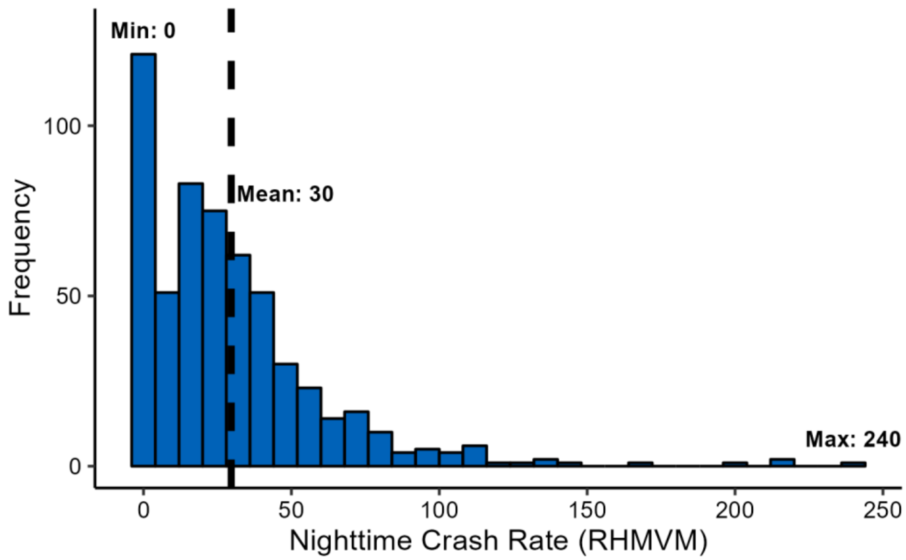


Figure 6-9 Histogram of RHMVM

6.4 Covariate Evaluation

An important part of the statistical analysis was determining which variables served as covariates. Mainly, statistical methods were used to determine the validity of potential covariates, but it is important to perform a sanity check on covariates as well to make sure they are reasonable. The covariates available in the road segment data were segment length, AADT, median type, and lane configuration. Other variables such as functional class were kept consistent in the data as only urban arterials were observed. Considering that segment length and AADT are used to calculate crash rates, these are strong choices for covariates if a crash count is used. Median type and lane configuration can also be considered as covariates as these might influence crashes, but not as strongly.

The two covariates with the strongest correlation without connecting the explanatory and response variables are AADT and segment length. Figure 6-10 and Figure 6-11 depict the distributions for AADT and segment length respectively to show whether the data are reasonable. AADT is normally distributed. This is expected because a random sample of arterial locations were chosen. Segment lengths are right-skewed which is reasonable because segments are much shorter in densely populated areas where there are more cross streets and changes in roadway characteristics.

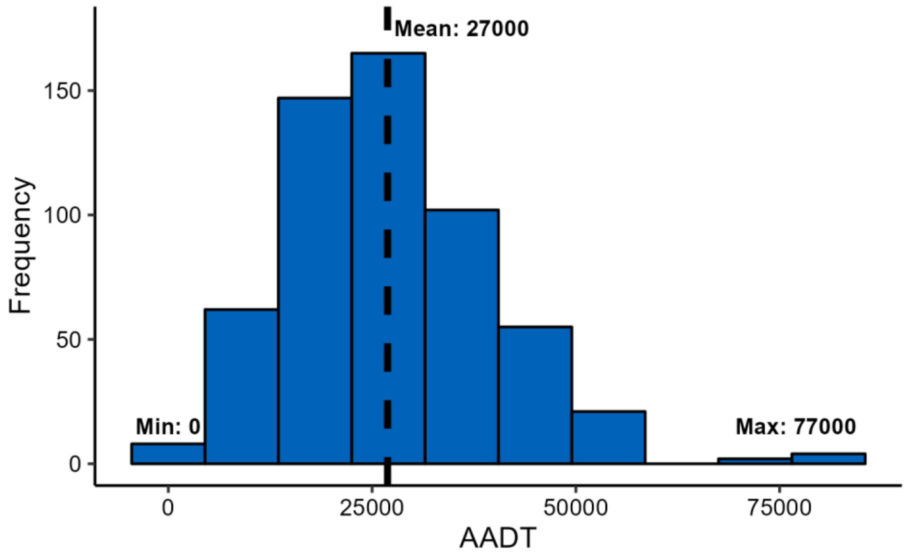


Figure 6-10 Histogram of AADT by Segment

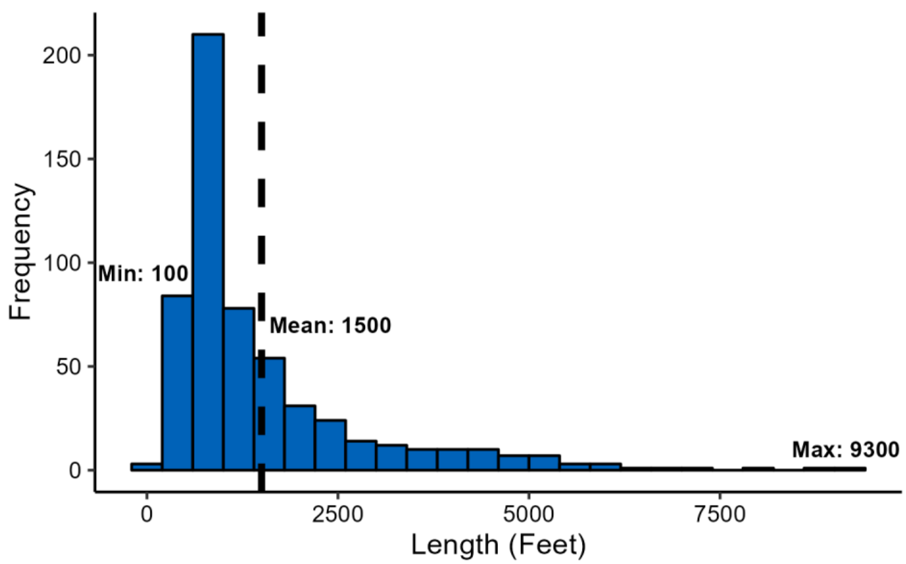


Figure 6-11 Histogram of Segment Lengths in Feet

Median type and lane configuration may also be considered as covariates, but only if they show a statistically significant relationship to crash response. These relationships are explored further in Section 7.3.3.

6.5 Summary

The data evaluation process ensured that the data collected are reasonable to use in analysis. A significant part of this effort was deciding which explanatory variables to use since these will define how CMFs are interpreted. Traditional lighting metrics such as average illuminance and maximum-over-minimum ratios were considered as well as some experimental metrics. Average illuminance, standard deviation of illuminance, max/avg illuminance, local max/min illuminance, and lighting frequency from illuminance were identified as reasonable explanatory variables to use in the analysis. Max/min illuminance spanning roadway segments was believed to be unreliable due to measurement error at low light levels. The data evaluation helped identify the benefits and drawbacks of each metric. Awareness of these benefits and drawbacks will be helpful when considering CMFs generated. The next chapter expounds on the evaluated variables with statistical analysis.

7.0 STATISTICAL ANALYSIS

7.1 Overview

The statistical analysis for this research can be separated into four steps: choosing the CMF study method, selecting variables, choosing a statistical model, and calculating one or more CMFs. The following sections describe these processes in detail.

7.2 CMF Study Selection

As shown in the flowchart produced by Gross et al. (2010) in Appendix A, there are three reasonable types of CMF studies that can be selected for this research. However, a cohort study was not considered since the treatment is not rare, so the other two are a typical cross-sectional study and a variation of a cross-sectional study called a case-control study. Both use data from a sample of similar roadways at the same point in time to draw conclusions about the explanatory variables, which in this case are light quality and quantity. Both studies were considered for their unique benefits, and an appropriate study was chosen as summarized in the following subsections.

7.2.1 Cross-Sectional Study

In their simplest form, cross-sectional studies take the results from simple linear regression to create a CMF. However, it is nearly impossible to find a sample of very similar roadways, so multiple variable regression is often used to account for differences in confounding variables across the data. This is often used in conjunction with a Poisson or negative binomial distribution to account for the rare and random nature of crashes. There is not a specific procedure to generate CMFs from this type of regression, but a generalized linear model (GLM) with a link function from Poisson or negative binomial regression can sometimes be used as an intermediate step for creating CMFs. Cross-sectional studies work well when data are sourced from locations with similar characteristics and when the treatment and crash type are not rare.

7.2.2 Case-Control Study

Case-control studies are a variation of cross-sectional studies, but they work in a fundamentally different way than typical cross-sectional studies. A case-control study applied to this research would separate sites based on whether a nighttime crash occurred or not, and then evaluate the presence of the treatment (e.g., lighting) within each group. The lighting metrics in the group with crashes are compared to the lighting metrics in the group with no crashes. A matching scheme can then be used to account for confounding variables. Case-control studies are effective when crash types are too rare to conduct a typical cross-sectional study because they do not need to account for the quantity of crashes at a location.

7.2.3 CMF Study Selection

It was decided to conduct this study as a typical cross-sectional study. Cross-sectional studies provide several benefits when conducted properly, including the ability to develop CMFunctions and the ability to empirically predict crashes. While before-after studies are typically preferred for CMF studies for their ability to normalize confounding factors, a well-conducted cross-sectional study can be just as effective at predicting crash effects despite the additional challenge of identifying covariates.

The major drawback with case-control studies is that they do not account for the quantity of crashes, only the presence of them. They also do not have a way to empirically demonstrate causality which means CMFs resulting from case-control studies require the assumption of causality. This type of study may be useful if the crash type is rare, but typical cross-sectional studies are preferred because they are more robust for proving the validity of a CMF.

7.3 Variable Selection

To an extent, the variables for this analysis were chosen in previous stages of this study. However, data were collected for a larger number of variables than were expected to be used. This allowed for variables to be filtered by statistical significance and by whether they were more intuitively reasonable than others. The following sections describe the variable selection process for potential explanatory variables, response variables, and covariates.

7.3.1 Explanatory Variable Selection

Metrics of average illuminance and uniformity were selected as potential explanatory variables. These were tested for statistical significance through an iterative process. Various cutoff values were iterated through each potential explanatory variable and used as inputs for the statistical model. Explanatory variables and cutoff values were then determined useful if the results of the statistical model were statistically significant. The coefficients of these statistically significant models were then used to calculate CMFs. Average illuminance, standard deviation, local max/min, and lighting frequency were chosen as explanatory variables for their respective CMFs.

7.3.2 Response Variable Selection

Total nighttime crashes were selected as response variables for the statistical model. Crashes with any of severity levels 2 through 5 were also considered, but these did not generate any useful CMFs from the iterative process mentioned in Section 7.3.1.

7.3.3 Covariate Selection

The researchers chose variables to account for covariance in the statistical model. Random Forests were used to assess which covariates were most important in predicting the response variable "number of nighttime crashes." A Random Forest is a machine learning model useful in classification and prediction. At each iteration (denoted as a "tree"), a random subset of explanatory variables is chosen from the dataset. A mathematical algorithm then selects the best way to "split" the data based on one of the explanatory variables (denoted as a "branch"). The data is continually split according to the best available explanatory variable until a full decision tree is made. Each new observation can then receive a predicted response value based on its explanatory variables and the given decision tree. The algorithm is called a "Random Forest" because many trees are created, each with a different random subset of explanatory variables. The algorithm also tracks how useful each explanatory variable is in the creation of the decision trees, assigning each variable a score based on importance. The importance scores of the variables are a useful method of variable selection when looking at a specific response (Genuer et al., 2010). Appendix C shows an example tree used in the Random Forest model.

Random Forests showed that “number of daytime crashes” was successful in predicting “number of nighttime crashes” and thus it was accounted for in the model as a covariate. Other important covariates were AADT and road segment length. Besides being significant in Random Forest analysis, these are important to crashes occurring on roadway segments specifically. This is helpful because intersection-related crashes were removed from the analysis to isolate segment crashes as explained in Section 5.2. Number of lanes was also determined through engineering judgment to be a covariate in the model as this is a major roadway characteristic. While number of lanes wasn’t successful at predicting crashes in the dataset with Random Forests, it is likely applicable outside of the dataset, and its confounding effects would not be addressed if it was excluded from the model. Thus, the covariates included in the model were as follows:

- Number of daytime crashes
- AADT
- Road segment length
- Number of lanes.

7.4 Statistical Model Selection

The data were collected in a cross-sectional study, thus, the method for finding CMFs involved multivariate regression with a log link function and an assumed negative binomial structure. Poisson and negative binomial regression are both appropriate for modeling discrete count data such as crash data, but the negative binomial structure was assumed because it is better suited for addressing overdispersion in the data (Berk and MacDonald, 2008; Hilbe, 2011).

To illustrate, the Poisson distribution was created such that its mean and variance are identical – thus, the dispersion, or variability, of the data increases as its expected value increases. For example, if the mean number of nighttime crashes was 10, then a Poisson model would assume the variance is also 10, or the standard deviation is 3.13 (the square root of 10). That means roughly 95 percent of the crash totals would be expected to fall between 4 and 16. Since the mean of the sample crash data is around 3.6, a Poisson model would assume a variance of 3.6 and standard deviation of 1.89. That means roughly 95 percent of the crash totals should fall between 0 and 7 to properly fit a Poisson distribution, and the probability of having a crash total above 10 would essentially be 0. Since this is not the case (there are crash totals well beyond 10) the variability of the data does not match the imposed variability of a Poisson distribution. Thus, the data is considered "overdispersed," and should be fit to a negative binomial model to account for larger variance. Furthermore, Figure 7-1 shows that the crash data are much closer to a negative binomial structure than a Poisson structure.

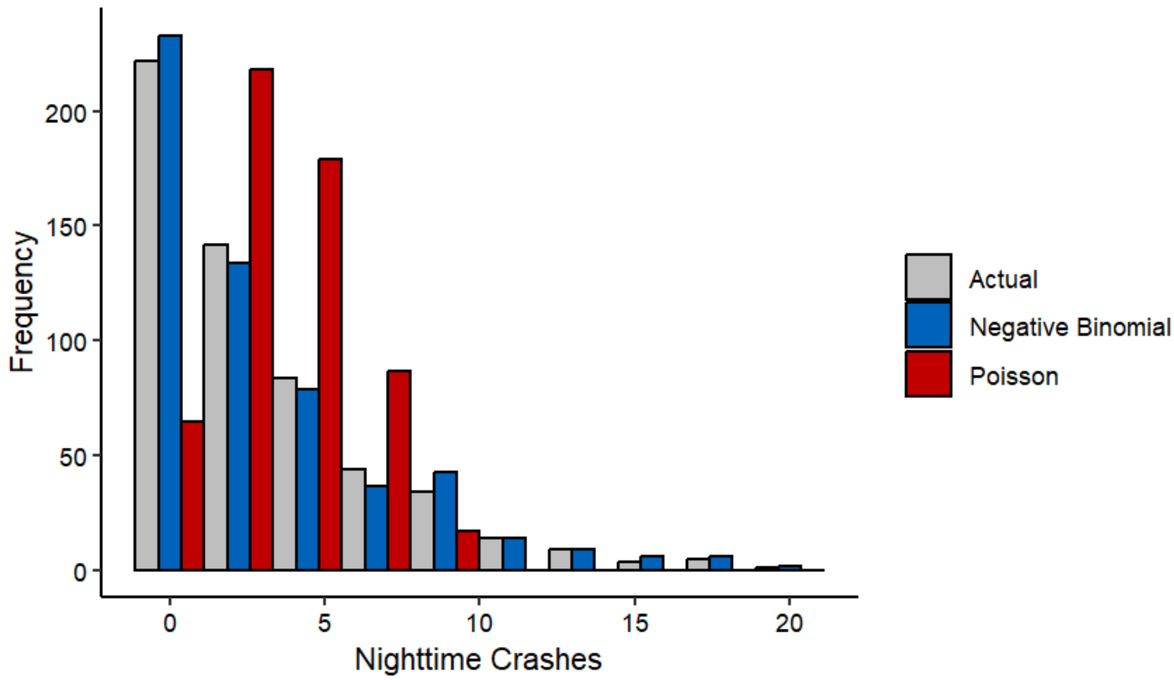


Figure 7-1 Comparison of Poisson and Negative Binomial Distribution

CMFs were then extracted from the model via coefficient values as long as they were statistically significant, and all other known covariates were accounted for as noted in Section 7.3.3. With the inclusion of important covariates in the model, each CMF is applicable to each road segment in the dataset, averaging over such variables as AADT, segment length, and number of lanes. Since each road segment in the data was classified as an urban arterial in Utah, CMFs can confidently be applied to urban arterials in Utah and road segments with similar characteristics.

The next step was to test different statistical methods of calculating CMFs using the negative binomial structure and covariates described previously. The following sections describe two methods used to calculate CMFs, including a bivariate with single cutoff method, and a single variable with two cutoffs method.

7.4.1 Bivariate with Single Cutoff

Since the goal of the statistical analysis was to determine the effect of illuminance on nighttime crashes, the explanatory variables included various illuminance metrics. Creating a binary indicator variable for the desired illuminance countermeasure makes the proposal of that countermeasure more interpretable: The effect of changing the metric from below some cutoff value to above that value can be quantified. This is easier than treating the metric as a continuous variable. It is also important that when looking at countermeasures such as uniformity metrics, some measure of total lighting should be included to account for any interaction. For this research, the average illuminance was used as the measure of total lighting.

An iterative process was used to vary the cutoffs for average illuminance and various other lighting metrics. Models were iterated using the negative binomial structure described previously. Significant relationships were identified from this process and the coefficients of these relationships were used to calculate CMFs. This process also helped to identify which explanatory variables yielded significant models as explained previously in Section 7.3.1.

Because of overdispersion, the response variable, “number of nighttime crashes,” approximately follows a negative binomial distribution as discussed previously. Negative binomial regression is a common approach in determining CMFs from cross-sectional studies. The log link function allows for covariates to have an exponential effect on the response. The resulting model structure is shown in Equation 7-1:

Equation 7-1: Negative Binomial Model for CMF Development

$$\mathrm{Log}(\mathrm{E}(Y_i|X)) = \beta_0 + \beta_1 DTC_i + \beta_2 AADT_i + \beta_3 L_i + \beta_4 N_i + \beta_5 X_{1i} + \beta_6 X_{2i} + \beta_7 X_{3i}$$

$$\mathrm{E}(Y_i|X) = \exp\{\beta_0 + \beta_1 DTC_i + \beta_2 AADT_i + \beta_3 L_i + \beta_4 N_i + \beta_5 X_{1i} + \beta_6 X_{2i} + \beta_7 X_{3i}\}$$

$$X_{1i} = \begin{cases} 0\ if\ M_i < \alpha \\ 1\ if\ M_i > \alpha \end{cases}, \quad X_{2i} = \begin{cases} 0\ if\ U_i < \gamma \\ 1\ if\ U_i > \gamma \end{cases}, \quad X_{3i} = \begin{cases} 0\ if\ M_i < \alpha\ or\ U_i < \gamma \\ 1\ if\ M_i > \alpha\ and\ U_i > \gamma \end{cases}$$

Where:

$\mathrm{E}(Y_i|X)$ is the expected crash outcome,

β_0 is the intercept,

X_{1i} is the indicator variable for M_i,

X_{2i} is the indicator variable for U_i,

$\beta_{1,2,3,4,5,6,7}$ are coefficient values,

DTC_i is the "number of daytime crashes,"

$AADT_i$ is the Annual Average Daily Traffic,

L_i is the road segment length in miles,

N_i is the number of lanes,

X_{3i} is the indicator variable for M_i and U_i interaction,

M_i is the average illuminance,

U_i is the value for uniformity,

α is the cutoff value for M_i,

γ is the cutoff value for U_i

Note: The indicator variable inequalities represent a lighting metric above or below a cutoff value (M_i<>α, Ui <>γ), with X_{3i} describing the interaction between the average illuminance (M_i) and uniformity (U_i). A CMF can be calculated from the change in the indicator variables (with coefficients) when one lighting metric crosses a cutoff value to either above or below that value.

Given this model, X_{1i} and X_{2i} are indicator variables for average illuminance (M_i) and uniformity (U_i), respectively. These indicator variables are active if values are above the cutoffs and 0 if they are below. However, the interaction indicator variable (X_{3i}) alters the effect of the other indicator variables because it is active when both average illuminance and uniformity are above their cutoffs. This means e^{β_6} is the multiplicative effect of the change in the uniformity metric from "below to above" the cutoff γ on the response variable for roads with average illuminance less than the cutoff α. Also, $e^{\beta_6 + \beta_7}$ is the multiplicative effect of the change in the uniformity metric from "below to above" the cutoff γ on the response variable for roads with average illuminance greater than the cutoff α. Confidence intervals can be created using the standard errors of the β coefficients at the specified level of significance.

Figure 7-2 depicts the multiplicative effect of changing lighting frequency from "below to above" a cutoff of 30 fluctuations per mile given an average illuminance less than 0.6 fc. Therefore, the change is represented by e^{β_6} from Equation 7-1 which produces a CMF of 1.22. The inverse of this ($e^{-\beta_6}$) produces a CMF of 0.82 which represents changing lighting frequency from "above to below" the cutoff. In other words, decreasing the frequency of lighting below 30 fluctuations per mile when light levels are low correlates to an 18 percent crash reduction. This example shows vehicle miles traveled plotted on the x-axis with total nighttime crashes plotted on the y-axis and includes the following assumptions:

- A uniformity cutoff of 30 fluctuations per mile for lighting frequency
- An average illuminance cutoff of 0.6 fc
- 13 daytime crashes
- A segment length of 0.3 miles.

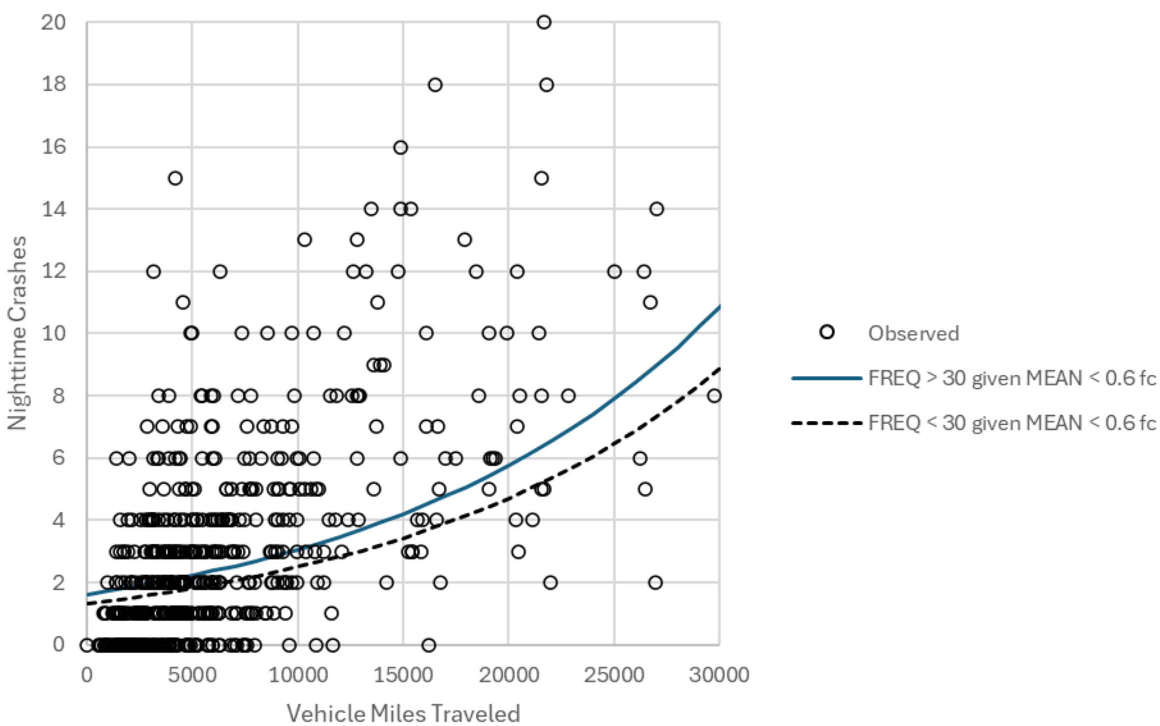


Figure 7-2 Crash Effect Using Single Cutoff Method

7.4.2 Single Variable with Two Cutoffs

The single-cutoff method allows for statistically significant CMFs to be identified through iteration on different lighting values, but the cutoffs end up being somewhat arbitrary without a practical justification for them. In practice, street lighting tends to be grouped into "low," "medium," and "high" bins. With a two-cutoff method, lighting variables are divided into these three groups. Rather than iteration, the cutoffs were determined from existing standards on lighting. This meant CMFs developed using this method were not necessarily as statistically significant as the single-cutoff method, but they were more directly applicable.

A Bayesian approach was used to identify the effect of changing from different average lighting bins – that is, the effect of moving from “low” to “medium” and/or “medium” to “high” average lighting. The Bayesian approach allows for easier computation of uncertainty, which is important for the two-cutoff method. This method does not rely as heavily on statistical significance. The same negative binomial model assumption was used for this method as was used for the single-cutoff method, but there was no interaction term for uniformity and the average lighting variable was incorporated as three separate indicator variables for the three different bins. The structure for this method is shown in Equation 7-2.

Equation 7-2: Model Adjustment for Low to Medium Lighting

$$\text{Log}(\text{E}(Y_i|X)) = \beta_1 \log(DTC_i) + \beta_2 \log(AADT_i) + \beta_3 L_i + \beta_4 N_i + \beta_5 + \beta_6 Med_i + \beta_7 High_i$$

$$\text{E}(Y_i|X) = \exp\{\beta_1 \log(DTC_i) + \beta_2 \log(AADT_i) + \beta_3 L_i + \beta_4 N_i + \beta_5 + \beta_6 Med_i + \beta_7 High_i\}$$

$$Med_i = \begin{cases} 0 \; if \; M_i < \gamma \; or \; M_i > \alpha \\ 1 \; if \; M_i > \gamma \; and \; M_i < \alpha \end{cases}, \qquad High_i = \begin{cases} 0 \; if \; M_i < \alpha \\ 1 \; if \; M_i > \alpha \end{cases}$$

Where:

$\text{E}(Y_i|X)$ is the expected crash outcome,
$\beta_{1,2,3,4,6,7}$ are coefficient values,
DTC_i is the “number of daytime crashes,”
$AADT_i$ is the Annual Average Daily Traffic,
L_i is the road segment length in miles,
N_i is the number of lanes,
M_i is the average illuminance,
β_5 is the intercept (i.e., the effect of being in the “low” category),
Med_i is the indicator variable for medium M_i (i.e., the difference between “medium” and “low” categories),
$High_i$ is the indicator variable for high M_i (i.e., the difference between “high” and “low” categories),
γ is the cutoff between the “low” and “medium” bins,
α is the cutoff between the “medium” and “high” bins

The effect of moving from the “low” bin to the “medium” bin is directly represented by the “medium” coefficient (and moving from the “medium” to the “low” bin is the inverse of that coefficient). In other words, e^{β_6} is the CMF of moving from “low” to “medium” lighting, while $e^{-\beta_6}$ is the CMF of moving from “medium” to “low” lighting.

To calculate the effect of moving from “medium” to “high” lighting, it is easiest to change the model slightly, making the “medium” lighting term into the intercept and changing the former intercept to a “low” lighting term. The final three coefficients of the model would then look as they do in Equation 7-3. In this case, e^{β_7} becomes the CMF of moving from “medium” to “high” lighting, and $e^{-\beta_7}$ becomes the CMF of moving from “high” to “medium” lighting.

Equation 7-3: Model Adjustment for Medium to High Lighting

$$\text{Log}(\text{E}(Y_i|X)) = \beta_1 DTC_i + \beta_2 AADT_i + \beta_3 L_i + \beta_4 N_i + \beta_5 Low_i + \beta_6 + \beta_7 High_{\boldsymbol{i}}$$

$$\text{E}(Y_i|X) = \exp\{\beta_1 DTC_i + \beta_2 AADT_i + \beta_3 L_i + \beta_4 N_i + \beta_5 Low_i + \beta_6 + \beta_7 High_{\boldsymbol{i}}\}$$

$$Low_i = \begin{cases} 0 \; if \; M_i > \gamma \\ 1 \; if \; M_i < \gamma \end{cases}, \qquad High_i = \begin{cases} 0 \; if \; M_i < \alpha \\ 1 \; if \; M_i > \alpha \end{cases}$$

Splitting the data into more than two categories makes it harder to find significant results; using more categories naturally places less data in each category, lowering the power of the tests. This is why a Bayesian approach was used instead. In using the Bayesian paradigm, distributions for each β_p parameter can be constructed. This provides a more holistic view of the likely values for the CMF rather than relying on a single point estimate.

Figure 7-3 depicts the effect of changing lighting from the “low” to the “medium” bin and from the “medium” to the “high” bin and vice versa. This chart plots vehicle miles traveled on the x-axis and nighttime crashes on the y-axis and includes the following assumptions:

- A low cutoff of 0.7 fc for average illuminance
- A high cutoff of 1.4 fc for average illuminance.
- 13 daytime crashes
- 0.3 miles of segment length
- 4 lanes.

Using these cutoffs, there is only a very slight difference between the "low," "medium," and "high" bins.

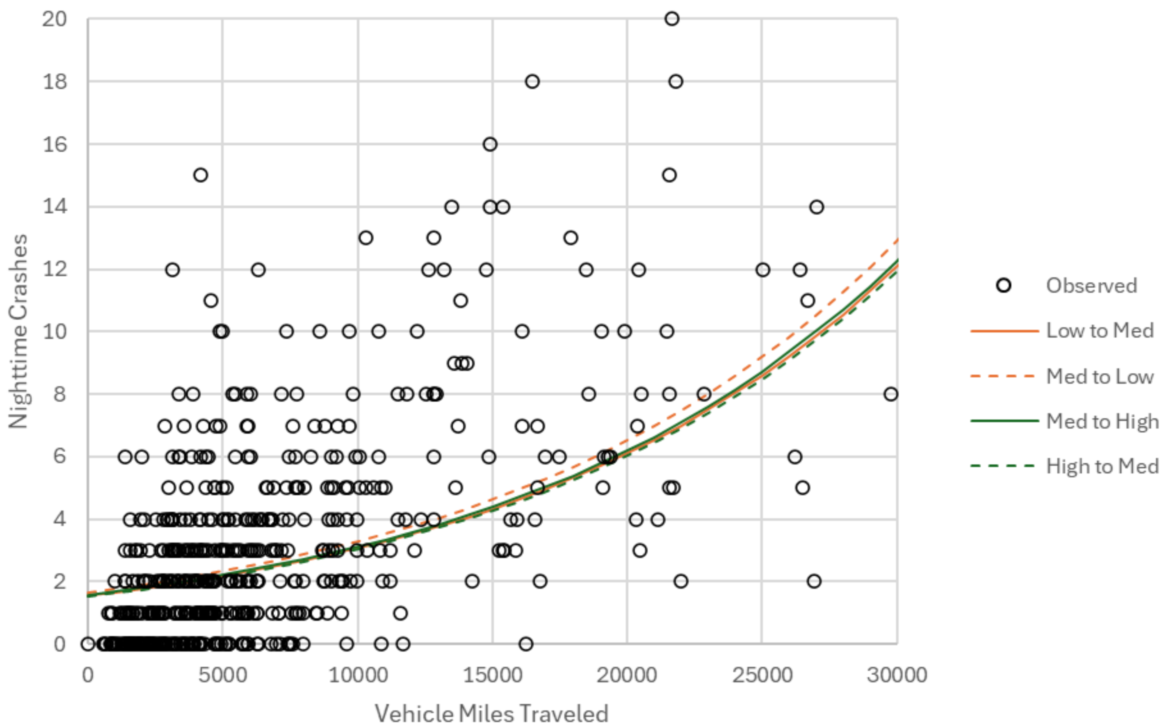


Figure 7-3 Crash Effect Using the Two-Cutoff Method

7.5 Results

The statistical methods described in Section 7.4 were used to generate seven CMFs: three for average illuminance, one for standard deviation, one for local max/min, and two for lighting frequency. The CMFs for average illuminance, except for one, were generated using the two-cutoff method described in Section 7.4.2 to ensure they would better represent lighting standards. The single-cutoff method was used for standard deviation, local max/min, and lighting frequency because these variables are less understood, and the single-cutoff method is better for exploratory analysis. Since the CMFs generated from this method are more statistically significant, these are preferred in the absence of additional knowledge.

The CMFs for average illuminance, standard deviation, local max/min, and lighting frequency shown in Table 7-1 were determined to be the most statistically significant using the single-cutoff method. An extended list of CMFs derived from the single-cutoff method is shown in Appendix D.

The CMFs for average illuminance shown in Table 7-2 show the expected crash response from changing light values between different levels defined by lighting standards using the two-cutoff method. All the CMFs shown are for total nighttime crashes.

Table 7-1 CMFs from the Single-Cutoff Method

Countermeasure	CMF	Standard Error	Confidence Interval	Significance Level
Increase Avg. Illuminance above 0.3 fc given Std. Dev. > 0.3 fc	0.73	-	(0.58, 0.92)	0.05 (95%)
Increase Std. Dev. above 0.6 fc given Avg. Illuminance < 0.9 fc	0.81	0.12	(0.67, 0.99)	0.15 (85%)
Increase Local Max/Min above 2.5 given Avg. Illuminance < 0.9 fc	0.85	0.08	(0.73, 0.98)	0.10 (90%)
Decrease Frequency below 30 fluctuations per mile given Avg. Illuminance below 0.6 fc	0.82	0.09	(0.67, 0.99)	0.05 (95%)
Increase Frequency above 30 fluctuations per mile given Avg. Illuminance above 0.6 fc	0.86	-	(0.66, 1.06)	0.10 (90%)

Table 7-2 CMFs from the Two-Cutoff Method

Average Illuminance Change (fc)	Expected Value of CMF	95% C.I. of CMF	Probability CMF > 1	Probability CMF < 1
0-0.7 (low) to 0.7-1.4 (med)	0.99	(0.84, 1.16)	0.43	0.57
0.7-1.4 (med) up to 1.4-1.9 (high)	0.94	(0.78, 1.12)	0.22	0.78

The CMFs for average illuminance in Table 7-2 do not show strong evidence of a crash response. The selected CMFs are slightly less than 1 for increasing between "low," "medium," and "high" lighting bins, but the confidence intervals include 1. This means that while there may be a very slight safety benefit from increasing street lighting, there may be cases when increased lighting is not beneficial and may even be detrimental to safety. The only exception was the 0.73

CMF generated from the single-cutoff method shown in Table 7-1 from increasing average illuminance above 0.3 fc given a standard deviation greater than 0.3 fc. Since these cutoffs are very low, this CMF is essentially implying that some light is generally better than no light.

CMFs for uniformity correlated to a safety benefit when values increased while average illuminance was low, whether uniformity was measured as standard deviation or local max/min. In this case, increased values indicate less uniformity because standard deviation and max/min are measures of variability and contrast. Therefore, less uniformity is preferred, at least when light levels are low. Likewise, CMFs show that less lighting frequency is preferred, at least at low light levels. However, the 0.86 CMF suggests that more frequency may be preferred at higher light levels.

7.6 Summary

CMFs were developed using a cross-sectional study as this was most suitable for the data collected and best for predicting crashes. Variables were selected which showed the most statistical significance in the model and which made the most sense for predicting safety effects of lighting. These include nighttime crash counts, average illuminance, standard deviation, max/min, local max/min, lighting frequency, daytime crashes, AADT, segment length, and number of lanes. The model chosen from these variables was a negative binomial regression because it is best for accounting for overdispersion in crash data.

Each CMF was created with a different model either using the single-cutoff methodology or the two-cutoff methodology depending on whether the CMF was developed for uniformity or average illuminance, respectively. The resulting CMFs show that average illuminance has minimal effect on safety, but less uniformity may be preferred for safety, at least at low light levels.

8.0 CONCLUSIONS

8.1 Summary

Even though there are fewer road users at night, there is a disproportionate number of severe crashes and severe VRU crashes which occur at night. Roadway lighting is commonly used to mitigate nighttime crashes, so the purpose of this research was to determine how much lighting and what lighting qualities are most effective at improving nighttime roadway safety.

CMFs were developed for this research to show the crash response from quantity and quality of street lighting. While most CMFs show the effect of the existence of lighting, this research is unique in that various lighting levels and conditions are defined to apply CMFs to. It was decided to collect horizontal illuminance data, meaning light which falls incident to the pavement, and 5 years of crash data filtered by nighttime conditions based on astronomical calculations. The light data were collected by attaching sensors to a moving vehicle which measured horizontal illuminance in every lane of each of the study roadways. The study roadways were then split into segments based on roadway characteristics, and lighting metrics were summarized by roadway segment.

Once the data were collected, lighting metrics related to average illuminance and uniformity were calculated. CMFs were then calculated using a negative binomial regression model which included nighttime crashes, daytime crashes, AADT, segment length, number of lanes, and lighting indicator variables. Different methodologies were used to determine the cutoffs for these indicator variables and the resulting coefficients for indicator variables were directly used to calculate CMFs. The two methodologies used were the single-cutoff method and the two-cutoff method described in Sections 7.4.1 and 7.4.2, respectively. Only CMFs which were statistically significant or reasonable were included in the results.

8.2 Findings

CMFs developed from this research did not match initial expectations. It is typically assumed that more lighting is generally better for reducing nighttime crashes, but the results of this research show that simply changing light levels does not have a significant impact on safety. These results are similar to observations made by van Schalkwyk et al. (2016) who noted that there is little empirical evidence for the general assumption that more lighting is better for safety. However, CMFs relating to lighting uniformity show that there may be some nuance to how lighting affects roadway safety. For example, the analysis indicated that less uniformity is preferred at low light levels while the opposite may be true at higher light levels. The following subsections explain the three findings from this research: that average illuminance is not well correlated to crashes, lighting uniformity is well correlated to crashes, and that lighting requirements should reflect the nuanced effects of lighting.

8.2.1 Average Illuminance is Not Well Correlated to Crashes

As shown in Section 7.5, the following CMFs were calculated for average illuminance using the two-cutoff method and the single-cutoff method:

- CMF = 0.99 to increase from “low” light (0-0.7 fc) to “medium” light (0.7-1.4 fc)
- CMF = 0.94 to increase from “medium” light to “high” light (1.4-1.9 fc)
- CMF = 0.73 to increase average illuminance from below to above 0.3 fc given a standard deviation above 0.3 fc.

The first two CMFs were generated using the two-cutoff method and were very close to 1. There also was not much statistical evidence that they were greater or less than 1. This means that a change in average illuminance is not expected to significantly change crash outcomes at the segment level. These CMFs showed a slight crash reduction correlated to increasing light levels, but they were not strong enough to use practically. This may indicate some survivorship bias (Wald, 1943) if lighting already exists in areas where it is most needed in Utah, but it is difficult to prove this with only a sample of the data. It may also be an example of Simpson’s Paradox (Pearl, 2022) which occurs when the overall trend varies from trends in subgroups in the data. The average illuminance CMF generated using the single-cutoff method indicates more of a crash reduction, but this is limited to conditions with standard deviations above 0.3 fc. Since the average illuminance cutoff for this CMF is very low, values below it essentially indicate no lighting. This CMF suggests that in comparison, some lighting is better than no lighting.

8.2.2 Lighting Uniformity is Well Correlated to Crashes

Lighting uniformity, which refers to how consistent light levels are along a roadway segment, including lighting frequency, which refers to how often light levels change, was used to evaluate the "quality" of lighting in Utah. CMFs were calculated for uniformity using the single-cutoff method to account for interaction with average illuminance. The following CMFs were generated for uniformity as explained in Section 7.5:

- CMF = 0.81 to increase standard deviation from below to above 0.6 fc given an average illuminance below 0.9 fc
- CMF = 0.85 to increase local max/min ratio from below to above 2.5 given an average illuminance below 0.9 fc
- CMF = 0.82 to decrease lighting frequency from above to below 30 fluctuations per mile given an average illuminance below 0.6 fc
- CMF = 0.86 to increase lighting frequency from below to above 30 fluctuations per mile given an average illuminance above 0.6 fc.

This research generally showed that less uniformity was correlated to reduced crashes at low light levels, but more uniformity may be preferred at higher light levels. This conclusion is reasonable because uniformity is not very helpful when light levels are low. Non-uniformity on a roadway with low overall lighting indicates that there are at least some locations on the segment with higher lighting. Similarly, higher lighting frequency may mean there are more light poles which serve as hazards to drivers if light levels are low. However, uniform lighting with frequently spaced light sources means drivers are better able to see obstacles in a higher lighting situation.

8.2.3 Lighting Requirements Should be Nuanced

Many lighting requirements apply to general roadway categories which may be a useful starting point, but requirements related to specific scenarios may be necessary to truly improve nighttime roadway safety in Utah. For example, the fact that there is some benefit to non-uniformity in low-lighting scenarios shows that lighting concentrated at specific locations may be preferred. These may be crosswalks, intersections, or other features where better lighting is required. Such implementations would be less costly and more effective for improving safety and would reduce the negative impacts of light pollution. However, there may still be a need for uniform lighting in settings where more lighting is expected such as urban centers.

8.3 Limitations and Challenges

The primary limitations for this project were associated with data collection as data were only collected for sample locations and not for the entire roadway network. There was also some error with the light data sensors used; however, GIS interpolation was used to address most of these errors. Other challenges came from having initial statistical results which did not correspond with expectations. Several statistical methods were tested to obtain logical results, given this challenge. Since this research was conducted as a cross-sectional study rather than a before-after study, statistical methods needed to account for confounding variables and doing so was also a challenge.

9.0 RECOMMENDATIONS AND IMPLEMENTATION

9.1 Recommendations

Based on the results of this research, it is recommended that UDOT focus on improving lighting at specific locations and improving lighting uniformity at locations with high light levels. The theory behind this recommendation is that lighting at night is more attractive to drivers' eyes than during the day which means important obstacles should be well-illuminated, but excess lighting may become a distraction to the driver unless there is enough uniform lighting to imitate daytime conditions. This recommendation is based on the CMFs developed for average illuminance and uniformity. Since CMFs for average illuminance were very close to 1, there is not much need to improve overall lighting in Utah beyond current levels. However, the CMFs for uniformity show that less uniformity at low light levels and more uniformity at high light levels are correlated to crash reductions. This supports the recommendations because non-uniformity at low light levels indicates that there are locations along the segment with significantly more light than the rest of the segments. These locations with more lighting may provide the necessary contrast for drivers to identify important obstacles such as pedestrians at crosswalks. For locations with high light levels, more uniformity may be required because there are significantly more obstacles at these locations (e.g., urban centers).

UDOT may use the CMFs developed in this study for detailed safety analysis, or to improve lighting standards in Utah. The CMFs summarized previously in Section 8.2 can be applied using HSM methodology. The cutoffs for these CMFs may also be used to set the following standards for lighting uniformity:

- For areas with an average horizontal illuminance of less than 0.9 fc, the standard deviation should be greater than 0.6 fc and the local max/min should be greater than 2.5
- For areas with an average horizontal illuminance of less than 0.6 fc, lighting should be spaced at least 250 ft apart
- For areas with an average horizontal illuminance greater than 0.6 fc, lighting should be spaced less than 250 ft apart.

9.2 Future Research

Research on street lighting is ongoing for many transportation agencies. There are many questions related to street lighting and safety which can still be answered. The following are possible follow-up research topics:

- Pedestrian Lighting: The safety impact of lighting on pedestrian walkways specifically could be explored. This could include sidewalks, crosswalks, and intersections. It is recommended that vertical and horizontal illuminance be collected for this. The current data collection system could be modified to be mounted to a scooter or other device which can travel on or adjacent to pedestrian walkways.

- Eye Tracking: An eye tracking device could be used on various drivers in different lighting conditions to determine if there is a correlation between eye movement and lighting. Eye movement has been used as a safety indicator in several road safety applications (Crundall and Underwood, 2011).

- Lighting Design Validity: AGi32 software is commonly used to design street lighting according to standards, but various factors may cause the actual light levels to vary from design. A study comparing actual light conditions with design conditions would be helpful for lighting design.

- Collect More Data: There were limited data for this study, but if more data were collected, the results might be more statistically valid. Alternatively, data collected at the same locations at a future date may allow researchers to perform a before-after study to provide more variety to the statistical methods used.

- Community Surveys: Collect data from communities on which lighting characteristics (e.g., glare, color temperature) contribute to improved comfort and perception of safety.

9.3 Implementation Plan

The purpose of the implementation plan for UDOT is to provide direction on steps needed to implement the technology or products developed under this contract; provide recommendation on staffing needs and resources; and list individuals and organizational roles and responsibilities recommended for implementation. This section outlines the requirements of the implementation plan.

9.3.1 Technology or Products Developed Under This Contract

The contract specifically called for the development of an experimentation methodology as well as CMFs generated to describe the safety effects of the quantity and quality of lighting on Utah roadways. Only the CMFs are expected to be implemented, but UDOT may also choose to implement the technology and data developed from this research in future research projects. Since CMFs are the primary product of this research, these should be implemented first using the following steps:

1. CMFs from this research should be integrated into UDOT's internal CMF database for use in future safety analysis. These primarily include the CMFs shown in Table 7-1 and Table 7-2 of this report. UDOT may also choose to include CMFs shown in Appendix D in their internal CMF database
2. CMFs may also be uploaded to the CMF Clearinghouse website for public use.
3. When applying CMFs in safety analysis, UDOT should follow HSM procedures.
4. Since CMFs in this research were developed using cutoff values, UDOT may choose to use these cutoff values in their lighting standards. Examples of this are included in Section 9.1.

If UDOT decides to conduct more research on street lighting they may use the data from this research or collect new data. The following steps may be taken to collect new data based on the experimentation developed for this research:

1. A mobile light data collection system may be installed on any vehicle with top mounting racks.

2. Individuals collecting the data should reference the procedures shown in Figure 5-1 to maintain safety and data consistency.
3. Data collectors should consult the user manuals for the equipment before collecting data. The photometric sensors may need to be recalibrated prior to data collection.
4. Raw data may need to be cleaned, interpolated, and summarized. The code and GIS workflow shown in Appendix E may be referenced for this step.

9.3.2 Staffing Needs and Resources

UDOT staff will need to integrate the CMFs from this research using their own procedures. If UDOT also desires to modify lighting standards based on this research, they will need to follow the proper administrative procedures for creating standards. It is recommended that these steps be performed by UDOT personnel as they will be more familiar with UDOT's internal structure.

9.3.3 Roles and Responsibilities

The UDOT Traffic and Safety Division should be responsible for implementing CMFs for UDOT's use.

REFERENCES

Abdel-Aty, M., Lee, C., Abuzwidah, M., Al-Arifi, S., and Ahmed, M. (2014). “Validation and Application of Highway Safety Manual (Part D) in Florida.” (Contract Number BDK78-977–14), Florida Department of Transportation.

Adrian, W. (1989). “Visibility of Targets: Model for Calculation.” *Lighting Research & Technology (1969)*, 21(4), 181-188.

American Association of State Highway and Transportation Officials (AASHTO) (2010). *Highway Safety Manual, 1st Ed. (No. 978-1-56051-477-0)*, Washington, DC.

American Association of State Highway and Transportation Officials (AASHTO) (2018). *Roadway Lighting Design Guide (No. 978-1-56051-725-2)*, Revision 7, Washington, DC.

American National Standards Institute and Illuminating Engineering Society (ANSI/IES) (2018). *Recommended Practice for Design and Maintenance of Roadway and Parking Facility Lighting*, Illuminating Engineering Society, New York, NY.

Arecchi, A., V., Messadi, T., and Koshel, R. J. (2007). *Field Guide to Illumination*, SPIE Press, Bellingham, WA.

Berk, R., and MacDonald, J. M. (2008). “Overdispersion and Poisson Regression.” *Journal of Quantitative Criminology*, 24(3), 269-284.

Bhagavathula, R., Gibbons, R., and Kassing, A. (2021). “Roadway Lighting’s Effect on Pedestrian Safety at Intersection and Midblock Crosswalks (No. FHWA-ICT-21-023).” US Department of Transportation, Federal Highway Administration, Washington DC.

Bhagavathula, R., Gibbons, R. B., and Edwards, C. J. (2015). “Relationship Between Roadway Illuminance Level and Nighttime Rural Intersection Safety.” *Transportation Research Record*, 2485(1), 8-15.

Box, P. C., and Alroth, W. A. (1971). “Relationship Between Illumination and Freeway Accidents.” *Journal of the Illuminating Engineering Society*, 66(5), 365-393.

Bullough, J. D., Donnell, E. T., and Rea, M. S. (2013). “To Illuminate or Not to Illuminate: Roadway Lighting as it Affects Traffic Safety at Intersections.” *Accident Analysis and Prevention*, 53, 65-77.

Carter, D., Srinivasan, R., Gross, F., and Council, F. (2012). "Recommended Protocols for Developing Crash Modification Factors: NCHRP 20-7 (314) Final Report." Transportation Research Board,Washington, DC.

Crundall, D., and Underwood, G. (2011). "Chapter 11 - Visual Attention While Driving: Measures of Eye Movements Used in Driving Research." *Handbook of Traffic Psychology*, B. E. Porter, ed., Academic Press, San Diego, CA, 137-148.

Donnell, E. T., Porter, R. J., and Shankar, V. N. (2010). "A Framework for Estimating the Safety Effects of Roadway Lighting at Intersections." *Safety Science*, 48(10), 1436-1444.

Edwards, C. J., and Gibbons, R. B. (2008). "Relationship of Vertical Illuminance to Pedestrian Visibility in Crosswalks." *Transportation Research Record*, 2056(1), 9-16.

Elvik, R. (1995). "Meta-Analysis of Evaluations of Public Lighting as Accident Countermeasure." *Transportation Research Record*, (1485), 112-123.

Elvik, R., and Vaa, T. (2004). *The Handbook of Road Safety Measures*, Elsevier, Amsterdam.

Florida Department of Transportation (FDOT) (2024). *Design Manual (No. 625-000-002) Section 231 - Lighting*.

Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. (2010). "Variable Selection Using Random Forests." *Pattern Recognition Letters*, 31(14), 2225-2236.

Gibbons, R. B., Guo, F., Medina, A., Terry, T., Du, J., Lutkevich, P., Corkum, D., and Vetere, P. (2014a). "Guidelines for the Implementation of Reduced Lighting on Roadways (No. FHWA-HRT-14-050)." US Department. of Transportation, Federal Highway Administration, Research, Development, and Technology, Turner-Fairbank Highway Research Center, McLean, VA.

Gibbons, R. B., Guo, F., Medina, A., Terry, T., Du, J., Lutkevich, P., and Li, Q. (2014b). "Design Criteria for Adaptive Roadway Lighting (No. FHWA-HRT-14-051)." US Department of Transportation, Federal Highway Administration, Research, Development, and Technology, Turner-Fairbank Highway Research Center, McLean, VA.

Gibbons, R. B., Meyer, J. E., and Edwards, C. J. (2018). "Development of a Mobile Measurement System for Roadway Lighting (No. 18-UR-062)." National Surface Transportation Safety Center for Excellence.

Gross, F., Persaud, B., and Lyon, C. (2010). "A Guide to Developing Quality Crash Modification Factors (No. FHWA-SA-10-032)." US Deptartment of Transportation, Federal Highway Administration, Washington, DC.

Hilbe, J. M. (2011). *Negative Binomial Regression*, Cambridge University Press, Cambridge.

Illuminating Engineering Society (IES) (1989). “IES Guide for Photometric Measurement of Roadway Lighting Installations.” *Journal of the Illuminating Engineering Society*, 18(2), 136-138.

Janoff, M. S. (1993). “The Relationship Between Small Target Visibility and a Dynamic Measure of Driver Visual Performance.” *Journal of the Illuminating Engineering Society*, 22(1), 104-112.

Johnson, M., Fabregas, A., Zhenyu, W., Katkoori, S., and Pei-Sung, L. (2014) “Embedded System Design of an Advanced Illumination Measurement System for Highways.” Institute of Electrical and Electronics Engineers, 579-586.

Keck, M. E. (2001). “A New Visibility Criteria for Roadway Lighting.” *Journal of the Illuminating Engineering Society*, 30(1), 84-89.

Kirley, B., Robison, K., Goodwin, A., Harmon, K., O’Brien, N., West, A., Harrell, S., Thomas, L., and Brookshire, K. (2023). *Countermeasures That Work: A Highway Safety Countermeasure Guide for State Highway Safety Offices (No. DOT HS 813 490)*, National Highway Traffic Safety Administration, Washington, DC.

LI-COR (2024a). “LI-210R Photometric Sensor.” <https://www.licor.com/env/products/light/photometric.html>. (Jul 29, 2024).

LI-COR (2024b). “LI-1500 light sensor logger.” <https://www.licor.com/env/products/light/light-logger.html>. (Jul 29, 2024).

Li, Q., Wang, Z., Li, M., Yang, R., Lin, P.-S., and Li, X. (2021). “Development of Crash Modification Factors for Roadway Illuminance: A Matched Case-Control Study.” *Accident Analysis and Prevention*, 159, 106279.

Meeus, J. H. (1991). *Astronomical Algorithms*, Willmann-Bell, Incorporated.

National Highway Traffic Safety Administration (NHTSA) (2024). “Pedestrians Traffic Safety Facts 2022 Data (No. DOT HS 813 590).” Washington, DC.

National Oceanic and Atmospheric Administration (NOAA) (2015). “Solar Calculation Details.” <https://gml.noaa.gov/grad/solcalc/calcdetails.html>. (Jul 31, 2023).

New York State Department of Transportation (NYSDOT) (1995). *Highway Design Manual Chapter 12 - Highway Lighting*, Revision 24.

Niaki, M. S. N., Fu, T., Saunier, N., Miranda-Moreno, L. F., Amador, L., and Bruneau, J.-F. (2016). "Road Lighting Effects on Bicycle and Pedestrian Accident Frequency: Case Study in Montreal, Quebec, Canada." *Transportation Research Record*, 2555(1), 86-94.

Pearl, J. (2022). "Comment: Understanding Simpson's Paradox." *Probabilistic and Causal Inference: The Works of Judea Pearl*, 399-412.

Pew, T., Warr, R. L., Schultz, G. G., and Heaton, M. (2020). "Justification for Considering Zero-Inflated Models in Crash Frequency Analysis." *Transportation Research Interdisciplinary Perspectives*, 8, 100249.

Revalize. 2021. AGi32, Version 20.

Rice, J., Hoon, K. Y., Zhihui, S., and Mohsen, J. P. (2020). "Evaluation of Light Reflectiveness of Modern Pavement: Standard Tungsten Incandescent and LED." *Journal of Transportation Engineering, Part B: Pavements*, 146(2), 04020007.

Sacchi, E., and Tayebikhorami, S. (2021). "Evaluating the Effectiveness of the Safety Improvement Program in Saskatchewan Using an Observational Before–After Study with the Full Bayes Approach." *Canadian Journal of Civil Engineering*, 48(12), 1706-1712.

Scott, P. P. (1980). "The Relationship Between Road Lighting Quality and Accident Frequency (No. TRRL LR929 Monograph)." Transport and Road Research Laboratory, Wokingham, Berkshire.

Suk, J. Y., and Walter, R. J. (2019). "New Nighttime Roadway Lighting Documentation Applied to Public Safety at Night: A Case Study in San Antonio, Texas." *Sustainable Cities and Society*, 46, 101459.

Texas Department of Transportation (TxDOT) (2018). *Highway Illumination Manual (No. 43 TAC 25.11)*.

Tomczuk, P., Chrzanowicz, M., Jaskowski, P., and Budzynski, M. (2021). "Evaluation of Street Lighting Efficiency Using a Mobile Measurement System." *Energies (Basel)*, 14(13), 3872.

U.S. Naval Observatory (USNO) (2024). "Dates of Primary Phases of the Moon." <https://aa.usno.navy.mil/data/MoonPhases>. (Jul 31, 2024).

Utah Department of Transportation (UDOT) (2021). *Roadway Lighting Design Guidelines*, Revision 7.

van Schalkwyk, I., Venkataraman, N., Shankar, V., Milton, J., Bailey, T., and Calais, K. (2016). "Evaluation of the Safety Performance of Continuous Mainline Roadway Lighting on Freeway Segments in Washington State." Washington State Department of Transportation, Olympia, WA.

Wald, A. (1943). "A Method of Estimating Plane Vulnerability Based on Damage of Survivors." *Statistical Research Group*, 432.

Wang, Z., Lin, P.-S., Chen, Y., Hsu, P. P., Ozkul, S., and Bato, M. (2017). "Safety Effects of Street Illuminance on Roadway Segments in Florida." *CUTR Faculty Journal Publications*.

Wanvik, P. O. (2009). "Effects of Road Lighting: An Analysis Based on Dutch Accident Statistics 1987–2006." *Accident Analysis and Prevention*, 41(1), 123-128.

Washington State Department of Transportation (WSDOT) (2023). *Design Manual (No. M 22-01.22) Section 1040 - Lighting*.

Wei, F., Wang, Z., Lin, P.-S., Hsu, P. P., Ozkul, S., Jackman, J., and Bato, M. (2016). "Safety Effects of Street Illuminance at Urban Signalized Intersections in Florida." *Transportation Research Record*, 2555(1), 95-102.

Yang, R., Wang, Z., Lin, P.-S., Li, X., Chen, Y., Hsu, P. P., and Henry, A. (2019). "Safety Effects of Street Lighting on Roadway Segments: Development of a Crash Modification Function." *Traffic Injury Prevention*, 20(3), 296-302.

Ye, X., Pendyala, R. M., Washington, S. P., Konduri, K., and Oh, J. (2009). "A Simultaneous Equations Model of Crash Frequency by Collision Type for Rural Intersections." *Safety Science,* 47(3), 443-452.

Yoomak, S., and Ngaopitakkul, A. (2018). "Optimisation of Lighting Quality and Energy Efficiency of LED Luminaires in Roadway Lighting Systems on Different Road Surfaces." *Sustainable Cities and Society*, 38, 333-347.

Zatari, A., Dodds, G., McMenemy, K., and Robinson, R. (2005). "Glare, Luminance, and Illuminance Measurements of Road Lighting Using Vehicle Mounted CCD Cameras." *LEUKOS: The Journal of the Illuminating Engineering Society*, 1(2), 85-106.

Zhao, J., Zhou, H., and Hsu, P. (2015). "Correlating the Safety Performance of Urban Arterials with Lighting." *Transportation Research Record*, 2482(1), 126-132.

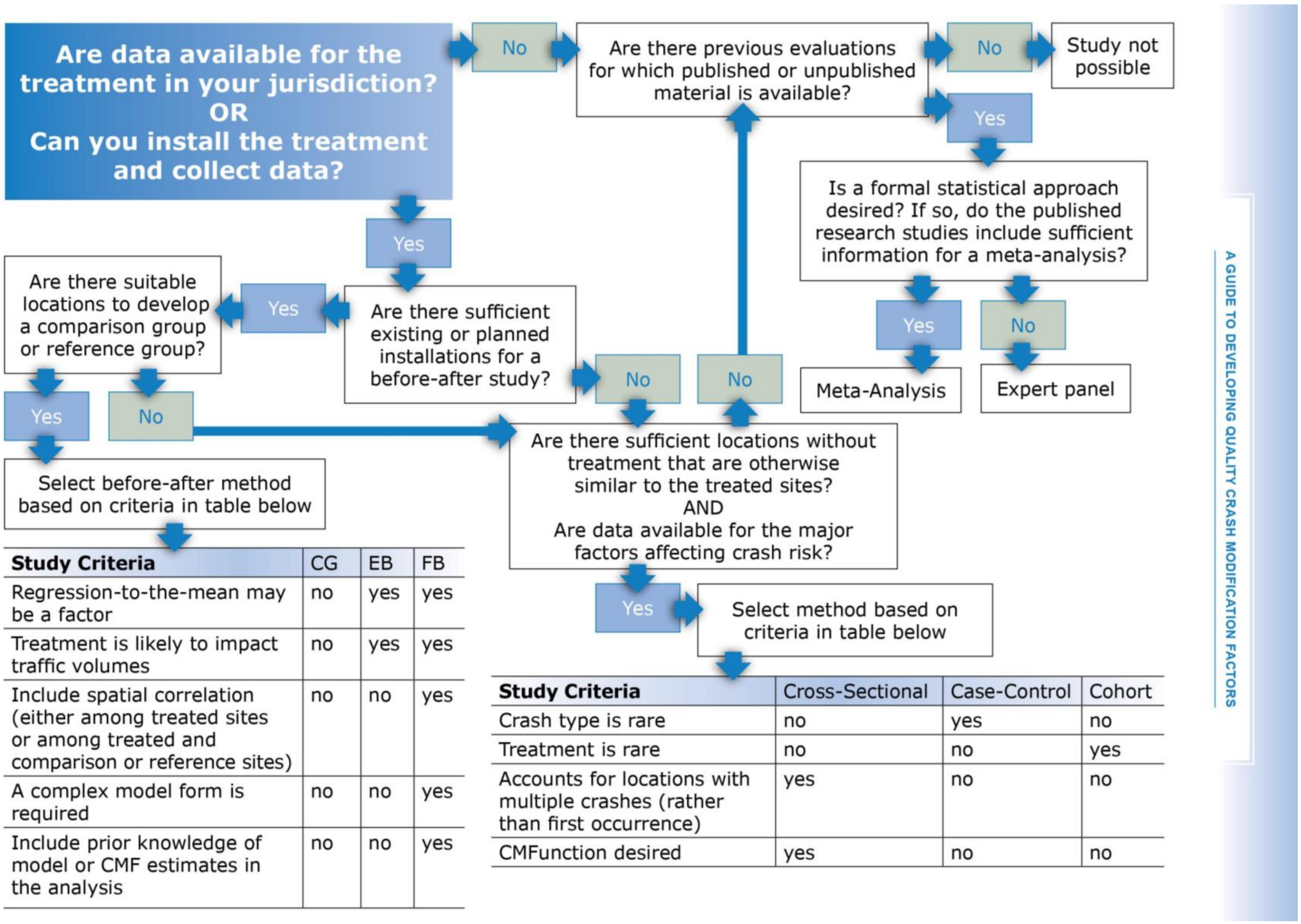
Zhou, H., Pirinccioglu, F., and Hsu, P. (2009). "A New Roadway Lighting Measurement System." *Transportation Research Part C: Emerging Technologies*, 17(3), 274-284.

Zimmer, R. A. (1988). “Mobile Illumination Evaluation System.” *Transportation Research Record*, (1172), 68-73.

APPENDIX A: FLOWCHART FOR CMF STUDY DESIGN SELECTION

This appendix includes the flowchart provided by Gross et al. (2010) for aiding researchers in study design selection for CMF studies. This chart was used to select a cross-sectional study for this research. The flow for this study is as follows:

- Are data available for the treatment in your jurisdiction? OR Can you install the treatment and collect data? (Yes, illuminance data were collected for this study.)
- Are there sufficient existing or planned installations for a before-after study? (No, historical and planned illuminance data are not available.)
- Are there sufficient locations without treatment that are otherwise similar to the treated sites? AND Are data available for the major factors affecting crash risk? (Yes, arterials with similar characteristics were selected and roadway data and crash context data were collected to account for other major factors affecting crash risk.)
- A cross-sectional study was chosen over a case-control or a cohort study because the treatment (lighting) is not rare, the crash type (nighttime crashes) isn’t particularly rare, and it was desired to account for locations with multiple crashes.



Study Criteria	CG	EB	FB
Regression-to-the-mean may be a factor	no	yes	yes
Treatment is likely to impact traffic volumes	no	yes	yes
Include spatial correlation (either among treated sites or among treated and comparison or reference sites)	no	no	yes
A complex model form is required	no	no	yes
Include prior knowledge of model or CMF estimates in the analysis	no	no	yes

Study Criteria	Cross-Sectional	Case-Control	Cohort
Crash type is rare	no	yes	no
Treatment is rare	no	no	yes
Accounts for locations with multiple crashes (rather than first occurrence)	yes	no	no
CMFunction desired	yes	no	no

Figure A-1 Flowchart for Study Design Selection (Gross et al., 2010)

APPENDIX B: SUITABLE LOCATIONS FOR DATA COLLECTION

This appendix includes a table of the full list of suitable locations for data collection. These were not all included in the study, but a random sample of locations was selected from this list. The full list is presented here to show which locations this research is most applicable to. Within the table, there are columns describing the route identification, beginning milepoint and ending milepoint for each suitable location as well as the length in miles. The number of subsegments shows how many segment breaks there are within each location. Segment breaks occur whenever there is a change in roadway characteristics. Segments created from segment breaks constituted the data points used in the analysis, so lighting and crash variables were summarized by each individual segment.

Table B-1 All Suitable Study Locations

Route ID	Number of Subsegments	Length (miles)	Beginning Milepoint	Ending Milepoint
0048PM	6	2.024	0.000	2.024
0048PM	10	1.770	2.025	4.622
0051PM	4	1.034	0.000	3.373
0052PM	13	4.117	0.000	4.117
0068PM	5	0.883	32.285	33.168
0068PM	7	2.562	38.025	40.587
0068PM	58	16.704	40.587	57.291
0068PM	8	2.240	57.291	59.531
0068PM	9	3.158	59.531	62.689
0071PM	6	2.894	0.000	2.894
0071PM	9	2.045	3.939	5.984
0071PM	31	9.743	5.984	15.727
0071PM	30	6.755	15.727	22.482
0074PM	17	5.075	0.000	5.075
0075PM	3	0.934	0.000	0.934
0077PM	8	2.467	6.636	9.103
0089PM	17	3.348	333.393	336.741
0089PM	23	9.308	336.741	346.049
0089PM	7	1.449	362.641	364.090
0089PM	12	2.822	364.090	366.912
0089PM	4	1.127	366.912	368.039
0089PM	54	13.901	368.039	381.940
0089PM	2	0.338	328.335	328.673
0089PM	7	1.244	328.673	329.917
0089PM	7	1.863	346.049	347.912
0089PM	7	1.403	347.912	349.315
0089PM	6	1.218	349.315	350.533
0089PM	8	2.881	350.533	353.414
0092PM	6	0.671	0.000	0.671
0092PM	4	1.370	5.231	6.601
0111PM	8	2.015	8.607	10.622
0114PM	3	0.712	0.000	0.712
0114PM	7	1.804	1.240	3.044
0114PM	2	1.823	3.044	4.867
0114PM	6	2.920	4.867	7.787
0115PM	5	1.655	0.000	1.655

Route ID	Number of Subsegments	Length (miles)	Beginning Milepoint	Ending Milepoint
0129PM	4	1.738	0.000	1.738
0129PM	7	5.599	1.738	7.337
0131PM	5	2.779	0.000	2.779
0145PM	9	1.237	5.641	6.878
0147PM	14	7.024	11.107	18.131
0151PM	18	4.310	0.000	4.310
0152PM	2	0.652	2.402	3.054
0154PM	8	2.151	14.289	16.440
0156PM	7	1.388	0.000	1.388
0171PM	6	2.961	0.000	2.961
0171PM	10	2.781	2.961	5.742
0171PM	19	5.100	5.742	10.842
0171PM	20	4.811	10.842	15.653
0172PM	24	9.220	0.000	9.220
0173PM	6	2.256	1.531	3.787
0173PM	11	3.918	3.787	7.705
0173PM	9	1.534	7.705	9.941
0175PM	11	4.905	0.000	4.905
0176PM	1	1.053	0.000	1.053
0186PM	2	0.318	0.000	0.318
0186PM	7	0.860	1.077	1.937
0186PM	26	5.122	1.937	7.059
0186PM	6	1.565	7.059	8.624
0189PM	10	1.520	0.133	1.653
0189PM	10	2.752	1.653	4.405
0194PM	3	1.347	1.709	3.056
0198PM	12	4.234	11.519	15.753
0198PM	12	3.039	3.511	6.550
0209PM	10	4.567	5.027	9.594
0209PM	18	4.517	9.594	14.111
0209PM	14	4.831	14.111	18.942
0241PM	7	1.555	0.000	1.555
0265PM	18	4.339	0.000	4.339
0266PM	5	0.906	0.000	0.906
0266PM	9	1.987	2.438	4.425
0266PM	8	1.129	4.425	5.554
0266PM	10	2.575	5.554	8.129

APPENDIX C: RANDOM FORESTS DECISION TREE EXAMPLE

A Random Forest is a machine learning model useful in classification and prediction. At each iteration (denoted as a “tree”), a random subset of explanatory variables is chosen from the dataset. A mathematical algorithm then selects the best way to “split” the data based on one of the explanatory variables (denoted as a “branch”). The data is continually split according to the best available explanatory variable until a full decision tree is made. Each new observation can then receive a predicted response value based on its explanatory variables and the given decision tree. The algorithm is called a “Random Forest” because many trees are created, each with a different random subset of explanatory variables. The algorithm also tracks how useful each explanatory variable is in the creation of the decision trees, assigning each variable a score based on importance. The importance scores of the variables are a useful method of variable selection when looking at a specific response (Genuer et al., 2010).

The example shown in Figure C-1 is just one tree, but it is important to remember that a Random Forest iterates through thousands of trees, using a random subset of the data each time. A Random Forest model was used to determine which variables were best at predicting nighttime crashes (i.e., which variables mathematically caused the most branches). Thus, the results of individual trees are meaningless since only the variables which cause branches are important and not the values the branches led to. This example is only meant to illustrate how Random Forests work but was not used in the analysis on its own.

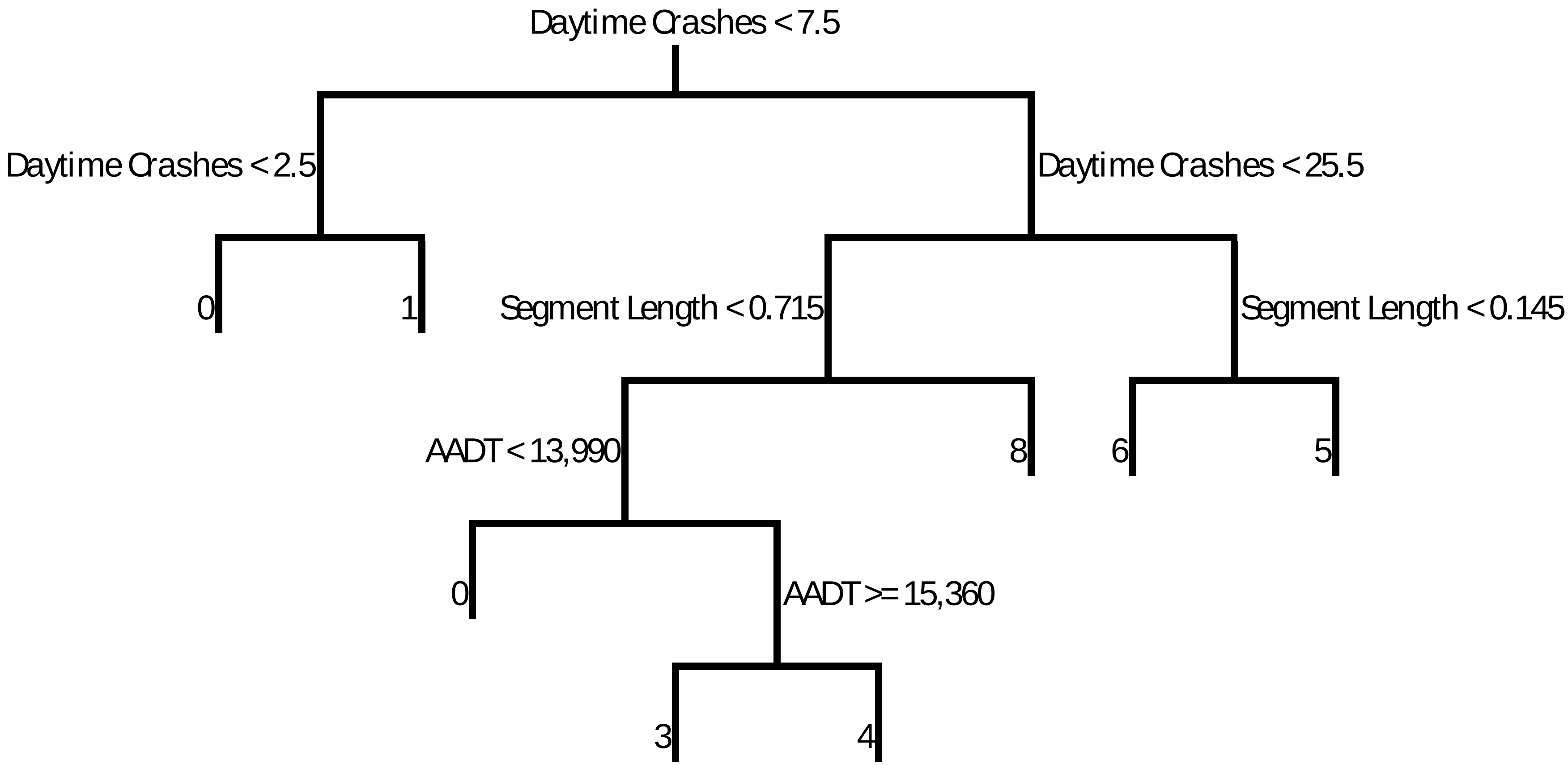


Figure C-1 Decision Tree Example from Random Forests

APPENDIX D: SINGLE-CUTOFF METHOD CMFs

This appendix includes additional CMFs developed by the single-cutoff method which were not necessarily considered useful for this research. However, these show that changing the cutoffs can change the CMFs quite a bit. Therefore, CMFs should only be applied at the cutoffs they were designated for.

Table D-1 Statistically Significant CMFs

Base Illuminance Property	Base Conditions	Countermeasure	Crash Type	CMF	Standard Error	Confidence Interval	Significance Level
Std. Dev. > 0.3 fc and < 1.9 fc	All	(Avg Illuminance < 0.3 fc) to (Avg Illuminance > 0.3 fc and < 1.9 fc)	Night	0.73	-	(0.58, 0.92)	0.05 (95%)
Avg Illuminance < 0.9 fc	All	(Std. Dev. < 0.6 fc) to (Std. Dev. > 0.6 fc and < 1.9 fc)	Night	0.81	0.12	(0.67, 0.99)	0.15 (85%)
Avg Illuminance < 0.9 fc	All	(max/min < 2.5) to (max/min > 2.5 and < 50)	Night	0.85	0.08	(0.73, 0.98)	0.1 (90%)
Avg Illuminance < 0.4 fc	All	(Std. Dev. < 0.3 fc) to (Std. Dev. > 0.3 fc and < 1.9 fc)	Night to Day Crash Ratio	0.90	0.07	(0.81, 0.99)	0.15 (85%)
Avg Illuminance < 1.0 fc	Segment Length between 0 and 0.3 miles	(Std. Dev. < 0.7 fc) to (Std. Dev. > 0.7 fc and < 1.9 fc)	Night	0.61	0.21	(0.37, 1.00)	0.1 (90%)
Avg Illuminance > 1.0 fc	Segment Length between 0.3 and 1 miles	(Std. Dev. > 0.7 fc and < 1.9 fc) to (Std. Dev. < 0.7 fc)	Night	0.65	0.15	(0.46, 0.92)	0.1 (90%)
Avg Illuminance < 0.9 fc	Segment Length between 0 and 0.3 miles	(max/min < 3.5) to (max/min > 3.5 and < 50)	Night	0.81	0.10	(0.67, 0.98)	0.1 (90%)
Avg Illuminance > 0.9 fc	Segment Length between 0 and 0.3 miles	(lighting frequency < 35 fluctuations/mile) to (lighting frequency > 35 fluctuations/mile)	Night	0.66	0.15	(0.47, 0.93)	0.1 (90%)

APPENDIX E: DATA PREPARATION

The following figures show the R code and GIS workflows used to prepare the data for this research. The figures are listed in the order they were executed as follows:

- Figure E-1 shows the code used to clean the raw data and prepare it for GIS analysis.
- Figure E-2 shows the GIS workflow used to interpolate and join light data to segments.
- Figure E-3 shows the code used to join segment and light data to crash data.

```r
library(tidyverse)
library(sf)

# Read in data
light <- read_tsv("data/light/LIGHT.txt", skip = 9) %>%
  filter(!is.na(Date) & Date != "Date") %>%
  select(-contains("..."), -DATAH, -CHK)

light2 <- read_tsv("data/light/History/LIGHT_20230801.txt", skip = 9) %>%
  filter(!is.na(Date) & Date != "Date") %>%
  select(-contains("..."), -DATAH, -CHK)

light <- rbind(light2, light) %>%
  mutate(
    Record = as.numeric(Record),
    LEFT = as.numeric(LEFT),
    RIGHT = as.numeric(RIGHT),
    GPS_LAT = as.numeric(GPS_LAT),
    GPS_LONG = as.numeric(GPS_LONG),
    GPS_ELEV = as.numeric(GPS_ELEV),
    GPS_NSATS = as.numeric(GPS_NSATS),
    GPS_HDOP = as.numeric(GPS_HDOP),
    LANE = as.numeric(LANE)
  )


# Create datetimes
datetimes <- as.POSIXct(strptime(paste0(light$Date," ",light$Time), "%Y-%m-%d
%H:%M:%S"))
light$datetime <- datetimes
light <- light %>% arrange(datetime)


# Add session numbers
light$Record <- as.integer(light$Record)
session <- 0
light$Session <- NA
for(i in 1:nrow(light)){
  if(light$Record[i] == 0){
    session <- session + 1
  }
  light$Session[i] <- session
}


# Convert lat/long to UTM meters
light <- light %>%
  st_as_sf(coords = c("GPS_LONG", "GPS_LAT"), crs = 4326, remove = FALSE) %>%
  st_transform(crs = 26912)
light <- light %>%
  mutate(UTM_X = unlist(map(light$geometry,1)),
         UTM_Y = unlist(map(light$geometry,2))) %>%
  st_drop_geometry()
```

Figure E-1 Data Cleaning Code

```r
# Calculate distances and trajectories between points
light$dist_prev <- NA
light$dist_nxt <- NA
light$traj <- NA
light$traj_prev <- NA
light$traj_nxt <- NA
for(i in 1:nrow(light)){
  session <- light$Session[i]
  nxt_session <- ifelse((i+1) <= nrow(light), light$Session[i+1], 0)
  prev_session <- ifelse((i-1) > 0, light$Session[i-1], 0)
  x <- light$UTM_X[i]
  y <- light$UTM_Y[i]
  prev_x <- light$UTM_X[i-1]
  prev_y <- light$UTM_Y[i-1]
  nxt_x <- light$UTM_X[i+1]
  nxt_y <- light$UTM_Y[i+1]
  if(prev_session == session){
    light$dist_prev[i] <- sqrt((x-prev_x)^2+(y-prev_y)^2)
    x_dist <- (x-prev_x)
    y_dist <- (y-prev_y)
    if(x_dist < 0){
      light$traj_prev[i] <- atan(y_dist/x_dist) + pi
    } else{
      light$traj_prev[i] <- atan(y_dist/x_dist)
    }
  }
  if(nxt_session == session){
    light$dist_nxt[i] <- sqrt((x-nxt_x)^2+(y-nxt_y)^2)
    x_dist <- (nxt_x-x)
    y_dist <- (nxt_y-y)
    if(x_dist < 0){
      light$traj_nxt[i] <- atan(y_dist/x_dist) + pi
    } else{
      light$traj_nxt[i] <- atan(y_dist/x_dist)
    }
  }
  if(prev_session == nxt_session){
    x_dist <- (nxt_x-prev_x)
    y_dist <- (nxt_y-prev_y)
    if(x_dist < 0){
      light$traj[i] <- atan(y_dist/x_dist) + pi
    } else{
      light$traj[i] <- atan(y_dist/x_dist)
    }
  }
}


# Convert measurements to numeric LUX and round
light$LEFT <- round(as.numeric(light$LEFT)*1000,4)
light$RIGHT <- round(as.numeric(light$RIGHT)*1000,4)
```

Figure E-1 (continued)

```r
# Average point clumps by 5 ft intervals
interval <- 5 / 3.2808399
light$flagged <- FALSE
for(i in 1:nrow(light)){
  # check if it's end of session
  if(!is.na(light$dist_nxt[i])){
    # check if distance between points is too small
    if(light$dist_nxt[i] < interval){
      # average measurements and location readings
      light$LEFT[i+1] <- mean(c(light$LEFT[i],light$LEFT[i+1]),na.rm=TRUE)
      light$RIGHT[i+1] <- mean(c(light$RIGHT[i],light$RIGHT[i+1]),na.rm=TRUE)
      light$UTM_X[i+1] <- mean(c(light$UTM_X[i],light$UTM_X[i+1]),na.rm=TRUE)
      light$UTM_Y[i+1] <- mean(c(light$UTM_Y[i],light$UTM_Y[i+1]),na.rm=TRUE)
      # recalculate distance to next
      if(!is.na(light$dist_nxt[i+1])){
        x_dist <- (light$UTM_X[i+2]-light$UTM_X[i+1])
        y_dist <- (light$UTM_Y[i+2]-light$UTM_Y[i+1])
        light$dist_nxt[i+1] <- sqrt(x_dist^2+y_dist^2)
      }
      # flag row for deletion
      light$flagged[i] <- TRUE
    }
  }
}
light <- light %>% filter(flagged == FALSE) %>% select(-flagged)


# Recalculate distances and trajectories (yes this is inefficient)
light$dist_prev <- NA
light$dist_nxt <- NA
light$traj <- NA
light$traj_prev <- NA
light$traj_nxt <- NA
for(i in 1:nrow(light)){
  session <- light$Session[i]
  nxt_session <- ifelse((i+1) <= nrow(light), light$Session[i+1], 0)
  prev_session <- ifelse((i-1) > 0, light$Session[i-1], 0)
  x <- light$UTM_X[i]
  y <- light$UTM_Y[i]
  prev_x <- light$UTM_X[i-1]
  prev_y <- light$UTM_Y[i-1]
  nxt_x <- light$UTM_X[i+1]
  nxt_y <- light$UTM_Y[i+1]
  if(prev_session == session){
    light$dist_prev[i] <- sqrt((x-prev_x)^2+(y-prev_y)^2)
    x_dist <- (x-prev_x)
    y_dist <- (y-prev_y)
    if(x_dist < 0){
      light$traj_prev[i] <- atan(y_dist/x_dist) + pi
    } else{
      light$traj_prev[i] <- atan(y_dist/x_dist)
    }
  }
```

Figure E-1 (continued)

```
  if(nxt_session == session){
    light$dist_nxt[i] <- sqrt((x-nxt_x)^2+(y-nxt_y)^2)
    x_dist <- (nxt_x-x)
    y_dist <- (nxt_y-y)
    if(x_dist < 0){
      light$traj_nxt[i] <- atan(y_dist/x_dist) + pi
    } else{
      light$traj_nxt[i] <- atan(y_dist/x_dist)
    }
  }
  if(prev_session == nxt_session){
    x_dist <- (nxt_x-prev_x)
    y_dist <- (nxt_y-prev_y)
    if(x_dist < 0){
      light$traj[i] <- atan(y_dist/x_dist) + pi
    } else{
      light$traj[i] <- atan(y_dist/x_dist)
    }
  }
}


# Pivot longer
light <- light %>%
  pivot_longer(cols = LEFT:RIGHT, names_to = "SENSOR", values_to = "LUX")


# Displace left and right sensor locations
offset <- 3 / 3.2808399
for(i in 1:nrow(light)){
  x <- light$UTM_X[i]
  y <- light$UTM_Y[i]
  if(!is.na(light$traj[i])){
    traj <- light$traj[i]
  } else if(!is.na(light$traj_prev[i])){
    traj <- light$traj_prev[i]
  } else if(!is.na(light$traj_nxt[i])){
    traj <- light$traj_nxt[i]
  } else{
    traj <- 0
  }
  # calculate perpendicular angle
  if(light$SENSOR[i] == "RIGHT"){
    traj <- traj - pi / 2
  } else if(light$SENSOR[i] == "LEFT"){
    traj <- traj + pi / 2
  } else{
    print("error: improper sensor name")
    break
  }
  # calculate x and y distances
  x_dist <- cos(traj) * offset
  y_dist <- sin(traj) * offset
```

Figure E-1 (continued)

```r
  # assign new x and y
  light$UTM_X[i] <- x + x_dist
  light$UTM_Y[i] <- y + y_dist
}


# Filter out unusable measurements (cap was left on)
light <- light %>%
  mutate(
    flagged = ifelse(
      SENSOR == "RIGHT" &
        datetime > as.POSIXct(strptime("2023-08-08 20:00:00", "%Y-%m-%d %H:%M:%S")) &
        datetime < as.POSIXct(strptime("2023-08-09 06:00:00", "%Y-%m-%d %H:%M:%S")),
      TRUE, FALSE
    )
  ) %>%
  filter(flagged == FALSE) %>%
  select(-flagged)


# Rename mislabeled route
light <- light %>%
  mutate(
    flagged = ifelse(
      datetime > as.POSIXct(strptime("2023-12-15 21:32:37", "%Y-%m-%d %H:%M:%S")) &
        datetime < as.POSIXct(strptime("2023-12-15 21:49:09", "%Y-%m-%d %H:%M:%S")),
      TRUE, FALSE
    ),
    ROUTE = ifelse(flagged == TRUE, 189, ROUTE)
  ) %>%
  select(-flagged)


# Calculate foot-candles from lux
light <- light %>% mutate(FC = LUX / 10.7639104167)


# Add geometry from lat/long
light <- light %>%
  st_as_sf(coords = c("UTM_X", "UTM_Y"), crs = 26912)


# Visualize
plot(light %>% filter(GPS_LAT != 0) %>% select(LUX, geometry))


# Export shapefile
write_sf(light, "data/output/light_summer.shp")
# write_sf(light, "data/output/light_winter.shp")


# Export csv
write_csv(light, "data/output/light_summer.csv")
# write_csv(light, "data/output/light_winter.csv")
```
Figure E-1 (continued)

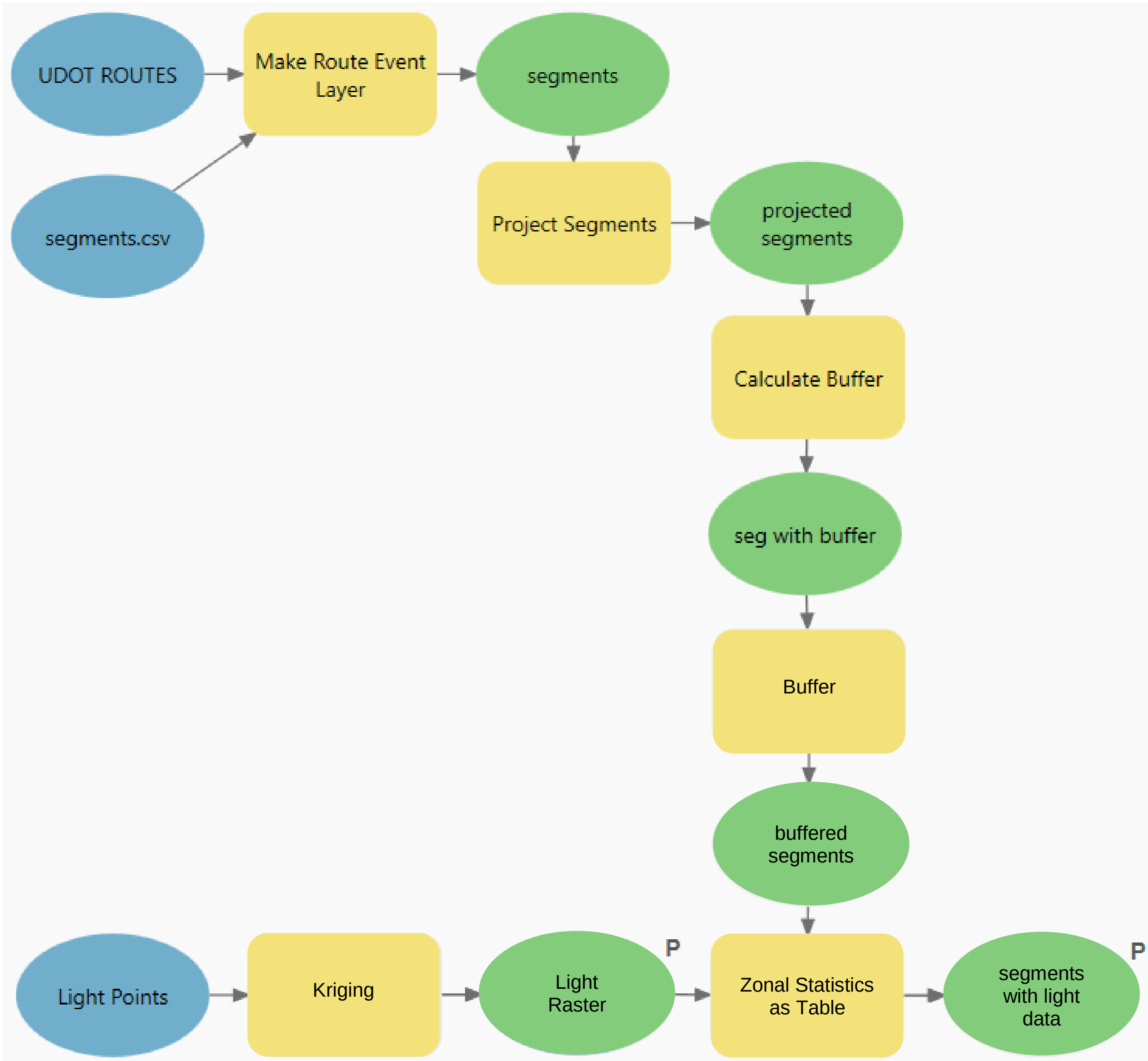


Figure E-2 Light Data Interpolation in GIS

```r
library(tidyverse)
library(sf)
library(openxlsx)
source("R/functions.R")



compile_light_data <- function(light_seg_stats,
                               seg_characteristics,
                               crash,
                               crash_light,
                               intersection,
                               unif,
                               seg_path,
                               crash_path){

# combine data
seg <- left_join(seg_characteristics, light_seg_stats, by = "Seg_ID")


# filter out low data segments
seg <- seg %>%
  mutate(low_data_flag = ifelse(COUNT < 30, TRUE, FALSE)) %>%
  filter(low_data_flag == FALSE) %>%
  select(-low_data_flag)


# fix seg length
seg <- seg %>% mutate(Seg_Length = EMP - BMP)


# Assign light to crashes
crash <- left_join(crash, crash_light, by = c("Crash.ID"="Crash_ID"))


# Create nighttime-only metrics
crash <- crash %>%
  mutate(
    UDOT_crashtype = case_when(
      Light.Condition == "Dark - Lighted" ~ "Nighttime",
      Light.Condition == "Dark - Not Lighted" ~ "Nighttime",
      Light.Condition == "Dark - Unknown Lighting" ~ "Nighttime",
      Light.Condition == "Daylight" ~ "Daytime",
      Light.Condition == "Dusk" ~ "Twilight",
      Light.Condition == "Dawn" ~ "Twilight",
      Light.Condition == "Unknown" ~ "Twilight"
      ),
    night_severity = paste0(Crash.Severity, " ", crashtype),
    UDOT_night_severity = paste0(Crash.Severity, " ", UDOT_crashtype),
    night_ped = ifelse(Pedestrian.Involved == "Y" & crashtype == "Nighttime", "Y",
"N"),
    UDOT_night_ped = ifelse(Pedestrian.Involved == "Y" & UDOT_crashtype ==
"Nighttime", "Y", "N")
  )
```

Figure E-3 Crash Data Joining Code

```r
# Separate Crashes into Intersection and Non-Intersection Related
crash_sf <- st_as_sf(crash %>% filter(!is.na(Longitude)), coords = c("Longitude",
"Latitude"), crs = 4326, remove = F) %>%
  st_transform(crash, crs = 26912) %>%
  select(crash_id=Crash.ID, intersection_related=Intersection.Involved,
long=Longitude, lat=Latitude) %>%
  mutate(long = as.numeric(long),
         lat = as.numeric(lat))
int_buff <- st_as_sf(intersection, coords = c("long_text", "lat_text"), crs = 4326,
remove = F) %>%
  st_transform(intersection, crs = 26912) %>%
  st_buffer(dist = intersection$Leg_Distan * 0.3048) %>% #buffer by FA
  select(Int_ID = UniqueID, long_int=long_text, lat_int=lat_text) %>%
  mutate(long_int = as.numeric(long_int),
         lat_int = as.numeric(lat_int))


# determine which crashes are within a functional area
crash_sf <- st_join(crash_sf, int_buff, join = st_within) %>%
  st_drop_geometry() %>%
  mutate(int_id = ifelse(intersection_related=="N",NA,Int_ID)) %>%
  select(-Int_ID) %>%
  arrange(crash_id)


# choose the closest intersection when there are overlaps
for(i in 1:nrow(crash_sf)){
  if(!is.na(crash_sf$int_id[[i]])){
    # get list of possible intersections
    id <- crash_sf$crash_id[[i]]
    ints <- which(crash_sf$crash_id==id)
    # check for duplicates and find closest intersection
    if(length(ints)>1){
      # assign crash lat long
      lat_crash <- crash_sf$lat[[i]]
      long_crash <- crash_sf$long[[i]]
      # find closest intersection
      closest_id <- 0
      closest_dist <- 100000
      start <- ints[1]
      end <- ints[length(ints)]
      for(j in start:end){
        # assign intersection lat long
        lat_int <- crash_sf$lat_int[[j]]
        long_int <- crash_sf$long_int[[j]]
        # calculate euclidean distance
        dist <- sqrt((lat_crash-lat_int)^2+(long_crash-long_int)^2)
        # check if closest
        if(dist < closest_dist){
          closest_dist <- dist
          closest_id <- crash_sf$int_id[[j]]
        }
      }
```

Figure E-3 (continued)

```r
      # assign closest intersection id to crash
      for(k in start:end){
        crash_sf$int_id[k] <- closest_id
      }
      # skip ahead to avoid redundancy
      i <- ints[-1]
    }
  }
}


# delete duplicates
crash_sf <- crash_sf %>%
  select(crash_id, int_id) %>%
  unique()


# join back to crash file
crash <- left_join(crash, crash_sf, by = c("Crash.ID" = "crash_id"))


# Filter "Intersection Related" Crashes
crash_seg <- crash %>% filter(is.na(int_id)) %>% select(-int_id)
crash_int <- crash %>% filter(!is.na(int_id))


# assign crashes to segments
# remove "M" from route name
seg$Route_ID <- substr(seg$Route_ID,1,5)
# identify segments for each crash
crash_seg$seg_id <- NA
for (i in 1:nrow(crash_seg)){
  rt <- crash_seg$Full.Route.Name[i]
  mp <- crash_seg$Milepoint[i]
  seg_row <- which(seg$Route_ID == rt &
                     seg$BMP < mp &
                     seg$EMP > mp)
  if(length(seg_row) > 0){
    crash_seg[["seg_id"]][i] <- seg$Seg_ID[seg_row]
  }
}
# delete unassigned crashes
crash_seg <- crash_seg %>% filter(!is.na(seg_id))
# save crash_seg with ids
write_csv(crash_seg, "data/temp/crash_seg.csv")
write_csv(crash_int, "data/temp/crash_int.csv")
write_csv(seg, "data/temp/seg_nocrash.csv")


# PICK UP FROM HERE IF NOTHING BEFORE HAS CHANGED
crash_seg <- read_csv("data/temp/crash_seg.csv")
seg <- read_csv("data/temp/seg_nocrash.csv")
```

Figure E-3 (continued)

```r
# Add uniformity to segments
seg <- left_join(seg, unif, by = "Seg_ID")


# Add crash Severity to segments by type
seg <- add_crash_attribute("night_severity", seg, crash_seg) %>%
  rename(Night.Severity.1 = `night_severity_No injury/PDO Nighttime`,
         Night.Severity.2 = `night_severity_Possible injury Nighttime`,
         Night.Severity.3 = `night_severity_Suspected Minor Injury Nighttime`,
         Night.Severity.4 = `night_severity_Suspected Serious Injury Nighttime`,
         Night.Severity.5 = `night_severity_Fatal Nighttime`,
         Day.Severity.1 = `night_severity_No injury/PDO Daytime`,
         Day.Severity.2 = `night_severity_Possible injury Daytime`,
         Day.Severity.3 = `night_severity_Suspected Minor Injury Daytime`,
         Day.Severity.4 = `night_severity_Suspected Serious Injury Daytime`,
         Day.Severity.5 = `night_severity_Fatal Daytime`,
         Twilight.Severity.1 = `night_severity_No injury/PDO Twilight`,
         Twilight.Severity.2 = `night_severity_Possible injury Twilight`,
         Twilight.Severity.3 = `night_severity_Suspected Minor Injury Twilight`,
         Twilight.Severity.4 = `night_severity_Suspected Serious Injury Twilight`,
         Twilight.Severity.5 = `night_severity_Fatal Twilight`)


# Add crash Severity to segments by UDOT type
seg <- add_crash_attribute("UDOT_night_severity", seg, crash_seg) %>%
  rename(UDOT.Night.Severity.1 = `UDOT_night_severity_No injury/PDO Nighttime`,
         UDOT.Night.Severity.2 = `UDOT_night_severity_Possible injury Nighttime`,
         UDOT.Night.Severity.3 = `UDOT_night_severity_Suspected Minor Injury
Nighttime`,
         UDOT.Night.Severity.4 = `UDOT_night_severity_Suspected Serious Injury
Nighttime`,
         UDOT.Night.Severity.5 = `UDOT_night_severity_Fatal Nighttime`,
         UDOT.Day.Severity.1 = `UDOT_night_severity_No injury/PDO Daytime`,
         UDOT.Day.Severity.2 = `UDOT_night_severity_Possible injury Daytime`,
         UDOT.Day.Severity.3 = `UDOT_night_severity_Suspected Minor Injury Daytime`,
         UDOT.Day.Severity.4 = `UDOT_night_severity_Suspected Serious Injury
Daytime`,
         UDOT.Day.Severity.5 = `UDOT_night_severity_Fatal Daytime`,
         UDOT.Twilight.Severity.1 = `UDOT_night_severity_No injury/PDO Twilight`,
         UDOT.Twilight.Severity.2 = `UDOT_night_severity_Possible injury Twilight`,
         UDOT.Twilight.Severity.3 = `UDOT_night_severity_Suspected Minor Injury
Twilight`,
         UDOT.Twilight.Severity.4 = `UDOT_night_severity_Suspected Serious Injury
Twilight`
         # UDOT.Twilight.Severity.5 = `UDOT_night_severity_Fatal Twilight`
         )
seg$UDOT.Twilight.Severity.5 = 0


# Add Segment Crash Years
seg <- add_crash_attribute("Year", seg,
                           crash_seg %>% filter(crashtype == "Nighttime"),
                           prefix = "Nighttime_") %>%
  rename(Nighttime_2018_Crashes = Nighttime_Year_2018,
```

Figure E-3 (continued)

```
         Nighttime_2019_Crashes = Nighttime_Year_2019,
         Nighttime_2020_Crashes = Nighttime_Year_2020,
         Nighttime_2021_Crashes = Nighttime_Year_2021,
         Nighttime_2022_Crashes = Nighttime_Year_2022)


# Add Segment Crash Attributes
seg <- add_crash_attribute("crashtype", seg, crash_seg)
seg <- add_crash_attribute("UDOT_crashtype", seg, crash_seg)


seg <- add_crash_attribute("MEAN", seg,
                           crash_seg %>% filter(crashtype == "Nighttime"),
                           return_stats = TRUE,
                           prefix = "crash_light_") %>%
  mutate(
    crash_light_max_MEAN = ifelse(crash_light_max_MEAN == 0, MAX,
crash_light_max_MEAN),
    crash_light_min_MEAN = ifelse(crash_light_min_MEAN == 0, MIN,
crash_light_min_MEAN),
    crash_light_avg_MEAN = ifelse(crash_light_avg_MEAN == 0, MEAN,
crash_light_avg_MEAN),
    crash_light_med_MEAN = ifelse(crash_light_med_MEAN == 0, MEDIAN,
crash_light_med_MEAN),
    crash_light_sd_MEAN = ifelse(crash_light_sd_MEAN == 0, STD, crash_light_sd_MEAN)
  )


seg <- add_crash_attribute("STD", seg,
                           crash_seg %>% filter(crashtype == "Nighttime"),
                           return_stats = TRUE,
                           prefix = "crash_light_") %>%
  mutate(
    crash_light_max_STD = ifelse(crash_light_max_STD == 0, STD, crash_light_max_STD),
    crash_light_min_STD = ifelse(crash_light_min_STD == 0, STD, crash_light_min_STD),
    crash_light_avg_STD = ifelse(crash_light_avg_STD == 0, STD, crash_light_avg_STD),
    crash_light_med_STD = ifelse(crash_light_med_STD == 0, STD, crash_light_med_STD),
    crash_light_sd_STD = ifelse(crash_light_sd_STD == 0, STD, crash_light_sd_STD)
  )


seg <- add_crash_attribute("Weather.Condition", seg,
                           crash_seg %>% filter(crashtype == "Nighttime"),
                           prefix = "Nighttime_")


seg <- add_crash_attribute("DUI.Involved", seg,
                           crash_seg %>% filter(crashtype == "Nighttime"),
                           prefix = "Nighttime_") %>%
  select(-Nighttime_DUI.Involved_N) %>%
  rename(Nighttime_DUI_Crashes = Nighttime_DUI.Involved_Y)
seg <- add_crash_attribute("DUI.Involved", seg,
                           crash_seg %>% filter(crashtype == "Daytime"),
                           prefix = "Daytime_") %>%
```

Figure E-3 (continued)

```r
  select(-Daytime_DUI.Involved_N) %>%
  rename(Daytime_DUI_Crashes = Daytime_DUI.Involved_Y)
seg <- add_crash_attribute("DUI.Involved", seg,
                           crash_seg %>% filter(crashtype == "Twilight"),
                           prefix = "Twilight_") %>%
  select(-Twilight_DUI.Involved_N) %>%
  rename(Twilight_DUI_Crashes = Twilight_DUI.Involved_Y)


seg <- add_crash_attribute("Drowsy.Driving.Involved", seg,
                           crash_seg %>% filter(crashtype == "Nighttime"),
                           prefix = "Nighttime_") %>%
  select(-Nighttime_Drowsy.Driving.Involved_N) %>%
  rename(Nighttime_Drowsy_Crashes = Nighttime_Drowsy.Driving.Involved_Y)
seg <- add_crash_attribute("Drowsy.Driving.Involved", seg,
                           crash_seg %>% filter(crashtype == "Daytime"),
                           prefix = "Daytime_") %>%
  select(-Daytime_Drowsy.Driving.Involved_N) %>%
  rename(Daytime_Drowsy_Crashes = Daytime_Drowsy.Driving.Involved_Y)
seg <- add_crash_attribute("Drowsy.Driving.Involved", seg,
                           crash_seg %>% filter(crashtype == "Twilight"),
                           prefix = "Twilight_") %>%
  select(-Twilight_Drowsy.Driving.Involved_N) %>%
  rename(Twilight_Drowsy_Crashes = Twilight_Drowsy.Driving.Involved_Y)


seg <- add_crash_attribute("Pedestrian.Involved", seg,
                           crash_seg %>% filter(crashtype == "Nighttime"),
                           prefix = "Nighttime_") %>%
  select(-Nighttime_Pedestrian.Involved_N) %>%
  rename(Nighttime_Pedestrian_Crashes = Nighttime_Pedestrian.Involved_Y)
seg <- add_crash_attribute("Pedestrian.Involved", seg,
                           crash_seg %>% filter(crashtype == "Daytime"),
                           prefix = "Daytime_") %>%
  select(-Daytime_Pedestrian.Involved_N) %>%
  rename(Daytime_Pedestrian_Crashes = Daytime_Pedestrian.Involved_Y)
seg <- add_crash_attribute("Pedestrian.Involved", seg,
                           crash_seg %>% filter(crashtype == "Twilight"),
                           prefix = "Twilight_") %>%
  select(-Twilight_Pedestrian.Involved_N) %>%
  rename(Twilight_Pedestrian_Crashes = Twilight_Pedestrian.Involved_Y)


seg <- add_crash_attribute("Bicycle.Involved", seg,
                           crash_seg %>% filter(crashtype == "Nighttime"),
                           prefix = "Nighttime_") %>%
  select(-Nighttime_Bicycle.Involved_N) %>%
  rename(Nighttime_Bicycle_Crashes = Nighttime_Bicycle.Involved_Y)
seg <- add_crash_attribute("Bicycle.Involved", seg,
                           crash_seg %>% filter(crashtype == "Daytime"),
                           prefix = "Daytime_") %>%
  select(-Daytime_Bicycle.Involved_N) %>%
  rename(Daytime_Bicycle_Crashes = Daytime_Bicycle.Involved_Y)
seg <- add_crash_attribute("Bicycle.Involved", seg,
```

Figure E-3 (continued)

```
                           crash_seg %>% filter(crashtype == "Twilight"),
                           prefix = "Twilight_") %>%
  select(-Twilight_Bicycle.Involved_N) %>%
  rename(Twilight_Bicycle_Crashes = Twilight_Bicycle.Involved_Y)


seg <- add_crash_attribute("VRU", seg,
                           crash_seg %>% filter(crashtype == "Nighttime"),
                           prefix = "Nighttime_") %>%
  select(-Nighttime_VRU_N) %>%
  rename(Nighttime_VRU_Crashes = Nighttime_VRU_Y)
seg <- add_crash_attribute("VRU", seg,
                           crash_seg %>% filter(crashtype == "Daytime"),
                           prefix = "Daytime_") %>%
  select(-Daytime_VRU_N) %>%
  rename(Daytime_VRU_Crashes = Daytime_VRU_Y)
seg <- add_crash_attribute("VRU", seg,
                           crash_seg %>% filter(crashtype == "Twilight"),
                           prefix = "Twilight_") %>%
  select(-Twilight_VRU_N) %>%
  rename(Twilight_VRU_Crashes = Twilight_VRU_Y)


# Calculate Night to Day crash ratio (set denominator to 0.5 if zero)
seg <- seg %>%
  mutate(
    Nighttime_Injury_Crashes = Night.Severity.3 + Night.Severity.4 +
Night.Severity.5,
    Daytime_Injury_Crashes = Day.Severity.3 + Day.Severity.4 + Day.Severity.5,
    Twilight_Injury_Crashes = Twilight.Severity.3 + Twilight.Severity.4 +
Twilight.Severity.5,
    Nighttime_Severe_Crashes = Night.Severity.4 + Night.Severity.5,
    Daytime_Severe_Crashes = Day.Severity.4 + Day.Severity.5,
    Twilight_Severe_Crashes = Twilight.Severity.4 + Twilight.Severity.5,

    ND_crash_ratio = crashtype_Nighttime / ifelse(crashtype_Daytime +
crashtype_Twilight == 0, 0.5, crashtype_Daytime + crashtype_Twilight),
    ND_injury_crash_ratio = Nighttime_Injury_Crashes / ifelse(Daytime_Injury_Crashes
+ Twilight_Injury_Crashes == 0, 0.5, Daytime_Injury_Crashes +
Twilight_Injury_Crashes),
    ND_severe_crash_ratio = Nighttime_Severe_Crashes / ifelse((Daytime_Severe_Crashes
+ Twilight_Severe_Crashes) == 0, 0.5, (Daytime_Severe_Crashes +
Twilight_Severe_Crashes)),
    ND_ped_crash_ratio = Nighttime_Pedestrian_Crashes /
ifelse((Daytime_Pedestrian_Crashes + Twilight_Pedestrian_Crashes) == 0, 0.5,
(Daytime_Pedestrian_Crashes + Twilight_Pedestrian_Crashes)),
    ND_bike_crash_ratio = Nighttime_Bicycle_Crashes / ifelse((Daytime_Bicycle_Crashes
+ Twilight_Bicycle_Crashes) == 0, 0.5, (Daytime_Bicycle_Crashes +
Twilight_Bicycle_Crashes)),
    ND_vru_crash_ratio = Nighttime_VRU_Crashes / ifelse((Daytime_VRU_Crashes +
Twilight_VRU_Crashes) == 0, 0.5, (Daytime_VRU_Crashes + Twilight_VRU_Crashes)),
```

Figure E-3 (continued)

```
    UDOT_Nighttime_Injury_Crashes = UDOT.Night.Severity.3 + UDOT.Night.Severity.4 +
UDOT.Night.Severity.5,
    UDOT_Daytime_Injury_Crashes = UDOT.Day.Severity.3 + UDOT.Day.Severity.4 +
UDOT.Day.Severity.5,
    UDOT_Twilight_Injury_Crashes = UDOT.Twilight.Severity.3 +
UDOT.Twilight.Severity.4 + UDOT.Twilight.Severity.5,
    UDOT_Nighttime_Severe_Crashes = UDOT.Night.Severity.4 + UDOT.Night.Severity.5,
    UDOT_Daytime_Severe_Crashes = UDOT.Day.Severity.4 + UDOT.Day.Severity.5,
    UDOT_Twilight_Severe_Crashes = UDOT.Twilight.Severity.4 +
UDOT.Twilight.Severity.5,

    UDOT_ND_crash_ratio = UDOT_crashtype_Nighttime / ifelse((UDOT_crashtype_Daytime +
UDOT_crashtype_Twilight) == 0, 0.5, (UDOT_crashtype_Daytime +
UDOT_crashtype_Twilight)),
    UDOT_ND_injury_crash_ratio = UDOT_Nighttime_Injury_Crashes /
ifelse((UDOT_Daytime_Injury_Crashes + UDOT_Twilight_Injury_Crashes) == 0, 0.5,
(UDOT_Daytime_Injury_Crashes + UDOT_Twilight_Injury_Crashes)),
    UDOT_ND_severe_crash_ratio = UDOT_Nighttime_Severe_Crashes /
ifelse((UDOT_Daytime_Severe_Crashes + UDOT_Twilight_Severe_Crashes) == 0, 0.5,
(UDOT_Daytime_Severe_Crashes + UDOT_Twilight_Severe_Crashes))
    # UDOT_ND_ped_crash_ratio = UDOT_Nighttime_Pedestrian_Crashes /
ifelse((UDOT_Daytime_Pedestrian_Crashes + UDOT_Twilight_Pedestrian_Crashes) == 0,
0.5, (UDOT_Daytime_Pedestrian_Crashes + UDOT_Twilight_Pedestrian_Crashes)),
    # UDOT_ND_bike_crash_ratio = UDOT_Nighttime_Bicycle_Crashes /
ifelse((UDOT_Daytime_Bicycle_Crashes + UDOT_Twilight_Bicycle_Crashes) == 0, 0.5,
(UDOT_Daytime_Bicycle_Crashes + UDOT_Twilight_Bicycle_Crashes)),
    # UDOT_ND_vru_crash_ratio = UDOT_Nighttime_VRU_Crashes /
ifelse((UDOT_Daytime_VRU_Crashes + UDOT_Twilight_VRU_Crashes) == 0, 0.5,
(UDOT_Daytime_VRU_Crashes + UDOT_Twilight_VRU_Crashes))
  )


# Convert NaN to zero (shouldn't be necessary anymore)
seg <- seg %>%
  mutate(
    ND_crash_ratio = ifelse(!is.finite(ND_crash_ratio), 0, ND_crash_ratio),
    ND_injury_crash_ratio = ifelse(!is.finite(ND_injury_crash_ratio), 0,
ND_injury_crash_ratio),
    ND_severe_crash_ratio = ifelse(!is.finite(ND_severe_crash_ratio), 0,
ND_severe_crash_ratio),
    ND_ped_crash_ratio = ifelse(!is.finite(ND_ped_crash_ratio), 0,
ND_ped_crash_ratio),
    ND_bike_crash_ratio = ifelse(!is.finite(ND_bike_crash_ratio), 0,
ND_bike_crash_ratio),
    ND_vru_crash_ratio = ifelse(!is.finite(ND_vru_crash_ratio), 0,
ND_vru_crash_ratio),

    UDOT_ND_crash_ratio = ifelse(!is.finite(UDOT_ND_crash_ratio), 0,
UDOT_ND_crash_ratio),
    UDOT_ND_injury_crash_ratio = ifelse(!is.finite(UDOT_ND_injury_crash_ratio), 0,
UDOT_ND_injury_crash_ratio),
    UDOT_ND_severe_crash_ratio = ifelse(!is.finite(UDOT_ND_severe_crash_ratio), 0,
UDOT_ND_severe_crash_ratio)
```

Figure E-3 (continued)

```r
# Delete NAs in light data
seg <- seg %>% filter(!is.na(MEAN))


# Assign segment data to crashes
crash_seg <- left_join(crash_seg, seg, by = c("seg_id" ="Seg_ID"), suffix =
c("_crash", "_seg"))


# Assign city data to segments
cities <- read_csv("data/seg_city.csv")
seg <- left_join(seg, cities, by = "Seg_ID")


# last-minute calcs
seg$RHMVM_Nighttime <- round(seg$crashtype_Nighttime * 100000000 / (seg$Seg_Length *
seg$AADT * 365 * 5))
# seg$RHMVM_Nighttime <- seg$crashtype_Nighttime * 100000000 / (seg$Seg_Length *
seg$AADT * 365 * 5)
seg$RHMVM_Daytime <- round((seg$crashtype_Daytime + seg$crashtype_Twilight) *
100000000 / (seg$Seg_Length * seg$AADT * 365 * 5))
seg <- seg %>%
  mutate(
    lane_group = case_when(Travel_Lan < 3 ~ "1-2 Lanes",
                           Travel_Lan < 5 ~ "3-4 Lanes",
                           Travel_Lan > 4 ~ "5-8 Lanes"),
    Median_Div = case_when(Median_Typ == "CONCRETE BARRIER OR BRIDGE" ~ TRUE,
                           Median_Typ == "DEPRESSED MEDIAN" ~ TRUE,
                           Median_Typ == "NO MEDIAN" ~ FALSE,
                           Median_Typ == "PAINTED MEDIAN" ~ TRUE,
                           Median_Typ == "RAILROAD" ~ TRUE,
                           Median_Typ == "RAISED MEDIAN" ~ TRUE,
                           Median_Typ == "TWO WAY LEFT TURN LANE" ~ TRUE,
                           Median_Typ == "UNDIVIDED" ~ FALSE),
    TWLTL = ifelse(Median_Typ == "TWO WAY LEFT TURN LANE", TRUE, FALSE),
    MEAN_FC = MEAN / 10.76391,
    STD_FC = STD / 10.76391
  )


# export data
write_csv(seg, seg_path)
write_csv(crash_seg, crash_path)


# return
return(seg)


}
```

Figure E-3 (continued)